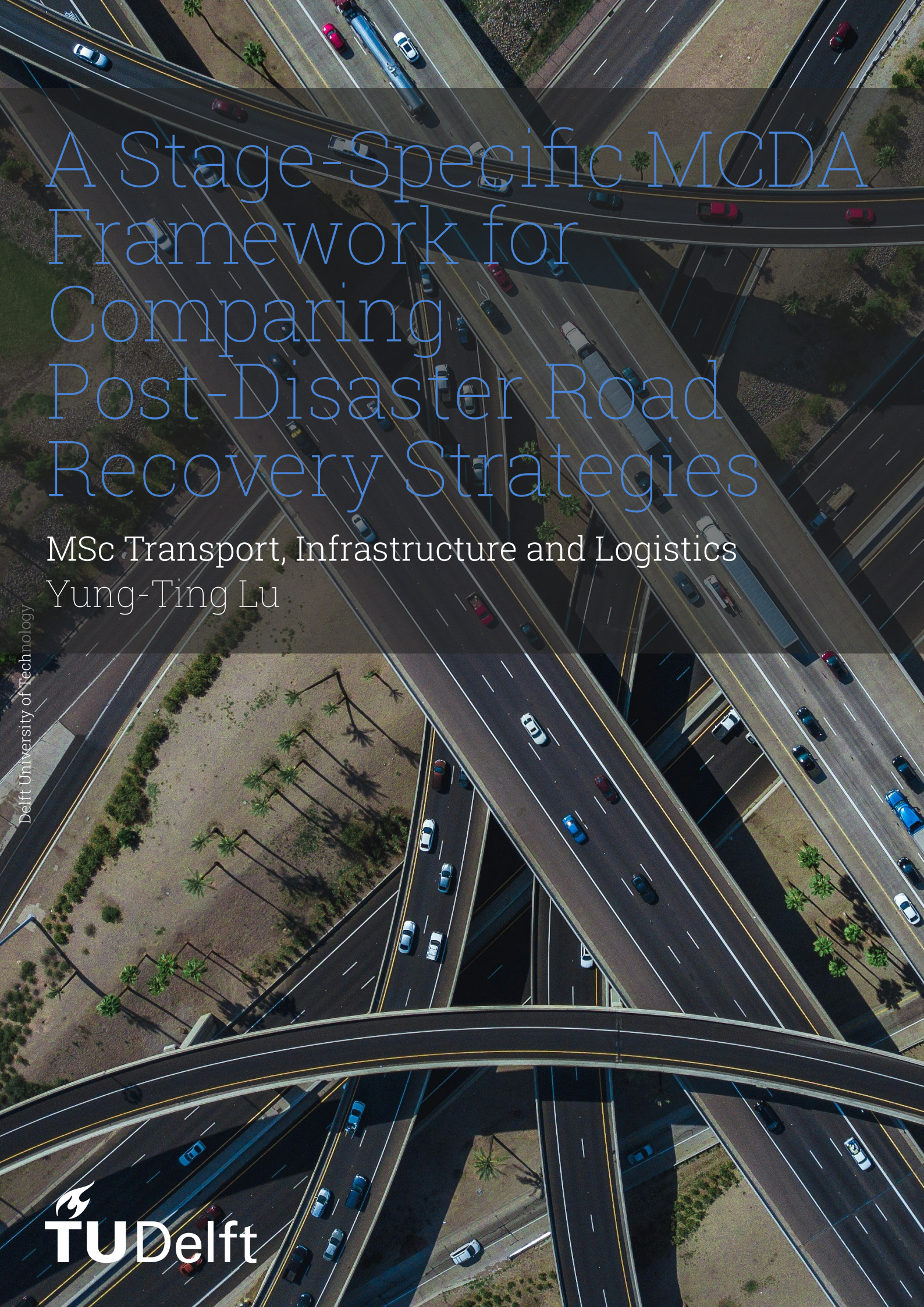


A Stage-Specific MCDA Framework for Comparing Post-Disaster Road Recovery Strategies

MSc Transport, Infrastructure and Logistics

Yung-Ting Lu

Delft University of Technology

TU Delft

A Stage-Specific MCDA Framework for Comparing Post-Disaster Road Recovery Strategies

by

Yung-Ting Lu

Student number: 6221149

Thesis committee: P.S.A. (Patrick) Stokkink (Chair)
S. (Srijith) Balakrishnan (Supervisor)
A.J. (Adam) Pel (Supervisor)
C. (Kees) Maat (Committee member)

Project duration: January 2026 – July 2026
Faculty: Faculty of Civil Engineering and Geosciences
Programme: MSc Transport, Infrastructure and Logistics

Cover photo by Jared Murray on Unsplash.

TU Delft

Preface

Looking back, I never imagined that this journey would become one of the most unforgettable experiences of my life. Completing this thesis marks the end of a very special chapter. My master's journey at TU Delft has challenged me, inspired me, and helped me grow in many ways, both academically and personally. Along the way, I was fortunate to meet many wonderful people who supported, encouraged, and believed in me. Without them, this thesis would not have been possible.

My sincere thanks go first to Patrick, the chair of my thesis committee, for always being so supportive throughout my thesis. I am especially grateful for his patience, his consistently quick feedback, and the help he gave me with both my research and the administrative process. Knowing that I could always rely on his support made this journey much easier.

I would also like to thank Srijith for always being supportive and for sharing many thoughtful suggestions throughout my thesis. His guidance, especially towards the end, helped me step back, see the bigger picture, and understand how the different parts of my research could come together more clearly.

My sincere thanks also go to Adam for his guidance and support throughout my thesis. I especially appreciated the time he took to discuss the simulation with me and the many useful suggestions he shared. His input helped me work through difficulties and gave me more confidence in my analysis.

I would also like to thank Pelin for helping me better understand MCDA and for guiding me through the survey design. Her advice gave me much more confidence in developing the survey, and her support was truly important in making this part of the research possible.

I would also like to thank Professor Kees Maat for taking the time to serve as the external member of my thesis committee and to evaluate my work.

I am deeply grateful to all 61 participants who completed my survey. Their time, willingness, and thoughtful responses made this research possible. This thesis could not have been completed without their contribution. I would also like to thank my friends and family who helped distribute the survey and reach more participants. Their help played an important role in making the data collection possible.

I would especially like to thank Owen for always being by my side throughout this journey. Thank you for taking care of me by driving, cooking, travelling with me, and reminding me to take a break whenever I needed one. Your care, encouragement, and constant support made even the most difficult days much easier, and I will always be grateful for everything you have done for me.

Finally, I would like to thank my family for supporting me throughout my studies in the Netherlands. Their constant love, encouragement, and genuine interest in my research gave me the strength to keep moving forward, especially during the most challenging moments. They always listened to my ideas, understood what I was working on, and believed in me even when I doubted myself. I could not have completed this journey without their trust and unwavering support.

To everyone who has been part of this journey, thank you. I will always treasure the memories, friendships, and experiences that I gained during my time at TU Delft. This thesis is not only the result of my own efforts, but also a reflection of the kindness and support I received from so many wonderful people.

Yung-Ting Lu
Delft, July 2026

Contents

Preface i

Preface i

Glossary of Key Terms vi

Summary viii

1 Introduction 1
- 1.1 Background and problem context . . . 1
- 1.2 Problem statement . . . 1
- 1.3 Research Aim and Questions . . . 2
- 1.4 Research approach . . . 3
- 1.5 Research scope . . . 3
- 1.6 Scientific and practical relevance . . . 4
- 1.7 Thesis outline . . . 4

2 Literature Review 6
- 2.1 Road recovery as a sequencing and prioritisation problem . . . 6
  - 2.1.1 Optimisation-based recovery scheduling . . . 6
  - 2.1.2 Topology- and centrality-based recovery strategies . . . 6
- 2.2 Resilience indicators for evaluating road recovery . . . 7
- 2.3 Stage-dependent recovery dynamics . . . 8
- 2.4 Stakeholder involvement in recovery decision-making . . . 8
- 2.5 MCDA-based decision-making approaches . . . 9
  - 2.5.1 Road recovery as a multi-criteria decision problem . . . 9
  - 2.5.2 MCDA applications and limitations in road and transport resilience . . . 9
  - 2.5.3 Weighting, aggregation, and simulation-based performance data . . . 10
- 2.6 Conclusion of the literature review and research gap . . . 10

3 Methodology 12
- 3.1 Overview of the stage-specific MCDA framework . . . 12
- 3.2 MCDA Decision Problem Formulation . . . 14
  - 3.2.1 Decision Objective and Analytical Contexts . . . 14
  - 3.2.2 Recovery Strategies . . . 14
  - 3.2.3 Evaluation Criteria . . . 15
  - 3.2.4 Recovery Stages . . . 16
  - 3.2.5 Stakeholder groups . . . 17
- 3.3 Best-Worst Method for Deriving Stakeholder Criterion Weights . . . 20
  - 3.3.1 BWM Formulations . . . 20
  - 3.3.2 Consistency Screening and Valid Weight Sets . . . 22
  - 3.3.3 Aggregation of Individual, Group-Level, and Overall Weights . . . 24
  - 3.3.4 Coding of Open-Ended Explanations . . . 25
- 3.4 Stakeholder Survey Design . . . 26
  - 3.4.1 Design requirements . . . 26
  - 3.4.2 How BWM shaped the questionnaire . . . 26
  - 3.4.3 Pilot test . . . 26
  - 3.4.4 The evaluation context shown to respondents . . . 27
  - 3.4.5 Questionnaire structure . . . 28
  - 3.4.6 Administration and ethics . . . 30
- 3.5 Simulation-Based Performance Evaluation Method . . . 30

3.5.1 Damage and Recovery Simulation Logic . . . 30
3.5.2 Recovery Trajectories and Performance Ratios . . . 31
3.5.3 Mapping Restoration Progress to Short- and Long-Term Windows . . . 31
3.5.4 Area-Above-Curve Deviation Scores . . . 32
3.6 Simple Additive Weighting for Strategy Ranking . . . 33
3.6.1 General SAW Formulation . . . 33
3.6.2 Application to Recovery Strategy Ranking . . . 34
3.7 Implementation of the cross-city and Delft applications . . . 35
3.7.1 Cross-City System-Level Implementation . . . 35
3.7.2 Delft Facility-Based Implementation . . . 36
3.7.3 Delft Supplementary Local Checks . . . 37
3.8 Sensitivity and Robustness Design . . . 38
3.8.1 Robustness check for operational-efficiency wording . . . 38
3.8.2 Sensitivity to stakeholder-group importance weights . . . 39
3.8.3 Sensitivity to the short-term and long-term window boundary . . . 39
3.8.4 Robustness comparisons across BWM configurations . . . 39
3.8.5 Delft boundary sensitivity analysis . . . 40

4 Stakeholder Priorities and Cross-City Strategy Rankings 41
4.1 Survey response profile and analysis samples . . . 41
4.1.1 Response filtering and valid BWM samples . . . 41
4.1.2 Respondent distribution by stakeholder group . . . 42
4.1.3 Respondent background and distribution channels . . . 42
4.2 Stakeholder priorities across recovery stages . . . 43
4.2.1 Overall patterns across stakeholder groups . . . 43
4.2.2 Stage-specific differences among the main stakeholder groups . . . 45
4.2.3 Respondent-level priority changes . . . 46
4.3 Reasoning behind stakeholder priorities . . . 47
4.3.1 Themes in most-important criterion explanations . . . 49
4.3.2 Themes in least-important criterion explanations . . . 49
4.4 Cross-city strategy ranking results . . . 50
4.4.1 Description of the analysed city networks . . . 50
4.4.2 Strategy rankings under 10% damage . . . 50
4.4.3 Strategy rankings under 30% damage . . . 51
4.4.4 Overall ranking pattern . . . 52
4.4.5 Similarity between single-indicator strategy orderings . . . 52
4.5 Differences in strategy rankings across stakeholder groups and recovery stages . . . 53
4.5.1 Differences across stakeholder groups . . . 53
4.5.2 Differences between short- and long-term recovery stages . . . 54
4.5.3 Linking the rankings to criterion weights and strategy performance . . . 54
4.6 Robustness and sensitivity checks . . . 55
4.6.1 Flagged operational-efficiency responses . . . 55
4.6.2 Short-term recovery stage boundary . . . 56
4.6.3 BWM configuration . . . 56
4.6.4 Stakeholder weight aggregation . . . 57

5 Delft Case Study 58
5.1 Facility-based separation of accessibility criteria . . . 58
5.2 Delft criterion-level recovery results . . . 60
5.2.1 AAC deviations under 10% damage . . . 60
5.2.2 AAC deviations under 30% damage . . . 61
5.2.3 Comparison between critical and routine accessibility . . . 61
5.3 SAW ranking results . . . 62
5.3.1 Main strategy rankings for key stakeholder groups . . . 62
5.3.2 Ranking results across all stakeholder groups . . . 64
5.3.3 Comparison with the cross-city ranking pattern . . . 64
5.4 Spatial restoration patterns . . . 64

5.4.1 Static edge priority order across strategies . . . 64
5.4.2 Spatial interpretation of the priority patterns . . . 65
5.4.3 Implications of the static priority order . . . 66
5.5 Delft boundary sensitivity results . . . 66
5.5.1 Buffer-based network extent comparison . . . 67
5.5.2 Sensitivity of betweenness and closeness priority orders . . . 68
5.5.3 Sensitivity of the remaining strategies . . . 68
5.5.4 Implications for boundary choices . . . 69
5.6 Local critical-service travel-time results . . . 69
5.6.1 Test scope and facility boundary . . . 69
5.6.2 Service-group results under 30% damage . . . 70
5.6.3 Damage-level comparison . . . 71
5.6.4 Interpretation relative to the SAW ranking . . . 72

6 Discussion **73**
6.1 Stakeholder priorities across recovery-stage contexts . . . 73
6.2 Betweenness as the most frequently top-ranked system-level strategy . . . 74
6.3 System-level rankings and local accessibility interpretation . . . 75
6.4 Methodological reflection on the evaluation design . . . 76
6.5 Practical interpretation of the framework . . . 77

7 Conclusion and Future Work **78**
7.1 Answering the research questions . . . 78
7.1.1 Stakeholder priorities across recovery stages and groups . . . 79
7.1.2 Using stakeholder-derived weights to rank recovery strategies . . . 79
7.1.3 Cross-city ranking patterns across stakeholders, stages, and damage levels . . . 79
7.1.4 Additional insights from the Delft case study . . . 80
7.1.5 Answer to the main research question . . . 80
7.2 Scientific contributions . . . 81
7.2.1 A stage-specific BWM-SAW framework for comparing recovery strategies . . . 81
7.2.2 Stage-specific stakeholder-priority elicitation . . . 81
7.2.3 Applying a system-level ranking to a city requires local context . . . 82
7.3 Practical implications . . . 82
7.4 Limitations . . . 83
7.4.1 Scope of the evaluation design . . . 83
7.4.2 Simplifications in damage and recovery modelling . . . 83
7.4.3 Stakeholder, indicator, and MCDA-related limitations . . . 83
7.4.4 Limitations of the Delft case study . . . 84
7.5 Future research . . . 84
7.5.1 Broader stakeholder engagement and validation . . . 84
7.5.2 Hazard-specific and adaptive recovery modelling . . . 84
7.5.3 Composite centrality-based recovery strategies . . . 85
7.5.4 Extension of the MCDA framework . . . 85
7.6 Closing remarks . . . 85

References **86**

A HREC approval **91**

B Declaration on the use of AI tools **93**

C Methodological supplements **94**
C.1 Worked example of the performance metrics . . . 94
C.2 Strategy-ordering similarity check . . . 98

D Survey questionnaire **99**

E Processed simulation outputs **112**
E.1 Processed deviation-score summaries . . . 112

F Cross-city ranking results **114**
F.1 Rank-one strategy distributions across cities 114
F.2 Plurality rank-one shares with Wilson 95% confidence intervals 116

G Sensitivity and robustness analyses **118**
G.1 BWM configuration sensitivity 118
G.2 Recovery-stage boundary sensitivity 119
G.3 Stakeholder group-weight sensitivity 121
G.4 Delft network-boundary sensitivity 122

H Scientific Paper **124**

Glossary of Key Terms

Resilience Resilience refers to the capacity of a transport system to withstand disruptions and recover functionality afterward, understood as an evolving trajectory spanning disruption and recovery stages[1].

Disaster A disaster refers to a natural hazard event, such as an earthquake, flood, or landslide, that causes physical damage to road infrastructure and disrupts road network functionality[2].

Road Recovery Practice Road recovery practice refers to the practical methods, processes, and judgments adopted by practitioners in post-disaster road recovery decision-making contexts[3].

Road Recovery Decision Road recovery decisions refer to the judgments and choices made by responsible road authorities to assess damage, prioritise repair actions, and allocate resources for the reconstruction of damaged road infrastructure following a disaster event[3].

Recovery Strategies Recovery strategies refer to alternative approaches for restoring failed road segments that reflect different prioritisation principles, such as targeting highly connected nodes or restoring globally central links[1]. In this research, these strategies are predefined and adopted without modification.

Recovery Planning Recovery planning is defined as the process of determining the sequence in which damaged road links are restored under resource and time constraints[1].

Restoration Schedules Restoration schedules refer to the sequencing and timing plans that determine the order and allocation of resources for repairing damaged road infrastructure following a disaster.

Road Recovery Implementation Road recovery implementation refers to the actual execution of post-disaster road recovery plans by operational actors and responsible authorities to restore road network accessibility.

Recovery Process The recovery process refers to the dynamic, stage-by-stage restoration of a road network's functionality following a disaster, in which priorities, resources, and network conditions evolve over time[1][4].

Performance metric A performance metric expresses the value of a network measure in a given recovery state as a fraction of its pre-disruption value ($P = M(G(t))/M(G)$), so that it ranges over $[0, 1]$, with 1 denoting full pre-disruption performance. Each of the four simulation criteria (operational efficiency, accessibility, spatial equity, and connectivity) is expressed in this relative, dimensionless form so that they share a common scale.

Recovery trajectory The sequence of performance-metric values $P(t)$ traced by a criterion as the network is progressively restored, from the disrupted state toward full recovery.

Deviation score A cost-type summary of a recovery trajectory within a recovery stage, computed as the area between the trajectory and the line of full recovery, normalised by the width of the stage window (area-above-the-curve, AAC) and clipped to $[0, 1]$. Clipping bounds the small fraction of values that fall outside this range, which can arise when a metric temporarily exceeds its pre-disruption level. Lower values indicate faster and more complete recovery within that stage.

Normalised score A deviation score rescaled across the competing strategies within a given scenario using max–min cost normalisation, giving a benefit-type score in $[0, 1]$ where 1 is the best-performing strategy on that criterion.

Decision matrix The table that serves as the input to the SAW aggregation, with the six recovery strategies as rows and the four criteria as columns. Its entries are the normalised scores r_{ij}, and a separate decision matrix is formed for each city, damage level, and recovery stage.

Composite score The Simple Additive Weighting (SAW) sum of a strategy's normalised scores across the criteria, weighted by the stakeholder criterion weights ($S_j = \sum_i w_i r_{ij}$). Higher composite scores indicate more preferred strategies.

Relationship between key terms. The terms above are closely related but refer to different aspects of post-disaster road recovery in this thesis. *Road recovery practice* describes the broader practical context in which recovery takes place. Within this context, *road recovery decisions* concern the choices made by responsible actors when assessing damage, setting priorities, and allocating resources. These decisions inform *recovery planning*, which leads to *restoration schedules* that specify the sequence and timing of repair actions. *Recovery strategies* provide the prioritisation principles used to guide these schedules. The execution of these schedules is referred to as *road recovery implementation*, while the overall dynamic development of restoration over time is described as the *recovery process*.

Relationship between the evaluation measures. These measures form a single pipeline that converts simulated network behaviour into a strategy ranking. For each criterion, the raw network measure is first expressed relative to its pre-disruption value as a *performance metric* $P(x)$, where x denotes restoration progress. Tracking this metric over the restoration process forms a *recovery trajectory*. Each trajectory is then summarised within a recovery window by a *deviation score*: the average gap between the trajectory and the pre-disruption performance baseline, where this baseline corresponds to $P(x) = 1$. The resulting deviation-score values are clipped to $[0, 1]$ to handle occasional values outside this range. Values below 0 can occur when a metric briefly exceeds its pre-disruption level, while values above 1 can occur when the trajectory deviates more than the window range allows. This is a cost-type quantity, so lower values are better. The deviation scores are then converted to *normalised scores* through max–min cost normalisation across strategies, which reverses the direction so that higher values are better. For each city, damage level, and recovery stage, these normalised scores are arranged into a *decision matrix* with strategies as rows and criteria as columns. Finally, the stakeholder weights are applied to this matrix to give a *composite score* per strategy, which is used to rank the strategies. The single change of preference direction, from cost-type deviation scores to benefit-type normalised scores, occurs at the normalisation step.

Summary

After a disaster, disruptions to the road network can slow emergency response, cut access to essential services, and disrupt daily mobility, logistics, and the recovery of other infrastructure sectors. Because resources, time, equipment, and personnel are limited, damaged roads cannot all be restored at once, and decision-makers must decide which roads, or which type of recovery strategy, to prioritise. This prioritisation is difficult, because recovery involves several competing objectives, different stakeholder groups may hold different priorities, and these priorities may differ between the short-term and long-term recovery contexts.

This study develops a stage-specific MCDA framework for comparing post-disaster road recovery strategies. Road recovery has been studied through recovery sequencing, resilience indicators, stage-dependent needs, stakeholder involvement, and MCDA-based decision support, but these strands are rarely combined to compare recovery strategies. In particular, stakeholder priorities are seldom translated into criterion weights that rank predefined strategies, and short-term and long-term priorities are seldom weighted separately. The framework closes this gap by deriving criterion weights separately for each stakeholder group and recovery stage, and combining them with simulated strategy performance.

The framework has two streams that are developed separately and then combined. The stakeholder stream uses the Best-Worst Method (BWM) to convert stakeholder judgements into criterion weights. Respondents compared five resilience criteria, operational efficiency, access to critical facilities, access to routine facilities, spatial equity, and connectivity, separately for a short-term temporary-repair context and a long-term reconstruction context. Individual weights were aggregated within each stakeholder group, and the eight groups were then combined with equal weight to obtain overall weights.

The simulation stream evaluates how each strategy performs. Five predefined centrality-based strategies and one random baseline are compared under controlled $10\,\%$ and $30\,\%$ random link-failure scenarios. For every road network, a fixed share of road links is removed at random, and the damaged links are then restored in the order defined by each strategy. Network performance is recalculated after each restoration step, producing a recovery trajectory for every strategy and criterion. Each trajectory is summarised into an area-above-curve deviation score, measured separately for a short-term and a long-term recovery window, where lower values indicate smaller losses relative to the pre-disruption baseline. Simple Additive Weighting (SAW) then combines the BWM criterion weights with the normalised deviation scores, producing a composite score and a strategy ranking for each road network, damage level, recovery-stage context, and stakeholder profile. The framework is demonstrated through two applications: a cross-city, system-level analysis across many road networks, and a city-level application in Delft that represents accessibility through facility-based measures.

The survey collected 115 responses, of which 61 were complete. After stage-specific consistency screening, 56 valid short-term responses and 56 valid long-term responses were retained, and 51 respondents provided valid weights in both contexts. The BWM results show a clear stage-dependent pattern. In the short-term context, the three main stakeholder groups all place the highest weight on access to critical facilities, at 0.442 for road users, 0.501 for the local community, and 0.374 for technical experts. This points to a shared emphasis on critical access during temporary repair. In the long-term context, the weights flatten and the highest-weighted criterion diverges across groups: access to critical facilities for road users (0.279), connectivity for the local community (0.227), and spatial equity for technical experts (0.244). These long-term leads are small, since the weights are much more evenly spread across the five criteria. Short-term priorities therefore concentrate on urgent access, while long-term priorities are more evenly distributed and differ across groups, with no single criterion clearly dominating.

A respondent-level analysis reinforces this pattern, using the 51 respondents with valid weights in both contexts. Long-term top priorities are not simply the short-term second priority moved up: only $17.2\,\%$ of long-term first-ranked criteria were ranked second in the short-term context, while $67.2\,\%$ were ranked

third, fourth, or fifth. The long-term criterion rankings were therefore generally not simple continuations of the short-term rankings.

In the cross-city, system-level analysis, betweenness is the most frequent first-ranked strategy for almost all stakeholder groups and both recovery contexts under the $10\,\%$ damage level. Closeness becomes most competitive under $30\,\%$ damage, particularly in some long-term contexts. For technical experts in this condition, betweenness still leads at $48.4\,\%$ of cities but closeness rises to $43.0\,\%$. For the emergency-responder group, which has only two valid long-term responses and is reported as exploratory, closeness has the largest first-ranked share. Betweenness is therefore the most frequently first-ranked system-level strategy, while closeness becomes a conditional alternative in some higher-damage, long-term contexts.

The influence of stakeholder-group differences on the rankings is moderate. A within-network comparison shows that all eight stakeholder profiles produce the same first-ranked strategy in about $50\,\%$ to $65\,\%$ of cases. Stakeholder weights therefore do not always change the first-ranked strategy, but they do change how far it leads.

The Delft case study applies the same framework at the city level. It separates accessibility into access to critical facilities and access to routine facilities, using 58 mapped critical-facility nodes and 364 routine-facility nodes, whereas the cross-city analysis uses a single aggregate measure. Under this more detailed representation, betweenness ranks first across the three main stakeholder groups, both damage levels, and both recovery contexts, while closeness consistently ranks second.

The supplementary local checks in Delft show that the system-level ranking does not by itself describe the selected origin-to-service outcomes examined in the local analysis. In a travel-time test from selected residential origins to the nearest critical services, closeness produces shorter median travel times than betweenness for all four service categories, emergency, medical, food, and water, under $30\,\%$ damage at $30\,\%$ restoration progress. These checks provide a road-user perspective rather than a second five-criterion ranking. The first-ranked aggregate strategy may therefore not provide the shortest travel times for every selected local trip.

The main contribution of this study is a stage-specific MCDA framework that brings stakeholder priorities, recovery-stage contexts, and simulation-based recovery performance into a single procedure for comparing recovery strategies, using BWM for stage- and group-specific weights and SAW to combine them with strategy performance. As an early-stage decision-support layer, it lets decision-makers compare different recovery-priority logics and see the trade-offs between competing objectives. More broadly, the findings support evaluating short-term and long-term recovery contexts separately and considering multiple stakeholder perspectives, because the relative importance of the criteria can differ across contexts and groups.

1

Introduction

1.1. Background and problem context

Disasters can have lasting and cumulative impacts on infrastructure systems. According to UNDRR and CRED, 7,348 major disaster events were recorded worldwide between 2000 and 2019, causing approximately 1.23 million deaths, affecting 4.2 billion people, and generating about USD 2.97 trillion in economic losses [5]. These impacts are shaped not only by the direct damage caused by a disaster, but also by how quickly infrastructure systems can recover afterwards. Transport systems are central to this process, because people, goods, repair crews, and emergency responders all depend on them to move. As a result, how quickly transport is restored shapes the pace of the wider recovery [6].

Road networks are particularly important within this broader transport context. They connect residents to work, schools, health care, food, and shelter, and they also support logistics, maintenance, and emergency response [7], [8], [9]. At the same time, road infrastructure is spread widely across space and is therefore exposed to natural hazards. A global multi-hazard risk assessment by Koks et al. estimates that around 27% of global road and railway infrastructure is exposed to at least one natural hazard [10]. When roads are damaged or blocked, the effects reach past the road network itself. The recovery of other infrastructure sectors, including water and energy supply, can also slow down [6].

Restoring road networks repairs damaged infrastructure and re-establishes access between affected communities and the services they depend on. The longer key road links stay out of service, the more the costs build up through delayed trips and lost business activity [6], [11], [12]. Road recovery that is both fast and well planned can therefore limit the wider social and economic impacts of a disaster. However, recognising the importance of road recovery does not by itself determine where recovery should begin.

1.2. Problem statement

A central challenge in post-disaster road recovery is deciding how repairs should be prioritised when it is not possible to restore all damaged road links at once. This is especially important after disasters, when several road links may be damaged or blocked while repair crews, equipment, budgets, and reliable damage information are limited. Transport recovery guidance therefore stresses the need to prioritise locations, damaged assets, and repair or reconstruction activities under limited resources [6].

For example, after a flood, a road authority may need to choose between reopening a major arterial road and repairing a smaller local access road. Reopening the arterial may improve network-wide traffic and freight movement, while repairing the local road may reconnect an isolated community to a hospital or emergency shelter. Emergency responders may favour the local road in the short term, whereas infrastructure managers may prefer the arterial because of its role in the wider network. This example reflects a broader question in recovery planning, which is what should guide the order in which damaged road links are restored. Which priority rule appears more suitable depends on the criterion being emphasised, the recovery stage being considered, and the stakeholder perspective

being represented.

Prioritisation is difficult because road recovery can pursue different objectives. One priority rule may be preferred because it restores access to essential services quickly, while another may be preferred because it distributes recovery benefits more evenly across the network. These objectives do not always lead to the same restoration order, so road recovery cannot be evaluated using a single performance measure [1]. Recovery planning therefore involves trade-offs among several resilience criteria.

The prioritisation problem is also stage-dependent. In the short term, decisions focus on emergency response, temporary access, evacuation, and the delivery of essential services. In the longer term, priorities may shift towards durable reconstruction, routine accessibility, network efficiency, spatial equity, and the recovery of normal travel patterns [13], [14]. The same priority rule may therefore be more suitable in one recovery stage than in another.

Recovery priorities are also shaped by the stakeholders involved in, or affected by, the recovery process. Post-disaster recovery usually involves several stakeholder groups with different responsibilities, resources, and priorities, and coordinating actors with different objectives is a recurring practical challenge [15]. These differences matter because the same recovery objective may not carry equal importance for all groups. Emergency responders may place greater emphasis on access to critical facilities, while infrastructure managers may focus more on the overall functioning of the road network.

Overall, post-disaster road recovery involves trade-offs among resilience criteria, recovery stages, and stakeholder perspectives. In practice, these trade-offs are often handled through experience rather than through formal decision-support tools [3]. In research, the importance of recovery criteria is often predefined by researchers rather than elicited from stakeholders, and priorities are rarely elicited separately for different recovery stages. As a result, existing evaluations rarely allow recovery strategies to be compared transparently for a specific stakeholder group and recovery stage. Chapter 2 examines this gap in more detail, and this thesis develops a framework to address it.

1.3. Research Aim and Questions

Building on this problem context, the aim of this thesis is to develop a stage-specific MCDA framework for comparing post-disaster road recovery strategies. The Best-Worst Method is used to derive stakeholder- and stage-specific criterion weights, and Simple Additive Weighting is used to combine these weights with simulated strategy performance to produce recovery strategy rankings.

The main research question is:

> *How can post-disaster road recovery strategies be prioritised through a stage-specific MCDA framework, given that resilience is multidimensional, stakeholders have different priorities, and recovery decisions evolve over time?*

This main question is addressed through four sub-questions:

1. How do stakeholder priorities for road-recovery resilience criteria differ across stakeholder groups and recovery stages?
2. How can stakeholder-derived and stage-specific criterion weights be used to rank predefined road recovery strategies based on simulated recovery performance?
3. How do the resulting strategy rankings vary across road networks, stakeholder groups, recovery stages, and disruption levels in the cross-city, system-level analysis?
4. How do recovery strategies compare in a detailed city-level application, using Delft as a case study?

Together, these sub-questions structure the framework developed in this thesis. The first establishes the stakeholder-priority input by examining how criteria priorities differ across groups and recovery stages. The second builds the core of the framework by combining these priorities with simulated performance through BWM and SAW. The third applies the framework as a cross-city, system-level analysis, and the fourth compares recovery strategies in a detailed city-level application, using Delft as a case study.

1.4. Research approach

This study follows a sequence that links stakeholder-priority elicitation to simulation-based strategy evaluation. The literature review defines the decision context and supports the methodological choices used in this setup, including the use of the Best-Worst Method for eliciting criterion weights, the distinction between stakeholder groups, and the definition of short-term and long-term recovery stages.

First, a stakeholder survey and the Best-Worst Method are used to derive criterion weights for different stakeholder groups and recovery stages. Second, simulation-based indicators are used to measure how six predefined road recovery strategies perform on the resilience criteria. The weights are then combined with the simulated performance indicators through Simple Additive Weighting to produce stakeholder- and stage-specific strategy rankings. Sensitivity and robustness analyses examine whether the resulting weights and rankings depend on key methodological assumptions.

The framework is demonstrated through two applications at different spatial granularities. The cross-city application examines how strategy rankings vary across road networks, disruption levels, stakeholder groups, and recovery stages. The Delft application applies the same evaluation logic within one urban network, where accessibility is represented through explicit facility destinations and supplemented by local accessibility checks. Table 1.1 summarises the method and the output for each sub-question.

Table 1.1: Method and output for each sub-question.

Sub-question	Method	Output
SQ1. How do stakeholder priorities for road-recovery resilience criteria differ across stakeholder groups and recovery stages?	Stakeholder survey and BWM	Stage- and group-specific criterion weights and their differences
SQ2. How can stakeholder-derived and stage-specific criterion weights be used to rank predefined road recovery strategies based on simulated recovery performance?	Simulation-based performance evaluation and SAW	A BWM-SAW procedure for ranking recovery strategies, forming the core of the framework
SQ3. How do the resulting strategy rankings vary across road networks, stakeholder groups, recovery stages, and disruption levels in the cross-city, system-level analysis?	Simulation and SAW (cross-city)	Cross-city, system-level ranking patterns
SQ4. How do recovery strategies compare in a detailed city-level application, using Delft as a case study?	Simulation, SAW, and supplementary local checks (Delft)	City-level strategy comparison, including local accessibility differences

1.5. Research scope

Post-disaster road recovery can be studied at different levels of detail. Some studies represent it as a scheduling or optimisation problem, sequencing damaged links or recovery projects under practical constraints such as budget, crew capacity, and traffic assignment effects [16], [17]. This thesis works at the level of strategy comparison.

The thesis does not model the governance process of post-disaster reconstruction, and it does not produce an implementable repair schedule with costs, crews, and construction durations. Recovery strategies are compared using simulated network performance together with stakeholder- and stage-specific criterion weights. Stakeholders provide importance weights for the evaluation criteria, and they do not evaluate or select the recovery strategies directly.

Six predefined road recovery strategies are used as controlled inputs to demonstrate the framework. Five of them are centrality-based and one is a random baseline. They are adopted from the recovery strategy set of Balakrishnan and Stokkink [1], and the reasons for using structure-based rules are set out in Section 3.2.2. The framework itself does not depend on this particular set. Another set of candidate

strategies could be compared in the same way, as long as their performance can be measured on the same criteria. These strategies are fixed restoration rules for damaged road links, and the restoration order does not change during a recovery simulation.

Short-term and long-term recovery are treated as two evaluation contexts. The study examines how criterion weights and strategy rankings differ between them, and it does not model a decision process that updates continuously over calendar time.

The framework is demonstrated through two applications at different spatial granularities:

- **Cross-city application:** multiple city networks are compared across disruption levels, stakeholder groups, and recovery stages to identify broad ranking patterns.
- **Delft application:** a single urban network is examined in greater detail, with facility-based accessibility measures used to examine more local recovery patterns.

The Delft case study complements the cross-city analysis with local detail. It is not a basis for drawing general conclusions from a single city.

Within this scope, the simulated 10% and 30% disruption levels provide controlled conditions for comparing recovery strategies. They represent specified network disturbance settings rather than specific hazard events. The cross-city findings should be interpreted within the city networks analysed here, and under the disruption conditions and modelling assumptions of this study.

1.6. Scientific and practical relevance

Scientific relevance. The scientific contribution of this study is a stage-specific MCDA framework for comparing predefined post-disaster road recovery strategies. Existing studies on road recovery and transport resilience have often treated stakeholder priorities, recovery stages, and strategy comparison separately. This framework uses BWM to derive criterion weights by stakeholder group and recovery stage, and uses SAW to combine these weights with simulation-based strategy performance. Stakeholder priorities and recovery stages therefore become direct inputs to the quantitative evaluation, allowing recovery strategies to be compared from the perspective of a given stakeholder group at a given stage of recovery.

Practical relevance. In practical terms, this study supports the early comparison of post-disaster road recovery strategies. It allows stakeholder priorities to be elicited separately for short-term and long-term recovery, so that strategy evaluation does not rely on one fixed set of criterion weights across the whole recovery process. The resulting rankings can help decision-makers screen and compare different recovery-priority logics before more detailed planning is undertaken. The framework should be used as a decision-support layer, and its results should be considered together with local accessibility information and the actual damage conditions of the city concerned.

1.7. Thesis outline

The remainder of this thesis is organised as follows.

Chapter 2 reviews the literature on post-disaster road recovery, transport resilience, recovery stages, stakeholder involvement, and MCDA methods, and identifies the gap that this study addresses.

Chapter 3 presents the stage-specific MCDA framework, including the definition of the decision problem, the stakeholder survey and BWM weight elicitation, the simulation-based evaluation of recovery performance, the SAW aggregation, and the implementation of the framework at the cross-city and Delft scales.

Chapter 4 presents the stakeholder-priority results and the cross-city, system-level strategy ranking results, covering differences across stakeholder groups, recovery stages, and damage levels.

Chapter 5 presents the Delft city-level case study, which compares the recovery strategies in a single urban network using a more detailed facility-based representation of accessibility and a set of supplementary local checks.

Chapter 6 discusses the meaning of the main findings, including the stage-dependent stakeholder priorities, the strategy ranking patterns, and the relationship between system-level and local accessibility.

Chapter 7 answers the research questions, summarises the scientific and practical contributions of the study, discusses the limitations, and outlines directions for future research.

2

Literature Review

2.1. Road recovery as a sequencing and prioritisation problem

Studies on post-disaster road and transportation-network recovery can be understood through two related modelling logics. Bhatia et al. [18] distinguish between optimisation-based approaches, which derive recovery sequences by optimising a specified objective function, and centrality-based recovery approaches, which use network-topological measures to prioritise components. Positioned against this distinction, the recovery strategies evaluated in this thesis are predefined centrality-based restoration rules.

2.1.1. Optimisation-based recovery scheduling

Optimisation-based recovery studies formulate post-disaster road recovery as a scheduling, sequencing, or routing problem. In these models, the recovery order is derived by optimising a specified objective function under constraints such as repair resources, crew availability, damaged-link conditions, or travel demand. The review by Almeida et al. [14] shows that this literature focuses mainly on models for road repair and restoration activities, including the scheduling of restoration tasks and the routing of repair crews.

Representative studies differ in their specific objectives and modelling assumptions. Zhang et al. [4] model road-bridge network recovery by optimising restoration sequences with respect to resilience-related performance indicators such as total recovery time and recovery rapidity. Kaviani et al. [19] develop a multi-objective scheduling model for regional road networks, incorporating passenger and freight traffic through indicators such as travel time and demand satisfaction. These studies show how optimisation models can translate network disruption into a recovery schedule that is evaluated through time-dependent network performance.

Other studies extend this optimisation logic by adding operational or contextual detail. Some models focus on specific disruption contexts, such as snow-related road disruption [20], while others integrate road recovery with emergency activities such as relief distribution [21]. Uncertainty in recovery times or system conditions has also been incorporated into road recovery models [22].

2.1.2. Topology- and centrality-based recovery strategies

A second modelling logic uses network structure to define recovery priorities. Instead of deriving a repair schedule through an optimisation model, topology- and centrality-based approaches rank damaged components according to their structural importance in the network. Bhatia et al. [18] describe this as a centrality-based recovery logic, in which measures such as degree, closeness, betweenness, and eigenvector centrality are used to generate prioritisation sequences. They also note that centrality-based approaches are intuitive and computationally efficient, but relatively static compared with optimisation-based approaches that can be adjusted to specific objectives or damage patterns.

This centrality-based logic appears in road-network recovery studies as a way to generate repair se-

quences from graph importance. Fan et al. [23] use a betweenness-centrality-based repair sequence as a baseline method for post-disaster road network recovery. More directly, Balakrishnan and Stokkink [1] define centrality-based recovery strategies by assigning priority scores to disrupted road links and restoring them according to the resulting order.

The present thesis is positioned within this topology- and centrality-based branch. It evaluates a fixed set of predefined recovery strategies instead of generating a new optimal repair schedule for each disruption scenario. Its contribution lies in combining these simulated strategy performances with stakeholder-derived criterion weights, so that the resulting rankings can be compared across stakeholder groups, recovery stages, city networks, and disruption levels.

2.2. Resilience indicators for evaluating road recovery

Resilience indicators provide a way to quantify how transport networks are affected by disruption and how they recover afterwards. In transport-infrastructure resilience literature, resilience metrics are used to assess the capacity of transport networks to withstand and absorb damage, recover after disruption, and adapt to future changes [24]. Such assessment requires a measure of infrastructure performance and its recovery over a defined time horizon, because disruption is commonly represented as a loss of functionality followed by gradual restoration [24]. In road-network recovery studies, this performance can be represented through indicators such as recovery time and recovery rapidity [4], travel-time-based performance [25], accessibility and unmet demand [26], and resilience indices derived from performance recovery curves [23].

These indicators play different roles depending on how the recovery problem is formulated. In optimisation-based studies, performance or resilience measures often enter the model as objective functions against which recovery sequences or restoration plans are optimised. Gokalp et al. [25], for example, identify repair sequences by minimising total system travel time over the repair horizon. Zhao and Zhang [26] formulate restoration planning using total travel time and unmet demand as resilience measures in a bi-objective optimisation framework. In the present thesis, the recovery strategies are predefined rather than generated through optimisation. The role of resilience indicators is therefore to evaluate and rank simulated strategy performance after the recovery sequences have been produced.

A key issue in resilience assessment is that different indicators make different aspects of recovery visible. A single-dimensional measure, such as efficiency or connectivity, describes only one part of network performance. The literature responds to this limitation with a wide range of indicator types. Serdar et al. [27] show that transportation resilience studies span a wide range of indicator types, including performance-based, connectivity, cost-based, statistical, and qualitative indicators. This variety points to something broader than measurement choice alone. Road-infrastructure resilience itself involves multiple dimensions, as Nickdoost et al. [28] show by representing it through technical, socioeconomic, and geospatial factors. The choice of indicator therefore shapes which aspect of resilience is assessed and which recovery outcomes are emphasised. For recovery strategy evaluation, this means that the same strategy may appear more or less effective depending on the indicator used. For example, a strategy assessed mainly through an aggregate efficiency indicator may be ranked highly because it restores overall network performance quickly, while still leaving some areas with poor access, unevenly distributed benefits, or weaker network connectedness.

Balakrishnan and Stokkink [1] demonstrate this problem by evaluating post-disaster road recovery strategies across operational efficiency, accessibility, spatial equity, and connectivity. They show that no single strategy performs best across all dimensions. They note, for example, that strategies aimed at improving efficiency may fail to guarantee equitable access, while strategies that improve connectedness may compromise overall system efficiency. The same strategy can therefore lead to different performance interpretations depending on the resilience dimension considered.

For the present thesis, these findings provide the basis for evaluating predefined recovery strategies through multiple resilience criteria rather than through a single aggregate performance measure. This multi-dimensional perspective allows the analysis to examine how strategy rankings change when different criteria are emphasised, and prepares the ground for incorporating stakeholder-derived criterion weights in the later evaluation framework.

2.3. Stage-dependent recovery dynamics

Beyond multi-dimensionality, post-disaster road recovery is widely recognised as a staged and time-dependent decision process, in which recovery objectives, constraints, and performance concerns may change over time [14]. This temporal characteristic has been treated in different ways across the recovery modelling literature, with studies differing in how they define recovery stages and how they connect these stages to evaluation criteria.

One common approach is to distinguish recovery stages by time scale, decision context, and intervention type. Kaviani et al. [16] describe post-event recovery as consisting of three sequential periods, namely a response recovery period lasting one to seven days (involving road closures, debris removal, and traffic diversions), a short-term recovery period of weeks to months, and a long-term recovery period focused on long-lasting reconstruction. Rey and Bar-Gera also distinguish short-term, mid-term, and long-term challenges, but apply a substantially longer time scale to long-term recovery, defining it as reconstruction scheduling over a planning horizon of around 12 to 24 months [17]. Other frameworks further introduce a pre-disaster mitigation stage to capture preparedness and robustness considerations [13]. These differences show that even when studies use similar terminology such as “long-term recovery,” the underlying time scales and intervention types are not consistently defined.

Studies also differ in how stages are linked to evaluation criteria. Some studies explicitly change the evaluation approach between stages. Zhang et al., for instance, assign different performance metrics to different stages, using robustness-related metrics for pre-disaster mitigation, emergency accessibility metrics for emergency response, and traffic-flow efficiency metrics for long-term recovery [13]. Other models acknowledge multiple recovery periods conceptually but apply a common set of indicators within a single modelled stage. For example, Kaviani et al. distinguish three post-event recovery periods in their conceptual background, but their optimisation model focuses on long-term recovery scheduling and evaluates network performance using two indicators, travel time resilience and demand satisfaction resilience, applied across vehicle classes rather than across stages [16]. Still others restrict the modelling scope to a single stage, most often long-term reconstruction, and use a single network-level objective such as total network travel time [17].

Despite these differences, the reviewed studies share a common limitation in how recovery priorities are determined across stages. Whether the same indicators are applied throughout the recovery process or different indicators are assigned to different stages, the choice of indicators and their relative importance is usually decided in advance by the researchers themselves, rather than elicited from stakeholders involved in or affected by the recovery process. As a result, when stage-dependent priorities are considered, they remain researcher-defined rather than stakeholder-derived.

2.4. Stakeholder involvement in recovery decision-making

The previous sections show that road recovery can be evaluated through multiple resilience criteria and that the relevance of these criteria may change across recovery stages. In practice, however, these criteria are not weighted in isolation. Their relative importance depends on the stakeholders involved in recovery decisions, the responsibilities they hold, and the groups affected by disruption.

Post-disaster recovery involves multiple stakeholder groups whose roles, priorities, and influence are not equal. Puri et al. [15] show that stakeholder-related challenges in post-disaster recovery commonly arise through organisational management problems and limited community engagement. In organisational terms, recovery may be weakened by poor coordination among actors with different goals, leading to duplicated work and inefficient resource use. At the community level, affected groups may have limited opportunities to participate in recovery decisions, and the perspectives of marginalised groups may be overlooked.

These participation and coordination issues are further complicated by the fragmented nature of post-disaster governance. Recovery activities are often distributed across multiple institutions operating under separate mandates, legal frameworks, and responsibilities [29]. In road recovery, this fragmentation can be seen in the aftermath of the Great East Japan Earthquake, where roads were restored through the uncoordinated efforts of national, prefectural, municipal, and private actors, resulting in disconnected road structures [30].

Decision-making practice also affects how recovery considerations are incorporated. Interviews with Australian practitioners involved in post-disaster road reconstruction show that social, economic, and environmental considerations are recognised in recovery decisions. However, these considerations are often incorporated through tacit knowledge, professional experience, and local knowledge, with limited use of systematic prioritisation methods [3]. The same study identifies a need for a framework that can validate current decision-making processes and support project prioritisation, while noting that such a framework should account for local conditions and input from multiple stakeholders.

2.5. MCDA-based decision-making approaches

The previous section shows that post-disaster road recovery decisions involve multiple stakeholders, competing priorities, and trade-offs that are not always represented through systematic evaluation methods. A decision-support approach is therefore needed that can connect recovery criteria, stakeholder priorities, and the performance of recovery alternatives. MCDA provides such a structure because it evaluates alternatives against multiple criteria and incorporates preference information into the ranking or selection process [31].

2.5.1. Road recovery as a multi-criteria decision problem

Road recovery can be interpreted as a multi-criteria decision problem because recovery strategies may perform differently depending on which objective is emphasised. Transport-sector recovery guidance identifies multiple recovery principles, including efficiency, equity and inclusiveness, access and mobility, and risk mitigation [6]. These objectives are not always aligned, since efficiency, equity, access, and risk mitigation can pull recovery priorities in different directions.

This multi-criteria nature is reinforced by resource constraints. Road network reconstruction requires limited budgets, work crews, equipment, and implementation capacity, so recovery projects or damaged links must be prioritised rather than restored simultaneously [17]. Together with the stakeholder considerations discussed in Section 2.4, this means that recovery evaluation requires clear criteria, stakeholder-based preference information, and a transparent way to combine criterion weights with strategy performance.

2.5.2. MCDA applications and limitations in road and transport resilience

MCDA offers a natural fit for these requirements, and has already been applied across several related post-disaster and infrastructure contexts. In a broader reconstruction setting, Mohammadnazari et al. [32] combine BWM-derived weights with multiple ranking methods to prioritise candidate reconstruction projects. This demonstrates the relevance of MCDA for post-disaster prioritisation, although the focus is on general construction projects rather than road recovery strategies.

Within road and transport resilience research, MCDA-related studies have more often been used to assess resilience conditions or prioritise resilience indicators. Morshed et al. [33] use AHP to validate an 8R resilience model for transportation infrastructure. In the context of flood resilience, Xu et al. [34] combine Delphi consultation, ISM, and ANP to identify and prioritise indicators influencing road network resilience to flooding. AHP-MCE is used by Mukesh et al. [35] to construct a Road Network Resilience Index. These studies show how expert or stakeholder judgement can be translated into criterion weights, but their evaluation targets are resilience indicators or infrastructure conditions rather than alternative road recovery strategies.

A more directly relevant contribution is provided by Zamanifar and Hartmann [36], who examine transportation network disaster recovery planning. They argue that the reliability of disaster recovery planning models depends on well-defined decision attributes, yet insufficient attention has been paid to establishing tenable attribute sets for transportation network recovery. Their study responds by developing a structured MCDA framework to identify a defensible set of decision attributes for transportation network recovery.

Taken together, these applications show that MCDA can support structured evaluation in disaster recovery and transport resilience contexts, but they leave several issues open for the present thesis. First, preference information is often elicited from experts or decision-makers and aggregated into a single set of weights, which limits the ability to compare priorities across stakeholder groups. Sec-

ond, weights are usually derived once and applied uniformly, even though recovery priorities may differ between short-term and long-term recovery stages. Third, many applications evaluate resilience conditions, indicators, or projects, while the ranking of predefined road recovery strategies remains less directly addressed. Addressing these issues requires a weighting method that can capture stakeholder- and stage-specific priorities, an aggregation method that links these weights to strategy performance, and performance data for the predefined recovery strategies.

2.5.3. Weighting, aggregation, and simulation-based performance data

This section addresses these requirements in turn, starting with the derivation of criterion weights. A central step in MCDA is deriving criterion weights that represent the relative importance of evaluation criteria. AHP and ANP are commonly used in road and transport resilience assessment [33], [34], [35]. These methods provide structured ways to translate judgement into weights, but standard AHP requires comparisons across all pairs of criteria, while ANP introduces additional complexity by allowing dependence and feedback among criteria [34], [37]. BWM offers a more manageable alternative because respondents compare all criteria with the most important and least important criteria. Rezaei [38] shows that BWM requires fewer comparison data and can lead to more consistent comparisons than AHP.

Deriving weights is not sufficient on its own, since these weights must also be combined with alternative performance. Aggregation methods differ in their decision logic. ELECTRE and PROMETHEE are outranking methods, while SAW and TOPSIS are value-function-based methods that generate scores for alternatives [39], [40]. SAW calculates the value of an alternative as the weighted sum of its evaluation ratings [39]. This provides a direct link between criterion weights, alternative performance, and final scores. SAW is also compensatory, meaning that weaker performance on one criterion can be offset by stronger performance on another [40]. For this reason, SAW rankings should be interpreted as rankings of overall weighted performance, not as evidence that a highly ranked strategy performs well on every individual criterion.

The final component is the performance data used to describe each recovery strategy. The resulting weights and aggregation method still require criterion values that indicate how each strategy performs across the selected recovery criteria. MCDA can incorporate quantitative, qualitative, model-based, or simulation-based indicators when evaluating alternatives [41]. Simulation outputs have been used as criteria values in decision-support contexts. Scholten et al. [42] combine modelled outcomes of water network rehabilitation strategies with preference-based weights to rank strategies under different future scenarios. Bellsolà Olba et al. [43] use simulation outputs as criteria values within a SAW-based evaluation of vessel traffic scenarios. These studies support the use of simulation-based performance values as MCDA input when alternatives need to be compared across multiple criteria.

2.6. Conclusion of the literature review and research gap

Overall, the reviewed literature shows that post-disaster road recovery has been studied from several complementary perspectives. In recovery modelling, damaged road links or recovery projects are sequenced through optimisation-based or topology-based approaches. Transport resilience research broadens the evaluation problem by measuring recovery performance through multiple indicators, including efficiency, accessibility, equity, and connectivity. Across stage-dependent recovery studies, recovery objectives and evaluation criteria may change over time. The stakeholder and recovery-practice literature further shows that prioritisation decisions are shaped by institutional responsibilities, coordination problems, local knowledge, and the priorities of different stakeholder groups. MCDA-based approaches provide a structured way to connect evaluation criteria, preference information, and alternative performance.

At the same time, these perspectives remain only partly connected. Existing road recovery and resilience evaluation methods do not consistently incorporate stakeholder priorities directly when comparing recovery strategies across multiple criteria. Stage-dependent studies recognise that recovery priorities may change over time, although the relative importance of criteria is often defined by researchers rather than elicited from stakeholders. MCDA studies offer tools for weighting and ranking, but existing applications in road and transport resilience have more often assessed resilience conditions, indicators, or projects.

This thesis addresses this gap by developing a stage-specific MCDA-based evaluation framework for predefined post-disaster road recovery strategies. The framework consists of two main steps. First, stakeholder judgements are elicited separately for the short-term and long-term recovery stages and converted into criterion weights using the Best-Worst Method. Second, Simple Additive Weighting is used to combine these weights with simulation-based performance scores and rank the predefined recovery strategies across multiple resilience criteria. The next chapter develops this framework in full and explains how the stakeholder weights and simulation-based performance scores are produced and combined.

3

Methodology

This chapter presents the stage-specific MCDA framework used to compare post-disaster road recovery strategies, and explains how it is implemented in this study. It first gives an overview of the framework and distinguishes the framework itself from the inputs and cases used to demonstrate it. It then develops each component in turn, from the decision problem and the stakeholder survey to the Best-Worst Method, the simulation-based performance evaluation, and the SAW combination step, before setting out the two applications and the sensitivity and robustness analysis.

3.1. Overview of the stage-specific MCDA framework

This study uses a stage-specific MCDA framework to rank six post-disaster road recovery strategies separately for the short-term and long-term recovery stages. The framework combines stakeholder priorities with simulated recovery performance through Simple Additive Weighting (SAW). It is designed to show how recovery strategy rankings change across different resilience criteria, stakeholder groups, and recovery stages.

The framework consists of two parallel streams that are developed separately and then combined in the SAW stage, as shown in Figure 3.1. The stakeholder stream derives criterion weights from stakeholder judgements using the Best-Worst Method (BWM). Stakeholders compare the relative importance of five recovery criteria separately for the short-term and long-term recovery stages, using the survey described in Section 3.4. Their responses are screened for consistency and aggregated into stakeholder-group and overall weights, following the procedure set out in Section 3.3. The output of this stream is a set of stage-specific criterion weights.

The simulation stream derives performance scores from road-network recovery simulations, and is described in Section 3.5. For each network and recovery strategy, edge importance scores are first calculated according to the strategy-specific priority rule. A disruption is then imposed by removing a subset of road links. Recovery then proceeds by restoring the damaged links according to the edge-importance order defined by each strategy. After each restoration step, network performance is recalculated, producing recovery trajectories for each strategy and criterion. These trajectories are summarised into area-above-curve (AAC) deviation scores for the short-term and long-term recovery windows. The output of this stream is a set of stage-specific performance scores for each strategy and criterion.

The two streams are combined through SAW to produce composite scores and stage-specific recovery-strategy rankings, as set out in Section 3.6. The main outputs of the framework are therefore the stage-specific criterion weights, the stage-specific performance scores, the SAW composite scores, and the recovery-strategy rankings.

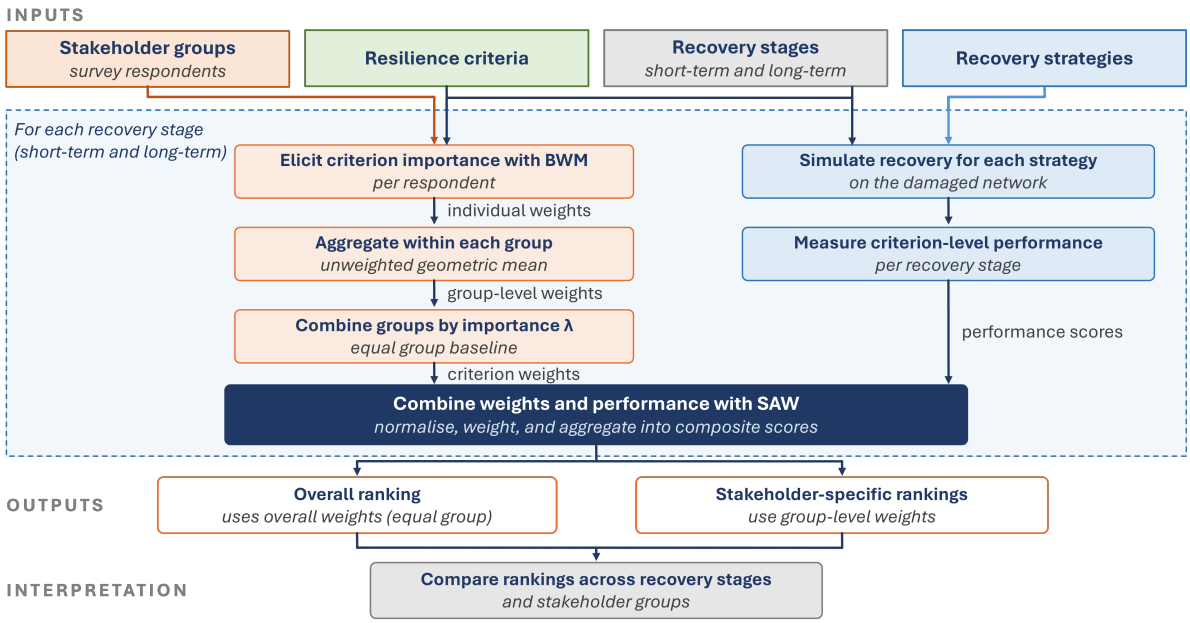


Figure 3.1: The stage-specific MCDA framework.

A distinction runs through the rest of this study between the framework itself and the inputs used to demonstrate it. The framework specifies how criterion weights, recovery stages, candidate strategies, and criterion-level performance are combined and compared to produce recovery-strategy rankings, and it does not depend on a particular set of candidate strategies or a particular spatial granularity. In this study, it is demonstrated using six predefined centrality-based strategies and random 10% and 30% disruption scenarios, which serve as controlled inputs that allow the strategies to be evaluated under consistent and comparable conditions. The same evaluation logic is applied at two spatial granularities, a cross-city system-level analysis that uses aggregate graph-based measures to identify broad ranking patterns across many networks, and a Delft city-level case study that uses more local data and adds a set of supplementary local checks. These strategies and applications are demonstration choices rather than fixed components of the framework, and the same framework could be applied to other strategies, other cities, or other spatial granularities. The implementation details of the two applications are described in Section 3.7.

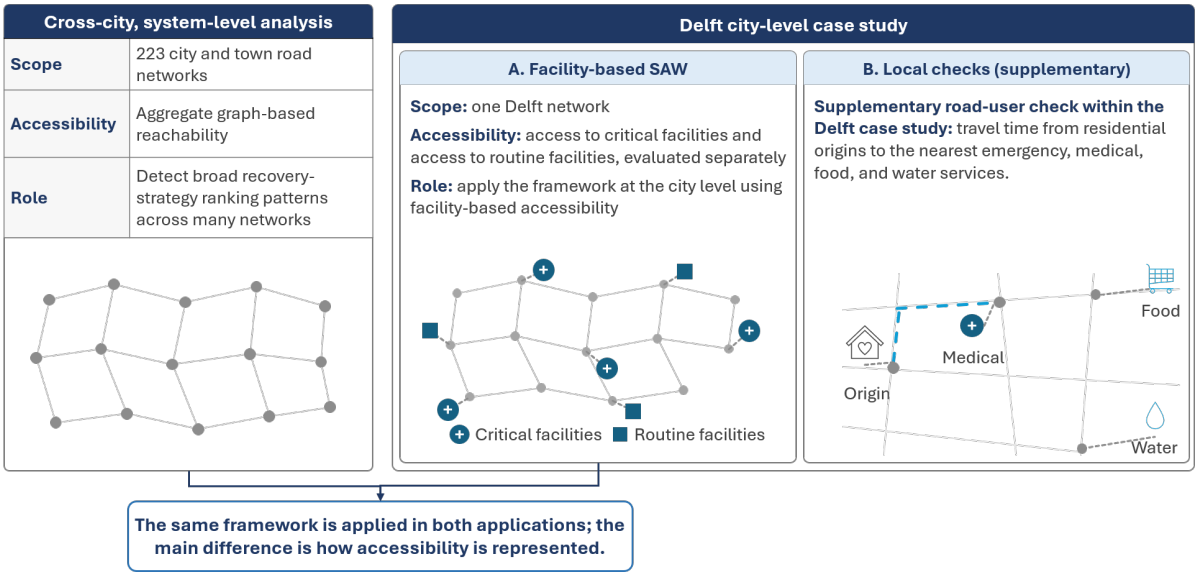


Figure 3.2: The two applications of the framework, differing in how accessibility is represented: aggregated across cities, and facility-based in Delft.

3.2. MCDA Decision Problem Formulation

3.2.1. Decision Objective and Analytical Contexts

This section defines the components of the recovery strategy prioritisation problem used in the MCDA framework: the recovery stages, candidate recovery strategies, evaluation criteria, and stakeholder groups. These components define the alternatives to be ranked, the criteria used to evaluate them, and the stakeholder perspectives through which the rankings are interpreted.

3.2.2. Recovery Strategies

This study evaluates six road recovery strategies, consisting of five centrality-based strategies and one random recovery baseline. These strategies are adopted from the recovery strategy set of Balakrishnan and Stokkink [1] and are used here because they are simple, reproducible, and transferable across many road networks.

Structure-based priority rules are suitable for this study because they can be applied under the information constraints of cross-city comparison. In practical recovery settings, predefined rules are often used to rank asset needs under limited resources and incomplete information [44]. Optimisation-based sequencing serves a different purpose: it can generate a locally implementable repair schedule, but it depends on detailed inputs such as repair costs, resource availability, and operational constraints [18]. Such information is not consistently available across the more than two hundred networks examined in this study. Moreover, the objective here is to compare how different restoration logics perform under common simulation settings, rather than to optimise a repair programme for any single city. Structure-based priorities are therefore better suited to the comparative purpose of the analysis.

The five centrality-based strategies reflect this structure-based logic in practice, with each strategy representing a different assumption about what makes a damaged road link important for recovery. Together, they cover global through-movement, attachment to highly connected nodes, local neighbourhood importance, proximity to the rest of the network, and connection to other important links. The random recovery strategy is included as a baseline, so that the added value of using network structure to guide restoration can be assessed. This random recovery baseline is distinct from the random edge-failure procedure used to generate disruption scenarios in the simulation dataset, which is described in Section 3.5.

In the simulations, each strategy is implemented as a rule for ordering damaged links. For every disrupted arc, the strategy computes a priority score, and all disrupted arcs are then sorted and restored according to that score [1]. The strategies therefore differ only in how they define restoration priority,

which allows their recovery consequences to be compared under the same disruption and evaluation settings.

Betweenness centrality (B) prioritises arcs that lie on a high proportion of shortest paths between node pairs in the network. These are the arterial links that carry much of the network's through-traffic, so restoring them early tends to improve network-wide travel performance.

Weighted degree centrality (D) ranks arcs by the centrality of the nodes they connect. The centrality of a node is the sum of its incoming and outgoing links, and the centrality of an arc is the product of the outgoing centrality of its origin node and the incoming centrality of its destination node. This prioritises links that connect highly connected nodes.

Nearest-neighbour edge centrality (NN) evaluates each arc using only the strength of its adjacent nodes and its own travel time. Because it relies on local neighbourhood information rather than global structure, it is computationally cheaper than the global measures and is suited to time-sensitive situations.

Closeness centrality (C) scores an arc by the average shortest-path distance of its incident vertices to all other vertices in the network. This prioritises links attached to nodes that are central, in the sense of being close to the rest of the network.

Eigenvector centrality (EV) scores an arc by the centrality of its adjacent arcs, defined through the dominant eigenvector of the weighted edge-adjacency matrix. It gives higher priority to arcs connected to other highly central arcs.

Random baseline (R) restores damaged arcs in a random order. Unlike the centrality-based strategies, it does not use network structure or any performance-related importance, and it is included as a reference against which the centrality-based strategies can be compared.

The strategies and the resilience criteria are not in one-to-one correspondence. The criteria defined in Section 3.2.3 measure how the network performs as restoration proceeds, while the strategies only determine which disrupted arcs are restored first. Because each centrality measure favours a different type of network structure, different strategies can improve different criteria at different rates. This motivates the multi-criteria ranking developed in this chapter, which compares the strategies across several resilience criteria.

3.2.3. Evaluation Criteria

The evaluation criteria define the recovery outcomes against which the six strategies are compared. They are derived from the multidimensional road-network recovery framework of Balakrishnan and Stokkink [1], which evaluates recovery across operational efficiency, accessibility, spatial equity, and connectivity. Because the accessibility dimension is separated into two criteria in the stakeholder survey, the survey contains five criteria while the cross-city simulation provides four metrics. This mapping is explained where the accessibility dimension is introduced below and in the SAW criterion mapping in Section 3.6.2.

These four simulation dimensions can be organised along a logic that moves from the state of the network itself, to what the network means for people, and finally to how that meaning is distributed.

Operational efficiency and connectivity are concerned with the performance of the network structure itself, without yet addressing who benefits from that performance. These structural measures matter because they capture the direct, physical consequence of a recovery strategy on the network.

Accessibility shifts the focus from the network itself to its practical use for people. The transport system matters because it connects the built environment with key community functions and services, and accessibility-based indicators can reveal how disruption consequences differ across geographical areas [45].

Spatial equity goes one step further, asking whether this overall accessibility is equally good everywhere. Infrastructure disruption and recovery outcomes may be unevenly distributed across places and social groups, a concern that has received growing attention in infrastructure resilience research [46]. These four dimensions therefore capture complementary recovery concerns, moving from network structure, to its overall usefulness to people, to how that usefulness is distributed, rather than

offering alternative measures of the same outcome.

Criteria and performance metrics

Operational efficiency. Operational efficiency reflects the ability of the road network to support movement between locations during recovery. Following Balakrishnan and Stokkink [1], it is measured as network efficiency based on inverse shortest-path travel times between node pairs. This calculation is used for two reasons. First, it fits what operational efficiency is meant to capture and reflects an established way of measuring network efficiency [47]. Shortest-path travel time is a direct measure of how hard it is to move between two locations, so it matches the concept being evaluated. Second, the metric can be recalculated at every stage of restoration, which makes it possible to track how efficiency changes as the network recovers and to compare this across different strategies. In this study, the metric should therefore be interpreted as graph-level movement efficiency, not as observed traffic flow or demand-weighted travel performance.

Accessibility. Accessibility refers to how easily destinations can be reached through the road network. In the cross-city simulation, accessibility is represented by a single aggregate graph-based metric. Following Balakrishnan and Stokkink [1], it is calculated as a closeness-weighted reachability measure, where each destination's pre-disruption closeness centrality is used as a proxy for the opportunities available at that location. This metric is suitable for the cross-city analysis because it can be calculated consistently across all city networks using only road-network data, without requiring local information on hospitals, schools, shops, or other facilities.

This simulation metric should be distinguished from the stakeholder survey criteria. In the survey, accessibility is separated into two criteria: access to critical services and access to routine activities. This separation reflects the idea that stakeholders may value emergency-related access and routine access differently, especially across short-term and long-term recovery stages. However, the cross-city dataset does not contain facility-specific destination data, so these two survey criteria are mapped to the same aggregate accessibility metric in the cross-city SAW analysis. The Delft case study addresses this limitation by calculating access to critical facilities and access to routine facilities separately using local facility data.

Spatial equity. Spatial equity captures whether accessibility losses are evenly distributed or spatially concentrated across the road network. For each disrupted state, accessibility is first computed for every network node, and the spatial autocorrelation of these node-level accessibility values is then measured using Moran's I with inverse-distance spatial weights [1]. Moran's I is used because the equity question is not only how much accessibility remains on average, but also whether low-accessibility outcomes cluster in particular parts of the network. Since the unit of analysis is the network node, the measure reflects how evenly accessibility is distributed across the road network's structural components, not across the people who live near them.

Connectivity. Connectivity refers to the extent to which the local structure of the road network remains interconnected after a disruption. It is measured as the average local directed clustering coefficient across all nodes, following Fagiolo [48]. This measure is used because road networks in this study are represented as directed graphs, and local clustering captures whether neighbouring links remain mutually connected as damage and recovery progress. In this study, connectivity therefore reflects how well small clusters of road links stay locally connected to each other during recovery.

Together, these evaluation dimensions and performance metrics define how the six recovery strategies are assessed. The processing steps used to convert the simulation outputs into strategy-level performance scores are described in Section 3.5. A worked example is provided in Section C.1 to illustrate how operational efficiency, accessibility, spatial equity, and connectivity are calculated.

3.2.4. Recovery Stages

Post-disaster road network recovery is widely recognised as a staged process, because recovery priorities, feasible interventions, and decision constraints change over time after a disaster. Reflecting this dynamic nature, researchers and practitioners commonly distinguish between short-term and long-term recovery phases [14], [49]. These phases involve different recovery objectives and constraints, and therefore require different evaluation emphases to reflect the priorities of each stage [13].

The short-term phase refers to the acute emergency response stage during or shortly after a hazard event [13]. During this phase, recovery activities mainly involve unblocking roads and repairing impassable roads and bridges to a minimally operational state. The primary objective is to resume emergency activities such as search-and-rescue, victim evacuation, and relief distribution [14]. Decision-making in this phase is typically constrained by severe time pressure, and the phase usually lasts several days, up to approximately seven days depending on the scale of the event [49], [50].

The long-term phase refers to the stage in which the impaired road network is gradually restored towards its pre-disaster functionality through more systematic repair and reconstruction. This phase includes debris collection and disposal, permanent repair, and the complete reconstruction of damaged infrastructure, and may extend over months or even years [14], [49]. Compared with the short-term phase, the long-term phase places greater emphasis on restoring overall network functionality and achieving a desirable recovery trajectory under resource constraints [13].

Based on this distinction, this study distinguishes between short-term and long-term recovery. In the stakeholder survey, this distinction was presented in a simplified form: short-term recovery was described as temporary repair within approximately seven days, while long-term recovery was described as reconstruction over several weeks to months. The translation of this distinction into progress-based simulation windows is described separately in the simulation processing section.

Table 3.1: Comparison of Short-term and Long-term Road Network Recovery

	Short-term Phase	**Long-term Phase**
Time Horizon	A few days, up to seven days depending on the scale of the event	Months or perhaps years
Operational Activities	Unblocking roads by pushing debris to the side and repairing impassable roads and bridges to a minimal operational state	Collection and transportation of detritus to recycling sites and disposal areas; complete reconstruction of damaged infrastructure
Primary Objective	Resume emergency activities using the road network, including search-and-rescue, victim evacuation, and relief distribution	Restore the impaired network gradually to its pre-disaster functionality level with the most desirable recovery trajectory
Decision Constraints	Severe time pressure	Availability of reconstruction resources

3.2.5. Stakeholder groups

Stakeholder identification is conducted to determine the key actors involved in road damage and post-disaster recovery, and to define the stakeholder role types used in the survey and subsequent aggregation of BWM weights. An initial stakeholder list is developed based on a review of relevant literature on post-disaster road recovery and road construction projects, where stakeholder involvement has been widely examined [16], [51], [52], [53], [54].

In this study, stakeholders are defined as actors who are either directly involved in decision making and implementation of post-disaster road recovery, or significantly affected by recovery outcomes. This includes decision-making authorities, infrastructure managers, technical experts, contractors, emergency responders, transport and logistics operators, local communities, and road users whose accessibility and socio-economic conditions depend on road network restoration.

Table 3.2: Stakeholder role types involved in road damage and post-disaster recovery

Stakeholder role type	Role in road recovery
Public authority / government agency	Strategic planning, policy guidance, coordination, and decision-making for post-disaster road recovery
Infrastructure manager	Management, inspection, maintenance, and restoration of damaged road infrastructure
Technical expert / researcher	Damage assessment, technical evaluation, risk assessment, and assessment of recovery options
Contractor / construction company	Physical repair, reconstruction, and on-site implementation of recovery measures
Emergency responder	Emergency response, traffic safety management, public safety, and access to critical services during recovery
Transport / logistics operator	Maintaining transport services, logistics operations, supply chains, and economic activities during recovery
Local community	Representing local accessibility needs, community disruption, and social impacts of road recovery
Road users significantly affected by road recovery	Experiencing direct impacts of road damage and recovery outcomes in daily travel and accessibility

The eight stakeholder role types in Table 3.2 are used in the questionnaire to classify respondents according to their primary relationship to post-disaster road recovery. An additional “Other” option is included in the survey to allow respondents to specify roles that do not fit the predefined categories. These responses can be reviewed during data processing and, where appropriate, reclassified into one of the predefined stakeholder role types.

Table 3.3 maps the survey role types to representative Dutch institutions and functions, drawing on the Rotterdam-Rijnmond Safety Region (VRR) Policy Plan 2023–2027 [55] as a concrete reference point. These examples illustrate how the role types can be interpreted within the Dutch institutional context, and are representative rather than comprehensive.

The stakeholder role types are validated by comparing academic literature with Dutch policy and practice-oriented documents. The academic literature, reviewed at the start of this section, provides the basis for identifying relevant stakeholder categories and their general roles. The Dutch policy and practice-oriented sources are then used to examine whether these categories are also relevant in the Dutch recovery governance context. This process helps ensure that the stakeholder role types are supported by both literature and practical context. The list is not intended to include every possible organisation involved in road recovery, but to provide a structured classification for survey design and stakeholder-based weight aggregation.

Table 3.3: Illustrative stakeholder examples in the Dutch context

Stakeholder role type	Entities and institutional descriptions
Public authority / government agency	• Ministry of Infrastructure and Water: National policy and strategic oversight • Province of South Holland: Regional governance and funding oversight • Municipalities: Local governance, crisis coordination, and resident support • General Board (AB): Budget approval and policy frameworks • Public Prosecution Service: Legal oversight and incident investigation
Infrastructure manager	• Rijkswaterstaat: Management of highways and tunnels • Water Boards: Recovery of water-related infrastructure and flood protection • Municipalities: Management and restoration of local roads
Technical expert / researcher	• DCMR Environmental Service: Environmental and chemical safety monitoring • Multi Intelligence Center: Real-time data and operational intelligence • KNMI: Weather-related risk assessment and early warnings • Advisory Council (ARC): Specialist advice on risk and impact reduction
Emergency responder	• Safety Region (VRR): Multi-agency disaster coordination • Police: Traffic control, site security, and investigation • Fire Brigade: Rescue, scene clearance, and hazardous material response • GHOR: Health impact assessment and medical coordination • USAR.NL: Technical rescue and collapsed-structure assessment • Joint Fire Brigade: Emergency support in industrial and port areas
Transport / logistics operator	• RET: Public transport continuity and alternative transport plans • Port Authority: Port accessibility and logistics continuity • Deltalinqs: Industrial supply chain and economic recovery interests
Local community	• Population Care: Community support and crisis communication • Local residents and businesses: Local accessibility, disruption, and recovery impacts
Road users significantly affected by road recovery	• Commuters and daily travellers: Delays, detours, and reduced reliability • Public transport passengers: Service disruption and accessibility changes • Drivers, cyclists, and pedestrians: Closures, damaged links, and safety risks

The illustrative Dutch examples in Table 3.3 do not include a specific contractor or construction company, as the table only lists stakeholder examples that could be clearly identified from the reviewed Dutch policy and practice-oriented sources [55]. This role type is nevertheless retained in the questionnaire, because contractors are directly involved in the physical implementation of road recovery measures, such as repair, reconstruction, and temporary traffic management. They may therefore provide useful insights into the feasibility, resource requirements, and implementation constraints of different recovery strategies.

3.3. Best-Worst Method for Deriving Stakeholder Criterion Weights

This section describes how stakeholder judgements are converted into criterion weights for the multi-criteria ranking procedure. The judgements are collected through a stakeholder survey, described in Section 3.4, in which respondents compare the importance of the five recovery criteria separately for the short-term and long-term recovery stages. This section focuses on the weighting method itself. The judgements are transformed into individual criterion-weight vectors using the Best–Worst Method (BWM), screened for consistency, and aggregated to produce stakeholder-group and overall weights for each recovery stage.

3.3.1. BWM Formulations

The stakeholder weights are derived using the Best–Worst Method (BWM), a pairwise-comparison method for eliciting criterion weights from decision-maker judgements [56]. In this study, the criteria are predefined based on the recovery evaluation criteria introduced in Section 3.2.3, and respondents evaluate the relative importance of the same five criteria separately for the short-term and long-term recovery stages. Two formulations are used to convert these comparisons into criterion weights, a linear formulation as the main method and an equivalent formulation as a robustness check.

For each respondent and recovery stage, the survey (see Section 3.4) provides a selected best criterion, a selected worst criterion, a Best-to-Others comparison vector, and an Others-to-Worst comparison vector. These inputs are converted into a five-element criterion-weight vector using the linear BWM formulation, which is used as the main weight-estimation method because it provides a unique optimal solution and avoids the multi-optimality issue associated with the original non-linear BWM model [56].

Alongside this main method, an equivalent BWM formulation is used as a supplementary calculation on the same comparison data. It estimates how much the original comparison vectors need to be adjusted to form a fully consistent BWM comparison system [57]. The equivalent BWM weights are used as a robustness comparison with the linear BWM weights, while the correction magnitudes are used in the consistency-screening procedure described in Section 3.3.2.

Linear BWM Formulation

This subsection defines the linear BWM formulation used to estimate the main individual criterion weights. Following Rezaei [56], the procedure consists of the following steps, applied separately for each respondent and recovery stage.

Step 1: Define the criteria. Let $C = \{c_1, c_2, \ldots, c_n\}$ denote the predefined set of criteria, where $n = 5$ in this study and c_j denotes criterion j. These are the five resilience criteria introduced in Section 3.2.3.

Step 2: Select the best and worst criteria. For each respondent and recovery stage, the respondent selects the most important (best) criterion $B \in C$ and the least important (worst) criterion $W \in C$.

Step 3: Construct the Best-to-Others comparison vector. The respondent rates the importance of the best criterion B over each of the other criteria on a 1–9 scale, forming the Best-to-Others vector

$$\mathbf{A}_B = (a_{B1}, a_{B2}, \ldots, a_{Bn}), \tag{3.1}$$

where a_{Bj} denotes the stated importance of the best criterion B over criterion c_j, and $a_{BB} = 1$ because the best criterion is compared with itself. On this scale, the defined preference levels correspond to the odd integers: 1 for equal importance, 3 for moderate importance, 5 for strong importance, 7

for very strong importance, and 9 for extreme importance. The even integers represent intermediate judgements between these levels.

Step 4: Construct the Others-to-Worst comparison vector. The respondent rates the importance of each criterion over the worst criterion W on the same 1–9 scale, forming the Others-to-Worst vector

$$\mathbf{A}_W = (a_{1W}, a_{2W}, \ldots, a_{nW})^{\mathsf{T}}, \tag{3.2}$$

where a_{jW} denotes the stated importance of criterion c_j over the worst criterion W, T denotes the transpose, and $a_{WW} = 1$ because the worst criterion is compared with itself.

Step 5: Solve for the optimal weights. The optimal criterion weights are obtained by solving the following optimisation problem:

$$\begin{aligned}
\min \quad & \xi_L && (3.3)\\
\text{s.t.} \quad & |w_B - a_{Bj} w_j| \le \xi_L, \quad j = 1, \ldots, n, && (3.4)\\
& |w_j - a_{jW} w_W| \le \xi_L, \quad j = 1, \ldots, n, && (3.5)\\
& \sum_{j=1}^{n} w_j = 1, && (3.6)\\
& w_j \ge 0, \quad j = 1, \ldots, n. && (3.7)
\end{aligned}$$

Here, w_j is the weight of criterion c_j, while w_B and w_W are the weights assigned to the best and worst criteria. The variable ξ_L represents the maximum absolute deviation from the linear BWM consistency conditions $w_B = a_{Bj} w_j$ and $w_j = a_{jW} w_W$. A smaller ξ_L indicates that the stated comparisons are more compatible with the estimated weight vector. Solving this model gives one weight vector for each respondent and recovery stage. These individual weight vectors are later screened for consistency and aggregated into stakeholder-group and overall weights.

This model was implemented in Python. The implementation was verified against BWM-Solver-4, the Excel solver published on the Best-Worst Method website [58], which implements the same linear formulation [56]. Both were given the same comparison vectors, and the resulting weights were compared before the code was applied to the survey data.

Equivalent BWM Formulation

A respondent's Best-to-Others and Others-to-Worst comparisons are not always mutually consistent. The equivalent BWM, proposed by Wu et al. [57], reformulates the original nonlinear BWM in terms of optimally modified comparison vectors. It finds modified Best-to-Others and Others-to-Worst vectors that satisfy the full BWM consistency condition while minimising the largest modification from the respondent's original comparisons. In this sense, the equivalent BWM preserves the consistency logic of the nonlinear BWM.

For problems with more than three criteria, the nonlinear BWM and the equivalent BWM may admit multiple optimal weight solutions. Since this study uses five criteria, multiple optimal modified vectors may exist. This study selects a single representative vector using the secondary objective function of Wu et al., described below, and treats the resulting equivalent BWM weights as a robustness reference alongside the primary linear BWM weights. The equivalent BWM is therefore used in two ways.

- First, the **size of the modification** is used as a consistency-related indicator for screening BWM responses, since it measures how far a response is from full consistency (Section 3.3.2).
- Second, the **modified-comparison weights** are used as a robustness comparison with the primary linear BWM weights, so that the strategy rankings can be checked against an alternative weight set derived from the consistency logic of the nonlinear BWM.

The notation from Section 3.3.1 is used here. Let $\mathbf{A}_B = (a_{B1}, \ldots, a_{Bn})$ and $\mathbf{A}_W = (a_{1W}, \ldots, a_{nW})^{\mathsf{T}}$ denote the respondent's original comparison vectors, and let $\tilde{\mathbf{A}}_B = (\tilde{a}_{B1}, \ldots, \tilde{a}_{Bn})$ and $\tilde{\mathbf{A}}_W = (\tilde{a}_{1W}, \ldots, \tilde{a}_{nW})^{\mathsf{T}}$ denote the modified vectors, where a tilde marks a modified value.

The modified vectors are found by solving the following problem, which minimises the largest modification ξ_E applied to any single comparison:

$$\begin{aligned}
\min \quad & \xi_E && (3.8)\\
\text{s.t.} \quad & |\tilde{a}_{Bj} - a_{Bj}| \leq \xi_E, \quad j = 1, \ldots, n, && (3.9)\\
& |\tilde{a}_{jW} - a_{jW}| \leq \xi_E, \quad j = 1, \ldots, n, && (3.10)\\
& \tilde{a}_{Bj}\tilde{a}_{jW} = \tilde{a}_{BW}, \quad j = 1, \ldots, n, && (3.11)\\
& \tilde{a}_{Bj} \geq 1, \quad j = 1, \ldots, n, && (3.12)\\
& \tilde{a}_{jW} \geq 1, \quad j = 1, \ldots, n. && (3.13)
\end{aligned}$$

The constraints have the following roles:

- the first two keep each modified value within ξ_E of the original, so the modification stays as small as possible;
- the third enforces full BWM consistency on the modified vectors, where $\tilde{a}_{BW}$ is the modified Best-to-Worst value;
- the last two keep the modified comparison values at least one, which preserves the BWM comparison direction where one represents equal importance.

The optimal value ξ_E is the size of the modification, and serves as the consistency measure used in the screening.

For five criteria, more than one optimal modified vector may satisfy this problem. To select a single representative vector, this study applies the secondary objective function proposed by Wu et al. [57], using its analytical closed-form solution (Theorem 6 in their work). After the optimal modified Best-to-Worst value is fixed, each non-best and non-worst criterion has its Best-to-Others and Others-to-Worst values adjusted by the same perturbation, obtained in closed form, so that full cardinal consistency is satisfied. The implementation additionally enforces a lower bound of one on the corrected comparison values, and verifies full cardinal consistency after correction; responses that no longer satisfy this condition are marked invalid and excluded.

The equivalent BWM weights are then computed from the resulting modified Others-to-Worst vector:

$$w_j^E = \frac{\tilde{a}_{jW}}{\sum_{k=1}^{n} \tilde{a}_{kW}}, \quad j = 1, \ldots, n, \tag{3.14}$$

where w_j^E is the equivalent BWM weight of criterion c_j and k is the summation index. The resulting weights sum to one by construction.

3.3.2. Consistency Screening and Valid Weight Sets

This subsection explains how individual BWM responses are evaluated for internal consistency and how the survey data are divided into a broader sample and a strict sample. Both samples are carried forward to the stakeholder-weight aggregation step in Section 3.3.3. The screening procedure uses the equivalent BWM described in Section 3.3.1, where the correction magnitudes indicate how far the original response deviates from a fully consistent BWM structure.

Before the perturbation indicators are calculated, responses are checked for two technical data-quality conditions. First, the corrected comparison vectors must satisfy the BWM consistency condition after the equivalent BWM correction. Second, the Best-to-Others vector must not contain a value of 1 between the selected best criterion and another non-best criterion, because this would make the selected best criterion indistinguishable from another criterion in the elicitation task. Responses that fail either condition are excluded from both the broader and strict samples and are reported separately.

For each response that passes these validity checks, three perturbation indicators are computed from the differences between the corrected and original comparison vectors.

- **Average perturbation** ($\mathrm{avg_pert}$): the mean of the absolute corrections $|\tilde{a}_{Bj} - a_{Bj}|$ and $|\tilde{a}_{jW} - a_{jW}|$ across all criteria.

- **Maximum perturbation** ($\mathrm{max_pert}$): the largest single correction across the same set.
- **Best-to-Worst gap** ($\mathrm{a_BW_gap}$): the discrepancy between the Best-to-Worst value implied by the Best-to-Others vector and that implied by the Others-to-Worst vector. A larger gap indicates that the two reference comparisons are less compatible with each other.

The retained responses are then classified into three perturbation grades.

- **Grade A**: $\mathrm{avg_pert} < 1.5$, $\mathrm{max_pert} < 2.0$, and $\mathrm{a_BW_gap} < 4$. The response requires only a small correction to reach consistency.
- **Grade B**: $\mathrm{avg_pert} < 2.0$, $\mathrm{max_pert} < 3.0$, and $\mathrm{a_BW_gap} < 4$. The response requires a moderate correction.
- **Grade C**: all other retained responses, including those with $\mathrm{a_BW_gap} \geq 4$.

These three grades and their thresholds are specific to this study. They are not taken from Wu et al. [57] or from any published consistency table. Instead, they are pragmatic cut-offs informed by the structure of the BWM preference scale and by the distribution of the survey data.

The maximum-perturbation thresholds are anchored to the 1–9 BWM scale. On this scale, the defined preference levels use the odd integers: 1 for equal importance, 3 for moderate importance, 5 for strong importance, 7 for very strong importance, and 9 for extreme importance. Moving from one defined preference level to the next, for example from 3 to 5 or from 5 to 7, changes the comparison value by two scale points. Grade A requires $\mathrm{max_pert} < 2.0$, meaning that no single comparison is shifted by more than one full defined preference level and that the qualitative structure of the response is preserved. Grade B relaxes this to $\mathrm{max_pert} < 3.0$, allowing the largest single correction to reach about one and a half defined preference levels. The average-perturbation thresholds, 1.5 and 2.0, provide an additional check on the overall size of the corrections. This prevents a response from receiving a high grade when many of its comparison values require noticeable correction.

The $\mathrm{a_BW_gap} \geq 4$ condition is applied as a separate grading condition for responses whose two Best-to-Worst reference comparisons are strongly incompatible. The largest possible discrepancy between the Best-to-Worst value implied by the Best-to-Others vector and the value implied by the Others-to-Worst vector is eight scale points, corresponding to a value of 1 against a value of 9. A gap of four is therefore half of this maximum possible discrepancy. This means that the two reference comparisons differ by half of the usable scale range. This cut-off is also consistent with the observed survey data: over 80% of the retained responses fall at a gap of three or below in both recovery stages, with 82% in the short-term stage and 84% in the long-term stage. Only a small upper tail remains at four and above.

Because the precise location of these cut-offs involves a degree of researcher judgement, their effect on the conclusions is addressed directly through the broader-versus-strict robustness comparison rather than through the threshold values themselves. This classification does not filter the main analysis: the main rankings are based on the broader sample, which includes all valid responses regardless of grade. The grades are used only to form the stricter Grade A–B subsample, which serves as a robustness reference to test whether excluding high-perturbation responses changes the rankings (Section 3.8.4).

Based on this classification, two samples are constructed for each recovery stage.

- The **broader sample** contains all responses that pass the validity checks, including Grades A, B, and C. It is the primary sample for the main rankings.
- The **strict sample** is the subset containing only Grade A and Grade B responses. It serves as a robustness reference for the main rankings.

The two samples are constructed separately for the short-term and long-term BWM data, because the perturbation indicators are computed independently for each stage and a respondent may receive a different grade in each stage.

The broader sample is used as the primary basis for two reasons. First, the survey targets a specialised population of stakeholders involved in post-disaster road recovery, and some stakeholder groups are small. Excluding Grade C responses could discard useful input and leave some groups with too few responses for group-level aggregation. Second, using the broader sample as the primary basis and

the strict sample as a robustness reference makes it possible to test directly whether high-perturbation responses materially affect the final rankings. This comparison is reported in Section 3.8.

3.3.3. Aggregation of Individual, Group-Level, and Overall Weights

This subsection describes how individual stakeholder weight vectors are computed for the broader and strict samples defined in Section 3.3.2, and how these individual weights are aggregated into stakeholder-group weights and overall weights. Two BWM formulations are used: the linear BWM [56] described in Section 3.3.1, and the equivalent BWM [57] described in Section 3.3.1.

The same five resilience criteria are evaluated in both recovery stages. Section C of the questionnaire collects short-term BWM responses, and Section D collects long-term BWM responses. Responses are processed at the individual level. For each respondent and recovery stage, the selected BWM formulation is applied to produce a five-element weight vector over the criteria.

Combining the two BWM formulations with the two samples gives four analysis configurations, summarised in Table 3.4. The *linear–broader* configuration applies the linear BWM formulation to the broader sample and is the primary configuration reported in this study. The other three configurations, *linear–strict*, *equivalent–broader*, and *equivalent–strict*, are reported as complementary analyses. The pairwise comparisons used to assess the robustness of the main rankings are described in Section 3.8.4.

Table 3.4: The four BWM weight-processing configurations used in this study.

Configuration	BWM formulation	Sample
linear–broader (primary)	Linear BWM	Broader sample (A + B + C)
linear–strict	Linear BWM	Strict sample (A + B only)
equivalent–broader	Equivalent BWM	Broader sample (A + B + C)
equivalent–strict	Equivalent BWM	Strict sample (A + B only)

The linear–broader configuration is the primary basis for the main rankings because it combines the standard BWM formulation with the broader sample that preserves the full range of stakeholder responses (Section 3.3.2). For each configuration, the BWM model is solved independently for every respondent and stage, yielding w_{ir}^{s}, the weight that respondent r assigns to criterion i in stage s. The four configurations differ only in which respondents are included (broader versus strict) and which solver produces w_{ir}^{s}. The aggregation procedure below is applied independently to each configuration.

Stakeholder groups are based on the eight role types in Section A of the questionnaire. Role types with sufficient valid responses are used for the main group-level aggregation. Role types with very few valid responses are retained only for exploratory descriptive analysis, because their group-level averages are sensitive to individual responses. The final grouping and the role types treated as exploratory are determined after data collection and reported in Chapter 4.

Group-level weights are obtained by aggregating individual BWM-derived weight vectors within each stakeholder group, separately for each recovery stage. Following group BWM applications that aggregate individual criteria weights using a weighted geometric mean method, this study uses the unweighted geometric mean because respondents within the same stakeholder group are treated as equally important [59]. The resulting geometric-mean values are then divided by their sum, so that the group-level criterion weights add up to one and can be combined with the simulation-based performance scores in the SAW stage.

For stakeholder group g, recovery stage s, and criterion i, the geometric mean of the individual weights is

$$\mathrm{GM}_{ig}^{s} = \left(\prod_{r \in R_g^s} w_{ir}^{s} \right)^{1/N_g^s}, \tag{3.15}$$

and the group-level weight is obtained by normalising across criteria:

$$\bar{w}_{ig}^{s} = \frac{\mathrm{GM}_{ig}^{s}}{\sum_{k=1}^{n} \mathrm{GM}_{kg}^{s}}, \tag{3.16}$$

where R_g^s is the set of valid respondents in stakeholder group g for stage s, N_g^s is the number of respondents in R_g^s, and w_{ir}^s is the individual criterion weight. The renormalisation step restores the unit-sum property after aggregation.

Overall weights are then obtained by combining the group-level weights using stakeholder-group importance weights λ_g. For recovery stage s and criterion i:

$$w_{i,\mathrm{overall}}^{s} = \sum_{g=1}^{G} \lambda_g \, \bar{w}_{ig}^{s}, \tag{3.17}$$

where G is the number of stakeholder groups and $\sum_{g=1}^{G} \lambda_g = 1$. Computing overall weights through group-level values, rather than averaging individual responses directly, prevents groups with more respondents from automatically dominating the overall result.

The values of λ_g are a research-design choice rather than a survey output, since the questionnaire does not elicit stakeholder-group importance separately. As a baseline, this study assigns equal importance to all groups, $\lambda_g = 1/G$. This provides a neutral reference point for interpreting the overall ranking when there is no empirical basis for assigning different importance levels to different stakeholder groups. Sensitivity to this choice is examined in Section 3.8.2.

3.3.4. Coding of Open-Ended Explanations

Alongside the closed weighting questions, the survey (see Section 3.4) asked each respondent to explain, in their own words, why they chose their most-important and their least-important criterion. These explanations were coded into themes to understand the reasoning behind the elicited weights. The analysis is exploratory and is meant to help interpret the weights rather than to test a hypothesis, and it is not part of the weight calculation itself. Responses were given in Mandarin and were translated into English before coding.

The themes were developed inductively, following the logic of qualitative content analysis [60]. In this approach, categories are derived from the material itself, and their frequencies can then be reported. A large language model (ChatGPT 5.5, OpenAI) was first used to read the translated responses and propose an initial set of themes. This AI-assisted step is consistent with recent applications of large language models in qualitative content analysis and inductive thematic analysis, where models are used to support the development of coding schemes, codes, or candidate themes from textual data [61, 62].

On reviewing the initial theme proposal, it became clear that the most-important and least-important answers needed to be treated differently, because a least-important answer explains why a criterion can be deprioritised rather than why it matters. The coding scheme was therefore reorganised into two separate families of codes, one for the most-important rationales (M1 to M8) and one for the least-important rationales (L1 to L11), and the themes were named on this basis.

Each response was then assigned to a single theme. The initial assignment and a confidence rating were produced by the model. The author reviewed every assignment against the original response and changed those that did not fit, with particular attention to the least confident cases. Of the 244 assignments, 35 were changed. Responses that gave no usable reason, or whose reason did not relate to the chosen criterion, were assigned to a separate category for unclear or unclassifiable responses.

The qualitative review also surfaced an unrelated but important issue in the closed-ended survey data. During the author's review of the individual theme assignments, the responses were also checked for a possible wording ambiguity in the operational-efficiency criterion (C1). This criterion was intended to represent network-level travel performance, consistent with the simulation metric based on inverse shortest-path travel times. However, the Mandarin wording referred to travel time and could therefore be interpreted by some respondents as their own individual trip time rather than as network-level

efficiency. Responses whose open-ended explanations clearly indicated this individual-trip-time interpretation were flagged for a separate robustness check. These flagged responses were retained in the main BWM sample and were only temporarily excluded in the sensitivity analysis described in Section 3.8.1.

The model's role was limited to supporting the initial organisation of themes and assignments. The author reviewed and finalised every assignment, following common practice in AI-assisted qualitative analysis, where large language models support human interpretation rather than replace it [62, 63]. However, the coding was conducted by a single coder, so the theme frequencies should be treated as exploratory evidence that helps explain the BWM weights.

3.4. Stakeholder Survey Design

The criterion weights in Section 3.3 come from a stakeholder survey. This section sets out what the survey had to achieve, the design choices made to achieve it, and the resulting questionnaire.

3.4.1. Design requirements

The survey had to meet four requirements.

First, it had to produce the comparisons that BWM needs for the five resilience criteria defined in Section 3.2.3.

Second, the short-term and long-term recovery stages had to be elicited separately, using the same five criteria in both, so that the resulting weights could be compared directly across stages.

Third, both stages had to be answered by the same respondent within one questionnaire, so that shifts in priority could be traced at the individual level rather than only between samples.

Fourth, the survey had to be completable by respondents with no knowledge of the simulation model. The stakeholder groups defined in Section 3.2.5 include road users and the local community, who are not technical experts, but whose priorities are a required input to the framework.

The fourth requirement pulls against the first three. The BWM comparison structure is abstract, and the division between short-term and long-term recovery is an analytical construct rather than everyday language. The central design problem was therefore how to make the task manageable for non-experts without weakening the elicitation itself.

3.4.2. How BWM shaped the questionnaire

The reasons for choosing BWM are set out in Section 2.5. Compared with AHP, BWM requires fewer comparisons and tends to produce more consistent judgements [38].

This property determines the size of the questionnaire. A full pairwise comparison of five criteria requires ten judgements. BWM requires only the best-to-others and others-to-worst vectors, which gives seven judgements per stage. Across two stages, this reduces the elicitation from twenty comparisons to fourteen. The reduction matters here because the intended respondents are practitioners and members of the public with limited time, and because long comparison tasks tend to cause fatigue, which in turn lowers the consistency of the judgements.

The BWM structure also fixes the order of the questions. Respondents first identify the most and the least important criterion, and only then give the numerical comparisons. This ordering also serves a methodological purpose. It fixes every later comparison to the two reference criteria that BWM requires, and it lets respondents form an overall judgement before assigning numbers.

BWM also yields a consistency measure for each response. Responses that fail the consistency checks are excluded from weight estimation. The screening thresholds and their justification are given in Section 3.3.2.

3.4.3. Pilot test

The questionnaire was piloted before the main distribution. Fourteen respondents took part, recruited through friends (ten) and experts (four). They covered road users (seven), technical experts and re-

searchers (four), and the local community (three), and were based in Taiwan, the Netherlands, China, Cameroon, and Indonesia.

The pilot feedback concentrated on three issues, shown in Table 3.5.

Table 3.5: Pilot feedback by category

Feedback category	Mentions
Distinction between short-term and long-term recovery	13
Clarity of the criterion definitions	6
Clarity of the recovery scenario	3

The three issues point to the same underlying problem. A definition can be correct and still be difficult to use at the moment a respondent has to make a judgement. The definitions were correct. The problem was where they appeared.

The questionnaire was revised in three ways.

First, the presentation of the two recovery stages was strengthened. Alongside the time horizon, the questionnaire now states the focus of each stage and shows images of typical recovery work. This is described in Section 3.4.4.

Second, the criterion definitions were moved. They were originally introduced once, in Section B. In the revised questionnaire, a short definition of each criterion appears next to the options in every BWM comparison, and a hint panel below the answer field explains what each rating level means. The principle behind this change is to bring the definition to the point where the judgement is made, rather than asking respondents to remember what they read in an earlier section.

Third, the wording of the recovery scenario was rewritten to make it easier to follow.

3.4.4. The evaluation context shown to respondents

Respondents need a clear context before they can make a BWM judgement. They have to know what situation they are in, and which recovery stage they are ranking the criteria for. This creates a design problem. The context has to be concrete enough for respondents to answer, but not so concrete that it gives them the answer.

What the questionnaire provides. The questionnaire presents a hypothetical road disruption scenario, so that all respondents answer with reference to the same concrete situation rather than in the abstract.

The two recovery stages are presented in a simplified form. Relative to the full definition in Table 3.1, the questionnaire keeps the time horizon and the focus of each stage, and adds images of typical recovery work. Short-term recovery is described as temporary repair within about seven days, focused on immediate needs, and is illustrated with steel plate road repair and a temporary bridge. Long-term recovery is described as permanent reconstruction over several weeks to months, and is illustrated with full resurfacing and reconstruction. Respondents can also expand a panel showing example recovery tasks for each stage. Figure 3.3 shows this screen.



Figure 3.3: How the two recovery stages were presented to respondents.

Note. The road recovery illustrations shown in the survey were generated using ChatGPT Images by OpenAI (2026) and subsequently selected, combined, labelled, and edited by the author.

What the questionnaire deliberately withholds. The questionnaire does not show the *primary objective* or the *decision constraints* rows of Table 3.1.

This is a deliberate elicitation choice. The two rows describe what each recovery stage is trying to achieve. In the short-term stage, for example, the stated objective is to resume search-and-rescue, evacuation, and relief distribution. That objective points directly at one of the five criteria, access to critical facilities (C2). A respondent who reads it before answering would be led towards C2 by the questionnaire itself, and the resulting weight would reflect the wording of the questionnaire rather than their own judgement. What each recovery stage should aim for is exactly what the BWM elicitation is meant to measure, so the questionnaire does not state it.

The questionnaire therefore describes what each stage *looks like*, in terms of duration and the kind of work involved, but not what each stage is *for*. This lets respondents distinguish between the two stages while leaving the priority judgement to them.

The same reasoning applies to the scenario. It is used only as framing. It is not the damage input to the simulation, which applies controlled disruption levels of $10\,\%$ and $30\,\%$ independently of the survey (Chapter D).

3.4.5. Questionnaire structure

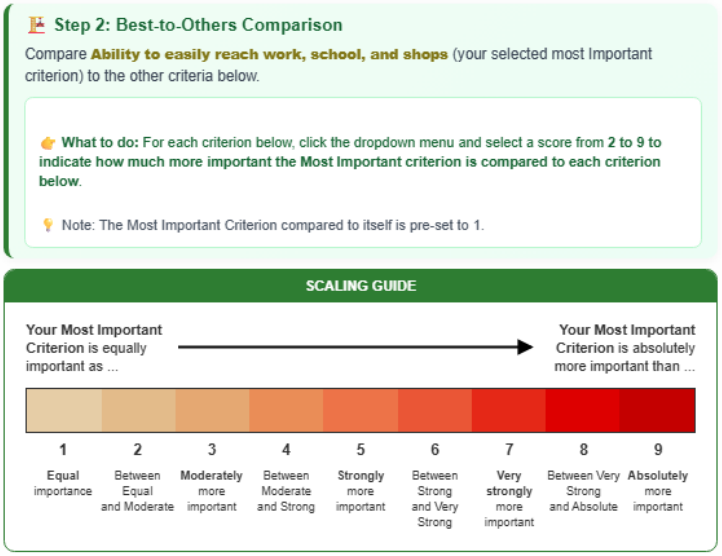
The questionnaire has four sections. The full version is provided in Chapter D.

Section A collects respondent background information, including demographic characteristics, stakeholder role type, years of working experience, and country of practice. The stakeholder role type classifies respondents by their main relationship to post-disaster road recovery, using the role types defined in Section 3.2.5. This supports the stakeholder-group analysis.

Section B presents the evaluation context: the hypothetical disruption scenario, the five resilience criteria, and the two recovery stages. As described in Section 3.4.4, the criterion definitions do not appear only here. A short version of each definition is repeated in every BWM comparison that follows.

Section C elicits priorities for the short-term recovery stage. Respondents first select the most and the least important criterion. They then compare the most important criterion against all others, and all criteria against the least important one, on a 1–9 scale, where 1 means equal importance and 9 means extreme importance. Figure 3.4 shows one of these comparison screens.

Section C also includes an open question asking respondents to explain briefly why they chose the most and the least important criterion. These answers serve two purposes. They allow the numerical weights to be read alongside the reasoning given for them, and they provide a check on whether the stated reasoning is consistent with the elicited weights. They are used in Section 4.2 to interpret the differences between stakeholder groups.



(a) Task instructions and scaling guide.



(b) Extract from the comparison screen.

Figure 3.4: The best-to-others comparison for the short-term recovery stage.

Section D repeats the same comparison tasks and the same open question for the long-term recovery stage.

The short-term stage is always presented before the long-term stage. This order follows the temporal logic of recovery and avoids asking respondents to reason about reconstruction before they have considered emergency repair. The order was not counterbalanced. The implications of this choice are discussed in Section 7.4.

3.4.6. Administration and ethics

The survey was administered online using Qualtrics. It was distributed through community groups, professional networks, professional email, and academic email. No quotas were set and no stratified sampling was applied. The resulting respondent distribution is reported in Section 4.1.

The study was approved by the TU Delft Human Research Ethics Committee before data collection began. The approval letter is included in Chapter A. Participation was voluntary and responses were collected anonymously.

3.5. Simulation-Based Performance Evaluation Method

This section describes how road-network recovery simulations are converted into performance scores for the six recovery strategies. The procedure has four steps:

1. simulate damage and strategy-based recovery;
2. convert raw network measures into recovery trajectories;
3. map restoration progress to short-term and long-term recovery windows;
4. summarise each recovery trajectory as an area-above-curve (AAC) deviation score.

The application of this procedure in the cross-city analysis and in the Delft facility-based case study is described in Section 3.7.

3.5.1. Damage and Recovery Simulation Logic

The simulation procedure evaluates how each recovery strategy affects network performance after a controlled disruption. It is used as a controlled comparison rather than as a prediction of a specific disaster event. For each network, damage level, and recovery strategy, the simulation first creates a disrupted network state and then restores the damaged links according to the priority order defined by the strategy. The remainder of this subsection describes how the disruption is generated, how recovery proceeds under each strategy, and how performance is measured and repeated across simulation runs.

The disruption scenarios are generated by randomly removing a fixed share of road links before recovery begins, following the design of Balakrishnan and Stokkink [1]. This represents a general level of network damage, without assuming any particular cause, spatial pattern, or hazard type. Because the same random-failure procedure is applied across all city networks, recovery strategies can be compared under a consistent disruption condition rather than one tailored to a specific city or event.

Recovery then proceeds according to the selected strategy. Before any damage occurs, each strategy calculates a priority score for every road link using the intact network. This score stays fixed throughout the recovery process and determines the order in which damaged links are restored, with higher-priority links restored first. The strategies therefore differ only in how they calculate this priority score, while the disrupted network state, damage level, and performance-evaluation procedure are kept the same across strategies.

After each restoration step, the network performance measures introduced in Section 3.2.3 are recalculated on the partially restored network. This produces a sequence of performance values for each strategy and criterion as restoration progresses. Because the disrupted links are selected at random, the simulation is repeated over 20 disruption iterations for each strategy and damage level. These repeated runs reduce dependence on any single random damage pattern and provide the recovery trajectories described in the next subsection.

3.5.2. Recovery Trajectories and Performance Ratios

At each restoration step, the raw network measures are recalculated on the partially restored network. Let $M_i(x)$ denote the raw value of performance metric i at restoration progress x, where x is the percentage of damaged links that has been restored. Let $M_i(0^-)$ denote the value of the same metric in the pre-disruption network. The performance ratio is defined as

$$f_i(x) = \frac{M_i(x)}{M_i(0^-)}. \tag{3.18}$$

A value of $f_i(x) = 1$ indicates that the network has reached its pre-disruption performance level for metric i. For most metrics, values below one indicate performance below the pre-disruption baseline. For spatial equity, the ratio is interpreted as a baseline-deviation measure, because the underlying Moran's I value can behave differently from direct performance measures. Expressing performance in ratio form makes recovery trajectories comparable across cities and metrics with different absolute scales.

Tracking $f_i(x)$ across successive restoration steps gives a recovery trajectory for each strategy, metric, city, damage level, and disruption iteration. The trajectory shows how quickly a strategy restores a given dimension of network performance as damaged links are progressively reinstated. These trajectories are the basis for the stage-specific deviation scores defined in Section 3.5.4.

Figure 3.5 illustrates this logic for operational efficiency in Civita di Bagnoregio under the 30% damage scenario. The horizontal axis shows restoration progress, and the vertical axis shows the efficiency performance ratio. Each line represents the mean trajectory across disruption iterations, and the shaded band shows the corresponding 95% confidence interval.

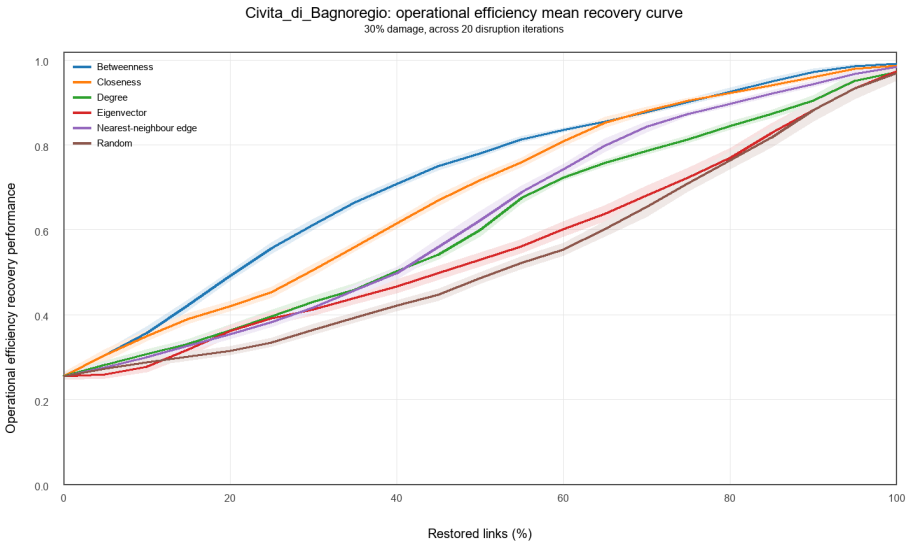


Figure 3.5: Mean efficiency recovery trajectories of the six recovery strategies for Civita di Bagnoregio under 30% damage. Each line is the mean over disruption iterations and the shaded band is its 95% confidence interval.

3.5.3. Mapping Restoration Progress to Short- and Long-Term Windows

The stakeholder survey defines short-term and long-term recovery using approximate calendar-time descriptions, while the simulation records recovery as restoration progress from 0% to 100% of damaged links restored. The simulation data do not include calendar-time information and do not distinguish between temporary repair and permanent reconstruction. A progress-based mapping is therefore required before the simulated recovery trajectories can be evaluated using the two recovery stages used in the survey.

In the main analysis, each recovery trajectory is divided into two windows using a 25% restoration-progress boundary. The short-term window covers 0% to 25% of restoration progress, and the long-

term window covers 25% to 100%. This boundary is used as a modelling approximation rather than as a universal definition of recovery time.

A key issue in this mapping is that the simulation data and the stakeholder survey define recovery stages in different ways. The simulation data are progress-based, because recovery is represented as a sequence of restored road links, whereas the survey is time-based, because respondents distinguish between short-term and long-term recovery using an approximate time horizon. Aydin et al. [64] is used as a reference because it reports both recovery progress and recovery time in an empirical road-recovery case.

The context and data used by Aydin et al. make it a useful reference for this mapping. Their study examines the rural road network of Sindhupalchok District in Nepal after the 2015 Gorkha earthquake, where earthquake-triggered landslides blocked road segments. The road network was modelled as a weighted graph with 457 nodes and 557 edges. By overlaying the road network with landslide data, the study identified 66 closed road segments, corresponding to approximately 12% of the network. This provides a contextually grounded reference for the light-damage scenario used in the present study. Recovery times in Aydin et al. are linked to landslide blockage severity, local and authority information, operational capacity, and the assumption of one recovery team, and are used in their simulation to calculate how long different recovery strategies take to reach 25%, 50%, 75%, and 100% recovery levels.

The most relevant reference point is Strategy 2 in Aydin et al., which restores road segments first by road hierarchy and then by proximity to the main resource centre. This strategy is suitable because it is not directly optimised by road-clearing time and is neither the fastest nor the slowest strategy in the 25% recovery results. In their results, Strategy 2 reaches the 25% recovery level in 144.51 hours, or approximately six days. This is close to the one-week short-term recovery horizon used in the stakeholder survey. The 25% restoration-progress threshold is therefore used as an operational approximation of the short-term recovery stage in the simulation. It is not treated as a universal rule that 25% recovery always equals one week, but as a context-informed mapping between a progress-based recovery measure and a time-based recovery horizon.

Because this mapping is based on a single empirical reference, and because recovery speed depends on hazard type, terrain, resource availability, and operational conditions, the threshold is treated as a modelling assumption. Its influence on the results is examined through the recovery-stage sensitivity analysis described in Section 3.8.3.

3.5.4. Area-Above-Curve Deviation Scores

Within each recovery window, the recovery trajectory is summarised using an area-above-curve (AAC) deviation score. The purpose of this score is to measure the average performance loss relative to the pre-disruption baseline over the window. A trajectory that remains close to the baseline has a small AAC deviation score, while a trajectory that stays far below the baseline has a larger score. Lower values therefore indicate better recovery performance.

For a recovery window $[a, b]$, the AAC deviation score is defined as

$$\mathrm{AAC}_w = \frac{\int_a^b \big(1 - f(x)\big)\,\mathrm{d}x}{b - a}, \tag{3.19}$$

where w indexes the recovery window, x denotes restoration progress, and $f(x)$ is the performance ratio at progress x, as defined in Section 3.5.2. The numerator measures the cumulative gap between the recovery trajectory and the pre-disruption baseline $f(x) = 1$. Dividing by the window length $b - a$ converts this area into an average deviation within the window, so that short-term and long-term scores can be compared despite their different window lengths.

In practice, $f(x)$ is observed only at the discrete restoration steps of the simulation. The integral in Equation (3.19) is therefore approximated using the trapezoidal rule between consecutive sampled points. If a recovery-window boundary falls between two simulated restoration steps, the value of $f(x)$ at the boundary is obtained by linear interpolation between the adjacent sampled points.

Most deviation scores fall within the expected range of $[0, 1]$. In the primary analysis, only 37 of 21,408 deviation scores, or about 0.17%, fall outside this range before clipping. These cases occur only for spatial equity and connectivity. Negative deviation scores can occur when a performance metric temporarily exceeds its pre-disruption baseline within a recovery window. Values above one occur only for spatial equity, because Moran's I is a spatial autocorrelation measure and can take negative values, making the gap from the pre-disruption baseline larger than one.

Given that the out-of-range values are rare, each deviation score is clipped to $[0, 1]$ before aggregation. This keeps all criteria on a common bounded scale for the SAW procedure. The clipped deviation scores are then averaged across disruption iterations for each strategy, city, damage level, criterion, and recovery window. These mean deviation scores form the simulation-side performance inputs for the SAW ranking procedure described in Section 3.6. Descriptive summaries of the clipped deviation scores are reported in Section E.1.

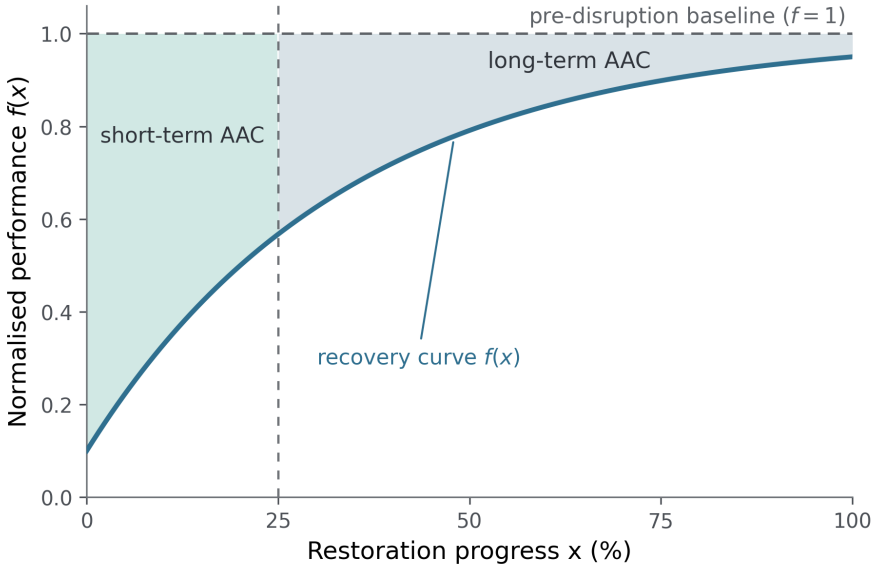


Figure 3.6: Schematic of the area-above-curve (AAC) deviation score. The dashed line represents the pre-disruption baseline $f(x) = 1$. The solid curve is an illustrative recovery trajectory. The shaded region is the deviation from the baseline within a recovery window. Dividing this area by the window length gives the AAC deviation score.

3.6. Simple Additive Weighting for Strategy Ranking

This section defines how stakeholder-derived criterion weights are combined with simulation-based recovery performance scores to rank the six recovery strategies. The section first presents the general Simple Additive Weighting (SAW) formulation and then explains how it is applied in this study.

3.6.1. General SAW Formulation

Simple Additive Weighting (SAW) is a value-function-based MCDA method in which each alternative is evaluated by a weighted sum of its criterion scores [39]. Let $A = \{A_1, A_2, \ldots, A_m\}$ denote the set of alternatives and $C = \{C_1, C_2, \ldots, C_n\}$ the set of evaluation criteria.

Before the weighted sum is calculated, the raw criterion values must be converted into comparable normalised scores. This step is necessary for two reasons. First, the criteria entering an MCDA problem are measured in different units and have different numerical ranges. Directly summing such raw values would let criteria with larger numerical scales dominate the composite score, regardless of their assigned weights, so normalisation places all criteria on a common scale. Second, criteria differ in their preference direction. For benefit criteria, larger values already indicate better performance, whereas for cost criteria, smaller raw values indicate better performance and must be transformed so that larger normalised values consistently represent better performance. After this step, all r_{ij} values can be

interpreted as benefit-oriented normalised scores.

After orientation and normalisation, the composite score of alternative A_j is

$$S_j = \sum_{i=1}^{n} w_i r_{ij}, \tag{3.20}$$

where r_{ij} is the normalised benefit-oriented score of alternative A_j on criterion C_i, w_i is the weight of criterion C_i, and $\sum_{i=1}^{n} w_i = 1$. Alternatives are ranked in descending order of S_j, with higher scores indicating better overall weighted performance. Because SAW is compensatory, weaker performance on one criterion can be partly offset by stronger performance on another criterion [40].

3.6.2. Application to Recovery Strategy Ranking

In this study, the alternatives A_j are the six recovery strategies. The normalised criterion scores r_{ij} are derived from simulation-based AAC deviation scores, and the criterion weights w_i are obtained from the BWM stakeholder-weighting procedure. The SAW calculation is applied separately for each city, damage level, recovery stage, stakeholder group, and weight-processing configuration.

Cost-Criterion Normalisation

The raw criterion scores used for SAW are the AAC deviation scores described in Section 3.5.4. Each deviation score measures the average gap between a recovery trajectory and the pre-disruption performance baseline within a recovery window. A higher deviation score therefore indicates worse recovery performance, so all criteria are treated as cost criteria before they enter the SAW calculation.

Although the AAC deviation scores are dimensionless and clipped to $[0, 1]$, different criteria may still occupy different parts of this range. To make the criteria comparable within each decision matrix, deviation scores are rescaled separately for each criterion across the six candidate strategies. The max–min normalisation for cost criteria follows [65], in which the strategy with the smallest deviation on a criterion receives a score of one and the largest receives zero:

$$r_{ij} = \frac{\max_k d_{ik} - d_{ij}}{\max_k d_{ik} - \min_k d_{ik}}, \tag{3.21}$$

where d_{ij} is the deviation score of strategy j on criterion i, and the maximum and minimum are taken over all candidate strategies k on that criterion. If all strategies have the same deviation score on a criterion, they are all assigned $r_{ij} = 1$.

Decision Matrices

The main analysis is carried out separately for each city, damage level, and recovery stage. For each combination, the simulation provides AAC deviation scores for each recovery strategy, criterion, and disruption iteration. These scores are first averaged over the 20 disruption iterations, giving one mean deviation score for each strategy on each of the four simulation criteria, operational efficiency, accessibility, spatial equity, and connectivity.

The mean deviation scores are then normalised using the cost-criterion normalisation defined in Equation (3.21). For each city, damage level, and recovery stage, this process produces one 6×4 decision matrix. The six rows represent the recovery strategies, the four columns represent the simulation criteria, and each entry r_{ij} is the normalised score of strategy j on criterion i. This 6×4 matrix applies to the cross-city analysis. In the Delft case study, the separation of critical and routine accessibility produces a 6×5 decision matrix, as described in Section 3.7.2.

Criterion Mapping

Before stakeholder weights can be applied, the five BWM survey criteria are mapped to the four simulation performance metrics, as summarised in Table 3.6. Operational efficiency, spatial equity, and connectivity each map to their corresponding simulation metric. In the cross-city analysis, the simulation provides a single aggregate accessibility metric. The weights for access to critical services and access to routine activities are therefore both applied to this same accessibility metric, so only their sum

affects the cross-city SAW score. This limitation is specific to the cross-city dataset and is addressed in the Delft case study, where critical- and routine-facility accessibility are modelled separately.

For the cross-city analysis, the SAW score of strategy j is therefore calculated as

$$S_j = w_1\, r_{\mathrm{eff},j} + (w_2 + w_3)\, r_{\mathrm{acc},j} + w_4\, r_{\mathrm{equity},j} + w_5\, r_{\mathrm{conn},j}, \tag{3.22}$$

where w_1 to w_5 are the stakeholder weights for the five BWM criteria, and $r_{\mathrm{acc},j}$ is the single aggregate accessibility score. For the Delft case study, where critical and routine accessibility are calculated separately, the SAW score is

$$S_j = w_1\, r_{\mathrm{eff},j} + w_2\, r_{\mathrm{crit},j} + w_3\, r_{\mathrm{rout},j} + w_4\, r_{\mathrm{equity},j} + w_5\, r_{\mathrm{conn},j}, \tag{3.23}$$

where $r_{\mathrm{crit},j}$ and $r_{\mathrm{rout},j}$ are the separate critical- and routine-accessibility scores.

Table 3.6: Mapping between BWM survey criteria and simulation performance metrics

BWM criterion	Simulation metric	Measurement
Operational efficiency	Operational efficiency	Inverse-shortest-path network efficiency
Access to critical services Access to routine activities	Accessibility	Closeness centrality-weighted reachability based on shortest-path travel times
Spatial equity	Spatial equity	Spatial autocorrelation, measured with Moran's I, of node-level accessibility
Connectivity	Connectivity	Average local directed clustering coefficient

Ranking and Tie Handling

After criterion mapping, the stakeholder weights are combined with the normalised simulation scores using the SAW formulation in Equation (3.20). For each decision matrix, the composite score S_j is calculated for every recovery strategy, and strategies are ranked in descending order of S_j. The highest score indicates the best overall stakeholder-weighted performance within that city, damage level, recovery stage, and stakeholder-weight configuration.

Tie handling follows the purpose of each analysis. When two or more strategies receive the same highest composite score, the case is treated as a tied rank-one outcome. In descriptive summaries of rank-one distributions, these ties are retained and reported as tied outcomes (Section 4.4). In analyses that require a unique top-ranked strategy, such as within-city stakeholder agreement and short–long stage switching, tied rank-one cases are excluded because no unambiguous preferred strategy can be assigned. Cases in which all six strategies receive identical composite scores are treated as non-differentiated cases and are also excluded from agreement-based comparisons. The number of excluded cases is reported in the relevant results sections.

3.7. Implementation of the cross-city and Delft applications

This section describes how the framework is implemented at the two scales. The cross-city implementation applies the simulation-processing procedure to a dataset covering many city road networks. The Delft implementation applies the same framework to a single urban network using more detailed facility-based accessibility and a set of supplementary local checks.

3.7.1. Cross-City System-Level Implementation

The cross-city analysis applies the simulation-processing procedure described above to the road-network recovery dataset of Balakrishnan and Stokkink [1]. The original dataset contains simulated recovery trajectories for 224 city and town road networks, damage levels, recovery strategies, and performance metrics.

In this study, one city is excluded because complete valid performance data are not available for all required combinations of strategy, damage level, criterion, and recovery stage. The final simulation-side analysis therefore uses 223 cities.

The main analysis uses the 10% and 30% damage scenarios. The 10% scenario is used as the lower-damage case because it is close to the empirical reference used for the recovery-stage mapping. Aydin et al. [64] report a landslide-induced road disruption case in which approximately 12% of road segments were disrupted, and restoring 25% of the disrupted segments under a moderate recovery strategy took about six days. The 30% scenario is included as the higher-disruption case because it falls within the moderate-disruption range where recovery-strategy effects are expected to become more visible. Song et al. [66] examine damage levels from 5% to 50% and report an inflection in the recovery curves around the 30%–40% damage range, where differences between recovery strategies become more pronounced. Together, the two scenarios allow the ranking procedure to be examined under a lower-disruption case connected to an empirical recovery reference and a higher-disruption case in which strategy-dependent recovery patterns are more likely to emerge.

For each city, damage level, and recovery strategy, the simulation output is based on 20 random damage iterations. In each iteration, a new set of damaged links is generated at the specified damage level, and the six recovery strategies are evaluated on comparable disrupted network states. The AAC deviation score is first calculated separately for each iteration and is then averaged across the 20 iterations. The resulting mean deviation score is used as the simulation-side performance value in the SAW decision matrix.

The resulting cross-city simulation output consists of four criterion-level deviation scores for each strategy: operational efficiency, accessibility, spatial equity, and connectivity. These scores form the simulation-side input to the SAW decision matrices described in Section 3.6.

3.7.2. Delft Facility-Based Implementation

The Delft case study applies the same simulation-processing logic as the cross-city analysis to a locally constructed Delft road network. The purpose is to test how the framework behaves when the aggregate accessibility representation used in the cross-city analysis is replaced by facility-based accessibility measures. In the cross-city analysis, access to critical services and access to routine activities are both represented by the same aggregate accessibility metric. In the Delft case study, these two criteria are calculated separately using local facility destinations.

The Delft implementation consists of six steps:

1. **Construct the local road-network model.** The Delft road network is derived from OpenStreetMap using the `OSMnx` package [67]. The drivable network is extracted with the `drive` filter, simplified topologically, and restricted to the largest connected component. The network is projected to a metric coordinate reference system and converted to a directed `igraph` object. Each edge is assigned a travel time based on edge length and speed. When an OSM `maxspeed` value is unavailable, speed is imputed from other edges of the same road type, with a fallback speed of $30\,\mathrm{km\,h^{-1}}$. Travel time is used as the edge weight in shortest-path calculations. The final Delft road network contains 2441 nodes and 5717 directed edges.
2. **Define facility destination sets.** Facility locations are extracted from OpenStreetMap and linked to their nearest road-network nodes. Two destination sets are constructed to match the two accessibility criteria used in the stakeholder survey. The critical-facility set represents services that support emergency response and basic post-disaster needs. It includes police stations, fire stations, hospitals, clinics, pharmacies, supermarkets, marketplaces, drinking-water points, and wells[68], [69], [70]. The routine-facility set represents non-emergency everyday activities, including schools and universities, workplaces represented by OSM `office` features, and retail represented by OSM `shop` features. Supermarkets are excluded from the routine set because they are already included in the critical-facility set. If multiple facilities from the same destination set are linked to the same road-network node, that node is counted once within that set. In Delft, 65 critical facilities are linked to 58 distinct road-network nodes, and 895 routine facilities are linked to 364 distinct road-network nodes. 32 nodes appear in both destination sets.
3. **Generate damage scenarios.** Damage is imposed by removing a fixed share of directed road

edges. The two main damage levels are $10\,\%$ and $30\,\%$. For each damage level, 20 random damage realisations are generated with a fixed random seed and reused across all recovery strategies.

4. **Apply recovery strategies.** Recovery proceeds in discrete steps of $10\,\%$ of the removed edge set. At each step, damaged edges are reinstated according to the priority order defined by the recovery strategy. The same six strategies as in the cross-city analysis are used, and the random recovery strategy also uses a fixed seed for reproducibility.
5. **Recalculate performance metrics.** Operational efficiency, spatial equity, and connectivity use the same graph-based definitions as in the cross-city analysis. Accessibility is calculated separately for the two Delft destination sets, producing one trajectory for access to critical facilities and one trajectory for access to routine facilities. Every road-network node is treated as an origin. Only the selected facility nodes are used as destinations, although other network nodes may still appear as intermediate nodes along shortest paths. Destination weights are fixed using their closeness centrality in the undisrupted Delft network.
6. **Construct Delft SAW inputs.** Each Delft recovery trajectory is converted into short-term and long-term AAC deviation scores using the procedure described in Section 3.5.4. The boundary between the short-term and long-term recovery stages is set at $\theta = 25\,\%$ of restoration progress, consistent with Section 3.2.4. The per-iteration deviation scores are averaged across the 20 damage realisations. The resulting matrix contains five criterion-level deviation scores for each recovery strategy, damage level, and recovery stage. These scores are then combined using the SAW procedure described in Section 3.6.

3.7.3. Delft Supplementary Local Checks

The Delft case study also includes three supplementary checks that use the local Delft network to interpret the facility-based SAW ranking. These checks are not additional criteria in the SAW calculation. Instead, they examine whether the ranking can be explained spatially, whether the centrality-based restoration orders are sensitive to the selected network boundary, and whether the system-level ranking is consistent with local critical-service access from a road-user perspective.

- **Edge-priority maps.** The edge-priority maps visualise the restoration order implied by each recovery strategy across the Delft road network. They show which parts of the network each strategy prioritises early in recovery, but they are not used as performance scores.
- **Boundary sensitivity analysis.** The boundary sensitivity analysis tests whether restoration priority patterns change when surrounding road-network context is added through buffer zones. This check is included because centrality-based restoration orders can be affected by the spatial boundary used to define the network. The detailed design is described in Section 3.8.5.
- **Local critical-service travel-time test.** The local travel-time test examines access from selected residential origins to the nearest reachable emergency, medical, food, and water services. It provides a road-user-side check on the two highest-ranked Delft strategies, but it does not replace the system-level SAW ranking. The setup of this test, including the selected origins, service groups, buffer network, and reachability calculation, is described in Section 3.7.3.

Local critical-service travel-time test

The local critical-service travel-time test provides a supplementary road-user-side check of the Delft SAW ranking. The test focuses on the two highest-ranked Delft strategies, betweenness and closeness, and examines whether they produce different travel-time outcomes from selected residential origins to nearby critical services. This test is not included as an additional SAW criterion and does not replace the system-level ranking.

The test uses the Delft plus $3\,\mathrm{km}$ buffer network selected from the boundary sensitivity analysis. Residential origins are kept inside the Delft municipal boundary, while reachable critical-service facilities within the selected buffer are included as possible destinations. This reflects a road-user perspective, because residents may use nearby services outside the municipal boundary if these services are reachable through the road network.

Thirty residential origins are selected from residential building locations inside Delft. Residential building features are converted to centroids and matched to their nearest road-network nodes. Duplicate matched nodes are removed, and only candidate road nodes inside the original Delft municipality are retained. From these candidates, 30 origins are selected using a greedy farthest-point method with a fixed random seed. This produces a spatially balanced set of origins across Delft, rather than concentrating the test in areas with many residential buildings.

Critical-service facilities are grouped into four service categories. The emergency group includes fire stations and police facilities. The medical group includes hospitals, clinics, and pharmacies. The food group includes supermarkets and marketplaces. The water group includes drinking-water points and wells. These facilities are extracted within the Delft plus $3\,\mathrm{km}$ buffer area and matched to their nearest road-network nodes.

For each residential origin and each service group, the travel time is calculated as the minimum shortest-path travel time to any reachable facility in that group. If no facility in the service group can be reached from the origin under a damaged or partially restored network, the case is recorded as unreachable. Median travel times are calculated only over reachable observations, while unreachable observations are reported separately as the unreachable share.

The test follows the same $10\,\%$ and $30\,\%$ damage levels used in the Delft simulation. The corresponding damaged networks and recovery sequences are reused, and travel-time outcomes are evaluated for the betweenness and closeness strategies. The test should be interpreted as a nearest-service travel-time analysis, not as a fixed-destination origin-destination analysis, because the nearest reachable facility may change during recovery.

3.8. Sensitivity and Robustness Design

The methodology described in the previous sections relies on several design choices whose influence on the results is examined here.

- **Operational-efficiency criterion interpretation.** The main analysis retains all valid BWM responses, including responses in which the open-ended explanation suggests a possible individual-travel-time interpretation of operational efficiency. A temporary exclusion of these flagged responses is tested in Section 3.8.1.
- **Stakeholder-group importance weights λ_g.** The main analysis uses equal weighting, $\lambda_g = 1/G$. Alternative schemes that emphasise specific stakeholder groups are tested in Section 3.8.2.
- **Short-term and long-term window boundary.** The main analysis uses a 25% restoration-progress boundary; a range of boundaries $\theta \in \{15\%, 20\%, 25\%, 30\%, 35\%\}$ is tested in Section 3.8.3.
- **BWM configuration.** The main analysis uses the *linear–broader* configuration; the remaining configurations from Table 3.4 are compared in Section 3.8.4.
- **Delft network boundary.** The Delft-only network is compared with buffered network extents to examine whether centrality-based restoration priority orders depend on the spatial boundary used to define the road network (Section 3.8.5).

Each check produces an alternative set of outputs for the part of the analysis it affects, such as stakeholder weights, composite scores, or strategy rankings. These outputs are then compared with the corresponding main-analysis results.

3.8.1. Robustness check for operational-efficiency wording

The operational-efficiency criterion (C1) was intended to represent network-level travel performance, consistent with the simulation metric based on inverse shortest-path travel times. The Mandarin wording used in the questionnaire referred to travel time, which may have led some respondents to interpret C1 as their own individual trip time.

To check whether this wording ambiguity affects the results, the open-ended explanations were reviewed to identify responses that clearly reflected an individual-trip-time interpretation. These responses were flagged and retained in the main analysis. As a robustness check, the stakeholder-group weights

were recomputed for the affected groups and recovery stages after temporarily excluding the flagged responses. The SAW rankings were then recalculated using these alternative weights and compared with the main-analysis rankings.

This check assesses whether the main strategy-ranking results depend on the inclusion of responses affected by the C1 wording ambiguity.

3.8.2. Sensitivity to stakeholder-group importance weights

The baseline analysis uses equal stakeholder-group weighting, $\lambda_g = 1/G$, as motivated in Section 3.3.3. This provides a transparent reference case, but it assumes that all stakeholder groups have the same importance in the aggregation. To test whether the overall strategy rankings depend on this assumption, the stakeholder-group aggregation is repeated under alternative group-weighting schemes.

In each alternative scheme, one stakeholder group with more than one valid response is assigned a larger aggregation weight. Groups with only one valid response are not assigned a dominant role, because increasing the weight of a single respondent would not provide a meaningful group-level sensitivity check. However, these groups remain included in the residual aggregation weight. When one group is assigned a dominance share p, the remaining weight $1 - p$ is distributed equally among the other stakeholder groups.

The dominance share p is varied over 0.2, 0.3, 0.4, 0.5, and 0.6. For each value of p, the overall criterion weights are recomputed and the SAW ranking is repeated for both recovery stages and for the 10% and 30% damage levels. The resulting rankings are then compared with the equal-group baseline.

3.8.3. Sensitivity to the short-term and long-term window boundary

The main analysis sets the boundary between the short-term and long-term windows at 25% restoration progress. To assess whether the results depend on this specific boundary, the deviation score calculation defined in equation (3.19) is repeated across a range of alternative boundaries, $\theta \in \{15\%, 20\%, 25\%, 30\%, 35\%\}$. For each value of θ, the short-term window is defined as 0% to θ restoration and the long-term window as θ to 100% restoration. The resulting clipped AAC deviation scores are then averaged by damage level, recovery stage, criterion, and strategy. These mean deviation scores are compared across θ values to assess whether the relative performance patterns are robust to the choice of recovery-stage boundary. This sensitivity check focuses on the criterion-level deviation scores rather than the final SAW ranks, because changing θ directly changes the recovery windows over which the AAC scores are computed.

3.8.4. Robustness comparisons across BWM configurations

As described in Section 3.3.3, criteria weights are derived under four BWM configurations: *linear–broader*, *linear–strict*, *equivalent–broader*, and *equivalent–strict*. The *linear–broader* configuration is the primary basis for the main analysis, while the other three are reported as complementary robustness analyses. Three pairwise comparisons assess the robustness of the derived criterion weights, each isolating one specific methodological choice.

Comparison 1: linear–broader versus linear–strict. This comparison isolates the effect of the consistency screening procedure under the linear BWM formulation. Both configurations use the same linear BWM formulation, so differences in the derived criterion weights reflect the exclusion of Grade C responses from the strict sample. If the criterion weight profiles are similar, the derived weights are not materially affected by excluding responses that required larger corrections to satisfy the consistency condition. A substantial divergence would instead suggest that including or excluding Grade C responses meaningfully affects the derived weights.

Comparison 2: linear–broader versus equivalent–broader. This comparison isolates the effect of the BWM formulation under the broader sample. Both configurations use the same set of valid responses, so differences in the derived criterion weights reflect the choice between the linear BWM formulation and the equivalent BWM formulation. If the criterion weight profiles are similar, the derived weights are insensitive to the choice of formulation. A substantial divergence would instead suggest that the formulation meaningfully affects the derived weights.

Comparison 3: equivalent–broader versus equivalent–strict. This comparison provides a second check on the consistency screening procedure, this time under the equivalent BWM formulation. Comparing it with Comparison 1 indicates whether the effect of the screening procedure is itself sensitive to the BWM formulation. If the weight profiles are similar in both Comparison 1 and Comparison 3, the screening procedure has a comparable effect under both formulations. If the two disagree, the impact of the screening procedure depends on the formulation and should be reported with this qualification.

For these robustness comparisons, the weights are compared at the all-respondent level rather than at the stakeholder-group level. For each configuration, individual criterion weights are first computed from each valid respondent's BWM responses, separately for the short-term and long-term recovery stages. The individual weights are then aggregated directly across all valid respondents using the geometric mean and renormalised to produce one overall criterion weight vector per configuration and recovery stage. This differs from the main analysis, where respondent weights are first aggregated within stakeholder groups and then combined using stakeholder-group importance weights. The all-respondent comparison does not preserve stakeholder-group structure, but it provides a robustness check that is not affected by uneven or small stakeholder-group sample sizes. The SAW rankings are then recomputed using the all-respondent weight vector for each configuration and compared with the main-analysis rankings.

3.8.5. Delft boundary sensitivity analysis

The boundary sensitivity analysis examines whether the Delft restoration priority patterns are sensitive to the spatial extent of the network model. This check is included because network centrality measures can be affected by the size and position of the model boundary [71]. This issue is relevant for the Delft case study because several recovery strategies use centrality-based priority rules.

The Delft municipality is treated as the core study area. Three buffered networks are created by adding $1\,\mathrm{km}$, $3\,\mathrm{km}$, and $5\,\mathrm{km}$ buffers around the Delft boundary. Together with the Delft-only network, these produce four network extents for comparison.

For each network extent, the relevant centrality scores and restoration priority orders are calculated on the full network, including the buffer area. The comparison is then restricted to road edges located inside the original Delft municipality. The buffer therefore provides computational context, while the evaluated restoration order remains focused on Delft core edges.

To make the maps comparable across network extents, the restoration priority bands are assigned after ranking only the Delft core edges within each extent. This means that each map shows the relative restoration priority of roads inside Delft, while still allowing surrounding roads to influence the centrality calculation. The results of this check are reported in Section 5.5.

4

Stakeholder Priorities and Cross-City Strategy Rankings

This chapter presents the results of the main analysis. It first describes the survey response profile and the construction of the valid analysis samples. It then reports the stakeholder priorities derived from the Best–Worst Method (BWM), including both the criterion weights and the reasoning expressed in the open-ended explanations. These stakeholder priorities are subsequently combined with simulated recovery performance through Simple Additive Weighting (SAW) to rank recovery strategies across many city road networks. The chapter closes with comparisons across stakeholder groups and recovery stages, followed by sensitivity and robustness results.

4.1. Survey response profile and analysis samples

This section reports the survey responses used in the stakeholder-priority analysis. It describes the filtering process from collected responses to valid BWM analysis samples, and summarises the distribution of respondents by stakeholder role, country of practice, age group, and distribution channel.

4.1.1. Response filtering and valid BWM samples

A total of 115 survey responses were collected. Of these, 61 responses were completed and retained for respondent-profile analysis. The BWM judgement data were then screened to identify responses that could be converted into valid criterion weights. This produced 56 valid broader BWM responses. The stricter BWM samples used for robustness checks are reported where those robustness results are presented. The response filtering process is summarised in Figure 4.1.

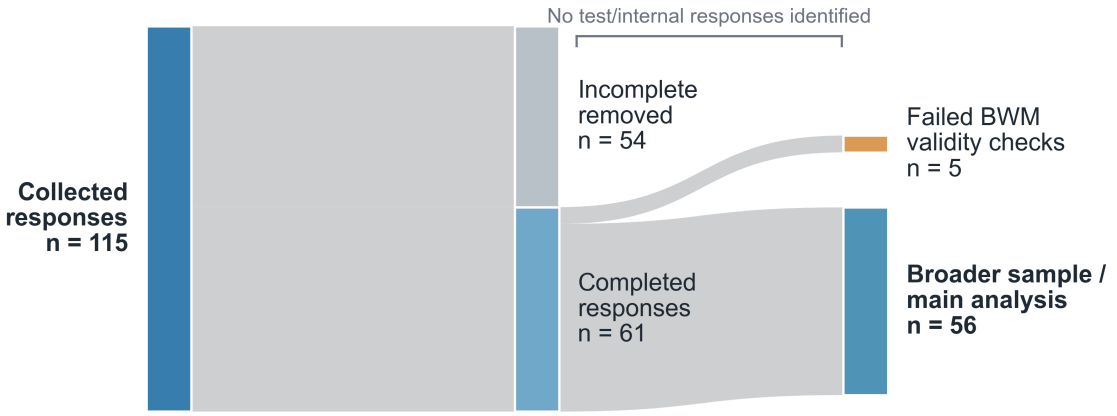


Figure 4.1: Filtering of survey responses from collection to the broader analysis sample. Of 115 collected responses, 54 incomplete responses were removed and no test or internal responses were identified, leaving 61 completed responses. A further 5 responses failed the BWM validity checks, giving a broader sample of 56 responses used as the basis for the main analysis.

4.1.2. Respondent distribution by stakeholder group

The completed survey sample is dominated by road users and local community respondents, which together account for more than two thirds of the completed responses. Technical experts and researchers form the third largest group, while emergency responders, transport and logistics operators, public authorities, infrastructure managers, and other respondents are represented by smaller exploratory samples. The distribution by stakeholder role type is shown in Table 4.1.

Table 4.1: Respondent distribution by stakeholder role type

Stakeholder role type	Respondents	Percentage
Road users significantly affected by road recovery	22	36.1%
Local community	19	31.1%
Technical expert / researcher	9	14.8%
Emergency responder	3	4.9%
Transport / logistics operator	3	4.9%
Public authority / government agency	3	4.9%
Infrastructure manager	1	1.6%
Other	1	1.6%
Total	61	100.0%

4.1.3. Respondent background and distribution channels

The respondents are predominantly based in Taiwan (40 responses, 66%), with smaller numbers from the Netherlands (6), China (5), and several other countries. In terms of age, the sample is concentrated in the middle and older age bands, with most respondents aged between 45 and 64. This geographic concentration reflects the distribution channels used, as most responses were collected through local community groups. The demographic distributions are summarised in Tables 4.2 and 4.3, and the distribution by distribution channel is given in Table 4.4.

Table 4.2: Distribution by country of practice

Country	n	%
Taiwan	40	65.6
Netherlands	6	9.8
China	5	8.2
United Kingdom	2	3.3
United States	2	3.3
Other	6	9.8
Total	61	100.0

Table 4.3: Distribution by age group

Age group	n	%
18–24	10	16.4
25–34	13	21.3
35–44	4	6.6
45–54	16	26.2
55–64	17	27.9
65 or above	1	1.6
Total	61	100.0

Table 4.4: Distribution by distribution channel

Distribution channel	Respondents	Percentage
Community groups	48	78.7%
Professional network	10	16.4%
Professional email	2	3.3%
Academic email	1	1.6%
Total	61	100.0%

4.2. Stakeholder priorities across recovery stages

This section addresses the first sub-question by presenting the stage-specific stakeholder priorities obtained from the BWM survey. It compares how stakeholders prioritise the five resilience criteria in the short-term and long-term recovery stages, and it uses the open-ended survey explanations to interpret the main patterns in the elicited weights. These weights are later combined with simulated strategy performance in the SAW step.

All weights reported in this section are group-level *linear-broader* weights, computed from the broader sample of 56 responses that pass the BWM validity screening (Section 3.3.2). This sample is smaller than the 61 completed questionnaires summarised in the response profile because responses failing the validity checks are excluded from weight estimation. Since the screening is applied separately to the short-term and long-term comparisons, the number of retained responses differs slightly between stages within some groups, as reflected in the per-stage sample sizes.

4.2.1. Overall patterns across stakeholder groups

Across the stakeholder groups, C2 receives the highest weight in most groups during the short-term recovery stage, showing the strong role of access to critical facilities in this stage. In the long-term recovery stage, the weights are more evenly distributed across the five criteria. Although C2 remains important in several groups, the relative weights of operational efficiency, access to routine facilities, spatial equity, and connectivity increase compared with the short-term stage. These overall patterns are reported in Table 4.5 and are followed by a more detailed stakeholder-specific comparison in the next subsection.

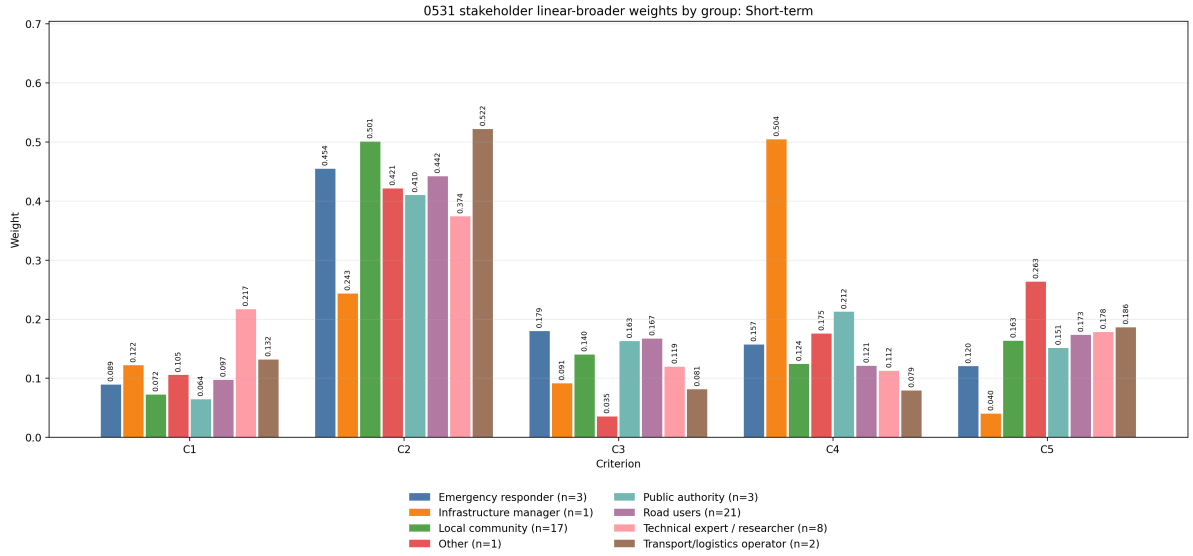


Figure 4.2: Group-level criterion weights for all eight stakeholder groups in the short-term stage (linear-broader configuration). Value labels give each weight; group sample sizes are shown in the legend. Because the BWM screening is applied separately by stage, the per-group sample sizes differ slightly from Figure 4.3.

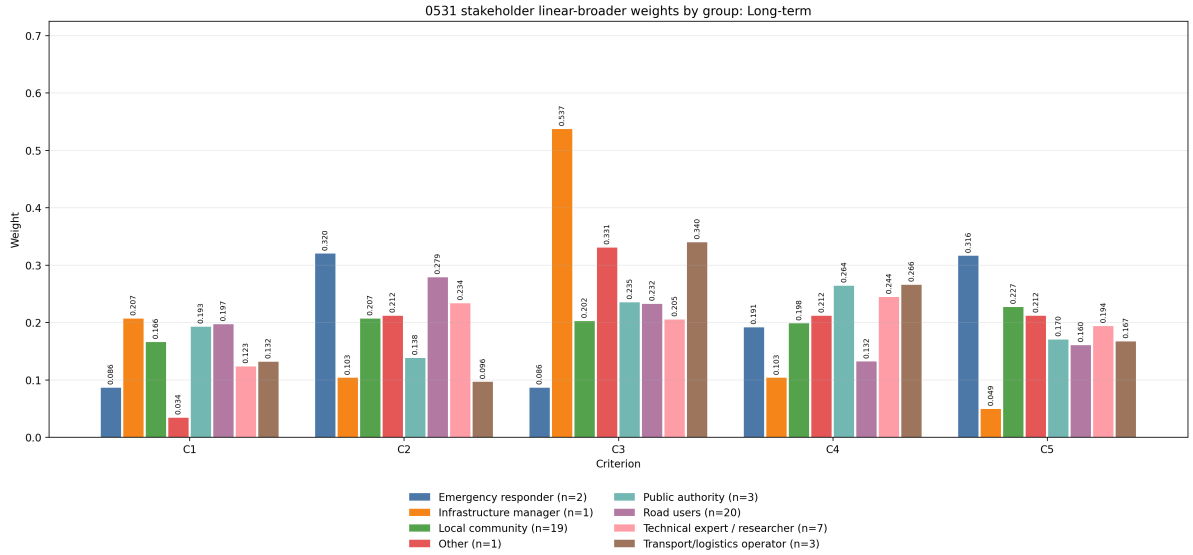


Figure 4.3: Group-level criterion weights for all eight stakeholder groups in the long-term stage (linear-broader configuration). Value labels give each weight; group sample sizes are shown in the legend. The single-respondent infrastructure-manager and *other* groups produce the extreme single-criterion values noted in the text.

These group-level weights must be interpreted with the underlying sample sizes in mind. Five of the eight stakeholder groups are based on three or fewer responses, and two groups, the infrastructure manager and *other* groups, contain only one respondent each. Their weights therefore reflect individual preferences rather than robust group-level patterns. This helps explain some of the most extreme values in Figure 4.3, such as the infrastructure manager's long-term routine-access weight of 0.537. For this reason, the detailed group-level interpretation focuses on the three stakeholder groups with sufficient sample sizes: road users, the local community, and technical experts or researchers (Section 3.2.5). The remaining groups are reported for completeness and treated as exploratory.

Table 4.5: Group-level criterion weights for the three main stakeholder groups (linear-broader configuration). The largest weight in each row is shown in bold. Weights for all eight groups, including the five exploratory groups, are shown in Figures 4.2 and 4.3. C1: operational efficiency; C2: accessibility to critical services; C3: accessibility to routine activities; C4: spatial equity; C5: connectivity.

Stage	Group	C1	C2	C3	C4	C5
Short-term	Road users ($n = 21$)	0.097	**0.442**	0.167	0.121	0.173
Short-term	Local community ($n = 17$)	0.072	**0.501**	0.140	0.124	0.163
Short-term	Technical experts ($n = 8$)	0.217	**0.374**	0.119	0.112	0.178
Long-term	Road users ($n = 20$)	0.197	**0.279**	0.232	0.132	0.160
Long-term	Local community ($n = 19$)	0.166	0.207	0.202	0.198	**0.227**
Long-term	Technical experts ($n = 7$)	0.123	0.234	0.205	**0.244**	0.194

4.2.2. Stage-specific differences among the main stakeholder groups

Among the three main stakeholder groups, C2, access to critical facilities, receives the highest short-term weight in all groups. The C2 weight is 0.442 for road users, 0.501 for the local community, and 0.374 for technical experts. For road users and the local community, this value is more than twice as large as any other criterion. For technical experts, C2 is also the highest-weighted criterion, although the gap between C2 and C1, operational efficiency, is smaller.

In the long-term stage, the C2 weight decreases in all three main groups, falling to 0.279 for road users, 0.207 for the local community, and 0.234 for technical experts. The long-term weight distributions are also flatter than in the short-term stage. This is most visible for the local community, whose long-term weights range from 0.166 to 0.227 across the five criteria.

The three main groups differ in their highest-weighted long-term criterion. Road users continue to assign the highest weight to C2 (0.279), while C3, access to routine facilities, becomes the second-highest criterion (0.232). The local community assigns its highest long-term weight to C5, connectivity (0.227), although the difference from C2, C3, and C4 is small. Technical experts assign their highest long-term weight to C4, spatial equity (0.244), followed by C2 (0.234).

Two group-level contrasts are particularly relevant. First, technical experts assign a higher short-term weight to C1 than road users and the local community. For technical experts, C1 is the second-largest short-term weight (0.217), compared with 0.097 for road users and 0.072 for the local community.

Second, the long-term weight assigned to C4 differs across the three main groups. Technical experts assign C4 their highest long-term weight (0.244), while road users assign C4 their lowest long-term weight (0.132). The local community lies between these two groups, with a long-term C4 weight of 0.198. The implications of this contrast are discussed further in the discussion chapter.

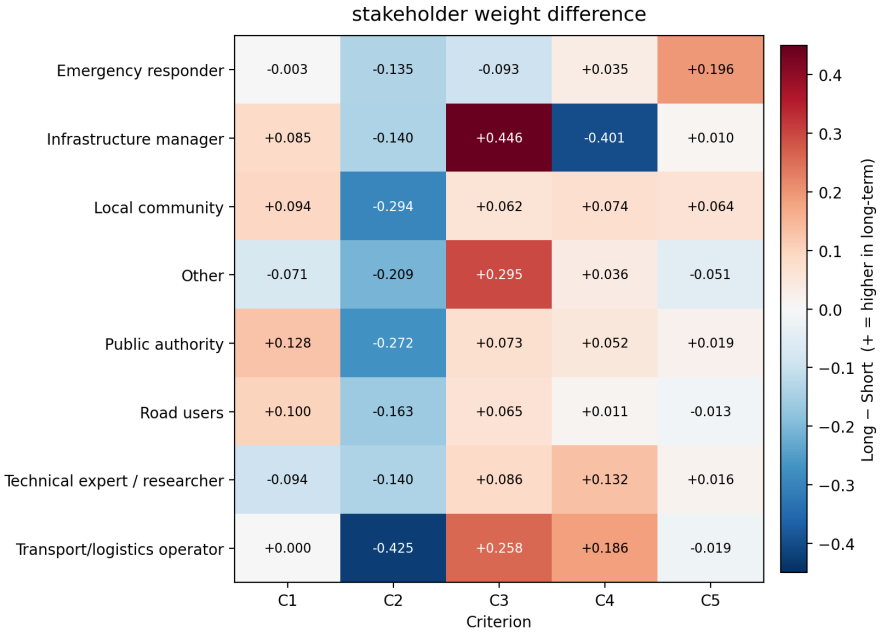


Figure 4.4: Stakeholder criteria weight differences by group (long-term minus short-term).

4.2.3. Respondent-level priority changes

The group-level comparison shows how average criterion weights differ between the short-term and long-term recovery stages. However, this aggregate comparison does not show whether respondents retained a largely fixed priority list, shifted their priorities sequentially, or reconsidered the ranking more broadly. To examine this, a paired respondent-level comparison was conducted using only respondents with valid linear BWM weights in both recovery stages.

Of the 61 completed respondents, 56 had valid short-term linear BWM weights and 56 had valid long-term linear BWM weights. These two valid sets were not identical. Five respondents were valid only in the short-term stage, and another five were valid only in the long-term stage. The paired respondent-level analysis is therefore based on 51 respondents.

For each respondent, the five criteria were ranked from most to least important, separately for the short-term and long-term stages. Where two or more criteria shared the same weight, the tied positions were handled using fractional assignment across the tied ranks. Figure 4.5 shows, for each short-term rank, how the corresponding criteria are distributed across the long-term ranks, expressed as row-normalised percentages. For example, of the criteria ranked first in the short-term stage, only 15.7% stayed first in the long-term stage, while 27.0% moved to second, 20.1% to third, and so on across the row. Each row sums to 100%, apart from small rounding differences. Because each rank carries the same total fractional weight in both stages, the columns also sum to approximately 100%, allowing the long-term ranks to be traced back to their short-term positions.

Two simple explanations can be examined using the transition matrix. First, respondents may have one fixed priority list that remains largely unchanged across the two recovery stages. If so, the percentages should be concentrated along the diagonal. Second, the short-term ranking may work like a to-do list. Once the first short-term priority is set aside, the second-ranked criterion would become the first long-term priority, and the remaining criteria would move up in the same way.

The matrix does not strongly support either explanation. Only 15.7% of the criteria ranked first in the short term remained first in the long term, and unchanged ranks were not consistently the most common transitions across the rows. Support for the to-do-list explanation is also limited. Only 17.2% of the long-term first-ranked cases came from criteria ranked second in the short term, while 67.2% came from criteria previously ranked third, fourth, or fifth.

These results suggest that respondents did not simply keep one fixed ranking or move the remaining

priorities up by one position. Instead, they reconsidered the relative importance of several criteria when moving from short-term to long-term recovery. This provides respondent-level evidence that recovery priorities are stage-dependent.

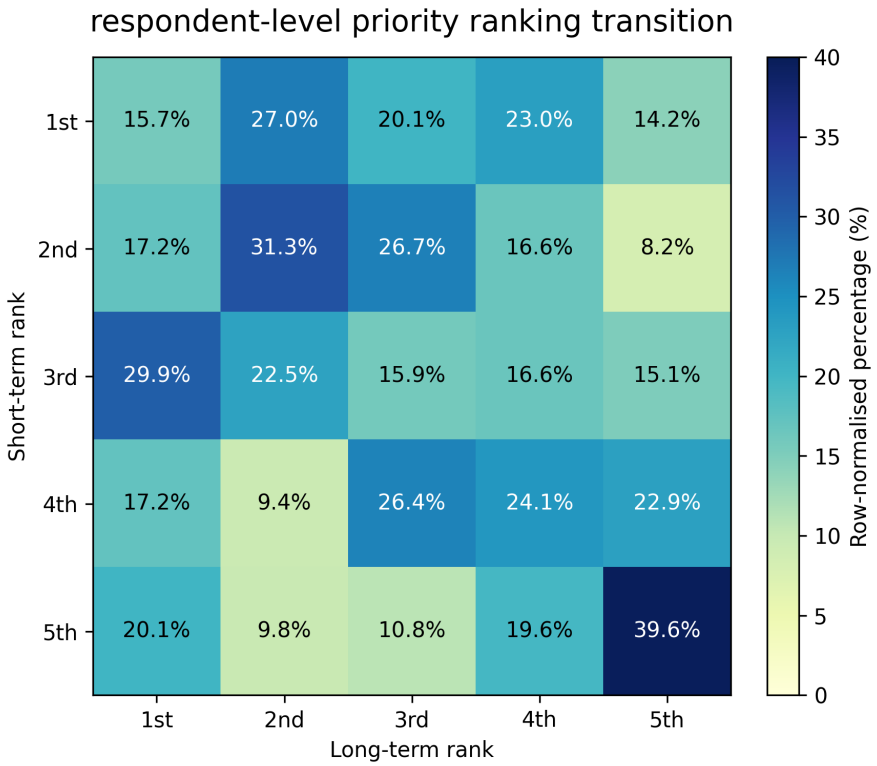


Figure 4.5: Transition matrix of respondent-level priority rankings from the short-term to the long-term recovery stage. Each row shows how criteria assigned to a given short-term rank are distributed across long-term ranks. Darker shading indicates a larger row-normalised percentage.

4.3. Reasoning behind stakeholder priorities

This section reports the open-ended explanations provided by respondents for their most-important and least-important criteria. The coding procedure is described in the methodology chapter; here, the focus is on the resulting themes and how they help interpret the elicited BWM weights. An overview of the identified themes and their frequencies across the short-term and long-term recovery windows is presented in Table 4.6.

Table 4.6: Themes identified in the open-ended criterion rationales

Code	Theme	Count
Short-term window: most-important rationale ($n = 61$)		
M1	Life and basic needs	45
M2	Emergency response and rescue access	7
M4	Return to normal daily life	4
M3	Connectivity and isolation prevention	3
M5	Network efficiency and public benefit	1
M6	Equity and inclusive recovery	1
Short-term window: least-important rationale ($n = 55$)		
L1	Lower priority than survival or basic needs	18
L5	Equity is difficult, vague, or lower-benefit	12
L3	Acceptable delay or lower urgency	9
L10	Ambiguous or subjective rationale	6
L2	Minimum access is enough; efficiency can wait	6
L6	Efficiency or travel-condition improvements are lower priority	1
L4	Alternative access or redundancy exists	1
L7	Limited population or wider impact	1
L8	Already handled in another phase	1
Long-term window: most-important rationale ($n = 61$)		
M4	Return to normal daily life	21
M5	Network efficiency and public benefit	10
M1	Life and basic needs	10
M3	Connectivity and isolation prevention	8
M6	Equity and inclusive recovery	6
M7	Resilience and long-term preparedness	3
M8	Ambiguous or unclear most-important rationale	2
M2	Emergency response and rescue access	1
Long-term window: least-important rationale ($n = 58$)		
L5	Equity is difficult, vague, or lower-benefit	12
L3	Acceptable delay or lower urgency	11
L8	Already handled in another phase	7
L2	Minimum access is enough; efficiency can wait	7
L6	Efficiency or travel-condition improvements are lower priority	6
L10	Ambiguous or subjective rationale	5
L1	Lower priority than survival or basic needs	4
L11	Lower priority than returning to normal life	2
L9	Context or feasibility constraint	2
L4	Alternative access or redundancy exists	1
L7	Limited population or wider impact	1

Each response was assigned to a single theme, using separate sets of codes for the most-important and least-important rationales. The themes were not fixed in advance but were developed from the responses themselves. An initial set of themes and assignments was drafted with the help of a large language model, and all assignments were then reviewed by the author against the original responses. The least-important rationale counts exclude responses flagged for the C1 wording issue. These responses are examined separately in the robustness check. The coding is exploratory and is intended to help interpret the elicited weights.

4.3.1. Themes in most-important criterion explanations

The themes identified in the most-important criterion explanations show how respondents justified their priorities in the short-term and long-term recovery windows. In the short-term window, life and basic needs was the dominant theme (M1, 45 of 61). Most of these responses were associated with access to critical facilities (C2, 42 of 45). The comments referred to basic physiological needs, food, water, medicine, hospital access, injured disaster victims, and maintaining a minimum living standard during the immediate recovery period. This indicates that respondents mainly understood short-term recovery through the need to secure essential access for survival and basic welfare.

A further pattern is that emergency response and rescue access appeared as a smaller short-term theme (M2, 7 of 61). Although this theme is closely related to the short-term disaster context, only a limited number of respondents explicitly mentioned emergency response, rescue operations, or reaching affected people. This pattern suggests that respondents more often described the goals of short-term recovery than the actions used to achieve them. Most respondents described what needed to be secured in the short term, such as life, basic needs, medical access, and survival. Fewer respondents described the specific actions used to secure these needs.

Alongside these two themes, the absence of an ambiguous or unclear theme in the short-term most-important rationales is also informative. It suggests that respondents found the most-important short-term choice relatively easy to justify. In this stage, the reasoning was strongly centred on immediate survival and essential access, which made the priority logic relatively explicit.

The long-term window shows a different pattern. The most frequent theme was return to normal daily life (M4, 21 of 61), compared with only 4 occurrences in the short-term window. The comments referred to restoring normal living order, returning to daily activity locations, supporting daily travel, and returning to pre-disaster conditions. M4 appeared mainly in explanations for access to routine facilities (C3, 13 responses), but it also appeared in explanations for operational efficiency (C1), spatial equity (C4), and connectivity (C5).

This shift from M1 to M4 as the dominant theme reflects a broader change in how respondents explained their priorities. Short-term priorities were mainly justified through access to critical facilities and basic survival needs, whereas long-term priorities reflected a wider concern with daily routines, network performance, connectivity, and fairness in recovery.

4.3.2. Themes in least-important criterion explanations

The patterns for the least-important criteria were interpreted from a slightly different angle, by first identifying which criteria were most often selected as least important and then examining the themes used to justify these choices.

Operational efficiency was frequently selected as the least-important criterion. In the short-term window, 28 respondents selected operational efficiency (C1) as the least-important criterion, and in the long-term window this number was 26. The explanations commonly referred to basic access being sufficient, efficiency being less urgent than access to critical facilities, or improvements in travel conditions being something that could wait. These comments describe C1 as secondary to maintaining minimum accessibility during recovery. A subset of C1-related comments also reflected the wording issue around the Mandarin term for travel time; these responses are not interpreted further here and are examined separately in the robustness check in Section 4.6.1.

For the reasons given for the least-important criterion, spatial equity was often described as difficult, vague, or lower-benefit (L5). This theme appeared 12 times in the short-term window and 12 times in the long-term window, mostly among respondents who selected spatial equity (C4) as the least-important criterion. Across the two stages, respondents referred to urgent needs, regional differences, population density, economic value, construction difficulty, different levels of disaster severity, and the difficulty of achieving complete fairness. Some comments described fairness as perspective-dependent or less urgent.

In the long-term window, another recurring reason was that some functions could be delayed or restored gradually. Acceptable delay or lower urgency (L3, 11 of 61) appeared mainly among respondents who selected connectivity (C5), operational efficiency (C1), or access to routine facilities (C3) as the least-

important long-term criterion. For connectivity, respondents described full network reconnection or broader connectivity improvements as something that could be restored gradually rather than immediately. For operational efficiency, the comments treated efficiency improvements as less urgent than maintaining access to necessary activities. For access to routine facilities, respondents indicated that some daily activities could be restored later in the recovery process.

4.4. Cross-city strategy ranking results

This section reports the SAW-based strategy rankings for the 223 cities included in the final simulation-side analysis. Building on the city-network description in Section 4.4.1, the analysis compares how the six recovery strategies perform across cities under different stakeholder-weight scenarios, damage levels, and recovery stages. For each city, stakeholder group, damage level, and recovery stage, the strategies are scored using the stage-specific BWM weights and the simulated recovery-performance scores, and are then ranked from highest to lowest overall score.

4.4.1. Description of the analysed city networks

Before presenting the results, this subsection briefly describes the road-network sample used as the basis for the simulation-side analysis. As described in Section 3.7.1, the original dataset contains 224 city and town road networks from Balakrishnan and Stokkink [1]. Kotor was excluded from the present analysis because complete valid performance data were not available for all required combinations of strategy, damage level, criterion, and recovery stage. The results in this chapter are therefore based on 223 cities.

The descriptive statistics reported below are based on the original 224-network sample and are used to provide context on the scale and diversity of the road networks. The original sample covers cities and towns from 156 countries across six continents. The networks were extracted from OpenStreetMap and were limited to at most 2500 road edges to manage computational complexity.

The analysed networks vary in size and topology. The number of road edges ranges from 14 to 2427, with a mean of 799.08 and a median of 657.50. The networks also differ in density, average degree, clustering, assortativity, average link length, and average link travel time. These differences are relevant for this chapter because the simulated recovery outcomes and strategy rankings are partly shaped by the structure of each city network.

Table 4.7 illustrates the variation in the original 224 road-network sample. In the strategy-ranking analysis, the same stakeholder weights are applied to every city. This use of common stakeholder weights is consistent with the survey design, which elicited general recovery priorities rather than city-specific preferences. As a result, city-level differences in the rankings come from differences in simulated strategy performance across networks. However, in practice, local priorities may vary with damage conditions, past disaster experience, accessibility problems, and dependence on particular facilities. The rankings should therefore be interpreted as results under a common set of stakeholder weights. This assumption is discussed further as a limitation in Section 7.4.

Table 4.7: Descriptive statistics of the analysed road network sample

Variable	Mean	SD	Minimum	Median	Maximum
Number of edges	799.08	637.01	14.00	657.50	2427.00
Density	0.02	0.03	0.00	0.01	0.21
Average degree	4.83	0.60	3.10	4.77	6.50
Global clustering coefficient	0.08	0.05	0.01	0.07	0.40
Average link length (m)	292.52	528.42	53.45	152.03	5397.43
Average link travel time (min)	30.93	57.76	4.66	16.32	647.69

4.4.2. Strategy rankings under 10% damage

This subsection reports the strategy-ranking results under the 10% damage scenario. This scenario represents a relatively limited disruption, where only a small share of road links is initially removed. The analysis examines whether the stakeholder-specific BWM weights lead to different preferred recovery

strategies across cities, or whether one strategy remains dominant across stakeholder groups and recovery stages.

Table 4.8: Rank-one strategy distribution across cities for 10% damage in the short-term window. Each cell reports the share of cities in which the strategy is ranked first. The most frequently selected strategy in each stakeholder column is shown in bold.

Strategy	Emergency responder	Infrastructure manager	Local community	Public authority	Road users	Technical experts	Transport operator	Other
Betweenness	**70.0%**	33.6%	**68.6%**	**65.9%**	**69.1%**	**70.9%**	**70.4%**	**59.2%**
Closeness	9.9%	**34.1%**	10.8%	13.0%	10.8%	9.9%	9.4%	17.5%
Degree	0.4%	0.9%	0.4%	0.4%	0.4%	0.0%	0.4%	0.4%
Eigenvector	2.2%	5.4%	2.7%	2.7%	2.2%	2.2%	2.2%	2.7%
Nearest-neighbour edge	0.0%	4.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.3%
Random	0.4%	4.9%	0.4%	0.9%	0.4%	0.0%	0.4%	1.8%

Table 4.9: Rank-one strategy distribution across cities for 10% damage in the long-term window. Each cell reports the share of cities in which the strategy is ranked first. The most frequently selected strategy in each stakeholder column is shown in bold.

Strategy	Emergency responder	Infrastructure manager	Local community	Public authority	Road users	Technical experts	Transport operator	Other
Betweenness	**55.2%**	**84.8%**	**64.6%**	**64.6%**	**81.2%**	**59.6%**	**61.4%**	**60.1%**
Closeness	30.9%	7.6%	24.7%	25.6%	11.2%	29.1%	27.8%	27.8%
Degree	0.4%	0.4%	0.4%	0.4%	0.4%	0.4%	0.4%	0.4%
Eigenvector	4.0%	1.3%	3.6%	2.7%	1.3%	3.6%	3.6%	4.0%
Nearest-neighbour edge	0.9%	0.0%	0.4%	0.0%	0.0%	0.0%	0.0%	0.4%
Random	2.7%	0.0%	0.4%	0.9%	0.0%	1.3%	0.9%	1.3%

At the 10% damage level, betweenness centrality is the plurality choice, defined here as the strategy with the largest share of rank-first cities, for every stakeholder group in the long-term stage, and for all groups except the single-respondent infrastructure-manager group in the short-term stage. This indicates that, under limited disruption, differences in stakeholder weights do not lead to major differences in the most frequently preferred strategy across cities. Although the exact share of cities differs between groups and stages, betweenness remains the most common rank-first strategy throughout the 10% damage scenario.

4.4.3. Strategy rankings under 30% damage

This subsection reports the strategy-ranking results under the 30% damage scenario. Compared with the 10% damage scenario, this setting tests whether a larger initial disruption produces more variation in the preferred recovery strategies across stakeholder groups and recovery stages.

Table 4.10: Rank-one strategy distribution across cities for 30% damage in the short-term window. Each cell reports the share of cities in which the strategy is ranked first. The most frequently selected strategy in each stakeholder column is shown in bold.

Strategy	Emergency responder	Infrastructure manager	Local community	Public authority	Road users	Technical experts	Transport operator	Other
Betweenness	**74.0%**	33.2%	**73.5%**	**67.7%**	**72.6%**	**74.0%**	**73.5%**	**57.0%**
Closeness	15.2%	**45.3%**	15.7%	21.1%	16.6%	15.2%	15.7%	29.1%
Degree	0.4%	1.8%	0.4%	0.4%	0.4%	0.4%	0.4%	1.8%
Eigenvector	1.8%	3.6%	1.8%	1.8%	1.8%	1.8%	2.2%	1.8%
Nearest-neighbour edge	0.0%	3.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%
Random	1.8%	5.4%	1.8%	1.8%	1.8%	1.8%	1.3%	2.2%

Table 4.11: Rank-one strategy distribution across cities for 30% damage in the long-term window. Each cell reports the share of cities in which the strategy is ranked first. The most frequently selected strategy in each stakeholder column is shown in bold.

Strategy	Emergency responder	Infrastructure manager	Local community	Public authority	Road users	Technical experts	Transport operator	Other
Betweenness	39.0%	**81.6%**	**51.6%**	**53.4%**	**73.1%**	**48.4%**	**51.6%**	**47.5%**
Closeness	**45.7%**	15.2%	40.8%	40.8%	22.9%	43.0%	42.2%	43.0%
Degree	2.7%	0.0%	1.3%	0.9%	0.0%	1.3%	0.4%	1.3%
Eigenvector	4.9%	0.4%	2.7%	0.9%	0.9%	3.6%	2.2%	4.0%
Nearest-neighbour edge	1.8%	0.0%	0.4%	0.4%	0.0%	0.4%	0.4%	0.4%
Random	2.7%	0.0%	0.0%	0.4%	0.0%	0.0%	0.0%	0.4%

Overall, betweenness remains the most frequently preferred strategy across stakeholder groups, damage levels, and recovery stages. The main departures from this pattern occur under the 30% damage scenario, where closeness becomes more competitive for some stakeholder groups, especially in the long-term stage. This is most visible in the 30% long-term disruption condition, as reported in Table 4.11. In that condition, the emergency-responder group switches to closeness, which is consistent with its high weight on spatial equity (C4), the criterion on which closeness performs best

The equal-group baseline weights introduced in Section 3.3.3 are also applied in the SAW step. Under this profile, betweenness is the preferred strategy in 57 to 69% of cities across the short-term and long-term recovery stages under 10% and 30% damage. This baseline is used as the reference case for the sensitivity analysis in Section 4.6.4.

4.4.4. Overall ranking pattern

Across the four damage-stage scenarios, betweenness centrality is the most common plurality strategy. Under 10% damage, it is the plurality choice for every stakeholder group in the long-term window and for all groups except the single-respondent infrastructure-manager group in the short-term window. Under 30% damage, betweenness also remains the most frequent plurality strategy for most groups. The exceptions are the infrastructure-manager group in the short-term window and the emergency-responder group in the long-term window, where closeness centrality is selected most often.

The main interpretation is therefore that betweenness provides the most stable rank-one pattern across the city sample, while closeness is a conditional complementary alternative. The clearest closeness case for the criterion-level interpretation occurs in the 30% long-term condition. In that setting, the emergency-responder group selects closeness most often. This follows from the combined weight it places on spatial equity (C4) and connectivity (C5), which together account for $50.7\,\%$ of its long-term weight and are the two criteria on which closeness records its smallest deviation scores (Table E.2). Technical experts show a similar but weaker tendency in the same condition, where betweenness remains first but closeness rises to 43.0% compared with 48.4% for betweenness.

4.4.5. Similarity between single-indicator strategy orderings

The overall ranking pattern identifies betweenness centrality as the most stable top-performing strategy and closeness centrality as the main alternative. A related question is whether the single-indicator strategies prioritise similar road links. If two strategies produce highly similar edge-priority orderings, comparing them may provide limited additional information.

Following the method described in Section C.2, Table 4.12 reports the mean Spearman rank correlation between the strategies' intact-network edge-priority orderings across the road networks included in the cross-city analysis. The city-level coefficients are averaged with equal weight for each city. The random baseline is excluded because it does not generate a fixed priority ordering. These correlations compare predefined priority rules rather than damage-specific restoration sequences.

Overall, the pairwise rank correlations are generally low, although one strategy pair shows substantial overlap. The mean off-diagonal correlation is 0.22, indicating that the strategies are not redundant on average and capture different notions of link importance. The strongest agreement is between nearest-neighbour edge and weighted degree ($\rho = 0.64$), which is consistent with both strategies relying on local, degree-related network properties. This result indicates substantial overlap between their intact-

Table 4.12: Mean Spearman rank correlation between the strategies' edge-priority orderings.

	Betw.	Near.-nb.	Close.	Eigen.	Degree
Betweenness	1.00	0.29	0.35	0.08	0.23
Nearest-neighbour	0.29	1.00	0.17	0.09	0.64
Closeness	0.35	0.17	1.00	−0.01	0.22
Eigenvector	0.08	0.09	−0.01	1.00	0.12
Weighted degree	0.23	0.64	0.22	0.12	1.00

network edge-priority orderings.

Betweenness and closeness show a weak-to-moderate positive rank correlation ($\rho = 0.35$), indicating limited overlap between their intact-network edge-priority orderings. This pattern is consistent with their different criterion-level strengths. Betweenness prioritises links that lie on many shortest paths, whereas closeness prioritises links connected to nodes with shorter travel-time distances to the rest of the network. Eigenvector centrality shows almost no rank-order association with the other measures (ρ between -0.01 and 0.12), although its distinct ordering does not translate into consistently better recovery performance.

These results support treating the single-indicator strategies as distinct structural prioritisation rules in the cross-city analysis. Most strategy pairs show limited overlap in their intact-network edge orderings. Nearest-neighbour edge and weighted degree show the greatest overlap, but the correlation alone does not establish that the two strategies are interchangeable. Betweenness and closeness provide a useful contrast for interpreting different restoration logics because they combine relatively limited rank-order overlap with different criterion-level strengths.

4.5. Differences in strategy rankings across stakeholder groups and recovery stages

This section compares the rank-one strategy choices across stakeholder groups and between the short-term and long-term recovery stages. The purpose is to assess whether different stakeholder priorities and recovery stages lead to different preferred strategies.

4.5.1. Differences across stakeholder groups

Although the stakeholder groups assign different weights to the criteria, these differences rarely change the top-ranked strategy. As Section 4.4 shows, betweenness centrality is the plurality choice for most stakeholder groups across the four scenarios. The exceptions are the single-respondent infrastructure-manager group in the short-term stage under both damage levels, and the emergency-responder group in the long-term stage under 30% damage, where closeness is selected most often.

To examine agreement at the city level, the rank-one choices of the eight groups are compared within each city for a given damage level and recovery stage. Cities in which the six strategies produced identical SAW scores for all eight groups are excluded, because no rank-one strategy can be assigned in those cases. The exclusion is applied separately for each recovery stage, leaving 185 and 211 cities for the short- and long-term stages under 10% damage, and 208 and 217 cities under 30% damage. A small number of cities also contain a tie between two strategies for an individual stakeholder group; these affect at most one city per scenario and are retained. Table 4.13 reports the percentage of remaining cities in which all eight groups select the same rank-one strategy. Agreement is moderate across all four scenarios, ranging from 50% to 65%. Betweenness centrality is the most commonly selected rank-one strategy in every scenario. Agreement is consistently higher in the long-term stage than in the short-term stage, and the highest agreement is observed under 10% damage in the long-term window, where almost two-thirds of cities have the same rank-one strategy across all eight stakeholder groups.

Table 4.13: Within-city comparison of stakeholder group choices across scenarios. Valid cities exclude, separately for each recovery stage, cities in which the six strategies receive identical SAW scores for all eight stakeholder groups.

Scenario	Valid cities	All groups same	Most selected rank-one strategy
10% short-term	185	49.73%	Betweenness
10% long-term	211	64.93%	Betweenness
30% short-term	208	50.96%	Betweenness, followed by closeness
30% long-term	217	55.30%	Betweenness, followed by closeness

4.5.2. Differences between short- and long-term recovery stages

This section examines whether the preferred recovery strategy changes between the short-term and long-term recovery stages. For each city and each stakeholder group, the strategy ranked first in the short-term stage is compared with the strategy ranked first in the long-term stage. The unit of analysis is therefore a city–stakeholder combination.

For each damage level, combinations with a tied first-ranked strategy in either stage are excluded, because no unique transition can be assigned. The remaining combinations are used to calculate the switching rate, defined as the share of city–stakeholder combinations for which the first-ranked strategy changes between stages. A higher switching rate indicates stronger stage dependence, meaning that the strategy preferred in the short-term stage is less likely to remain preferred in the long-term stage. Table 4.14 reports the switching rate separately for each stakeholder group and damage level.

Table 4.14: Stage-dependent changes in rank-one strategy by stakeholder group. Each cell reports the number and percentage of included city–stakeholder pairs for which the rank-one strategy differs between the short- and long-term recovery stages. Cases with tied rank-one strategies in either stage are excluded.

Stakeholder group	10% damage	30% damage
Infrastructure manager	100/185 (54.1%)	119/207 (57.5%)
Emergency responder	65/185 (35.1%)	97/208 (46.6%)
Other	56/185 (30.3%)	73/207 (35.3%)
Technical expert / researcher	54/185 (29.2%)	78/208 (37.5%)
Transport/logistics operator	52/185 (28.1%)	79/208 (38.0%)
Public authority / government agency	50/185 (27.0%)	70/207 (33.8%)
Local community	47/185 (25.4%)	76/208 (36.5%)
Road users	27/185 (14.6%)	40/208 (19.2%)

The results show that stage-dependent switching differs across stakeholder groups. The infrastructure manager group has the highest switching rate under both damage levels, with the first-ranked strategy changing in 54.1% of included city–stakeholder combinations under 10% damage and 57.5% under 30% damage. This result should be interpreted cautiously, as both the infrastructure manager group and the other category are represented by a single respondent. Among the multi-respondent groups, emergency responders show the strongest stage dependence, with switching rates of 35.1% under 10% damage and 46.6% under 30% damage. Road users are the most stable group, with switching rates of 14.6% and 19.2%, respectively.

For every stakeholder group, the switching rate is higher under 30% damage than under 10% damage. This indicates that the difference between short- and long-term recovery stages becomes more important as disruption severity increases.

4.5.3. Linking the rankings to criterion weights and strategy performance

These patterns follow from the interaction between the criterion weights in Section 4.2 and the performance of each strategy on those criteria. Betweenness centrality is the lowest-deviation strategy for accessibility (C2) and operational efficiency (C1) in both stages and at both damage levels (Section 4.6.2). Since C2 carries the largest short-term weight in every main group (Table 4.5), betweenness

remains the plurality choice for these groups regardless of the smaller differences in their other weights. This explains the high short-term agreement in Table 4.13.

Betweenness loses ground only when weight moves towards spatial equity (C4), the criterion on which closeness centrality performs best. This occurs mainly in the long-term stage under 30% damage, where the group weights flatten. Technical experts, who give C4 their highest long-term weight (0.244), shift furthest towards closeness, where betweenness falls to 48.4% and closeness rises to 43.0% (Table 4.11). Road users keep C2 highest in both stages, stay with betweenness in 73.1% of cities in the same scenario, and switch least often (Table 4.14).

4.6. Robustness and sensitivity checks

This section reports the robustness and sensitivity checks defined in the methodology chapter. The checks examine whether the main results remain stable under selected methodological assumptions and potential data interpretation issues. They are not intended as a formal MCDM sensitivity analysis in which all criterion weights are systematically varied. Instead, each check targets a specific part of the analysis: the interpretation of the operational-efficiency criterion, the short-term recovery-stage boundary, the BWM weight-processing configuration, and the stakeholder-group aggregation weights.

The first and third checks are treated as robustness checks, because they test whether the results change under alternative data screening or weight-processing choices. The second and fourth checks are treated as sensitivity checks, because they vary a modelling boundary or aggregation-weight assumption. The results are compared with the main analysis in terms of criterion weights, deviation-score patterns, and strategy rankings.

4.6.1. Flagged operational-efficiency responses

As described in Section 3.8.1, a small number of responses were flagged because their open-ended explanations suggested an individual-travel-time interpretation of the operational-efficiency criterion (C1). Within the broader sample, six short-term responses were flagged, comprising three from the local community, two from road users, and one from the public authority/government agency group. Three long-term responses were also flagged, all from the local community. One further long-term response indicated the same interpretation but fell outside the broader sample and was therefore not included in the re-aggregation.

The top-ranked strategy did not change in any case. The betweenness centrality strategy remained ranked above the closeness centrality strategy throughout the local-community and road-user checks. For the public authority/government agency group in the short-term stage, excluding the flagged response changed the weights from 0.064 to 0.059 for C1, from 0.410 to 0.397 for C2, from 0.163 to 0.186 for C3, from 0.212 to 0.240 for C4, and from 0.151 to 0.118 for C5. The overall weight pattern remained similar, with C2 remaining the highest-weighted criterion. The numerical changes are larger than in the other checks because this group contains only three valid short-term responses before exclusion. Since temporarily excluding the flagged responses did not change the strategy rankings, all flagged responses were retained in the final analysis.

Table 4.15: Criterion weights before and after temporarily excluding the flagged operational-efficiency responses.

	Local community (short-term)		Local community (long-term)		Road users (short-term)	
Criterion	Before	After	Before	After	Before	After
C1	0.072	0.077	0.166	0.199	0.097	0.101
C2	0.501	0.494	0.207	0.213	0.442	0.428
C3	0.140	0.135	0.202	0.190	0.167	0.172
C4	0.124	0.126	0.198	0.185	0.121	0.120
C5	0.163	0.169	0.227	0.214	0.173	0.179

C1 denotes the operational-efficiency criterion. Weights are from the linear configuration. Sample sizes are $n = 17 \rightarrow 14$ for the local community in the short-term stage, $n = 19 \rightarrow 16$ for the local community in the long-term stage, and $n = 21 \rightarrow 19$ for road users in the short-term stage. The public authority/government agency check is reported in the text because it is based on only three valid short-term responses before exclusion. Columns may not sum to exactly one because of rounding.

4.6.2. Short-term recovery stage boundary

This sensitivity check examines whether the main findings depend on the 25% restoration boundary used to separate the short- and long-term recovery stages. The full sensitivity plots are reported in Section G.2.

Changing the boundary affects the magnitude of the mean deviation scores, but not the main criterion-level strategy patterns. Across the tested boundary values, betweenness remains the lowest-deviation strategy for operational efficiency and accessibility at both damage levels and in both recovery stages. Closeness remains the lowest-deviation strategy for spatial equity across all tested boundaries, damage levels, and recovery stages. For connectivity, closeness remains lowest in the short-term stage at both damage levels and in the long-term stage under 10% damage, while random remains lowest in the long-term stage under 30% damage.

Overall, the sensitivity check indicates that the 25% boundary affects the absolute size of the deviation scores but does not change the lowest-deviation strategy within each criterion, damage level, and recovery stage. This supports the use of the 25% boundary in the main analysis while acknowledging that it is a modelling assumption.

4.6.3. BWM configuration

The fourth robustness check examines whether the strategy rankings depend on the BWM configuration used to derive the criterion weights. Following the all-respondent comparison procedure described in Section 3.8.4, the SAW rankings were recomputed under the four BWM configurations: linear–broader, linear–strict, equivalent–broader, and equivalent–strict.

The all-respondent criterion weights vary only slightly across configurations. In the short-term stage, all four configurations produce the same criterion ranking, C2 > C5 > C3 > C4 > C1. In the long-term stage, the main difference is that linear–broader ranks C2 slightly above C3, whereas the other three configurations rank C3 slightly above C2. The largest pairwise criterion-weight difference is approximately 0.015 in the short-term stage and approximately 0.019 in the long-term stage.

These differences do not alter the main SAW ranking pattern. Across all four configurations, both recovery stages, and both damage levels, betweenness centrality remains the plurality rank-one strategy. No configuration changes the plurality winner relative to the linear–broader primary configuration. The share of cities in which betweenness centrality is the rank-one strategy ranges from 67.7% to 69.1% under 10% short-term damage, from 70.4% to 71.3% under 10% long-term damage, from 71.3% to 71.7% under 30% short-term damage, and from 58.3% to 59.6% under 30% long-term damage.

Overall, the main strategy-ranking result is stable across the tested BWM configurations. This suggests that the dominance of betweenness centrality is not driven by the inclusion of Grade C responses or by the choice between the linear and equivalent BWM formulations.

4.6.4. Stakeholder weight aggregation

This final sensitivity check reports the sensitivity of the strategy rankings to the stakeholder-group importance weights. The procedure follows Section 3.8.2, where each stakeholder group with more than one valid response is assigned a larger aggregation weight in turn. In this dataset, the single-respondent groups are infrastructure managers and *other*; these groups are therefore not tested as dominant groups but remain included in the residual aggregation weight.

For each recovery stage and damage level, the rankings obtained under the alternative group-weighting schemes are compared with the equal-group baseline.

The resulting overall criterion weights change gradually as the dominance share increases. The changes are smaller in the short-term recovery stage, where most stakeholder groups place their highest weight on access to critical facilities (C2). The changes are larger in the long-term recovery stage, where the group-level weights are more diverse. Figure 4.6 illustrates this pattern for the public authority/government agency group. As the aggregation share of this group increases, the overall criterion weights move gradually away from the equal-group baseline, but no abrupt change in the weight profile is observed.

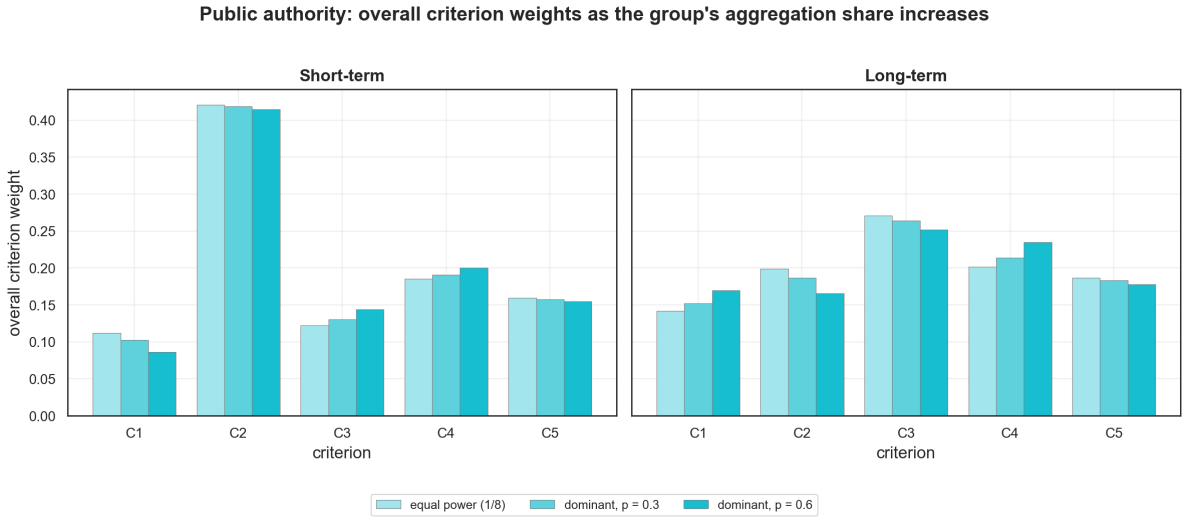


Figure 4.6: Overall criterion weights under equal-group aggregation and under public authority/government agency dominance shares of 0.3 and 0.6, for the short- and long-term recovery stages.

After examining how the aggregation weights affect the overall criterion weights, the sensitivity check was extended to the resulting SAW rankings. The aim is to assess whether increasing the aggregation weight of one stakeholder group changes the plurality rank-one strategy across cities.

The aggregation sensitivity results show that the main ranking pattern is stable across all tested stakeholder weighting schemes. In every scenario, betweenness centrality remains the plurality rank-one strategy for each of the six multi-respondent stakeholder groups tested as the dominant group, identical to the equal-group baseline. The two single-respondent groups, infrastructure managers and the other category, were not used as dominant groups in this check. The share of cities in which betweenness centrality is the rank-one choice ranges from 47% to 77%.

The 30% damage long-term scenario is the most sensitive to stakeholder weighting: the betweenness share ranges from 46.6% to 66.8% depending on which stakeholder group dominates and at what level. When the emergency-responder group is assigned a dominance share of 0.6, the betweenness share reaches its minimum of 46.6%, although closeness centrality does not overtake it as the plurality choice. The other three scenarios show narrower betweenness ranges (65.9 to 69.5% for 10% short-term, 62.3 to 77.1% for 10% long-term, and 68.6 to 72.6% for 30% short-term).

5

Delft Case Study

This chapter presents the Delft case study, the city-level application of the framework introduced in Chapter 3. The cross-city analysis in Chapter 4 compares recovery strategies across many city networks using aggregate performance metrics. The Delft case study applies the same stakeholder-weighted evaluation logic to a single urban network, where accessibility can be represented in more local detail. Its purpose is to examine how the recovery strategies compare at the city level when accessibility is modelled through explicit facility destinations and when selected local road-user outcomes are examined.

The Delft case study has two complementary roles:

- **Facility-based strategy ranking.** The same recovery strategies, BWM-derived stakeholder weights, AAC scoring logic, and SAW aggregation procedure are applied to the Delft road network. The main difference from the cross-city analysis is that the aggregate accessibility proxy is replaced by two facility-based measures. Access to critical facilities corresponds to C2, while access to routine facilities corresponds to C3. This allows the case study to examine how the strategies compare when accessibility is represented more explicitly.
- **Supplementary local road-user checks.** These checks are kept separate from the SAW criteria. They examine restoration-order patterns, boundary sensitivity, and travel times from selected residential origins to reachable emergency, medical, food, and water services. This provides a road-user perspective on the Delft ranking results and helps assess whether the system-level rankings remain meaningful for specific local accessibility outcomes.

The Delft results are therefore used to support local interpretation of the framework. They show how the same evaluation logic works when more detailed local accessibility information is added to the Delft network, and how the resulting rankings can be interpreted in a specific urban context

5.1. Facility-based separation of accessibility criteria

As described in Section 3.7.2, the Delft case study separates the two accessibility criteria used in the stakeholder survey. Access to critical facilities corresponds to C2, while access to routine facilities corresponds to C3. This differs from the cross-city analysis, where both criteria are represented by the same aggregate accessibility metric.

In the Delft implementation, facility records are first linked to their nearest road-network nodes. The mapped nodes are then used as destinations in the accessibility calculations. The points shown in Figure 5.1 therefore represent mapped road-network nodes, rather than the original facility records. Multiple facilities may be represented by the same network node. The Delft case study uses 58 mapped critical-facility nodes and 364 mapped routine-facility nodes.

Figure 5.1 provides a reference map for interpreting the Delft case study results. It shows the Delft road network and the mapped facility nodes used to calculate the two facility-based accessibility criteria.

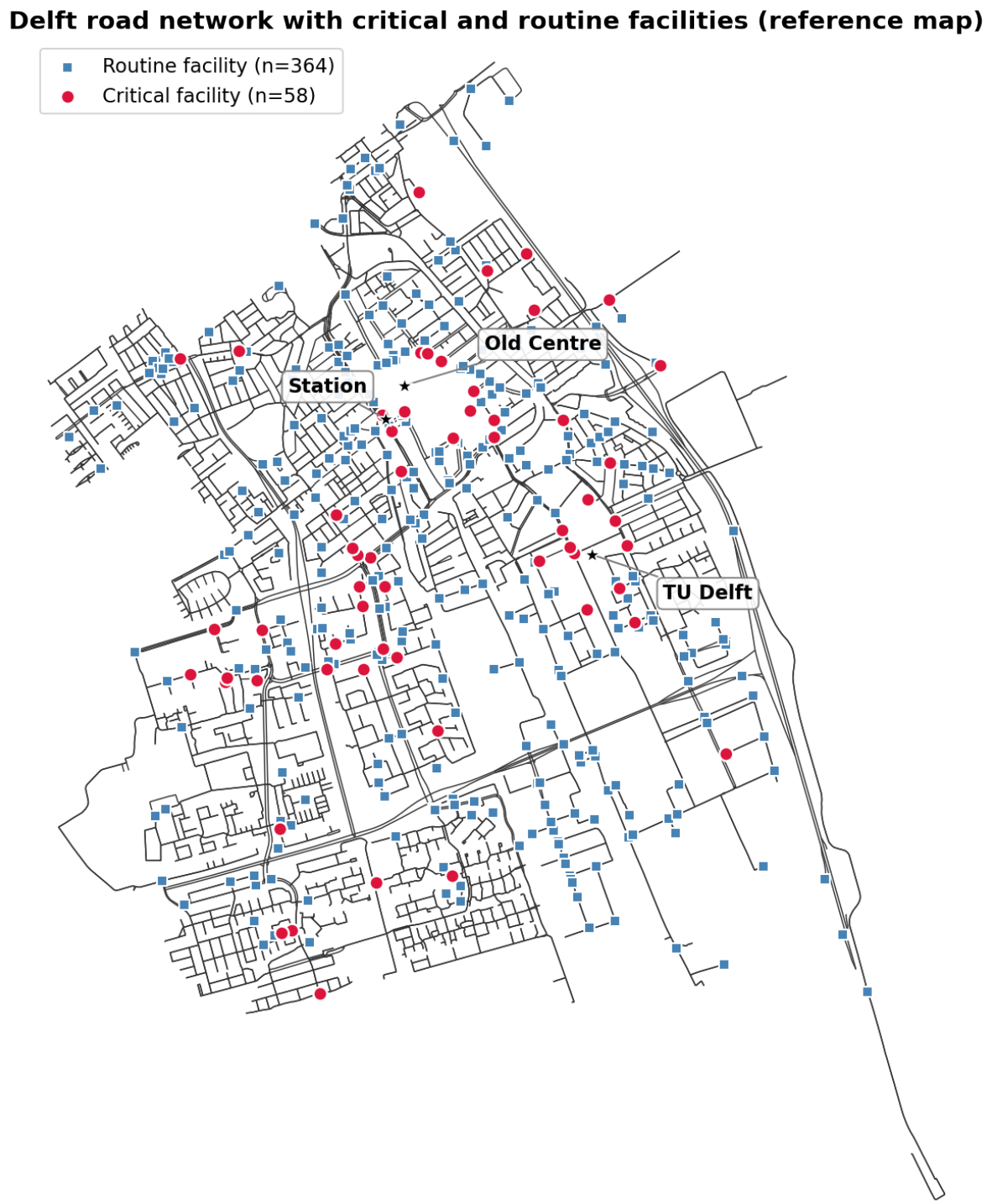


Figure 5.1: Delft road network with mapped facility nodes used in the accessibility calculations. The figure shows 58 critical-facility nodes and 364 routine-facility nodes after linking original facility records to the nearest road-network nodes.

5.2. Delft criterion-level recovery results

This section reports the Delft recovery results at the criterion level. The focus is on the two accessibility criteria that are separated in the Delft case study: access to critical facilities (C2) and access to routine facilities (C3). The aim is to compare how the six recovery strategies perform when these two accessibility criteria are evaluated using separate facility destination sets. The other three criteria, operational efficiency (C1), spatial equity (C4), and connectivity (C5), are also reported because they form part of the SAW decision matrix.

Each recovery curve is summarised by its area-above-curve (AAC) deviation, as defined in Section 3.5.4. The AAC deviation is calculated separately for the short-term and long-term recovery stages, with lower values indicating faster recovery. The values reported in Tables 5.1 and 5.2 are mean AAC deviations over 20 damage realisations. These criterion-level deviation values are later used as inputs for the SAW ranking in Section 5.3. The corresponding recovery curves are shown in Figures 5.2 and 5.3, where the lines show mean recovery ratios and the shaded bands show $95\,\%$ confidence intervals across the 20 damage realisations.

5.2.1. AAC deviations under 10% damage

Table 5.1: Criterion-level AAC deviations under $10\,\%$ damage in Delft. Lower values indicate faster recovery, and the best mean value in each column is shown in bold. Values are means over 20 damage realisations. C2 and C3 are calculated using facility-based destination sets.

	Short-term					Long-term				
Strategy	Eff.	Crit.	Rout.	Equity	Conn.	Eff.	Crit.	Rout.	Equity	Conn.
Betweenness	**0.235**	**0.232**	**0.182**	**0.291**	0.068	**0.040**	**0.038**	**0.028**	**0.074**	0.026
Closeness	0.315	0.280	0.271	0.397	**0.065**	0.098	0.070	0.070	0.130	**0.021**
Weighted degree	0.383	0.391	0.346	0.340	0.083	0.188	0.200	0.187	0.269	0.050
Eigenvector	0.385	0.381	0.331	0.308	0.077	0.209	0.214	0.171	0.119	0.055
Nearest-neighbour	0.379	0.386	0.343	0.344	0.081	0.160	0.171	0.154	0.198	0.038
Random	0.357	0.356	0.314	0.352	0.074	0.142	0.130	0.130	0.180	0.029

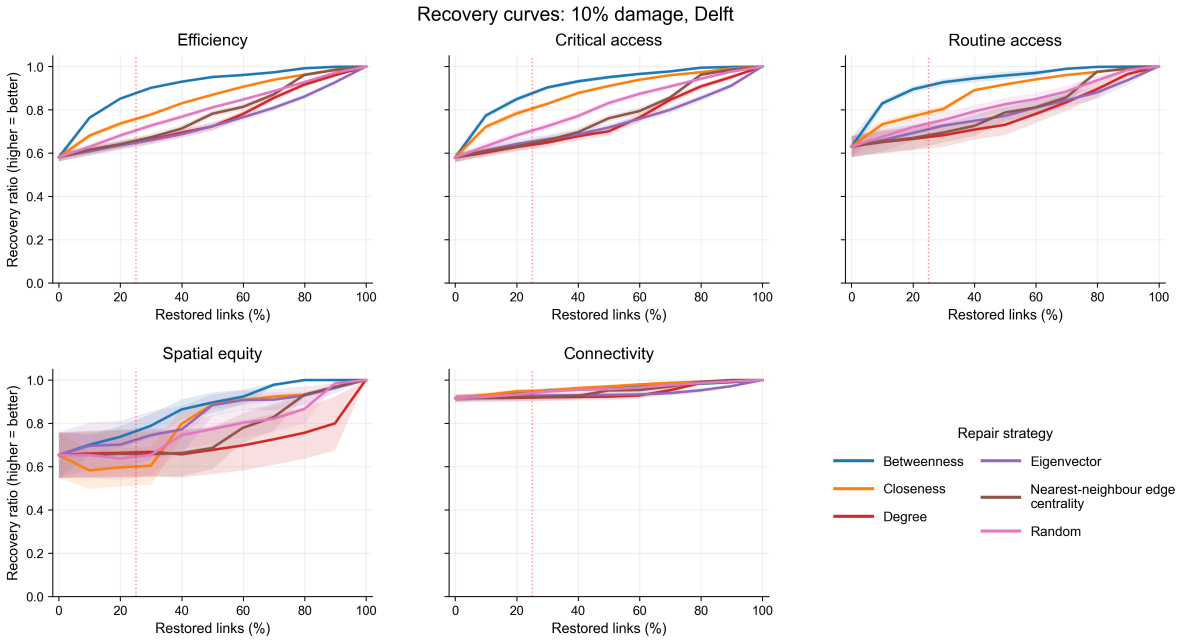


Figure 5.2: Delft recovery curves under $10\,\%$ damage, by criterion. Lines show mean recovery ratios over 20 damage realisations, with shaded bands indicating $95\,\%$ confidence intervals. Higher values indicate better recovery. The dashed line marks the $\theta = 25\,\%$ boundary between the short-term and long-term recovery stages.

5.2.2. AAC deviations under 30% damage

Table 5.2: Criterion-level AAC deviations under 30 % damage in Delft. Lower values indicate faster recovery, and the best mean value in each column is shown in bold. Values are means over 20 damage realisations. C2 and C3 are calculated using facility-based destination sets.

	Short-term					Long-term				
Strategy	Eff.	Crit.	Rout.	Equity	Conn.	Eff.	Crit.	Rout.	Equity	Conn.
Betweenness	**0.621**	**0.622**	**0.563**	**0.547**	0.252	**0.127**	**0.124**	**0.096**	**0.181**	0.085
Closeness	0.762	0.700	0.808	0.703	**0.246**	0.272	0.203	0.206	0.220	**0.077**
Weighted degree	0.877	0.899	0.916	0.712	0.312	0.486	0.514	0.580	0.492	0.176
Eigenvector	0.879	0.884	0.882	0.598	0.296	0.564	0.575	0.536	0.285	0.200
Nearest-neighbour	0.880	0.900	0.924	0.733	0.307	0.442	0.469	0.463	0.411	0.133
Random	0.869	0.887	0.905	0.703	0.290	0.435	0.422	0.487	0.402	0.118

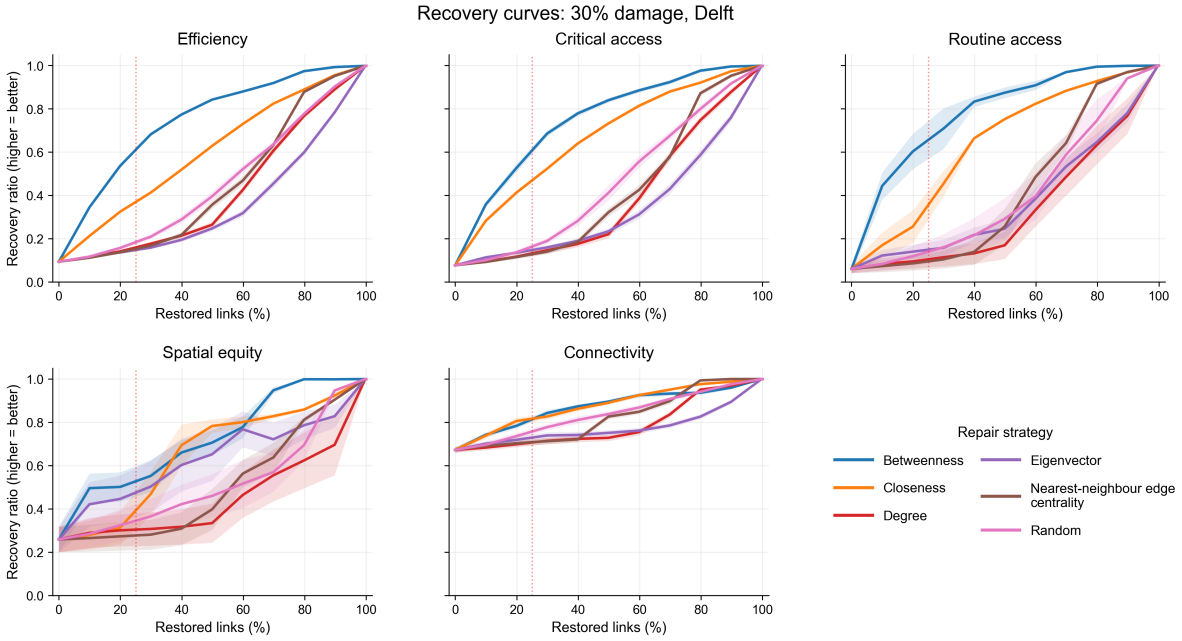


Figure 5.3: Delft recovery curves under 30 % damage, by criterion. Lines show mean recovery ratios over 20 damage realisations, with shaded bands indicating 95 % confidence intervals. Higher values indicate better recovery. The dashed line marks the $\theta = 25\,\%$ boundary between the short-term and long-term recovery stages.

5.2.3. Comparison between critical and routine accessibility

For both access to critical facilities and access to routine facilities, betweenness centrality has the lowest AAC deviation across both damage levels and both recovery stages. Closeness centrality is the second-best strategy for the two accessibility criteria in all cases. The remaining strategies, namely weighted degree, eigenvector, nearest-neighbour edge, and random, generally show higher accessibility deviations.

Although the strategy ordering is similar for the two accessibility criteria, the deviation values differ between critical and routine accessibility. For betweenness, the short-term routine-accessibility deviation is lower than the short-term critical-accessibility deviation under both damage levels. Under 10 % damage, the routine-accessibility deviation is 0.182, compared with 0.232 for critical accessibility. Under 30 % damage, the routine-accessibility deviation is 0.563, compared with 0.622 for critical accessibility. For closeness, the opposite pattern appears under 30 % damage in the short-term stage, where the critical-accessibility deviation is lower than the routine-accessibility deviation, with 0.700 compared with 0.808.

These results show that separating the two destination sets reveals differences between critical and

routine accessibility that would not be visible if both criteria were represented by a single aggregate accessibility measure. The other three criteria are included in Tables 5.1 and 5.2 because they are used together with C2 and C3 in the SAW ranking. Since C1, C4, and C5 are calculated in the same way as in the cross-city analysis, the interpretation in this section focuses mainly on the two facility-based accessibility criteria.

5.3. SAW ranking results

This section combines the five Delft criterion-level deviation values with the BWM stakeholder weights using the SAW procedure. Unlike Section 5.2, which focuses on criterion-level recovery patterns, this section evaluates the six recovery strategies using all five criteria together. The aim is to examine whether the Delft case study produces the same main ranking pattern as the cross-city analysis when access to critical facilities and access to routine facilities are evaluated separately.

5.3.1. Main strategy rankings for key stakeholder groups

The main comparison focuses on the three stakeholder groups with sufficient responses: road users ($n = 22$), the local community ($n = 19$), and technical experts and researchers ($n = 9$). The ranking uses the linear–broader BWM weight configuration as the primary weight set, as described in Section 3.3.3.

Because the SAW scores are normalised separately for each damage level, recovery stage, and stakeholder group, they should be compared only between strategies within the same column of each table. Higher scores indicate better relative performance within that specific scenario and stakeholder perspective.

Table 5.3: SAW scores for baseline restoration strategies in Delft from the road users' perspective. Scores are shown by damage level and recovery stage. Higher values indicate better relative performance within each column.

	10 %		30 %	
Strategy	Short-term	Long-term	Short-term	Long-term
Betweenness	**0.97**	**0.98**	**0.99**	**0.99**
Closeness	0.61	0.78	0.61	0.82
Weighted degree	0.07	0.07	0.02	0.10
Eigenvector	0.20	0.12	0.17	0.11
Nearest-neighbour edge	0.10	0.30	0.01	0.30
Random	0.28	0.47	0.11	0.34

Table 5.4: SAW scores for baseline restoration strategies in Delft from the local community perspective. Scores are shown by damage level and recovery stage. Higher values indicate better relative performance within each column.

	10 %		30 %	
Strategy	Short-term	Long-term	Short-term	Long-term
Betweenness	**0.97**	**0.97**	**0.99**	**0.99**
Closeness	0.61	0.80	0.62	0.84
Weighted degree	0.07	0.07	0.02	0.10
Eigenvector	0.20	0.17	0.17	0.15
Nearest-neighbour edge	0.10	0.33	0.01	0.32
Random	0.28	0.50	0.11	0.37

Table 5.5: SAW scores for baseline restoration strategies in Delft from the technical experts and researchers' perspective. Scores are shown by damage level and recovery stage. Higher values indicate better relative performance within each column.

	10 %		30 %	
Strategy	Short-term	Long-term	Short-term	Long-term
Betweenness	**0.96**	**0.97**	**0.99**	**0.99**
Closeness	0.60	0.79	0.60	0.84
Weighted degree	0.06	0.06	0.02	0.09
Eigenvector	0.19	0.21	0.16	0.18
Nearest-neighbour edge	0.10	0.32	0.01	0.31
Random	0.28	0.49	0.11	0.36

Across all Delft SAW scenarios, betweenness centrality remains the top-ranked strategy. This holds for both damage levels and both recovery stages across the three main stakeholder groups. Closeness centrality consistently ranks second. The remaining strategies receive substantially lower scores, with random and nearest-neighbour edge generally occupying the middle positions depending on the recovery stage, while weighted degree and eigenvector are usually among the lowest-ranked strategies.

The dominance of betweenness centrality is also reflected in the score differences. Across the three stakeholder groups, its SAW score is close to the maximum in every scenario, ranging from 0.96 to 0.99. The gap between betweenness and the second-ranked closeness strategy is approximately 0.35–0.39 in the short-term stage and about 0.14–0.20 in the long-term stage. This suggests that betweenness-based restoration is particularly advantageous during the early recovery phase, while its advantage narrows later in the restoration process.

5.3.2. Ranking results across all stakeholder groups

The same ranking pattern is observed when all stakeholder groups are considered. Road users, the local community, and technical experts and researchers all rank betweenness centrality first in every damage level and recovery stage. Closeness centrality is also ranked second by each of these three groups.

The five smaller stakeholder groups are treated as exploratory because they each include only one to three respondents. Their results point in the same direction, with betweenness centrality ranked first throughout. Some variation appears in the second-ranked strategy, but this variation is not interpreted further because of the small group sizes.

5.3.3. Comparison with the cross-city ranking pattern

The Delft results are consistent with the main ranking pattern observed in the cross-city analysis. In both analyses, betweenness centrality is the most stable top-ranked strategy across damage levels, recovery stages, and stakeholder perspectives. The Delft case study therefore supports the interpretation that betweenness-based restoration performs well under the stakeholder-weighted evaluation framework.

The Delft results also show that closeness centrality remains the main alternative to betweenness. Its SAW scores are clearly lower than those of betweenness in the short-term stage, but the gap becomes smaller in the long-term stage. This pattern is consistent with the criterion-level results in Section 5.2, where closeness performs well for connectivity and becomes more competitive later in the recovery process.

The separation of access to critical facilities and access to routine facilities does not change the top-ranked strategy in Delft. Betweenness remains first even when C2 and C3 are calculated using separate facility destination sets. However, the separation improves the interpretation of the results because it shows that critical and routine accessibility do not always recover in exactly the same way. The Delft case study therefore arrives at the same first-ranked strategy as the cross-city analysis, while adding local detail about how the accessibility criteria contribute to that ranking.

5.4. Spatial restoration patterns

The SAW ranking in Section 5.3 reports the weighted recovery performance of the six strategies. This section provides a spatial interpretation of the strategy rules by showing how each strategy translates into an edge priority order across the Delft road network. The purpose is to show which parts of the network each strategy tends to prioritise early. This section supports the interpretation of the SAW results and is not used as an additional performance measure. Because the priority orders are calculated before damage is imposed, this section also discusses how the static priority design affects the interpretation of the Delft results.

5.4.1. Static edge priority order across strategies

A key modelling choice is that the edge priority order is calculated on the intact Delft road network before damage is imposed. Each strategy assigns an importance score to every link in the full network using the strategy definitions in Section 3.2.2. These scores define a fixed priority order. In each damage

realisation, only the damaged links are eligible for restoration, and these damaged links are restored according to their position in the predefined priority order.

The same priority order is therefore kept across damage realisations. This design keeps the six strategies comparable across random damage realisations and represents restoration rules that could be prepared before a specific disruption is observed. It also avoids turning each damage realisation into a separate adaptive optimisation problem. The Delft results should therefore be interpreted as a comparison of static network-based restoration heuristics, rather than as damage-specific optimal restoration plans.

Figure 5.4 visualises the fixed priority order by colouring all Delft links according to their priority band, from the earliest-priority links (Band 1, dark red) to the latest-priority links (Band 8, blue). The bands divide each strategy's full edge priority ranking into eight equal groups. They should be read as a visual summary of the precomputed priority order, not as the $10\,\%$ restoration steps used in the simulation. The black points mark the 58 mapped critical-facility nodes.

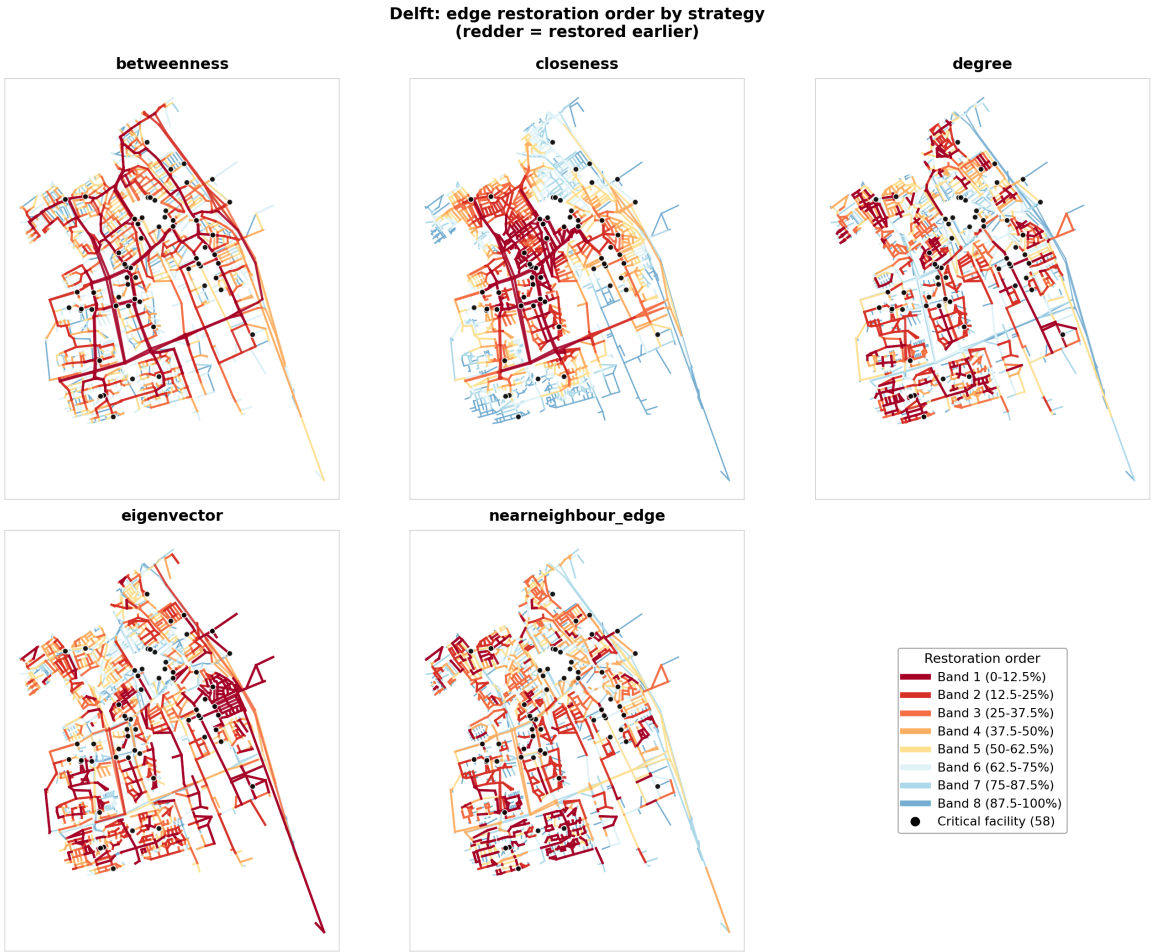


Figure 5.4: Static edge priority order in Delft by strategy. Links are coloured by priority band, from earliest priority (Band 1, dark red) to latest priority (Band 8, blue). The bands divide each strategy's full edge priority ranking into eight equal groups. Black points mark the 58 mapped critical-facility nodes.

5.4.2. Spatial interpretation of the priority patterns

Figure 5.4 shows that the betweenness and closeness strategies produce different early-priority patterns. Under the betweenness centrality strategy, early priority is given to links that connect different parts of Delft, including links between the city centre and surrounding areas. This follows the definition of edge betweenness, where links receive higher scores when they lie on many shortest paths between node pairs.

Under the closeness centrality strategy, early priority is more concentrated in the central part of the network. This follows from the closeness-based rule, where links connected to nodes with shorter average travel times to the rest of the network receive higher priority. The contrast between these two strategies is useful because both strategies perform relatively well in the Delft SAW results, but they do so through different spatial restoration patterns.

The other strategies show less network-wide early-priority patterns. Weighted degree and nearest-neighbour edge give early priority to more local road fragments, while eigenvector centrality concentrates early priority in a dense part of the network. The random strategy does not produce a clear spatial pattern.

These spatial patterns help interpret the SAW results in Section 5.3. Betweenness performs best in the Delft SAW ranking, and its priority map shows that it restores links connecting different parts of the network early. This pattern is consistent with its strong performance across the accessibility and efficiency criteria, especially under higher damage levels. Closeness also performs well, particularly in the long-term stage, and its more centralised early-priority pattern helps explain why it remains the main alternative to betweenness.

5.4.3. Implications of the static priority order

The use of a static edge priority order is an important modelling assumption. It represents a predefined restoration rule that is prepared from the intact network before a specific damage pattern is observed. This can reflect pre-disaster planning, where transport agencies prepare priority corridors or restoration rules before disruption occurs. However, actual post-disaster restoration is likely to include adaptive decisions once the location and severity of damage are known.

Compared with an adaptive restoration strategy, the static approach has less information. An adaptive strategy could recompute priorities after each damage realisation, or after each restoration step, and could therefore target links that are most useful for the specific damaged network. The static approach does not do this. It prioritises links according to their structural importance in the intact network and then restores only those priority links that are damaged in a given realisation.

This assumption may influence the results. It may favour strategies such as betweenness and closeness when their intact-network priority order identifies links that remain important across many damage realisations. At the same time, it may underestimate the performance of strategies that would benefit from reacting to the realised damage pattern. For example, a static priority order may give high priority to a link that is structurally important in the intact network but less critical in a particular damaged network. An adaptive strategy could instead redirect restoration resources to links that reconnect isolated areas or improve access to specific facilities.

The Delft results should therefore be interpreted as the performance of predefined network-based restoration heuristics under repeated random damage. They do not represent the best possible restoration sequence for each damage realisation. This design supports a controlled comparison between strategies, because all strategies are evaluated under the same damage realisations and the same predefined priority logic. The trade-off is that the results provide a benchmark for static planning rules rather than a full representation of adaptive post-disaster decision-making. This limitation is returned to in Section 7.4, where the implications of using predefined restoration rules are discussed.

5.5. Delft boundary sensitivity results

This section reports the boundary sensitivity results for the Delft case study. As described in Section 3.8.5, the analysis examines whether the centrality-based restoration priority patterns within Delft change when surrounding road-network context is added to the Delft municipal network. This check is included because centrality measures can be affected by the spatial extent used to define the road-network model.

Delft is kept as the core study area, while $1\,\mathrm{km}$, $3\,\mathrm{km}$, and $5\,\mathrm{km}$ buffers are used as additional computational context for the centrality calculations. For each network extent, centrality scores are calculated on the full network, including the buffer area. The comparison then remains focused on road edges inside the original Delft boundary. Only Delft core edges are visualised and ranked, so the maps show how

the relative restoration priority within Delft changes when surrounding road-network context is included.

This section focuses on the road-network boundary effect. It does not test the inclusion of facilities within the buffer zone, because the restoration priority orders examined here are based on network structure rather than facility destinations. The facility-boundary issue is considered separately in Section 5.6, where reachable facilities within the selected buffer are included from a road-user perspective.

5.5.1. Buffer-based network extent comparison

Figure 5.5 shows the four network extents used in the boundary sensitivity analysis: the Delft-only network, the Delft network with a $1\,\mathrm{km}$ buffer, the Delft network with a $3\,\mathrm{km}$ buffer, and the Delft network with a $5\,\mathrm{km}$ buffer. The Delft-only network represents the municipal study area used in the main Delft case study. The buffered networks add surrounding road-network context while keeping Delft as the area of comparison.

The purpose of this comparison is to examine whether the restoration priority order assigned to roads inside Delft changes when roads outside the municipal boundary are included in the centrality calculation. A stable pattern across the four extents suggests that the priority order is not strongly affected by the selected network boundary. A changing pattern suggests that the strategy is more sensitive to surrounding road-network context.

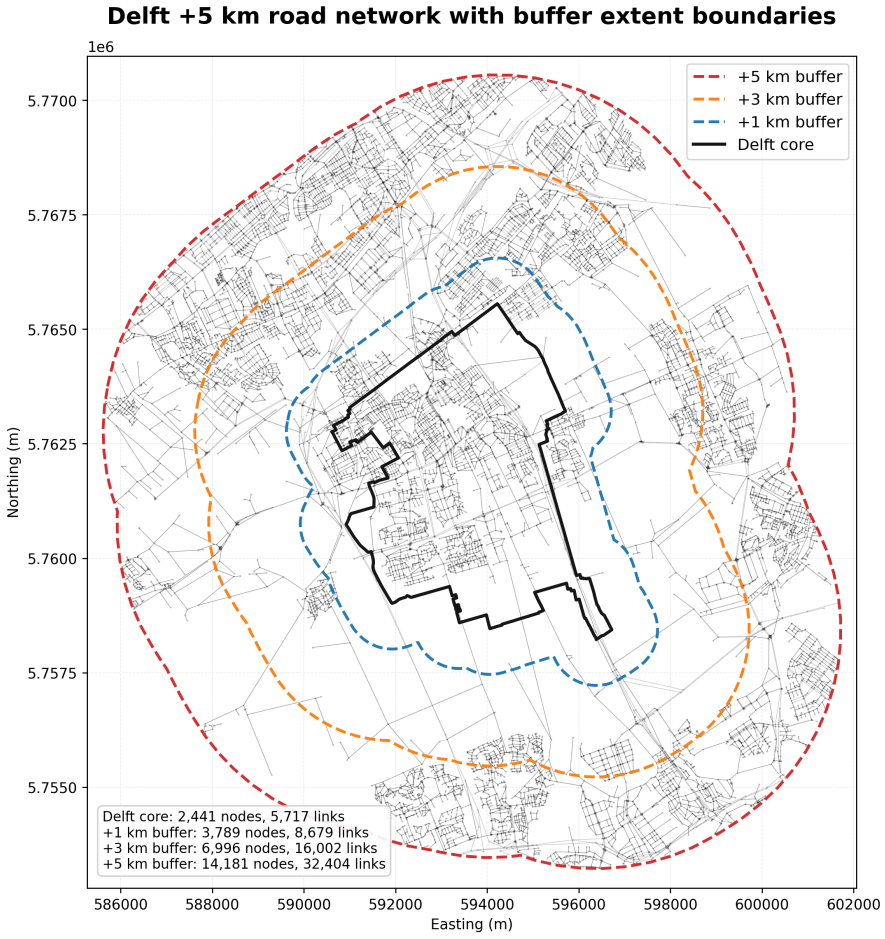


Figure 5.5: Network extents used for the Delft boundary sensitivity analysis. Delft municipality is treated as the core study area, while $1\,\mathrm{km}$, $3\,\mathrm{km}$, and $5\,\mathrm{km}$ buffers are added to provide surrounding road-network context for the centrality calculations.

5.5.2. Sensitivity of betweenness and closeness priority orders

The boundary sensitivity results show that the influence of the network boundary differs by strategy. The betweenness-based restoration order changes less across the four network extents than the closeness-based restoration order. As shown in Figure 5.6, the main high-priority roads inside Delft remain broadly similar when the network is expanded from Delft only to the $1\,\mathrm{km}$, $3\,\mathrm{km}$, and $5\,\mathrm{km}$ buffer networks. This suggests that, within the tested buffer range, the betweenness-based restoration pattern is relatively stable with respect to the selected network boundary.

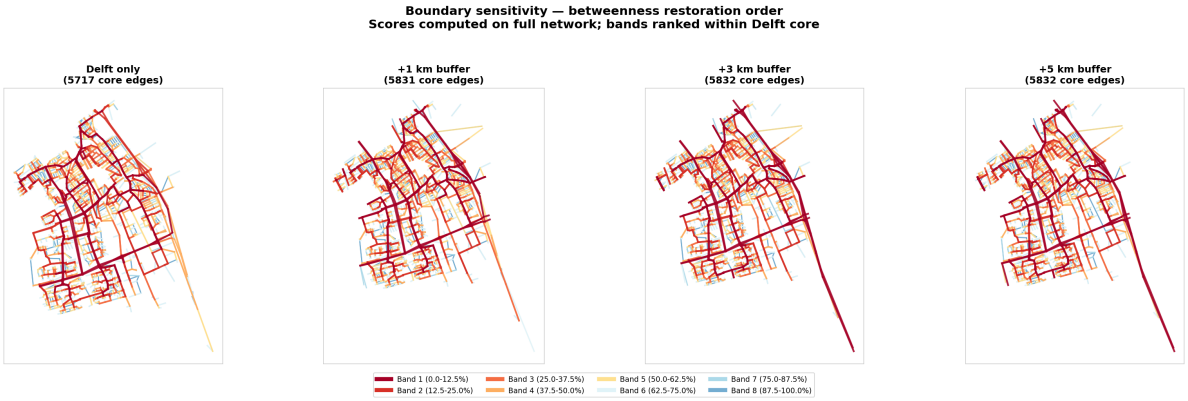


Figure 5.6: Boundary sensitivity of the betweenness-based restoration order. Centrality scores are calculated on the full network extent, while only Delft core edges are shown and ranked.

The closeness-based restoration order shows stronger sensitivity to the boundary definition, as shown in Figure 5.7. The Delft-only closeness pattern differs from the buffered patterns, especially when moving from no buffer to the $1\,\mathrm{km}$ and $3\,\mathrm{km}$ buffer networks. This is expected because closeness depends on the travel-time distance from a node to all other nodes in the analysed network. When surrounding roads are added, the set of nodes used in the calculation changes, and the relative position of Delft roads within the wider network also changes. The pattern changes less between the $3\,\mathrm{km}$ and $5\,\mathrm{km}$ buffers than between the Delft-only and smaller-buffer cases.

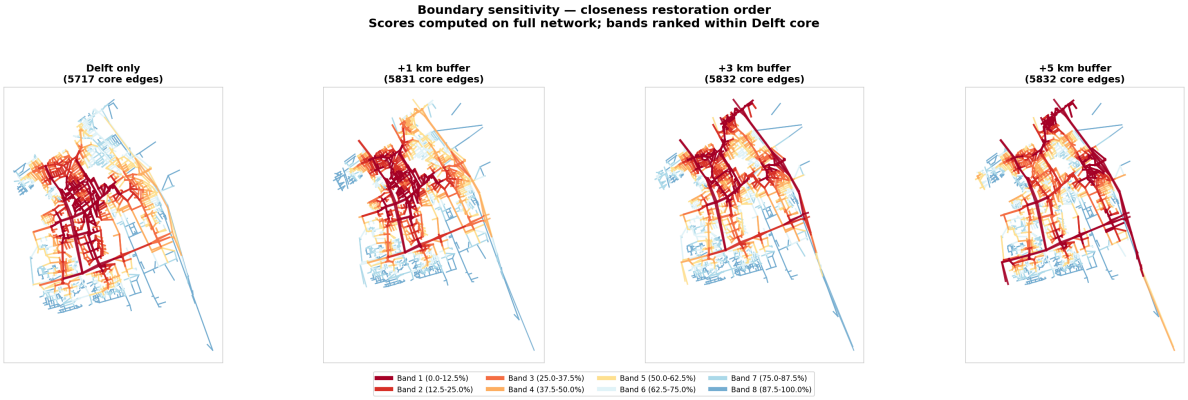


Figure 5.7: Boundary sensitivity of the closeness-based restoration order. Centrality scores are calculated on the full network extent, while only Delft core edges are shown and ranked.

5.5.3. Sensitivity of the remaining strategies

The remaining strategies show lower to moderate boundary sensitivity. The weighted-degree and nearest-neighbour edge strategies change mainly between the Delft-only network and the buffered networks, but their patterns become more similar once a buffer is added. Eigenvector centrality shows moderate sensitivity, although the main spatial structure remains visible across the tested network extents. The random strategy is not interpreted as a boundary-sensitive spatial pattern because its priority order is not based on network centrality.

These results indicate that the boundary effect is most relevant for strategies whose scores depend on distances or paths across the wider network. Closeness is therefore more sensitive to the selected network extent than betweenness. This matters for interpretation because a narrowly defined municipal network may change the relative importance of some roads when the surrounding road context is omitted.

5.5.4. Implications for boundary choices

The boundary sensitivity results show that the selected network extent affects restoration priority orders differently across strategies. The betweenness-based priority order remains relatively stable across the tested extents, while the closeness-based priority order changes more clearly when surrounding road-network context is added. This indicates that the betweenness pattern observed in the Delft-only network is less dependent on the municipal boundary. The closeness pattern requires more caution, because closeness depends strongly on the set of nodes included in the network.

These results suggest that the network boundary should be selected according to the purpose of the analysis. A Delft-only network provides a clear administrative study area and is useful when the analysis needs to remain strictly within the municipal boundary. However, it can omit nearby roads that influence shortest paths and centrality values inside Delft. A buffered network is more suitable when the analysis depends on shortest-path relationships near the municipal edge or when Delft is interpreted as part of a connected local road system.

Among the tested extents, the Delft plus $3\,\mathrm{km}$ buffer provides a practical balance. The priority patterns change noticeably when moving from the Delft-only network to the buffered networks, especially for closeness. The differences between the $3\,\mathrm{km}$ and $5\,\mathrm{km}$ buffer results are smaller. The $3\,\mathrm{km}$ buffer therefore adds surrounding road-network context while keeping the analysis focused on the Delft urban area and avoiding the higher computational cost and broader spatial scope of the $5\,\mathrm{km}$ network.

A similar boundary choice applies to facilities. If the analysis is interpreted as a municipal system-level evaluation, restricting facilities to the Delft core keeps the destination set aligned with the Delft case-study boundary. If the analysis is interpreted from a road-user perspective, facilities just outside the municipal boundary may also be relevant, because residents can use nearby services beyond Delft when these services are reachable. The appropriate facility boundary therefore depends on whether the analysis represents administrative service provision or practical user access.

In this thesis, the boundary sensitivity analysis in this section is interpreted as a test of the road-network boundary. It informs the choice of network extent by showing how centrality-based priority orders respond to added surrounding road context. The use of the selected $3\,\mathrm{km}$ buffer in the subsequent local critical-service travel-time test is consistent with the purpose of the supplementary local checks. These checks are intended to provide a road-user-side interpretation of the five-criterion SAW ranking, so it is reasonable to include reachable services just outside the municipal boundary when these services may be relevant for Delft residents.

5.6. Local critical-service travel-time results

This section reports the local critical-service travel-time results from a road-user perspective. The test is supplementary to the Delft SAW ranking. It does not replace the system-level ranking, because it focuses only on selected residential origins and critical-service destinations. Instead, it examines whether the two highest-ranked Delft strategies, betweenness and closeness, produce different local access outcomes when travel time to nearby critical services is considered.

The analysis separates the aggregated critical-accessibility criterion (C2) into four service groups: emergency, medical, food, and water services. For each residential origin and each service group, the travel time is measured to the nearest reachable facility in that group. Median travel times are calculated over reachable observations, while cases where no facility can be reached are reported separately as unreachable shares.

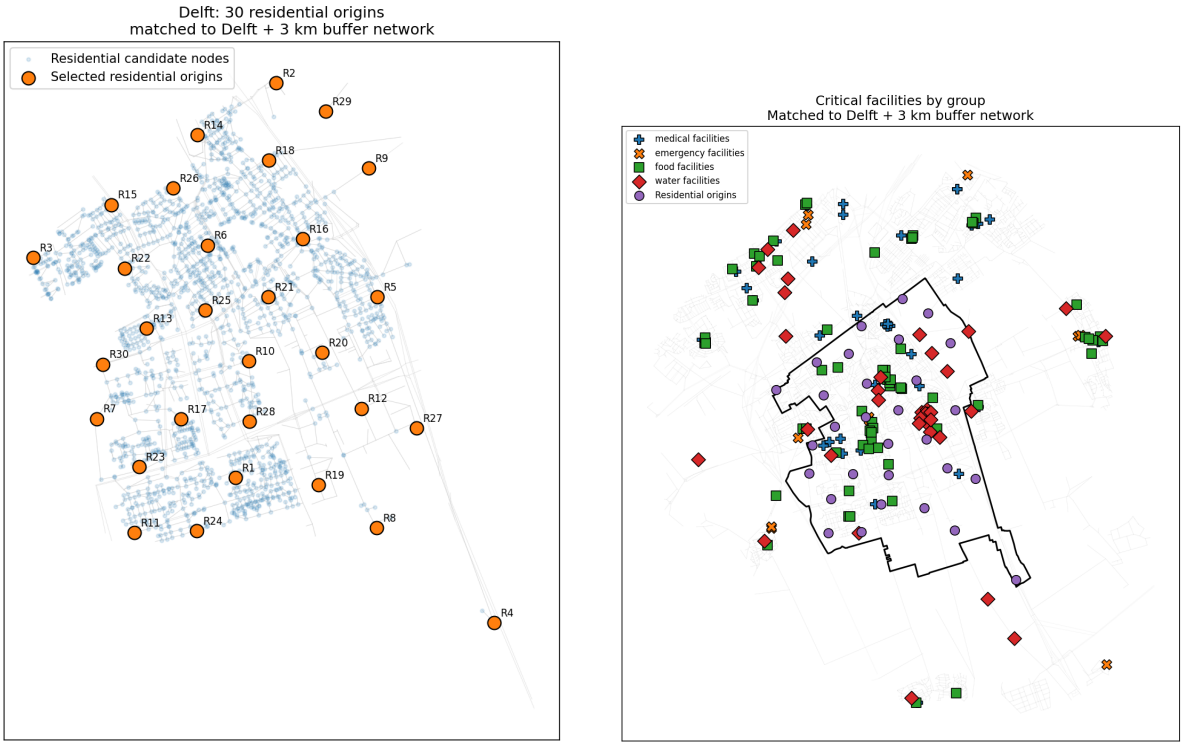
5.6.1. Test scope and facility boundary

The local access test uses the Delft plus $3\,\mathrm{km}$ buffer network selected in Section 5.5. The residential origins remain inside the Delft municipal boundary, while reachable facilities within the selected buffer

are included as possible destinations. This differs from the boundary sensitivity analysis in Section 5.5, which only examined the effect of road-network extent on centrality-based restoration orders.

Including facilities within the selected buffer is appropriate for this road-user-side test because residents may use nearby emergency, medical, food, or water services outside the municipal boundary if these services are reachable. The test therefore treats the $3\,\mathrm{km}$ buffer as both surrounding road-network context and a local service catchment. The comparison remains focused on Delft residents because all origins are located inside Delft.

Figure 5.8a shows the selected residential origins used in the test, and Figure 5.8b shows the critical-service facilities included within the Delft plus $3\,\mathrm{km}$ buffer.



(a) Thirty spatially balanced residential origins.

(b) Critical facilities grouped into emergency, medical, food, and water services.

Figure 5.8: Origins and destinations used in the Delft case study, matched to the Delft plus $3\,\mathrm{km}$ buffer road network.

5.6.2. Service-group results under 30% damage

The detailed service-group results are first reported for the $30\,\%$ damage level at $30\,\%$ recovery progress. This recovery progress value is used because it is the first available simulation output after the $25\,\%$ recovery-stage boundary defined in the main analysis.

Figure 5.9 shows the median travel time from the residential origins to the nearest reachable critical facility within each service group. Across all four critical-service groups, the closeness-based strategy produces lower median travel times than the betweenness-based strategy. The largest difference is found for emergency services, where the median travel time is 3.548 minutes under betweenness and 3.022 minutes under closeness, a difference of 0.526 minutes.

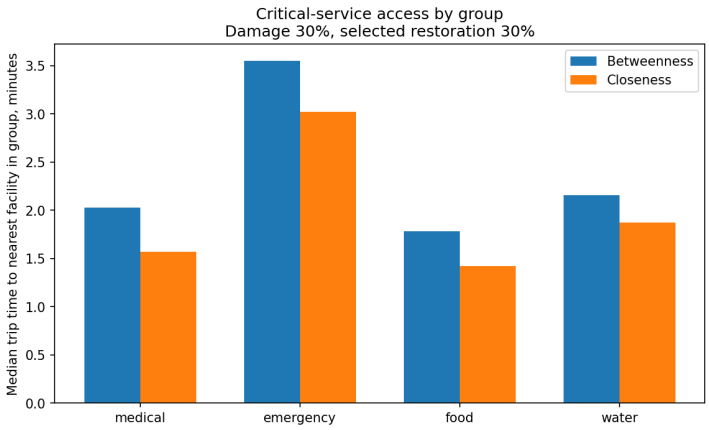


Figure 5.9: Median travel time to the nearest reachable critical facility by service group under the $30\,\%$ damage level when $30\,\%$ of damaged links had been restored.

Figure 5.10 shows the unreachable share by critical-service group. The unreachable share represents cases where no facility in the same service group can be reached from a residential origin under the damaged and partially restored network. The difference between the two strategies is small for emergency and food access. For medical and water access, the closeness-based strategy produces a lower unreachable share. The clearest difference is found for water access, where the unreachable share is 26.3% under betweenness and 23.0% under closeness.

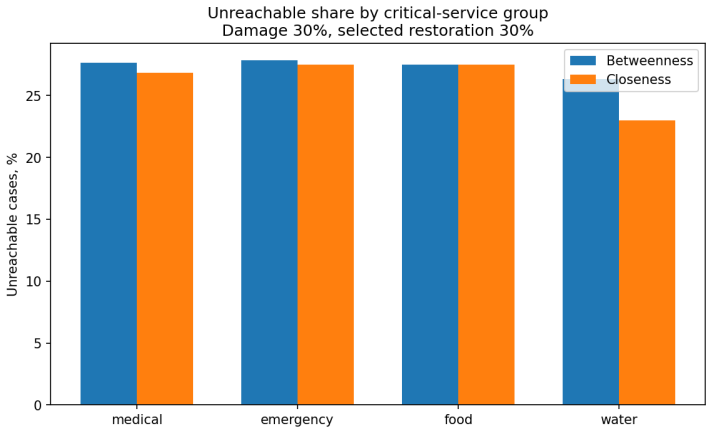


Figure 5.10: Share of unreachable residential-origin to critical-service-group cases when $30\,\%$ of damaged links had been restored.

5.6.3. Damage-level comparison

The $10\,\%$ and $30\,\%$ damage levels are compared using the same residential origins, critical-service destinations, Delft plus $3\,\mathrm{km}$ buffer network, recovery strategies, and recovery progress step. This comparison shows how the difference between betweenness and closeness changes when the initial damage level increases.

Table 5.6: Comparison of median local critical-service travel time and unreachable share at $30\,\%$ recovery progress under $10\,\%$ and $30\,\%$ damage levels.

Damage level	Service group	Median trip time (min)			Unreachable share (%)	
		Betw.	Close.	Difference Betw. – Close.	Betw.	Close.
$10\,\%$	Emergency	3.508	3.403	0.105	7.5	5.3
$10\,\%$	Food	1.728	1.578	0.150	7.5	5.3
$10\,\%$	Medical	1.914	1.742	0.172	7.3	5.3
$10\,\%$	Water	1.830	1.830	0.000	7.0	4.2
$30\,\%$	Emergency	3.548	3.022	0.526	27.8	27.5
$30\,\%$	Food	1.783	1.419	0.364	27.5	27.5
$30\,\%$	Medical	2.026	1.568	0.458	27.7	26.8
$30\,\%$	Water	2.160	1.874	0.286	26.3	23.0

The median travel-time differences between the two strategies are smaller under the $10\,\%$ damage level than under the $30\,\%$ damage level. Under $30\,\%$ damage, closeness produces lower median travel times for all four critical-service groups. Under $10\,\%$ damage, the differences are smaller, and both strategies produce the same median travel time for water access. The unreachable share is also much lower under $10\,\%$ damage because fewer road links are initially damaged.

5.6.4. Interpretation relative to the SAW ranking

The local critical-service travel-time test gives a different view from the Delft SAW ranking. The SAW ranking evaluates strategies using five weighted criteria and shows that betweenness performs best overall in Delft. The local travel-time test focuses only on selected residential origins and nearest reachable critical services. Under this narrower road-user perspective, closeness produces shorter median travel times and, for some service groups, lower unreachable shares.

This result does not overturn the main SAW ranking. Instead, it shows that a strategy with the strongest system-level performance may not always produce the shortest travel time for selected local origin-service pairs. The Delft case study therefore illustrates why the supplementary local checks are useful: they help interpret the system-level ranking from a road-user perspective and reveal local access patterns that are not visible in the aggregated SAW score.

6

Discussion

This chapter discusses three main findings from the stakeholder survey, the cross-city analysis, and the Delft application:

- **Stakeholder priorities are stage-dependent.** Short-term recovery is dominated by access to critical facilities, whereas long-term recovery involves a broader balance between accessibility, equity, connectivity, and efficiency (Section 6.1).
- **Betweenness is the most frequently top-ranked system-level recovery strategy under the tested settings.** Across the cross-city simulations and the Delft SAW analysis, betweenness is most often ranked first, although this result remains conditional on the tested damage levels, stakeholder weights, recovery stages, and static priority-order assumptions (Section 6.2).
- **SAW rankings and selected local accessibility outcomes can favour different strategies.** In Delft, betweenness ranks first in the SAW evaluation, while the local road-user checks show that closeness can provide shorter travel times for selected origin-to-service trips. These analyses answer different recovery questions and provide complementary interpretations (Section 6.3).

The chapter then reflects on the evaluation design in Section 6.4, before discussing the practical relevance of the framework for road recovery decision-making in Section 6.5.

6.1. Stakeholder priorities across recovery-stage contexts

Respondents evaluated the criteria differently in the two recovery-stage contexts. In the short-term stage, access to critical facilities (C2) receives the highest weight in all three main stakeholder groups, ranging from 0.374 for technical experts to 0.501 for the local community (Table 4.5). In the long-term stage, the weight distributions become flatter, and no single criterion dominates.

Existing accessibility research explains why access to essential services becomes critical when a network is disrupted. Esmalian et al. examine disaster-related disruption in access to grocery stores [72], and Yuan et al. show that flooding can reduce access to dialysis centres, with medical-access disruptions continuing beyond the flood period itself [73]. These studies establish the consequences of losing such access. The present results add what stakeholders make of it: respondents placed critical-facility access above every other criterion in the short-term context, and their open-ended responses repeatedly referred to life safety, medical access, food, water, and essential services.

Zhang et al. also distinguish between recovery stages. Emergency response is evaluated through the number of independent pathways between the affected population and critical facilities, while long-term recovery is evaluated through independent pathways weighted by pre-disaster average daily traffic [13]. Their stage distinction is defined through researcher-selected indicators, whereas the distinction in this study emerges from stakeholder-derived weights. Both approaches nevertheless treat short-term and long-term recovery as distinct evaluation contexts.

The two approaches differ in how long-term recovery is represented. Zhang et al. place greater emphasis on pathways associated with pre-disaster traffic demand. Other restoration studies evaluate network recovery through total system travel time [25], or through total travel time and unmet demand [26]. The stakeholder-derived weights in this study do not point in a single direction on this question. Technical experts assign spatial equity their highest long-term weight (0.244), and the local community assigns it 0.198, while road users assign it their lowest long-term weight (0.132) and place more weight on access to critical and routine facilities (Table 4.5). **Alongside traffic-related network performance, the distribution of recovery benefits carries substantial weight for some stakeholder groups, and the groups do not agree on how much.**

The elicited weights are stage-specific stated preferences. They show how respondents prioritised recovery criteria under the survey scenarios, not how a specific city would prioritise recovery in practice. Once the damaged network, the affected communities, and the available repair resources are known, these priorities may change. **The differences between the two recovery-stage contexts indicate that one fixed set of stakeholder weights may not represent both short-term and long-term recovery priorities.**

6.2. Betweenness as the most frequently top-ranked system-level strategy

The results identify betweenness centrality as the most consistently top-ranked system-level recovery strategy under the tested settings. Across the cross-city simulations, betweenness is most often ranked first across stakeholder groups, recovery-stage contexts, and disruption levels. Betweenness also ranks first in the aggregate Delft SAW analysis under a more detailed representation of accessibility.

This finding extends the simulation-based recovery analysis by Balakrishnan and Stokkink, who evaluate road-network recovery across operational efficiency, accessibility, spatial equity, and connectivity, and show that recovery performance is multidimensional. In their analysis, no single recovery strategy consistently performs best across all dimensions [1]. The present thesis adds stakeholder-derived criterion weights to this performance evidence. It shows that the betweenness advantage remains visible in the aggregate SAW ranking, while the criterion-level results and selected stakeholder-weighted cases still reveal trade-offs with closeness.

The recovery strategies and stakeholder weights play different roles in the framework. The strategies define link-restoration orders based on individual network indicators, while the stakeholder weights determine how the resulting criterion-level performance is evaluated in the SAW aggregation. The final ranking should therefore be read as an evaluation outcome rather than a stated stakeholder choice. It reflects how the performance profile of betweenness is rewarded under the stakeholder-weighted evaluation.

The criterion-level deviation summaries help explain this pattern. Across all four cross-city settings reported in Tables E.1 and E.2, betweenness has the lowest mean deviation for operational efficiency and the cross-city accessibility criterion. This indicates the smallest average losses from the pre-disruption baseline before SAW normalisation and stakeholder weighting are applied. This performance profile follows from the logic of betweenness centrality, which gives priority to links that lie on many shortest paths. Restoring such links early can reconnect separated parts of the network and improve travel performance across many origin-destination relations. This mechanism is also consistent with vulnerability-oriented road-network studies showing that betweenness centrality is closely related to global-efficiency-based vulnerability [74]. The strategy is therefore strongly rewarded when accessibility-related criteria and operational efficiency carry substantial weight. **The consistent aggregate ranking performance of betweenness is mainly explained by its strength on accessibility and operational efficiency.**

The stage-dependent results further clarify why betweenness is usually rewarded in the aggregate SAW ranking. In the cross-city analysis, the weights on access to critical facilities (C2) and access to routine facilities (C3) are both applied to the same aggregate accessibility metric, so only their sum affects the SAW score (Section 3.6.2). In the short-term stage this sum is large, ranging from 49.3% for technical experts to 64.1% for the local community (Table 4.5). Roughly half to two thirds of the

stakeholder weight therefore falls on the single indicator on which betweenness performs best. In the long-term stage the same sum falls to between 40.9% and 51.1%, and the weight distributions become flatter across the five criteria, which makes the ranking more sensitive to criteria where betweenness is less dominant. This sensitivity is most visible in the 30% long-term scenario reported in Table 4.11. In this scenario, technical experts assign their highest weight to spatial equity (C4), and the betweenness plurality narrows to 48.4%, compared with 43.0% for closeness. Closeness also becomes the plurality choice for the emergency-responder group, which has only two valid long-term responses and should be read as exploratory. **The closeness results show that aggregate SAW rankings can make criterion-level trade-offs and stakeholder-weight sensitivity less visible.**

A broader distribution of weight benefits an alternative strategy only when that strategy performs relatively better on the criteria receiving more weight. This is also visible in the Delft aggregate SAW results, where betweenness performs strongly across most of the weighted criteria. The local interpretation of the Delft results is discussed in Section 6.3.

The strategy-ordering correlation supports the interpretation that betweenness and closeness represent distinct restoration logics. These two strategies warrant closer comparison because betweenness is the most frequently top-ranked strategy, while closeness is its main alternative in the cross-city results. Their link-priority orderings are only moderately correlated ($\rho = 0.35$), which indicates that they do not simply prioritise the same links. Their criterion-level strengths also differ. In the cross-city criterion-level results, betweenness performs better on aggregate accessibility and operational efficiency, while closeness has the lowest deviation on spatial equity and generally performs well on connectivity. The two strategies therefore have contrasting criterion strengths rather than being interchangeable variants of the same centrality-based approach. These contrasting strengths may provide a basis for future composite strategy design, which is discussed further in Section 7.5.3.

The betweenness finding is therefore a system-level result within the tested settings. It reflects the analysed networks, disruption levels, stakeholder-weight profiles, recovery-stage contexts, and fixed restoration orders used in this thesis. Translating these analytical rankings into operational repair planning requires additional local information, as discussed in Section 6.5.

6.3. System-level rankings and local accessibility interpretation

The Delft application provides a complementary city-level perspective on aggregate strategy rankings. In the cross-city analysis, accessibility is represented through an aggregate network-level measure. In Delft, access to critical facilities (C2) and access to routine facilities (C3) are represented separately using explicit facility destinations. This allows the two forms of accessibility to be interpreted separately rather than as a single aggregate network property.

In the aggregate Delft SAW results, betweenness ranks first across the main stakeholder groups, both disruption levels, and both recovery-stage contexts (Tables 5.1 and 5.2). At the criterion level in Delft, betweenness performs strongly on operational efficiency, access to critical facilities, access to routine facilities, and spatial equity, while closeness performs better on connectivity.

The local critical-service travel-time test adds a different result. It focuses on selected residential origins and measures travel time to the nearest reachable facility within each critical-service group. Under this narrower road-user perspective, closeness produces shorter median travel times than betweenness for all four critical-service groups at the 30% damage level. Under the 10% damage level, the differences are smaller, and both strategies produce the same median travel time for water access Table 5.6. The strategy ranked first in the aggregate SAW evaluation therefore does not necessarily provide the shortest travel times for the selected origin-to-facility trips.

This difference arises because the facility-based SAW analysis and the local checks answer different questions. The SAW analysis evaluates strategy performance across the full set of framework criteria, combining operational efficiency, access to critical facilities, access to routine facilities, spatial equity, and connectivity with stakeholder-derived weights. The local checks instead focus on selected residential origins and their nearest reachable critical services. The two analyses therefore provide complementary views: betweenness performs best in the aggregate SAW evaluation, while closeness performs better for some selected local accessibility outcomes. **The Delft results show that aggre-**

gate strategy rankings and selected local accessibility outcomes can favour different strategies because they represent different evaluation questions.

The arterial-versus-local dilemma introduced in Section 1.2 is reflected in the Delft results. A strategy that performs well at the system level and a strategy that provides better access for selected residents answer different recovery questions and should therefore be evaluated from different perspectives.

These local accessibility outcomes are, however, tied to the spatial setting in which they are measured. The boundary sensitivity analysis in Section 5.5 shows that this dependence differs by strategy. The betweenness-based restoration order within Delft remains broadly stable when surrounding road-network buffers are added, while the closeness-based order changes more clearly, because closeness depends on travel-time distances to all nodes in the analysed network. The selected local accessibility results for closeness should therefore be interpreted alongside this boundary sensitivity, since its restoration order changes more clearly than that of betweenness when the analysed network extent is expanded. This comparison is limited to restoration priority orders and does not quantify how the boundary choice would affect the SAW rankings or the travel-time results themselves. Extending the boundary test to the full evaluation chain could therefore be a useful direction for future work.

For this reason, the system-level ranking should be treated as a strategic comparison rather than as a complete description of local road-user access. When the framework is applied to a specific city, the aggregate SAW ranking should be complemented with local accessibility checks that reflect the relevant facility locations, residential origins, and network boundary. **System-level recovery rankings are useful for strategic comparison, but their local meaning depends on facility geography, selected residential origins, and the spatial boundary used to define the road network.**

6.4. Methodological reflection on the evaluation design

This thesis compares recovery strategies under controlled conditions rather than reproducing a specific hazard event. To interpret the results, it is necessary to separate the methodological choices that were varied from the design assumptions that were held fixed. The sensitivity analyses examined three methodological variations: the BWM solver, the consistency sample, and the wording of C1. The main weight patterns and ranking conclusions remained stable across these checks. The stakeholder weights were not compared with observed recovery decisions, and they should be read as stated preferences elicited under the survey scenarios, as discussed in Section 6.1. The rest of this section examines two design assumptions that were held fixed: the random link-failure model and static restoration orders.

The 10% and 30% disruption levels provide controlled conditions for comparing recovery strategies. Damaged links are selected randomly, which evaluates all strategies under a common disruption rule but does not reproduce the spatial pattern of any particular hazard, as described in Section 3.5. Real hazard impacts can be spatially structured, and flood-focused resilience studies show that flood effects on road-network serviceability vary substantially across urban zones [35]. Damage concentrated in one area may change the strategy ranking, because recovery performance may then depend more directly on which links are needed to restore access to the affected area. A flood-focused single-case study illustrates this possibility, finding that a recovery order based on critical trip paths outperformed a static betweenness-based order in the analysed network [75]. Because damage is generated randomly in the present analysis, it cannot show whether a critical-trip-path-based order would perform better under concentrated damage. Testing that would require a design that represents spatially structured damage explicitly.

Static restoration orders are the second fixed assumption. Each strategy determines its restoration order from the pre-disruption network, which keeps the repair logic constant across disruption scenarios and makes the strategies directly comparable. Recalculating centrality as the network changes can produce substantially different priority orders, and this difference can be particularly large for betweenness because it depends on the global structure of the network [76]. The betweenness pattern reported in this thesis therefore applies to fixed restoration orders. An adaptive variant that re-ranks the remaining damaged links during recovery is a relevant extension of the same centrality-based recovery logic, and is left for future work.

Taken together, these design choices clarify the type of claim that the results can support. The analysis identifies ranking patterns that recur across the analysed networks, recovery-stage contexts, stakeholder-weight profiles, and tested methodological variations. These patterns hold within a design based on random link failure and fixed restoration orders. Spatially structured damage and adaptive restoration are alternative modelling choices that may produce different ranking behaviour, and comparing them would show how damage representation and restoration logic shape recovery-strategy rankings.

6.5. Practical interpretation of the framework

The practical relevance of the framework lies in its ability to support early-stage comparison of road recovery prioritisation logics. In post-disaster recovery, decision-makers need to balance multiple objectives, and the framework makes the trade-offs between these objectives explicit by showing how strategy rankings change under different stakeholder priorities, disruption scenarios, and city networks. Whether this comparison is useful in practice depends on how the stakeholder weights and the recovery strategies relate to real recovery decisions. The stakeholder weights and the recovery strategies each correspond to a recognisable part of recovery practice, while the execution of repairs remains outside the framework.

The stakeholder weights correspond to the staged nature of recovery decisions. In the BWM survey, respondents weighted the recovery criteria separately for a short-term and a long-term recovery scenario. This stage framing has a practical basis, as road-recovery guidance distinguishes emergency repairs that restore essential traffic immediately after a disaster from permanent repairs that restore roads to pre-disaster conditions [77]. As discussed in Section 6.1, these weights should be interpreted as stage-specific stated preferences rather than as fixed priorities for every real recovery case.

The recovery strategies, in turn, can be understood as simplified representations of recovery priorities that also appear in practice. A travel-time-weighted betweenness strategy is closely related to the practical logic of reopening main corridors first. In urban road networks, betweenness centrality is concentrated on a small number of roads, and this concentration reflects the hierarchical structure of the road network [78]. Empirically, the streets at the top of such network hierarchies have been observed to carry around 80% of measured traffic flow, and the same top streets carry betweenness values that are orders of magnitude higher than those of other streets [79]. Betweenness computed at larger spatial scales also tends to concentrate on higher-class roads that support longer-distance movement [80]. Therefore, ordering road recovery by travel-time-weighted betweenness can be interpreted as an abstract version of the practical rule of restoring major roads and transport corridors first.

Another practical priority enters the framework in a different way. Reopening access to hospitals and other critical facilities is a central goal in road-recovery guidance [77]. In this framework, this priority is represented as an evaluation criterion, C2, rather than as a recovery strategy. It therefore affects which strategy receives a higher score, instead of directly defining the order in which roads are restored.

These practical interpretations do not extend to the execution of repairs. The six strategies rank the order in which links are restored, while real repair campaigns are also shaped by repair capacity and by how responsibilities are distributed across road authorities [64], [77].

The framework should therefore be understood as a decision-support layer rather than a direct operational scheduling model. Its value is to compare recovery logics and reveal trade-offs under controlled assumptions.

7

Conclusion and Future Work

This thesis developed a stage-specific MCDA framework for comparing post-disaster road recovery strategies. The central issue in post-disaster road recovery is how recovery should be prioritised when different recovery criteria, stakeholder perspectives, and recovery contexts may lead to different evaluation outcomes. In this study, the Best-Worst Method was used to derive stakeholder- and stage-specific criterion weights, while Simple Additive Weighting was used to combine these weights with simulation-based recovery performance scores. The framework was applied to six predefined recovery strategies. The analysis was carried out at two scales: a cross-city, system-level analysis to examine broad ranking patterns across many urban networks, and a detailed Delft city-level application using facility-based accessibility and local road-user checks.

The thesis addresses a gap in the road recovery and transport resilience literature. Existing studies have examined recovery sequencing, network performance, multidimensional resilience criteria, and MCDA methods, but these elements are less often connected in a framework that combines stakeholder-derived priorities, recovery-stage contexts, and the ranking of predefined recovery strategies. This study connects stakeholder priorities, stage-specific recovery evaluation, and simulated strategy performance, so that road recovery strategies can be compared across multiple criteria, recovery stages, and stakeholder perspectives.

Three main conclusions follow from this framework. First, **stakeholder priorities differ between the short-term and long-term recovery contexts.** In short-term recovery, the three main stakeholder groups assign clear importance to access to critical facilities. In long-term recovery, the criterion weights become more dispersed, and the highest-weighted criterion differs across stakeholder groups. Second, under the tested random-failure scenarios, static priority-order assumptions, and SAW aggregation framework, **betweenness centrality is the most frequently top-ranked system-level recovery strategy.** This pattern appears in both the cross-city analysis and the Delft facility-based SAW analysis. Third, **the Delft case study shows that system-level rankings require local accessibility interpretation.** Betweenness remains first in the system-level SAW ranking, while the local road-user checks show that closeness performs better for selected origin-to-service travel-time outcomes. The recovery strategy rankings in this thesis should therefore be understood as analytical results that support strategic comparison under controlled assumptions. In practical use, these rankings should be interpreted together with local accessibility checks and operational information.

7.1. Answering the research questions

This section returns to the research questions introduced in Chapter 1. It answers the four sub-questions in sequence and then addresses the main research question. Each answer is based on the main findings from the previous chapters, with the focus placed on the overall conclusions instead of repeating all details from the results tables.

7.1.1. Stakeholder priorities across recovery stages and groups

The answer to the first sub-question, ***How do stakeholder priorities for road-recovery resilience criteria differ across stakeholder groups and recovery stages?***, is that stakeholder priorities differ between the two recovery-stage contexts, and that group differences become more visible in long-term recovery than in short-term recovery. In short-term recovery, the three main stakeholder groups all assign the highest importance to access to critical facilities. Road users, the local community, and technical experts assign weights of 0.442, 0.501, and 0.374 to C2, respectively. This shows that, in the temporary repair context, the three main groups share a clear emphasis on critical access, even though they represent different roles.

In long-term recovery, the criterion weights become more dispersed, and the highest-weighted criterion differs across the three main stakeholder groups. Road users still assign the highest weight to C2, with a value of 0.279. For the local community, connectivity receives the highest weight, with a value of 0.227, although the differences among the five criteria are small. For technical experts, spatial equity receives the highest weight, with a value of 0.244. This indicates that, in the reconstruction context, priority differences between stakeholder groups become more visible than in short-term recovery.

The paired respondent-level results further support this stage-dependent interpretation. Based on the 51 respondents with valid BWM weights in both the short-term and long-term stages, the change from short-term to long-term priorities was not a simple sequential shift in which the second priority moves up to first. Only 17.2% of the long-term first-ranked cases came from criteria ranked second in the short-term stage, whereas 67.2% came from criteria ranked third, fourth, or fifth. This indicates that respondents did not carry over a stable priority order and simply move it down by one position. Instead, they reassessed the relative importance of the criteria under the long-term context. Therefore, the overall answer to the first sub-question is that short-term recovery shows a shared emphasis on access to critical facilities, while long-term recovery presents a more dispersed and more group-specific priority pattern, formed with reference to the specific recovery context. This supports the treatment of criterion weights by both stakeholder group and recovery stage in this study.

7.1.2. Using stakeholder-derived weights to rank recovery strategies

The answer to the second sub-question, ***How can stakeholder-derived criterion weights be used to rank predefined road recovery strategies based on simulated recovery performance?***, is that stakeholder-derived criterion weights can be translated into recovery strategy rankings through the BWM-SAW framework. BWM converts respondents' criterion comparisons in the short-term temporary repair and long-term reconstruction contexts into criterion weights. The simulation stream then generates recovery trajectories for each predefined recovery strategy and converts these trajectories into area-above-curve deviation scores. These deviation scores represent the average performance loss during the recovery process relative to the pre-disruption baseline. Lower values indicate better recovery performance.

In the SAW step, the simulation-based AAC deviation scores are converted into benefit-oriented normalised scores and then combined with the BWM-derived criterion weights. This gives each recovery strategy a stakeholder-weighted composite score. Since the SAW calculation is conducted separately by city, damage level, recovery stage, and stakeholder group, the framework produces multiple ranking results instead of reducing the analysis to one overall ranking. This design allows the study to examine whether the same set of predefined recovery strategies leads to different ranking outcomes under different stakeholder priorities and recovery-stage contexts. The framework therefore converts stakeholder- and stage-specific priorities into comparable strategy rankings, without reducing the analysis to a single overall result.

7.1.3. Cross-city ranking patterns across stakeholders, stages, and damage levels

The answer to the third sub-question, ***How do the resulting strategy rankings vary across road networks, stakeholder groups, recovery stages, and disruption levels in the cross-city, system-level analysis?***, is that betweenness centrality is the most frequently top-ranked system-level strategy, while ranking differences still vary by damage level, recovery stage, and stakeholder weights. In the 10% damage scenario, betweenness is the strategy most often ranked first for all stakeholder groups

in long-term recovery, and for all groups except the single-respondent infrastructure-manager group in short-term recovery. This indicates that, under the lower disruption level, different stakeholder weights do not substantially change the strategy most frequently selected as rank one.

In the 30% damage scenario, betweenness remains the most frequently top-ranked in most conditions, while closeness centrality becomes more competitive. This pattern is most visible in the 30% long-term condition. For the emergency responder group, the rank-one distribution shifts towards closeness, which ranks first in 45.7% of cities compared with 39.0% for betweenness. Spatial equity (C4) and connectivity (C5) together account for 50.7% of this group's long-term weight, and closeness performs better than betweenness on both criteria. This group has only two valid long-term responses, so the result should be treated as exploratory. For technical experts in the same condition, closeness also becomes close to betweenness, although the ranking pattern does not fully reverse. Betweenness still ranks first in 48.4% of cities, while closeness ranks first in 43.0%.

The ranking differences between stakeholder groups should therefore be interpreted as moderate, rather than as strong disagreement. The within-city comparison shows that different stakeholder groups select the same rank-one strategy in about 50% to 65% of cases. This indicates moderate agreement between stakeholder groups, but not a high level of consistency. The stage-switching analysis also shows that the rank-one strategy can change between short-term and long-term recovery, and that this switching is more visible under the 30% damage condition. Therefore, the overall answer to the third sub-question is that betweenness is most often ranked first in the cross-city analysis, but closeness becomes an important conditional alternative under higher damage levels and in the long-term recovery stage. In most cases, different stakeholder weights do not fully change the top-ranked strategy, but they affect how competitive closeness becomes in specific conditions.

7.1.4. Additional insights from the Delft case study

The fourth sub-question asks, ***How do recovery strategies compare in a detailed city-level application, using Delft as a case study?*** Examining the patterns in more detail at the city level yields two main insights: betweenness also ranks first in the Delft facility-based SAW analysis, while the local road-user checks reveal that selected origin-to-service outcomes can favour a different strategy. The Delft case study separates access to critical facilities and access to routine activities through facility-based accessibility measures. In the Delft SAW analysis, betweenness ranks first across the three main stakeholder groups, two damage levels, and two recovery stages, while closeness consistently ranks second.

However, the Delft criterion-level results also show that critical accessibility and routine accessibility are not identical recovery outcomes. When the two facility-based accessibility measures are calculated separately, differences can be observed in their AAC deviation values and recovery patterns. This means that the aggregate accessibility metric is suitable for cross-city comparison, but it can hide local recovery differences between destination types. The Delft case study therefore does more than arrive at the same first-ranked strategy as the cross-city analysis. It also shows why the representation of accessibility matters for local interpretation.

The local road-user checks further show that the system-level best strategy is not necessarily the best strategy for all local accessibility outcomes. Under the 30% damage condition, closeness performs better for the selected origin-to-service median travel-time outcomes across all four critical service categories. These checks provide a local accessibility perspective rather than another five-criterion SAW evaluation. Therefore, the overall answer to the fourth sub-question is that the detailed city-level application returns the same top-ranked strategy in the Delft SAW ranking, while also showing that system-level rankings should be complemented by local accessibility checks, especially when the concern is access from specific residential origins to critical services.

7.1.5. Answer to the main research question

The answer to the main research question,

> ***How can post-disaster road recovery strategies be prioritised through a stage-specific MCDA framework, given that resilience is multidimensional, stakeholders have different priorities, and recovery decisions evolve over time?***

is that post-disaster road recovery strategies can be prioritised by combining stage-specific stakeholder weights with simulated strategy performance. BWM provides criterion weights that are specific to stakeholder group and recovery stage, while SAW combines these weights with normalised simulation-based performance scores and converts them into recovery strategy rankings. This procedure allows road recovery strategies to be compared under multidimensional resilience criteria, different stakeholder priorities, and different recovery-stage contexts.

These rankings should be understood as analytical results that support strategic comparison under controlled assumptions. They show the relative performance of different recovery-priority logics across multiple criteria, recovery stages, and stakeholder perspectives. They also help decision-makers see the difference between system-level performance and local accessibility interpretation. The results of this study show that, under the tested random-failure scenarios, static priority-order assumptions, and SAW aggregation framework, betweenness centrality is the most frequently top-ranked system-level strategy. However, closeness centrality becomes more competitive in higher-damage long-term contexts and in local origin-to-service travel-time outcomes.

Therefore, the overall answer to the main research question is that road recovery strategies are better compared through the integration of stakeholder-specific weights, stage-specific evaluation windows, and simulation-based strategy performance than through a single performance indicator or a fixed priority rule. The framework is intended for strategic comparison rather than operational scheduling, and its rankings still need to be interpreted together with local accessibility checks, hazard-specific damage information, repair resources, and institutional constraints. Within this scope, the framework offers decision-makers a structured way to compare recovery-priority logics and to see where system-level performance and local accessibility interpretation point in different directions.

7.2. Scientific contributions

The following subsections describe the contribution in three respects: the framework itself as a method, the case for eliciting stakeholder priorities separately by recovery stage, and what a system-level ranking requires before it can support a decision in a specific city.

7.2.1. A stage-specific BWM-SAW framework for comparing recovery strategies

The central methodological contribution of this thesis is a structure that combines stakeholder priorities with simulation-based recovery performance to compare recovery strategies. The Best-Worst Method converts stakeholder judgements into criterion weights that are specific to each stakeholder group and recovery stage. Simple Additive Weighting then combines these weights with the performance scores of each strategy on the evaluation criteria to produce recovery strategy rankings. Through this structure, stakeholder priorities and differences between recovery stages are no longer treated only as background considerations but become inputs to a quantitative evaluation. This is a concrete difference from existing methods, since current road recovery and transport resilience methods rarely use stakeholder-derived priorities directly to compare recovery strategies across several criteria.

This evaluation logic is not tied to a single spatial granularity, nor to any particular set of candidate strategies. In principle, the same structure can be used for other cities, hazards, evaluation criteria, recovery strategies, or data settings. In this thesis, it is demonstrated using six predefined centrality-based strategies, random 10% and 30% disruption scenarios, and two applications at the cross-city and Delft scales. Applying the framework to other settings has clear preconditions. The performance of each strategy on the selected criteria must be measurable, and applying the framework to real-world recovery strategies or hazard-specific disruptions would require further specification and validation, which is left for future work.

7.2.2. Stage-specific stakeholder-priority elicitation

Stakeholder priorities are stage-dependent, and eliciting them separately by recovery stage reveals differences that a fixed set of weights would conceal. This approach has methodological value, because respondents evaluated the criteria differently in the two contexts, and the paired respondent-level analysis shows that the long-term priorities were reassessed rather than obtained by moving the short-term order down by one position (see Section 4.2.3).

This indicates that, for future post-disaster transport recovery studies, eliciting priorities separately by recovery stage is worth considering, rather than assuming that priorities remain fixed throughout recovery.

7.2.3. Applying a system-level ranking to a city requires local context

Local context enters the analysis at two distinct points. A system-level ranking fixes both of them in advance, without reference to any particular city.

The first point is the construction of the restoration order. A centrality measure that depends on the global structure of a network changes when the extent of that network changes, so the order a strategy produces is not independent of where the boundary is drawn. Choosing that boundary is a local decision, and different strategies are not equally sensitive to it. The Delft boundary sensitivity analysis shows how far this can go. Restoration orders were recalculated for the Delft-only network and for the 1 km, 3 km, and 5 km buffer networks, and the closeness-based order changes more clearly across the four extents than the betweenness-based order, which is consistent with closeness depending on travel-time distances to all nodes in the analysed network. This comparison is limited to restoration orders and does not recompute the SAW scores.

The second point is the evaluation of the outcome. A SAW score is a weighted aggregate across the criteria and across the network, so it does not represent any single outcome at any particular location. A strategy can lead the aggregate while another performs better on a narrower outcome that a city cares about. In Delft, betweenness remains first in the five-criterion SAW ranking, while under 30% damage at 30% recovery progress, closeness produces shorter median travel times from the selected residential origins to the nearest reachable facility in all four critical-service groups.

The ranking is designed to compare strategies across networks, and that comparison is made without local information. Before a system-level ranking is used to support a decision in a specific city, the local context has to be supplied: the extent of the network to be analysed, the origins and destinations that matter, and the outcomes the city is trying to protect. The checks used in this thesis are one way of supplying it.

7.3. Practical implications

In practical use, the framework does three things for the stakeholders involved in a recovery decision.

It makes local stakeholder priorities visible. When the framework is applied to a particular city, criterion weights can be elicited from the stakeholders involved in that recovery context. Short-term emergency repair and long-term reconstruction are assessed separately so that stage-specific differences in the relative importance of the recovery criteria can be represented. This avoids assuming that different cities, stakeholder groups, or recovery stages share one set of weights. Before recovery strategies are compared, planners can therefore identify which recovery objectives local stakeholder groups prioritise in similar ways and where their priorities differ. The weights used in this thesis demonstrate this stage-specific elicitation and evaluation process.

It shows what those priorities imply for strategy evaluation. The framework combines stakeholder-derived weights with the simulated performance of each strategy. Strategy rankings can therefore be compared across stakeholder groups, recovery-stage contexts, and damage levels. These comparisons show which strategies are most frequently ranked first across the tested settings and which alternatives become more competitive under particular conditions. Even when different weighting profiles produce the same first-ranked strategy, the score margins and criterion-level performance can show whether the result is clear or whether another strategy is close behind. The framework can therefore support the comparison and initial screening of recovery-priority logics before more detailed planning begins.

It gives stakeholder disagreement a traceable source. Because the framework links criterion weights, criterion-level strategy performance, and final SAW rankings, differences in the evaluation can be traced back to their sources. When different stakeholder groups produce different rankings, the analysis can show which criteria contributed to that difference. The same applies when the first-ranked strategy remains unchanged but an alternative becomes more competitive under another weighting profile. This

makes trade-offs among the selected recovery criteria, such as accessibility, spatial equity, connectivity, and operational efficiency, easier to recognise and discuss. Stakeholder discussion can therefore focus on which recovery objectives carry the most weight and whether a different weighting of those objectives would change the evaluation outcome, rather than on the ranking alone. The discussion can also extend beyond the framework itself, to whether the criteria being used are appropriate for the local recovery context.

For city-level use, the system-level SAW ranking should be interpreted alongside local accessibility checks. A strategy may rank first in the overall SAW evaluation while another strategy provides shorter travel times from particular residential areas to hospitals or other essential facilities. When the two analyses favour different strategies, neither result is wrong, because they answer different evaluation questions. The difference reveals a trade-off between overall weighted performance and accessibility for specific locations, and it gives stakeholder discussion a concrete focus: how system-wide recovery should be balanced against access from particular residential areas to essential facilities.

The practical value of the framework lies in making the path from stakeholder priorities to strategy rankings visible, so that the resulting trade-offs can be compared and discussed explicitly.

7.4. Limitations

These limitations define the scope within which the findings should be interpreted. They describe the assumptions and settings under which the study was carried out, rather than weakening the reliability of its results.

7.4.1. Scope of the evaluation design

This thesis uses a controlled comparison of recovery strategies under common assumptions. Its aim is to identify broad ranking patterns across many networks and different stakeholder weights, rather than to reproduce a specific hazard event. It also does not produce an operational reconstruction plan that could be carried out directly, and the execution of the actual repair work is outside its scope.

7.4.2. Simplifications in damage and recovery modelling

For the damage side, disruption is represented by random link failure, and the 10% and 30% damage levels are controlled comparison scenarios. Real hazard impacts can be spatially correlated, since damage from floods, earthquakes, landslides, or storm surge often concentrates in specific areas, and this is not directly modelled.

For the recovery side, the study makes several simplifications to the recovery process:

- Each strategy uses a static priority order computed before the disruption, rather than an adaptive order updated once the actual damage is known.
- The survey describes short-term and long-term recovery in terms of temporary repair and reconstruction, while the simulation represents them through progress and performance trajectories, so the two do not correspond exactly.
- The simulation does not directly model repair duration, construction methods, contracting, crew availability, cost, or resource allocation.

7.4.3. Stakeholder, indicator, and MCDA-related limitations

For the stakeholder data, the BWM weights are stated preferences rather than observed institutional decisions. The sample is uneven across stakeholder groups, some of the groups are very small or represented by a single respondent, and the respondents are geographically concentrated, mainly among Taiwan-based participants. The same set of stakeholder weights is also applied across all cities.

The elicitation design also places a limit on how the stage comparison can be read. The two recovery stages were always presented in the same order, with short-term before long-term. Because this order was not counterbalanced, part of the observed shift in priorities may reflect the order in which respondents answered, rather than the recovery stage itself.

For the indicators, access to critical facilities (C2) and access to routine activities (C3) share a single aggregate accessibility metric in the cross-city analysis, so only their combined weight affects the cross-city SAW ranking. The Delft case study addresses this by calculating the two accessibility criteria separately, but only for a single city.

For the MCDA method, SAW is compensatory, so strong performance on one criterion can offset weak performance on another. The rankings should therefore be read as overall weighted performance, rather than as evidence that a strategy performs well on every individual criterion.

7.4.4. Limitations of the Delft case study

The Delft case study is a single-city local case study. It provides a more detailed local reading rather than a basis for generalisation to other cities. The selection of facilities, the selection of residential origins, and the choice of network boundary all affect the local outcomes. These local checks are supplementary and are not part of the five-criterion SAW ranking itself. They are measured through network accessibility, treating each residential origin as an equally weighted point, and do not account for the population size of residential locations or individual-level factors such as vehicle ownership.

7.5. Future research

The framework and findings of this thesis point to several directions for future research. These directions extend from the scope, assumptions, and results of this study, and can help further develop the use of MCDA for post-disaster road recovery.

7.5.1. Broader stakeholder engagement and validation

The first direction is to broaden and deepen stakeholder engagement. The BWM weights in this study were derived from a survey, but the sample was uneven across stakeholder groups and geographically concentrated. Future research could include a larger and more balanced sample, covering more emergency managers, road authorities, infrastructure operators, logistics operators, and residents. This would help develop stakeholder-priority inputs that are more balanced across different groups.

Future studies could also combine BWM with interviews, workshops, or participatory planning. These methods could provide a deeper understanding of how stakeholders form their priorities in specific recovery contexts. Further research could also compare stated preferences from surveys with actual or historical recovery decisions. This would help examine whether the priorities derived from stakeholder surveys are consistent with the priorities used in real recovery decision-making.

7.5.2. Hazard-specific and adaptive recovery modelling

The second direction is to make the evaluation closer to specific hazards and realistic recovery processes. This can be developed in three ways.

First, future research could replace random link failure with hazard-specific and spatially correlated damage inputs. For example, flood, earthquake, landslide, or storm surge scenarios could be combined with exposure and vulnerability data. This would make it possible to examine how strategy rankings change when damage is spatially concentrated rather than randomly distributed across the network.

Second, future research could develop adaptive recovery strategies. In this thesis, the six recovery strategies are based on static priority orders calculated before the disruption. Future studies could allow the repair order to be updated after the actual damage is known, or recalculated after each repair step. This would better reflect situations in which decision-makers adjust recovery actions as new information becomes available.

Third, future research could include more operational constraints and planning conditions in the evaluation. This thesis focuses on how restoration priority orders affect network performance. Future work could include repair duration, crew availability, repair cost, road ownership, and road hierarchy. Including these factors would make the framework closer to practical recovery planning.

7.5.3. Composite centrality-based recovery strategies

The third direction is to develop composite centrality-based recovery strategies. The criterion-level results in this thesis suggest that betweenness and closeness represent complementary recovery logics. Their link-priority orders show only a moderate level of correlation, with $\rho = 0.35$. Betweenness appears to perform strongly in relation to accessibility and operational efficiency, while closeness appears to perform strongly in relation to spatial equity and connectivity.

This complementarity makes betweenness and closeness promising candidates for future composite centrality-based recovery strategies. Such a composite strategy could combine indicators that are both competitive and complementary, instead of including all centrality measures without distinction. Future research could also develop stage-specific composite scores. For example, betweenness could be used to capture the accessibility and operational-efficiency logic, while closeness could be used to add dimensions that betweenness reflects less directly.

7.5.4. Extension of the MCDA framework

The fourth direction is to extend the MCDA method itself. This thesis uses SAW because it is transparent and easy to interpret. However, SAW is a compensatory method, meaning that strong performance on one criterion can offset weaker performance on another criterion. Future research could compare SAW with other aggregation methods, including non-compensatory or partially compensatory approaches. This would help test whether the strategy rankings change under different aggregation assumptions.

Future research could also include uncertainty analysis for both stakeholder weights and simulation outputs. Stakeholder weights may vary depending on sample composition, survey design, or recovery context. Simulation outputs may also change under different assumptions about damage, repair speed, or network conditions. Testing these uncertainties would help show how sensitive the rankings are to changes in the inputs, and would make the interpretation of MCDA results more robust.

7.6. Closing remarks

The framework developed in this thesis helps make the trade-offs between stakeholder priorities, recovery stages, and recovery-performance dimensions clearly visible. Under controlled assumptions, it provides a stage-specific way to compare post-disaster road recovery strategies. The resulting rankings should be understood as a starting point for recovery decision support, rather than as a final answer.

Looking ahead, we hope that this framework can serve as a basis for future related studies by helping stakeholder priorities and recovery stages be incorporated more systematically into road recovery evaluation. It can also support decision-makers in practice by helping them more clearly see and weigh the trade-offs between multiple recovery objectives.

References

[1] S. Balakrishnan and P. S. A. Stokkink, “Trade-offs in road network recovery: Analyzing resilience, efficiency, and connectivity from user perspective,” *Transportation Research Procedia*, 2026, World Conference on Transport Research (WCTR 2026), Toulouse.

[2] M. Zamanifar and S. M. Seyedhoseyni, “Recovery planning model for roadways network after natural hazards,” *Natural Hazards*, vol. 87, no. 2, pp. 699–716, 2017.

[3] A. Gajanayake, T. Khan, and G. Zhang, “Post-Disaster Reconstruction of Road Infrastructure: Decision making processes in an Australian context,” en, *European Journal of Transport and Infrastructure Research*, vol. 20, no. 1, pp. 1–16, Feb. 2020, ISSN: 1567-7141. DOI: 10.18757/ejtir.2020.20.1.3919. Accessed: Feb. 1, 2026. [Online]. Available: https://journals.open.tudelft.nl/ejtir/article/view/3919.

[4] W. Zhang, N. Wang, and C. Nicholson, “Resilience-based post-disaster recovery strategies for road-bridge networks,” *Structure and Infrastructure Engineering*, vol. 13, no. 11, pp. 1404–1413, Nov. 2017, _eprint: https://doi.org/10.1080/15732479.2016.1271813, ISSN: 1573-2479. DOI: 10.1080/15732479.2016.1271813. Accessed: Feb. 1, 2026. [Online]. Available: https://doi.org/10.1080/15732479.2016.1271813.

[5] United Nations Office for Disaster Risk Reduction and Centre for Research on the Epidemiology of Disasters, *The human cost of disasters: An overview of the last 20 years (2000–2019)*, Accessed 2026-07-09, 2020. [Online]. Available: https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019.

[6] M. J. Benavidez and A. M. Mortlock, “Transport sector recovery: Opportunities to build resilience,” World Bank Group, Global Facility for Disaster Reduction, and Recovery, Washington, DC, Disaster Recovery Guidance Series, Oct. 2018. [Online]. Available: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/624471541689229858/transport-sector-recovery-opportunities-to-build-resilience.

[7] K. Lucas, “Transport and social exclusion: Where are we now?” *Transport policy*, vol. 20, pp. 105–113, 2012.

[8] P. Roberts, S. Kc, and C. Rastogi, “Rural access index: A key development indicator,” Tech. Rep., 2006.

[9] A. Rohr, P. Priesmeier, K. Tzavella, and A. Fekete, “System criticality of road network areas for emergency management services—spatial assessment using a tessellation approach,” *Infrastructures*, vol. 5, no. 11, p. 99, 2020.

[10] E. E. Koks et al., “A global multi-hazard risk analysis of road and railway infrastructure assets,” *Nature communications*, vol. 10, no. 1, p. 2677, 2019.

[11] P. Cartes, T. E. Navarro, A. C. Giné, and E. A. Binet, “A cost-benefit approach to recover the performance of roads affected by natural disasters,” *International Journal of Disaster Risk Reduction*, vol. 53, p. 102 014, 2021.

[12] F. Wei, E. Koc, N. Li, L. Soibelman, and D. Wei, “A data-driven framework to evaluate the indirect economic impacts of transportation infrastructure disruptions,” *International Journal of Disaster Risk Reduction*, vol. 75, p. 102 946, 2022.

[13] W. Zhang, N. Wang, C. Nicholsonc, and M. H. Tehrani, “A stage-wise decision framework for transportation network resilience planning,” *arXiv preprint arXiv:1808.03850*, 2018.

[14] L. S. Almeida, F. Goerlandt, and R. Pelot, “Trends and gaps in the literature of road network repair and restoration in the context of disaster response operations,” *Socio-Economic Planning Sciences*, vol. 84, p. 101 398, 2022.

[15] A. Puri, M. Elkharboutly, and N. A. Ali, “Identifying major challenges in managing post-disaster reconstruction projects: A critical analysis,” *International Journal of Disaster Risk Reduction*, vol. 107, p. 104 491, 2024, ISSN: 2212-4209. DOI: https://doi.org/10.1016/j.ijdrr.2024.104491. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221242092400253X.

[16] A. Kaviani, R. G. Thompson, A. Rajabifard, and M. Sarvi, “A model for multi-class road network recovery scheduling of regional road networks,” en, *Transportation*, vol. 47, no. 1, pp. 109–143, Feb. 2020, ISSN: 1572-9435. DOI: 10.1007/s11116-017-9852-5. Accessed: Feb. 1, 2026. [Online]. Available: https://doi.org/10.1007/s11116-017-9852-5.

[17] D. Rey and H. Bar-Gera, “Long-term scheduling for road network disaster recovery,” *International journal of disaster risk reduction*, vol. 42, p. 101 353, 2020.

[18] U. Bhatia, L. Sela, and A. R. Ganguly, “Hybrid method of recovery: Combining topology and optimization for transportation systems,” *Journal of Infrastructure Systems*, vol. 26, no. 3, p. 04 020 024, 2020.

[19] A. Kaviani, R. G. Thompson, and A. Rajabifard, “Improving regional road network resilience by optimised traffic guidance,” *Transportmetrica A: Transport Science*, vol. 13, no. 9, pp. 794–828, 2017.

[20] J. Wang and H. Liu, “Snow removal resource location and allocation optimization for urban road network recovery: A resilience perspective,” en, *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 1, pp. 395–408, Jan. 2019, ISSN: 1868-5145. DOI: 10.1007/s12652-018-0717-3. Accessed: Feb. 1, 2026. [Online]. Available: https://doi.org/10.1007/s12652-018-0717-3.

[21] J. García-Alviz, G. Galindo, J. Arellana, and R. Yie-Pinedo, “Planning road network restoration and relief distribution under heterogeneous road disruptions,” *OR Spectrum*, vol. 43, no. 4, pp. 941–981, 2021.

[22] J. Lin, Q. Hu, W. Lin, and M. Tan, “Risk and resilience-based restoration optimization of transportation infrastructures under uncertainty,” *PLoS one*, vol. 19, no. 8, e0308138, 2024.

[23] X. Fan, X. Zhang, X. Wang, and X. Yu, “A deep reinforcement learning model for resilient road network recovery under earthquake or flooding hazards,” *Journal of Infrastructure Preservation and Resilience*, vol. 4, no. 1, p. 8, 2023.

[24] S. A. Argyroudis, “Resilience metrics for transport networks: A review and practical examples for bridges,” in *Proceedings of the institution of civil engineers-bridge engineering*, Thomas Telford Ltd, vol. 175, 2022, pp. 179–192.

[25] C. Gokalp, P. N. Patil, and S. D. Boyles, “Post-disaster recovery sequencing strategy for road networks,” *Transportation research part B: methodological*, vol. 153, pp. 228–245, 2021.

[26] T. Zhao and Y. Zhang, “Transportation infrastructure restoration optimization considering mobility and accessibility in resilience measures,” *Transportation Research Part C: Emerging Technologies*, vol. 117, p. 102 700, 2020.

[27] M. Z. Serdar, M. Koç, and S. G. Al-Ghamdi, “Urban transportation networks resilience: Indicators, disturbances, and assessment methods,” *Sustainable Cities and Society*, vol. 76, p. 103 452, 2022.

[28] N. Nickdoost, M. J. Shooshtari, J. Choi, D. Smith, and Y. Abdelrazig, “A composite index framework for quantitative resilience assessment of road infrastructure systems,” *Transportation research part D: transport and environment*, vol. 131, p. 104 180, 2024.

[29] K. Tierney, “Disaster governance: Social, political, and economic dimensions,” *Annual review of Environment and Resources*, vol. 37, pp. 341–363, 2012.

[30] A. Cho, “Post-tsunami recovery and reconstruction: Governance issues and implications of the great east japan earthquake,” *Disasters*, vol. 38, no. s2, s157–s178, 2014.

[31] S. Greco, R. Słowiński, and J. Wallenius, “Fifty years of multiple criteria decision analysis: From classical methods to robust ordinal regression,” *European Journal of Operational Research*, vol. 323, no. 2, pp. 351–377, 2025.

[32] Z. Mohammadnazari, M. Mousapour Mamoudan, M. Alipour-Vaezi, A. Aghsami, F. Jolai, and M. Yazdani, “Prioritizing post-disaster reconstruction projects using an integrated multi-criteria decision-making approach: A case study,” *Buildings*, vol. 12, no. 2, p. 136, 2022.

[33] S. A. Morshed, M. Arafat, S. Mokhtarimousavi, S. S. Khan, and K. Amine, “8R Resilience Model: A stakeholder-centered approach of disaster resilience for transportation infrastructure and network,” *Transportation Engineering*, vol. 4, p. 100 058, Jun. 2021, ISSN: 2666-691X. DOI: 10.1016/j.treng.2021.100058. Accessed: Feb. 1, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666691X21000142.

[34] W. Xu, Y. Zhang, D. Proverbs, and Z. Zhong, “Enhancing the resilience of road networks to flooding,” English, *International Journal of Building Pathology and Adaptation*, vol. 42, no. 2, pp. 213–236, 2024, ISSN: 23984708. DOI: 10.1108/IJBPA-09-2021-0120. Accessed: Feb. 1, 2026. [Online]. Available: https://www.proquest.com/docview/3032844826/abstract/763E7719807140A3PQ/1.

[35] M. M.S., Y. B. Katpatal, and D. S. Londhe, “Measurement of city road network resilience in hazardous flood events,” *International Journal of Disaster Resilience in the Built Environment*, vol. 15, no. 2, pp. 274–288, Nov. 2022, ISSN: 1759-5908. DOI: 10.1108/IJDRBE-11-2021-0155. Accessed: Feb. 1, 2026. [Online]. Available: https://doi.org/10.1108/IJDRBE-11-2021-0155.

[36] M. Zamanifar and T. Hartmann, “Decision attributes for disaster recovery planning of transportation networks; a case study,” *Transportation research part D: transport and environment*, vol. 93, p. 102 771, 2021.

[37] R. W. Saaty, “The analytic hierarchy process—what it is and how it is used,” *Mathematical modelling*, vol. 9, no. 3-5, pp. 161–176, 1987.

[38] J. Rezaei, “Best-worst multi-criteria decision-making method,” *Omega*, vol. 53, pp. 49–57, 2015, ISSN: 0305-0483. DOI: https://doi.org/10.1016/j.omega.2014.11.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0305048314001480.

[39] F. Ciardiello and A. Genovese, “A comparison between topsis and saw methods,” *Annals of Operations Research*, vol. 325, no. 2, pp. 967–994, 2023.

[40] S. Schär, E. Pohl, and J. Geldermann, “Analysing the compensatory properties of the outranking approach promethee,” *Journal of Multi-Criteria Decision Analysis*, vol. 32, no. 2, e70013, 2025.

[41] C. Macharis and A. Bernardini, “Reviewing the use of multi-criteria decision analysis for the evaluation of transport projects: Time for a multi-actor approach,” *Transport policy*, vol. 37, pp. 177–186, 2015.

[42] L. Scholten, A. Scheidegger, P. Reichert, M. Mauer, and J. Lienert, “Strategic rehabilitation planning of piped water networks using multi-criteria decision analysis,” *Water research*, vol. 49, pp. 124–143, 2014.

[43] X. B. Olba, W. Daamen, T. Vellinga, and S. P. Hoogendoorn, “Multi-criteria evaluation of vessel traffic for port assessment: A case study of the port of rotterdam,” *Case Studies on Transport Policy*, vol. 7, no. 4, pp. 871–881, 2019.

[44] Global Facility for Disaster Reduction and Recovery (GFDRR), “Transport sector recovery: Opportunities to build resilience,” World Bank Group, Washington, DC, Tech. Rep., 2020.

[45] J. L. Machado-León and A. Goodchild, “Review of performance metrics for community-based planning for resilience of the transportation system,” *Transportation Research Record*, vol. 2604, no. 1, pp. 44–53, 2017.

[46] N. Coleman, X. Li, T. Comes, and A. Mostafavi, “Weaving equity into infrastructure resilience research and practice: A decadal review and future directions,” *arXiv preprint arXiv:2310.19194*, 2023.

[47] V. Latora and M. Marchiori, “Efficient behavior of small-world networks,” *Physical review letters*, vol. 87, no. 19, p. 198 701, 2001.

[48] G. Fagiolo, “Clustering in complex directed networks,” *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 76, no. 2, p. 026 107, 2007.

[49] S. Yan and Y.-L. Shih, “A time-space network model for work team scheduling after a major disaster,” *Journal of the chinese institute of engineers*, vol. 30, no. 1, pp. 63–75, 2007.

[50] S.-S. Liu, M. F. A. Arifin, W. T. Chen, and Y.-H. Huang, “Emergency repair scheduling model for road network integrating rescheduling feature,” *Applied Sciences*, vol. 11, no. 4, p. 1447, 2021.

[51] M. Mojtahedi and B. L. Oo, “The impact of stakeholder attributes on performance of disaster recovery projects: The case of transport infrastructure,” *International Journal of Project Management*, vol. 35, no. 5, pp. 841–852, 2017.

[52] R. J. Yang, Y. Wang, and X.-H. Jin, “Stakeholders’ attributes, behaviors, and decision-making strategies in construction projects: Importance and correlations in practice,” *Project Management Journal*, vol. 45, no. 3, pp. 74–90, 2014.

[53] C. Macharis, L. Turcksin, and K. Lebeau, “Multi actor multi criteria analysis (MAMCA) as a tool to support sustainable decisions: State of use,” *Decision Support Systems*, vol. 54, no. 1, pp. 610–620, Dec. 2012, ISSN: 0167-9236. DOI: 10.1016/j.dss.2012.08.008. Accessed: Feb. 1, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167923612002047.

[54] Y. Chang, S. Wilkinson, E. Seville, and R. Potangaroa, “Resourcing for a resilient post-disaster reconstruction environment,” *International Journal of Disaster Resilience in the Built Environment*, vol. 1, no. 1, pp. 65–83, 2010.

[55] Veiligheidsregio Rotterdam-Rijnmond, “Beleidsplan 2023–2027: Samenwerken aan zorg en veiligheid,” Dutch, Veiligheidsregio Rotterdam-Rijnmond, 2023.

[56] J. Rezaei, “Best-worst multi-criteria decision-making method: Some properties and a linear model,” *Omega*, vol. 64, pp. 126–130, 2016.

[57] Q. Wu, X. Liu, L. Zhou, J. Qin, and J. Rezaei, “An analytical framework for the best–worst method,” *Omega*, vol. 123, p. 102 974, 2024.

[58] J. Rezaei, *BWM-Solver-4*, Excel solver, Best-Worst Method website, Accessed [12/07/2026]. [Online]. Available: https://bestworstmethod.com/software/.

[59] B. Srdjevic, Z. Srdjevic, K. M. Reynolds, M. Lakicevic, and S. Zdero, “Using analytic hierarchy process and best–worst method in group evaluation of urban park quality,” *Forests*, vol. 13, no. 2, p. 290, 2022.

[60] P. Mayring, “Qualitative content analysis,” *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [On-line Journal], http://qualitative-research.net/fqs/fqs-e/2-00inhalt-e.htm*, vol. 1, Jun. 2000.

[61] R. Bijker, S. S. Merkouris, N. A. Dowling, and S. N. Rodda, “Chatgpt for automated qualitative research: Content analysis,” *Journal of medical Internet research*, vol. 26, e59050, 2024.

[62] S. De Paoli, “Performing an inductive thematic analysis of semi-structured interviews with a large language model: An exploration and provocation on the limits of the approach,” *Social Science Computer Review*, vol. 42, no. 4, pp. 997–1019, 2024.

[63] J. Gao et al., “Collabcoder: A lower-barrier, rigorous workflow for inductive collaborative qualitative analysis with large language models,” in *Proceedings of the 2024 CHI conference on human factors in computing systems*, 2024, pp. 1–29.

[64] 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*, vol. 31, pp. 832–843, 2018.

[65] N. Vafaei, R. A. Ribeiro, and L. M. Camarinha-Matos, “Assessing normalization techniques for simple additive weighting method,” *Procedia Computer Science*, vol. 199, pp. 1229–1236, 2022.

[66] Y. Song, K. Wu, and D. Liu, “A two-stage simulation analysis of uncertain road damage on the urban emergency delivery network,” *PLoS one*, vol. 17, no. 5, e0267043, 2022.

[67] G. Boeing, “Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks,” *Computers, environment and urban systems*, vol. 65, pp. 126–139, 2017.

[68] C. Bach, A. K. Gupta, S. S. Nair, and J. Birkmann, “Critical infrastructures and disaster risk reduction,” *National Institute of Disaster Management and Deutsche Gesellschaft für internationale Zusammenarbeit GmbH (GIZ), New Delhi*, vol. 72, 2013.

[69] A. Kocatepe, M. B. Ulak, G. Kakareko, E. E. Ozguven, S. Jung, and R. Arghandeh, “Measuring the accessibility of critical facilities in the presence of hurricane-related roadway closures and an approach for predicting future roadway disruptions,” *Natural Hazards*, vol. 95, no. 3, pp. 615–635, 2019.

[70] S. A. Enderami, E. Sutley, J. Helgeson, L. Dueñas-Osorio, M. Watson, and J. W. van de Lindt, “Measuring post-disaster accessibility to essential goods and services: Proximity, availability, adequacy, and acceptability dimensions,” *Journal of Infrastructure Preservation and Resilience*, vol. 5, no. 1, p. 12, 2024.

[71] J. Gil, “Street network analysis “edge effects”: Examining the sensitivity of centrality measures to boundary conditions,” *Environment and Planning B: Urban Analytics and City Science*, vol. 44, no. 5, pp. 819–836, 2017.

[72] A. Esmalian, N. Coleman, F. Yuan, X. Xiao, and A. Mostafavi, “Characterizing equitable access to grocery stores during disasters using location-based data,” *Scientific reports*, vol. 12, no. 1, p. 20 203, 2022.

[73] F. Yuan, H. Farahmand, R. Blessing, S. Brody, and A. Mostafavi, “Unveiling dialysis centers’ vulnerability and access inequality during urban flooding,” *Transportation Research Part D: Transport and Environment*, vol. 125, p. 103 920, 2023.

[74] A. Furno, N.-E. El Faouzi, R. Sharma, V. Cammarota, and E. Zimeo, “A graph-based framework for real-time vulnerability assessment of road networks,” in *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*, IEEE, 2018, pp. 234–241.

[75] M. Z. Serdar and S. G. Al-Ghamdi, “Resilience-oriented recovery of flooded road networks during mega-sport events: A novel framework,” *Frontiers in Built Environment*, vol. 9, p. 1 216 919, 2023.

[76] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, “Attack vulnerability of complex networks,” *Physical review E*, vol. 65, no. 5, p. 056 109, 2002.

[77] Federal Highway Administration, *Emergency Relief Manual (Federal-Aid Highways)*, Updated May 31, 2013, Office of Infrastructure, Office of Program Administration, May 2013.

[78] S. Lämmer, B. Gehlsen, and D. Helbing, “Scaling laws in the spatial structure of urban road networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 363, no. 1, pp. 89–95, 2006.

[79] B. Jiang, “Street hierarchies: A minority of streets account for a majority of traffic flow,” *International Journal of Geographical Information Science*, vol. 23, no. 8, pp. 1033–1048, 2009.

[80] K. Yamaoka, Y. Kumakoshi, and Y. Yoshimura, “Local betweenness centrality analysis of 30 european cities,” in *Urban informatics and future cities*, Springer, 2021, pp. 527–547.

[81] S. Oldham, B. Fulcher, L. Parkes, A. Arnatkevičiūtė, C. Suo, and A. Fornito, “Consistency and differences between centrality measures across distinct classes of networks,” *PloS one*, vol. 14, no. 7, e0220061, 2019.

[82] E. B. Wilson, “Probable inference, the law of succession, and statistical inference,” *Journal of the American Statistical Association*, vol. 22, no. 158, pp. 209–212, 1927.

A

HREC approval

Date 04-May-2026
Correspondence hrec@tudelft.nl

Human Research Ethics Committee TU Delft
(http://hrec.tudelft.nl)

Visiting address
Jaffalaan 5 (building 31)
2628 BX Delft

Postal address
P.O. Box 5015 2600 GA Delft
The Netherlands

Ethics Approval Application: An Adaptive, Stakeholder-Informed Multi-Criteria Decision Analysis Framework for Post-Disaster Road Recovery
Applicant: Lu, Yung-Ting

Dear Yung-Ting Lu,

It is a pleasure to inform you that your application mentioned above has been approved.

Thanks very much for your submission to the HREC which has been conditionally approved. Please note that this approval is subject to your ensuring that the following condition/s is/are fulfilled:

Please find a survey platform that is GDPR-compliant. Contact your supervisor or faculty Data Steward for suggestions on this (Qualtrics is the default, but not only, option).

In addition to any specific conditions or notes, the HREC provides the following standard advice to all applicants:

• In light of recent tax changes, we advise that you confirm any proposed remuneration of research subjects with your faculty contract manager before going ahead.

• Please make sure when you carry out your research that you confirm contemporary covid protocols with your faculty HSE advisor, and that ongoing covid risks and precautions are flagged in the informed consent - with particular attention to this where there are physically vulnerable (eg: elderly or with underlying conditions) participants involved.

• Our default advice is not to publish transcripts or transcript summaries, but to retain these privately for specific purposes/checking; and if they are to be made public then only if fully anonymised and the

transcript/summary itself approved by participants for specific purpose.

• Where there are collaborating (including funding) partners, appropriate formal agreements including clarity on responsibilities, including data ownership, responsibilities and access, should be in place and that relevant aspects of such agreements (such as access to raw or other data) are clear in the Informed Consent.

Good luck with your research!

Sincerely,

B

Declaration on the use of AI tools

Generative AI tools were used during the preparation of this thesis. This declaration sets out how they were used and what remained the author's own work.

Writing. Large language models (ChatGPT, OpenAI; Claude, Anthropic) were used to improve the grammar, clarity, and structure of the English text. All arguments, claims, and interpretations are the author's own, and every AI-suggested revision was reviewed and edited by the author before inclusion.

Questionnaire design. AI tools were used to assist with the visual presentation of the questionnaire, including the layout and formatting of the recovery-stage descriptions and the BWM comparison screens. The questionnaire content, wording, structure, and elicitation design were decided by the author.

Questionnaire images. The illustrative images used to explain the five resilience criteria in the questionnaire were generated using AI image-generation tools. They were included to help non-expert respondents understand the criteria, and the author checked that each image was consistent with the criterion definition it accompanied. The photographs illustrating the two recovery stages are not AI-generated and are attributed to their original sources.

Analysis code. The simulation and analysis pipeline was based on existing code provided for this project, and AI tools were used to assist in adapting it to the requirements of this study. AI tools were also used to assist in writing the BWM weight-estimation code. The outputs of the linear BWM implementation were verified against BWM-Solver-4, the Excel solver published on the Best-Worst Method website, before the code was applied to the survey data. The author specified all analysis procedures, reviewed the resulting code, and verified its outputs against the methods described in Chapter 3.

Qualitative coding. A large language model was used to propose an initial set of themes from the open-ended survey responses. This use is described in full in Section 3.3.4, including the literature supporting AI-assisted qualitative content analysis and the author's review of every theme assignment.

AI tools were not used to generate research results, to fabricate or alter data, or to write the substantive arguments of this thesis.

C

Methodological supplements

C.1. Worked example of the performance metrics

To illustrate how the four simulation performance metrics are computed and interpreted, this example applies them to a small network. Figure C.1 shows the network in its pre-disruption state (a) and after a single bridging link, C–D, is removed (b). For clarity, the example uses an undirected network of five nodes; in the cross-city analysis, road directionality is preserved where present in the OSM-derived networks, but the principle of each metric is the same. Edge labels denote travel times.

Each metric is first reported as a performance ratio $P = M(G(t))/M(G)$, where $M(G)$ is the pre-disruption metric value and $M(G(t))$ is the value after the disruption. These ratios are shown here before the later AAC clipping step used in the SAW aggregation. Values above one therefore indicate that the metric value is higher after disruption than before disruption, not necessarily that the network has become more resilient overall.

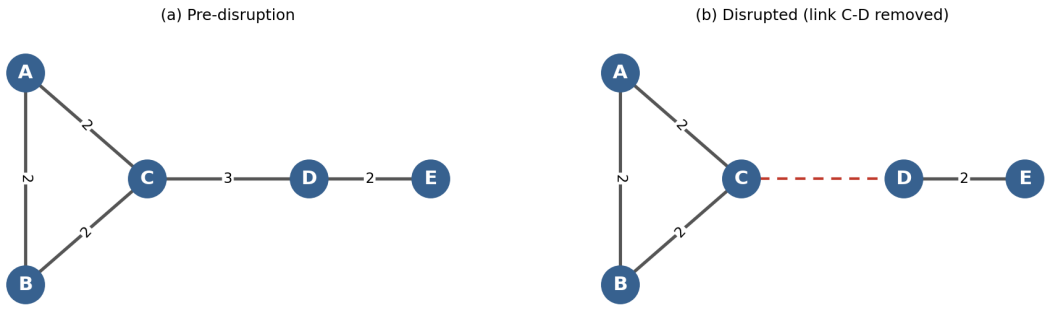


Figure C.1: Example network used to illustrate the performance metrics, before (a) and after (b) removal of the bridging link C–D. Edge labels are travel times. Removing C–D disconnects $\{A, B, C\}$ from $\{D, E\}$.

Table C.1: Performance metrics computed on the example network of Figure C.1. The first two columns give the raw metric values before and after removal of link C–D; the final column gives the performance ratio $P = M(G(t))/M(G)$, before the later AAC clipping step.

Metric	$M(G)$ (pre-disruption)	$M(G(t))$ (disrupted)	P
Operational efficiency	0.322	0.200	0.62
Accessibility	0.340	0.212	0.62
Spatial equity (Moran's I)	0.078	0.087	1.11
Connectivity	0.467	0.600	1.29

Operational efficiency. Operational efficiency is measured as network efficiency based on inverse shortest-path travel times. For a network G, it is defined as

$$E(G) = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}}, \tag{C.1}$$

where d_{ij} is the shortest-path travel time between nodes i and j. If two nodes are unreachable, their contribution is set to zero. This means that the metric rewards short travel times while also penalising disconnected node pairs.

In the pre-disruption network, all node pairs are reachable. Computing the inverse shortest-path travel times over all node pairs gives

$$M(G) = E(G) = 0.322. \tag{C.2}$$

After link C–D is removed, the network is split into two disconnected components. Node pairs that lie in different components are no longer reachable and therefore contribute zero to the efficiency sum. The disrupted efficiency is

$$M(G(t)) = E(G(t)) = 0.200. \tag{C.3}$$

The operational-efficiency performance ratio is therefore

$$\begin{aligned} P_{\text{Efficiency}} &= \frac{M(G(t))}{M(G)} \\ &= \frac{0.200}{0.322} \\ &= 0.62. \end{aligned} \tag{C.4}$$

This shows that removing the bridging link reduces operational efficiency because several node pairs become unreachable and no longer contribute to network efficiency.

Accessibility. Accessibility is calculated in two steps. First, the accessibility of each origin node is computed by summing the weighted contribution of all reachable destination nodes. The weight of each destination is its pre-disruption closeness centrality C_r, and this weight is divided by the shortest-path travel time from the origin node i to the destination node r:

$$A_i = \sum_{r \neq i} \frac{C_r}{\sigma_{ir}}, \tag{C.5}$$

where A_i is the accessibility of origin node i, C_r is the pre-disruption closeness centrality of destination node r, and σ_{ir} is the shortest-path travel time from node i to node r.

For example, for node A in the pre-disruption network, the reachable destinations are B, C, D, and E. Therefore, the accessibility of A is

$$\begin{aligned} A_{\text{A}} &= \frac{C_{\text{B}}}{\sigma_{\text{AB}}} + \frac{C_{\text{C}}}{\sigma_{\text{AC}}} + \frac{C_{\text{D}}}{\sigma_{\text{AD}}} + \frac{C_{\text{E}}}{\sigma_{\text{AE}}} \\ &= \frac{0.250}{2} + \frac{0.333}{2} + \frac{0.267}{5} + \frac{0.190}{7} \\ &= 0.372. \end{aligned} \tag{C.6}$$

The same calculation is repeated for all other origin nodes. The network-level accessibility is the average of the node-level accessibility values:

$$\begin{aligned} M(G) &= \frac{A_{\mathrm{A}} + A_{\mathrm{B}} + A_{\mathrm{C}} + A_{\mathrm{D}} + A_{\mathrm{E}}}{5} \\ &= \frac{0.372 + 0.372 + 0.377 + 0.306 + 0.271}{5} \\ &= 0.340. \end{aligned} \tag{C.7}$$

After the disruption, the same procedure is used, but only reachable destinations are included. For example, node A can still reach B and C, but it can no longer reach D and E. Its disrupted accessibility is therefore

$$\begin{aligned} A_{\mathrm{A}}(t) &= \frac{C_{\mathrm{B}}}{\sigma_{\mathrm{AB}}} + \frac{C_{\mathrm{C}}}{\sigma_{\mathrm{AC}}} \\ &= \frac{0.250}{2} + \frac{0.333}{2} \\ &= 0.292. \end{aligned} \tag{C.8}$$

Repeating this calculation for all nodes and averaging the results gives

$$M(G(t)) = 0.212. \tag{C.9}$$

The accessibility performance value is therefore

$$\begin{aligned} P_{\mathrm{Accessibility}} &= \frac{M(G(t))}{M(G)} \\ &= \frac{0.212}{0.340} \\ &= 0.62. \end{aligned} \tag{C.10}$$

This shows that accessibility decreases after the bridge link C–D is removed, but it does not collapse completely because nodes can still reach destinations within their own disconnected component.

Connectivity. Connectivity is measured using the average local clustering coefficient. In this paragraph, C_i denotes the local clustering coefficient of node i. It measures how many links actually exist among the neighbours of a node compared with the maximum number of links that could exist among those neighbours:

$$C_i = \frac{e_i}{k_i(k_i - 1)/2}, \tag{C.11}$$

where k_i is the number of neighbours of node i, and e_i is the number of links that actually exist among those neighbours. If a node has fewer than two neighbours, its local clustering coefficient is set to zero because no neighbour pair can be formed.

Before the disruption, node C has three neighbours: A, B, and D. Among these three neighbours, there are three possible neighbour links: A–B, A–D, and B–D. However, only A–B actually exists. Therefore, the local clustering coefficient of node C is

$$C_{\mathrm{C}} = \frac{1}{3}. \tag{C.12}$$

The same calculation is repeated for the other nodes. Nodes A and B both have fully connected neighbours, so $C_{\mathrm{A}} = 1$ and $C_{\mathrm{B}} = 1$. Nodes D and E have no connected neighbour pairs, so $C_{\mathrm{D}} = 0$ and $C_{\mathrm{E}} = 0$. The network connectivity before the disruption is therefore

$$\begin{aligned} M(G) &= \frac{C_{\mathrm{A}} + C_{\mathrm{B}} + C_{\mathrm{C}} + C_{\mathrm{D}} + C_{\mathrm{E}}}{5} \\ &= \frac{1 + 1 + \frac{1}{3} + 0 + 0}{5} \\ &= 0.467. \end{aligned} \tag{C.13}$$

After the link C–D is removed, node C has only two neighbours left: A and B. Since A and B are connected to each other, the neighbours of node C are now fully connected. Therefore, the local clustering coefficient of node C increases from $1/3$ to 1:

$$C_{\mathrm{C}}(t) = 1. \tag{C.14}$$

The local clustering coefficients after the disruption are therefore $C_{\mathrm{A}}(t) = 1$, $C_{\mathrm{B}}(t) = 1$, $C_{\mathrm{C}}(t) = 1$, $C_{\mathrm{D}}(t) = 0$, and $C_{\mathrm{E}}(t) = 0$. The disrupted network connectivity is

$$\begin{aligned} M(G(t)) &= \frac{C_{\mathrm{A}}(t) + C_{\mathrm{B}}(t) + C_{\mathrm{C}}(t) + C_{\mathrm{D}}(t) + C_{\mathrm{E}}(t)}{5} \\ &= \frac{1 + 1 + 1 + 0 + 0}{5} \\ &= 0.600. \end{aligned} \tag{C.15}$$

The connectivity performance value is therefore

$$\begin{aligned} P_{\mathrm{Connectivity}} &= \frac{M(G(t))}{M(G)} \\ &= \frac{0.600}{0.467} \\ &= 1.29. \end{aligned} \tag{C.16}$$

This increase does not mean that the whole network becomes better connected. It happens because local clustering measures local cohesion among neighbouring nodes, rather than the ability of all nodes to reach each other. After the disruption, the network is split into two disconnected parts, but the remaining neighbours of node C are more tightly connected. Therefore, connectivity should be interpreted together with efficiency and accessibility.

Spatial equity. Spatial equity is measured using Moran's I of the node-level accessibility values. Moran's I measures spatial autocorrelation, meaning whether nearby nodes have similar accessibility values. In this example, the metric is used to show whether the accessibility impact of a disruption is spatially concentrated.

The Moran's I value is calculated as

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(A_i - \bar{A})(A_j - \bar{A})}{\sum_i (A_i - \bar{A})^2}, \tag{C.17}$$

where n is the number of nodes, A_i and A_j are the accessibility values of nodes i and j, $\bar{A}$ is the mean accessibility value, and w_{ij} is the spatial weight between nodes i and j. A higher Moran's I means that nodes with similar accessibility values are more spatially clustered.

Before the disruption, the accessibility values are relatively evenly distributed across the network. The Moran's I value is

$$M(G) = 0.078. \tag{C.18}$$

After the link C–D is removed, the network is split into two parts. The left cluster, $\{\mathrm{A}, \mathrm{B}, \mathrm{C}\}$, still retains relatively high accessibility, while the right cluster, $\{\mathrm{D}, \mathrm{E}\}$, experiences a larger accessibility loss. As a result, high-accessibility nodes and low-accessibility nodes become more spatially separated. The Moran's I value therefore increases to

$$M(G(t)) = 0.087. \tag{C.19}$$

The spatial-equity performance value is therefore

$$\begin{aligned} P_{\mathrm{Equity}} &= \frac{M(G(t))}{M(G)} \\ &= \frac{0.087}{0.078} \\ &= 1.11. \end{aligned} \tag{C.20}$$

This does not mean that the network becomes more equitable after the disruption. Rather, it shows that Moran's I increases, meaning that accessibility values become more spatially clustered. In this example, the right side of the network is affected more strongly than the left side, so the disruption increases the spatial concentration of accessibility loss. In the main analysis, such raw performance trajectories are converted into AAC deviation scores, and the resulting deviation scores are clipped to $[0, 1]$ only before entering the SAW aggregation.

Summary. The same disruption produces different responses across the four metrics: operational efficiency and accessibility both fall to about 62% of their pre-disruption values, spatial autocorrelation increases, and the local clustering coefficient also increases. These increases do not mean that the network has improved overall; rather, they reflect the specific definitions of Moran's I and local clustering. The example therefore illustrates why the four metrics must be interpreted together before they are combined in the SAW analysis.

C.2. Strategy-ordering similarity check

Each recovery strategy uses its edge-importance scores to define a restoration order. Therefore, two strategies may lead to similar recovery behaviour if they assign similar priority rankings to the same links. To examine this, a strategy-ordering similarity check was conducted for the single-indicator strategies.

For each city, the edge-importance scores of each strategy were calculated on the intact road network using the definitions in Section 3.2.2. The resulting edge orderings were then compared pairwise using Spearman rank correlation. The correlations were calculated for each city and then averaged across the 220 distinct city networks. The random baseline was excluded because it does not define a structural edge ordering.

The use of 220 distinct city networks reflects the removal of duplicated city-name variants from the 223 valid network instances used in the cross-city analysis.[1]

This check is based on intact-network orderings and is independent of the damage scenario. Spearman rank correlation is used because the restoration strategies depend on the ranking of edge scores rather than on their absolute values. The edge scores are also measured on different scales and may have skewed distributions. Rank correlation is therefore more appropriate than Pearson correlation because it is invariant to monotonic rescaling and is commonly used when comparing centrality-based rankings [81].

[1] Three cities appear in the dataset under two name variants each (Kutná Hora, Lüderitz, and Žabljak), so the 223 network instances represent 220 distinct cities. Restricting the cross-city analysis to these 220 distinct cities changes the mean area-above-curve deviation scores by at most 0.002 and does not change which strategy attains the lowest mean deviation on any criterion.

D

Survey questionnaire

This appendix presents the survey instrument used for stakeholder-priority elicitation. The hypothetical disruption scenario shown in the questionnaire was used only to provide respondents with a concrete post-disaster road recovery context. It was not used as the damage input for the simulation analysis. The simulation analysis separately applies controlled disruption levels of 10% and 30% to compare recovery strategies under different levels of network damage.

TU Delft

English

Introduction

Road Recovery Survey

Master's Thesis Research | **TU Delft**

🌐 **Language / 語言:** You can switch to **Chinese (Traditional)** via the dropdown menu in the **top right corner**.

100

⏳ **10–15 Mins** | 🔒 **Anonymous** | 🌍 **Impactful**

Dear Participant, thank you for supporting this research! We aim to examine how **post-disaster road recovery strategies** can be prioritised, given that resilience is multidimensional, stakeholders have different priorities, and recovery decisions evolve across different recovery stages. Your expertise helps develop a decision-making framework for planning.

Conducted by **Yung-Ting Lu**, Master's student at TU Delft, Faculty of Civil Engineering and Geosciences, MSc Transportation, Infrastructure and Logistics, under the supervision of **Patrick Stokkink**.

You are invited to participate because your professional role or experience relates to post-disaster road recovery, transport infrastructure, emergency management, or related decision-making processes.

> **What You Will Be Asked To Do:**
>
> - Provide general background information about your professional role and organisational context.
> - Indicate the relative importance of different resilience criteria across two recovery stages (short-term and long-term) using a structured comparison approach (the Best–Worst Method).

> Method).

The data will be used for academic research purposes, including analysis and reporting in a Master's thesis and possible academic publications.

> **Your Rights:** Participation is entirely voluntary and **completely anonymous**. No personally identifiable information (IP, name, email) is collected. You may exit the survey at any time without providing a reason and without any negative consequences. Because the survey is anonymous, it will not be possible to remove individual responses after submission.

As with any online activity, the risk of a breach is always possible. To the best of our ability, your answers in this study will remain confidential. All data will be stored securely and handled in accordance with TU Delft data protection and privacy guidelines.

By clicking **"Next"**, you confirm you have read and understood the information above and voluntarily agree to participate in this study.

What will you do in this survey?

In this survey, you will evaluate what should be prioritised in road recovery after a disaster.

Below is what you will be asked to do:

1. **Provide** your background information
2. **Read** the disaster scenario, recovery stages, and criteria introduction
3. **Evaluate** the resilience criteria at both recovery stages

👉 There are no right or wrong answers — just your judgement.

Section A: Respondent Profile

PART 1 OF 4

Section 1: Respondent Profile

To kick things off, we would like to understand your **professional background and organisational context**. This ensures the insights we gather represent a diverse range of expertise in road transport network recovery.

What is your age?

- Under 18
- 18–24
- 25–34
- 35–44
- 45–54
- 55–64
- 65 or above

101

What is your gender?

- Male
- Female
- Non-binary
- Prefer not to say

Which of the following best describes your primary role in relation to road recovery after disasters?

Since you chose "Other",

What is your primary role in relation to road recovery?

How many years of experience do you have in your professional field?

- Less than 2 years
- 2–5 years
- 6–10 years
- More than 10 years

In which country do you primarily work or operate?

Type to search...

Based on your experience, which type(s) of disaster have you been involved with or affected by?

1022

CONTEXT: A disaster refers to any event that causes significant disruption to transport, infrastructure, or daily activities.

💡 Please select all that apply.

- ☐ Earthquake
- ☐ Flood (including heavy rainfall events)
- ☐ Landslide
- ☐ Typhoon / Hurricane
- ☐ Snowfall / severe winter weather
- ☐ Other:
- ☐ I have not experienced any of the above

Section B: Event Description

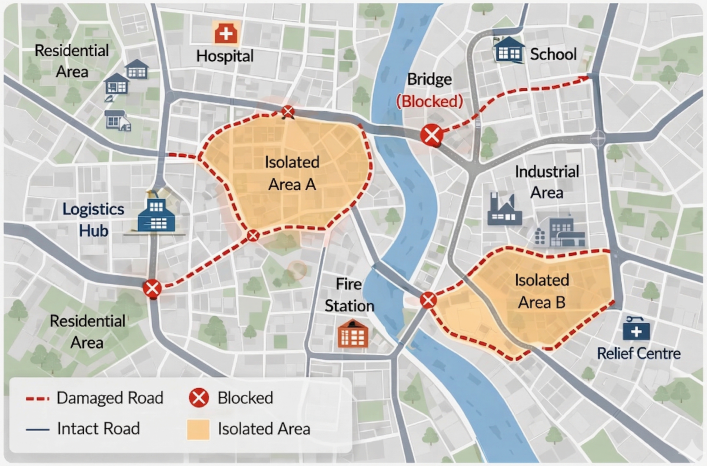
1. Profile | 2. Event Context | 3. Short-Term | 4. Long-Term

PART 2 OF 4

Section 2: Disaster Scenario

Imagine that a major disaster has struck a large urban area.

- **Widespread Damage:** Around 40% of major roads and bridges are structurally damaged or blocked by debris, making them **impassable**.
- **Community Isolation:** Two major regions (Isolated Area A and Isolated Area B) within the city are **completely cut off** from the rest. As a result:
  - Residents cannot reach the **hospital** for medical care.
  - Children cannot access the **school**.
  - **Emergency vehicles** cannot enter the isolated districts.
  - **Relief supplies** cannot be delivered.
  - The **logistics hub** cannot distribute goods.
  - Residents cannot evacuate or receive outside assistance.
- **Limited Resources:** There is not enough budget, equipment, or personnel to fix everything, so decisions must be made about which roads to repair first.



Note: Illustrative example of road network damage.

2 stages recovery

Two Stages of Road Recovery

After a major disaster, road recovery typically happens in two stages with different priorities.

In the following questions, you will evaluate priorities for each recovery stage.

STAGE I

Short-Term Recovery

⏳ ≤ 7 days

Focus is on **acute and immediate needs** through **temporary repairs**.



STAGE II

Long-Term Recovery

🗓️ Weeks to Months

▼ Show example recovery tasks

Focus shifts to **secondary priorities** through **permanent reconstruction**.



▼ Show example recovery tasks

👉 Click "Next" when you are ready to begin.

Criteria Description

Almost Ready to Start

Please review the **5 recovery criteria** below.

💡 You can review these anytime by clicking the **hint box** at the top of the page.

1. Operational Efficiency

How quickly people can travel across the road network, including commuting to work, reaching services, and completing daily trips.



💡 **Example:** Repairing the **main road** connecting the **city centre** first, so that most residents can **get to work and school faster, and do daily activities**, even if some smaller side streets remain blocked.

2. Ability to Access Critical Services

How easy it is to reach places like hospitals, food, and water when needed.



💡 **Example:** Prioritizing roads to **relief centers and hospitals** to ensure urgent needs are met.

3. Ability to Access Routine Activities

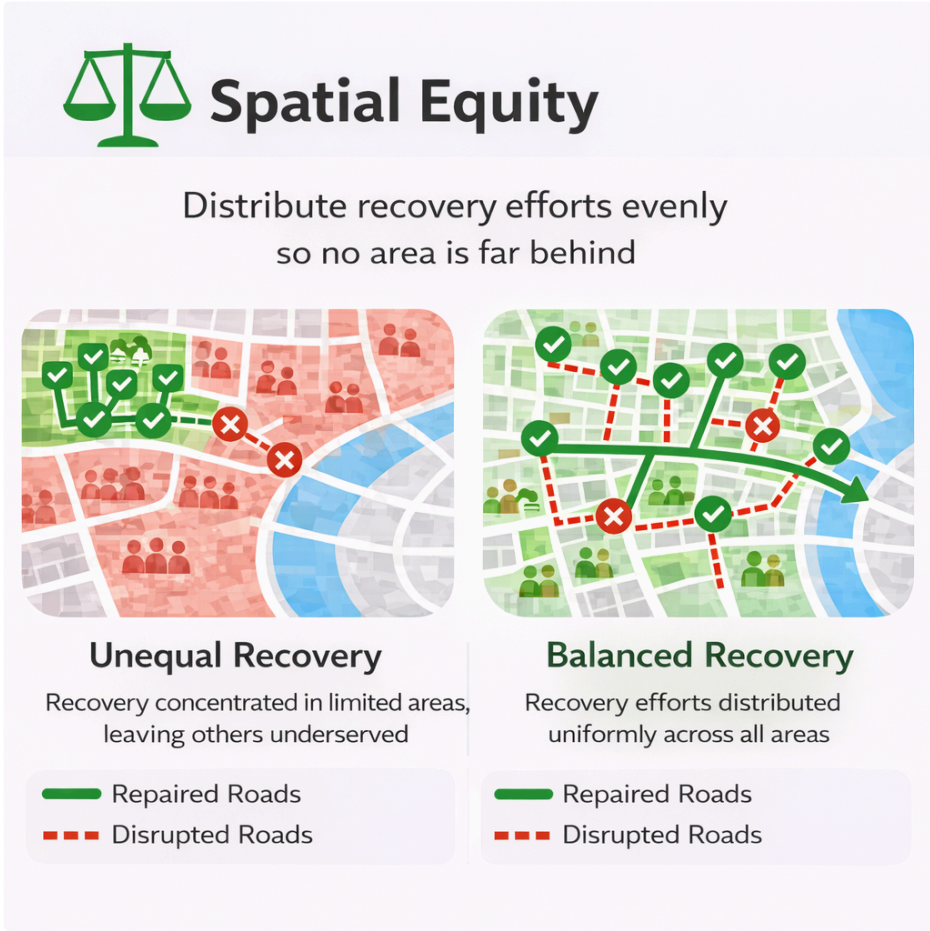
How easy it is to reach work, school, shops, and daily destinations.



💡 **Example:** Repairing roads so people can **go to work, school, and shops** easily.

4. Spatial Equity

How fairly resources, services, or opportunities are restored across all areas.



💡 **Example:** Even though the **city centre** is easier to repair, resources are also allocated to **rural areas** to ensure no region is left behind in the recovery process.

5. Connectivity

Rejoining network segments to prevent isolation.

Connectivity

How well different areas stay **connected** across the network



💡 **Example:** Restoring road connections between the **industrial area**, **logistics hub**, and **residential areas**, so that goods can be delivered **from factories to warehouses** and eventually reach **consumers**.

🚀 Ready to start your evaluation?

You will complete **3 short steps**:

1. Choose most & least important criteria
2. Compare your top criteria to others
3. Compare others to the least criteria

Click **Next** to begin

short term recovery

1. Profile | 2. Event Context | 3. Short-Term | 4. Long-Term

PART 3 OF 4

Section 3: Short-Term Recovery

🚀 Task: Evaluate Criteria for the First 7 Days

In the first 7 days after the disaster, only **temporary repairs** are possible.
Think about the most **urgent needs** during this period.

✨ You will be asked to:

1. select the most important criterion
2. select the least important criterion
3. compare the criteria based on their importance

You will be asked to compare the following 5 criteria:

1. Efficiency:	2. Ability to Access Critical Services:	3. Ability to Access Routine Activities:	4. Spatial Equity:	5. Connectivity:
Reducing **overall** travel time in the network via temporarily repaired roads	Reach **critical places** (e.g., hospitals, food supply points, water stations) via temporarily repaired roads	Reach **daily destinations** (e.g., work, school) via temporarily repaired roads	Ensuring **all areas** recover access equally via temporarily repaired roads	Reconnecting **isolated** areas via temporarily repaired roads

*Please keep the **short-term context (≤ 7 days)** in mind for this section.*

Step 1: Which is the **MOST important** criterion for road recovery during the **Short-Term Stage**?

- Operational Efficiency (reducing overall travel time in the network)
- Ability to easily reach places like hospitals, food, and water when needed
- Ability to easily reach work, school, and shops
- Spatial Equity (ensuring all areas recover access equally)
- Connectivity (reconnecting isolated areas to the rest of the network)

Reasoning for Your Priority

Based on the criteria you consider **most important for the Short-Term Recovery Stage**, please explain the reasoning behind your choices.

Your insights help us understand the logic behind disaster recovery decision-making.

Step 1: Which is the **LEAST important** criterion for road recovery during the **Short-Term Stage**?

- Operational Efficiency (reducing overall travel time in the network)
- Ability to easily reach places like hospitals, food, and water when needed
- Ability to easily reach work, school, and shops
- Spatial Equity (ensuring all areas recover access equally)
- Connectivity (reconnecting isolated areas to the rest of the network)

Reasoning for Your Priority

Based on the criteria you consider **least important for the Short-Term Recovery Stage**, please explain the reasoning behind your choices.

Your insights help us understand the logic behind disaster recovery decision-making.

⏳ Current Evaluation: Short-Term Recovery

Focus: Acute and immediate needs (Up to 7 days).

Step 2: Best-to-Others Comparison

Compare **${q://QID27/ChoiceGroup/SelectedChoices}** (your selected most Important criterion) to the other criteria below.

👉 **What to do:** For each criterion below, click the dropdown menu and select a score from **2 to 9 to indicate how much more important the Most Important criterion is compared to each criterion below**.

Note: The Most Important Criterion compared to itself is pre-set to 1.

SCALING GUIDE

Your Most Important Criterion is equally important as ... → **Your Most Important Criterion** is absolutely more important than ...

1	2	3	4	5	6	7	8	9
Equal importance	Between Equal and Moderate	**Moderately** more important	Between Moderate and Strong	**Strongly** more important	Between Strong and Very Strong	**Very strongly** more important	Between Very Strong and Absolute	**Absolute** more important

	Most Important Criterion: ${q://QID27/ChoiceGroup/SelectedChoices}
Operational Efficiency (reducing overall travel time in the network)	
Ability to easily reach places like hospitals, food, and water when needed	
Ability to easily reach work, school, and shops	
Spatial Equity (ensuring all areas recover access equally)	
Connectivity (reconnecting isolated areas to the rest of the network)	

103

🪜 Step 3: Others-to-the-Worst Comparison

Compare each criterion below with **${q://QID26/ChoiceGroup/SelectedChoices}** (your selected least important criterion).

⚠️ **NEW TASK: Different direction from Step 3!**

👉 **What to do:** For each criterion below, select a score (1–9) from the dropdown menu to indicate **how much more important each criterion is compared to the least important criterion.**

💡 Note: The least important criterion compared to itself is pre-set to 1.

SCALING GUIDE

Each criterion is equally important as the Least Important one → **Each criterion** is absolutely more important than the Least Important one

1	2	3	4	5	6	7	8	9
Equal importance	Between Equal and Moderate	**Moderately** more important	Between Moderate and Strong	**Strongly** more important	Between Strong and Very Strong	**Very strongly** more important	Between Very Strong and Absolute	**Absolute** more important

	Least Important Criterion: ${q://QID26/ChoiceGroup/SelectedChoices}
Operational Efficiency (reducing overall travel time in the network)	
Ability to easily reach places like hospitals, food, and water when needed	
Ability to easily reach work, school, and shops	
Spatial Equity (ensuring all areas recover access equally)	
Connectivity (reconnecting isolated areas to the rest of the network)	

Block 8

Congratulations! You have successfully completed the Short-Term Recovery stage! 🥳🎉

You will now move on to the **Long-Term Recovery Stage** (weeks to months after the disaster).

In the short-term stage, you made temporary repairs. Now, you will focus on **permanent reconstruction**.

You will do the **same 3 steps** as before, but keeping this different long-term goal in mind.

⚠️ **Note:** Your answers may differ from the Short-Term stage as the focus has shifted. However, you may also select the same criteria if you believe they remain the most important.

Click **"Next"** when you are ready.

long term recovery

109

1. Profile | 2. Event Context | 3. Short-Term | **4. Long-Term**

PART 4 OF 4

Section 4: Long-Term Recovery

🏗️ Task: Evaluate Criteria for the Weeks to Months

In the **weeks to months** after the disaster, roads are **permanently reconstructed.**
Think about the long-term needs during this period.

✨ You will be asked to:

1. select the most important criterion
2. select the least important criterion
3. compare the criteria based on their importance

In the short-term stage, roads were temporarily repaired so they could be used. Now, in the long-term stage, the goal is to rebuild the road network properly for long-term use.

Consider these criteria in the Long-Term context:

1. Efficiency:	2. Ability to Access Critical Services:	3. Ability to Access Routine Activities:	4. Spatial Equity:	5. Connectivity:
Reducing **overall** travel time in the network via permanently reconstructed roads	Reach **critical places** (e.g., hospitals, food supply points, water stations) via permanently reconstructed roads	Reach **daily destinations** (e.g., work, school) via permanently reconstructed roads	ensuring **all areas** recover access equally via permanently reconstructed roads	Reconnecting **isolated** areas via permanently reconstructed roads

*Please keep the **long-term context** in mind for this section.*

Step 1: Among the four main criteria below, which do you consider the **MOST important** for road recovery during the **Long-Term Recovery Stage**?

- ◯ **Operational Efficiency (reducing overall travel time)**
- ◯ **Ability to easily reach places like hospitals, food, and water when needed**
- ◯ **Ability to easily reach work, school, and shops**
- ◯ **Spatial Equity (ensuring all areas recover access equally)**
- ◯ **Connectivity (reconnecting isolated areas to the rest of the network)**

🏗️ Reasoning for Your Priority

Based on the criteria you consider **most important for the Long-Term Recovery Stage**, please explain the reasoning behind your choices.

Your insights help us understand the logic behind disaster recovery decision-making.

Step 1: Among the four main criteria below, which do you consider the **LEAST important** for road recovery during the **Long-Term Recovery Stage**?

- Operational Efficiency (reducing overall travel time)
- Ability to easily reach places like hospitals, food, and water when needed
- Ability to easily reach work, school, and shops
- Spatial Equity (ensuring all areas recover access equally)
- Connectivity (reconnecting isolated areas to the rest of the network)

Reasoning for Your Priority

Based on the criteria you consider **least important for the Long-Term Recovery Stage**, please explain the reasoning behind your choices.

Your insights help us understand the logic behind disaster recovery decision-making.

Current Evaluation: Long-Term Recovery

Focus: Secondary priorities (Weeks to Months).

Step 2: Best-to-Others Comparison

Compare **${q://QID63/ChoiceGroup/SelectedChoices}** (your selected MOST important criterion) to the other criteria below.

What to do: For each criterion below, click the dropdown menu and select a score from **2 to 9 to indicate how much more important the MOST important criterion is compared to each criterion below**.

Note: The **MOST important vs. itself** is pre-set to **1**.

SCALING GUIDE

Your MOST Important Criterion is equally important as ... → **Your MOST Important Criterion** is absolutely more important than ...

1	2	3	4	5	6	7	8	9
Equal importance	Between Equal and Moderate	**Moderately** more important	Between Moderate and Strong	**Strongly** more important	Between Strong and Very Strong	**Very strongly** more important	Between Very Strong and Absolute	**Absolutely** more important

	MOST Important Criterion: ${q://QID63/ChoiceGroup/SelectedChoices}
Operational Efficiency (reducing overall travel time)	
Ability to easily reach places like hospitals, food, and water when needed	
Ability to easily reach work, school, and shops	
Spatial Equity (ensuring all areas recover access equally)	

	MOST Important Criterion: ${q://QID63/ChoiceGroup/SelectedChoices}
Connectivity (reconnecting isolated areas to the rest of the network)	

Step 3: Others-to-the-Worst Comparison

Compare each criterion below with **${q://QID66/ChoiceGroup/SelectedChoices}** (your selected LEAST Important criterion).

⚠️ **NEW TASK: Different direction from Step 3!**
👉 **What to do:** For each criterion below, select a score (1–9) from the dropdown menu to indicate **how much more important each criterion is compared to the LEAST Important criterion.**
💡 **Note:** The LEAST Important criterion compared to itself is fixed at **1**.

SCALING GUIDE

Each criterion is equally important as the LEAST Important one → **Each criterion** is absolutely more important than the LEAST Important one

1	2	3	4	5	6	7	8	9
Equal importance	Between Equal and Moderate	**Moderately** more important	Between Moderate and Strong	**Strongly** more important	Between Strong and Very Strong	**Very strongly** more important	Between Very Strong and Absolute	**Absolutely** more important

	LEAST Important Criterion: ${q://QID66/ChoiceGroup/SelectedChoices}
Operational Efficiency (reducing overall travel time)	
Ability to easily reach places like hospitals, food, and water when needed	

	LEAST Important Criterion: ${q://QID66/ChoiceGroup/SelectedChoices}
Ability to easily reach work, school, and shops	
Spatial Equity (ensuring all areas recover access equally)	
Connectivity (reconnecting isolated areas to the rest of the network)	

missing critiera

Final Thoughts

After considering both the Short-Term and Long-Term Recovery stages, **do you think any important criterion was missing?**

If yes, please describe it briefly and explain why you think it is important. *(Optional)*

Powered by Qualtrics

E

Processed simulation outputs

E.1. Processed deviation-score summaries

This appendix reports descriptive summaries of the clipped AAC deviation scores used as simulation-side inputs to the SAW aggregation. The scores are averaged over the 223 cities and the 20 disruption iterations for each strategy, criterion, damage level, and recovery window. Lower values indicate smaller deviations from the pre-disruption performance baseline on the corresponding criterion. These tables document the intermediate processed data before SAW normalisation and stakeholder-weighted aggregation, and should therefore not be interpreted as final strategy rankings.

Table E.1: Processed mean AAC deviation scores by strategy and criterion, averaged over the 223 cities and 20 disruption iterations, in the short- and long-term windows at 10% damage. Lower values indicate smaller deviations from the pre-disruption performance baseline. These values are intermediate inputs before SAW normalisation and stakeholder-weighted aggregation.

Strategy	Operational efficiency	Accessibility	Spatial equity (Moran's I)	Connectivity (clustering)
Short-term				
Betweenness	0.170	0.178	0.151	0.062
Closeness	0.203	0.210	0.138	0.057
Degree	0.230	0.241	0.174	0.066
Eigenvector	0.227	0.235	0.168	0.061
Nearest-neighbour edge	0.228	0.239	0.173	0.066
Random	0.227	0.236	0.163	0.062
Long-term				
Betweenness	0.060	0.064	0.079	0.031
Closeness	0.086	0.088	0.068	0.029
Degree	0.117	0.130	0.134	0.046
Eigenvector	0.115	0.119	0.106	0.035
Nearest-neighbour edge	0.105	0.113	0.116	0.039
Random	0.112	0.117	0.094	0.031

Table E.2: Processed mean AAC deviation scores by strategy and criterion, averaged over the 223 cities and 20 disruption iterations, in the short- and long-term windows at 30% damage. Lower values indicate smaller deviations from the pre-disruption performance baseline. These values are intermediate inputs before SAW normalisation and stakeholder-weighted aggregation.

Strategy	Operational efficiency	Accessibility	Spatial equity (Moran's I)	Connectivity (clustering)
Short-term				
Betweenness	0.473	0.477	0.312	0.231
Closeness	0.531	0.529	0.253	0.213
Degree	0.584	0.590	0.381	0.231
Eigenvector	0.591	0.594	0.373	0.222
Nearest-neighbour edge	0.585	0.592	0.379	0.231
Random	0.588	0.592	0.362	0.218
Long-term				
Betweenness	0.139	0.145	0.156	0.130
Closeness	0.205	0.199	0.101	0.108
Degree	0.286	0.304	0.277	0.135
Eigenvector	0.292	0.297	0.216	0.116
Nearest-neighbour edge	0.259	0.271	0.235	0.117
Random	0.289	0.295	0.194	0.095

F

Cross-city ranking results

F.1. Rank-one strategy distributions across cities

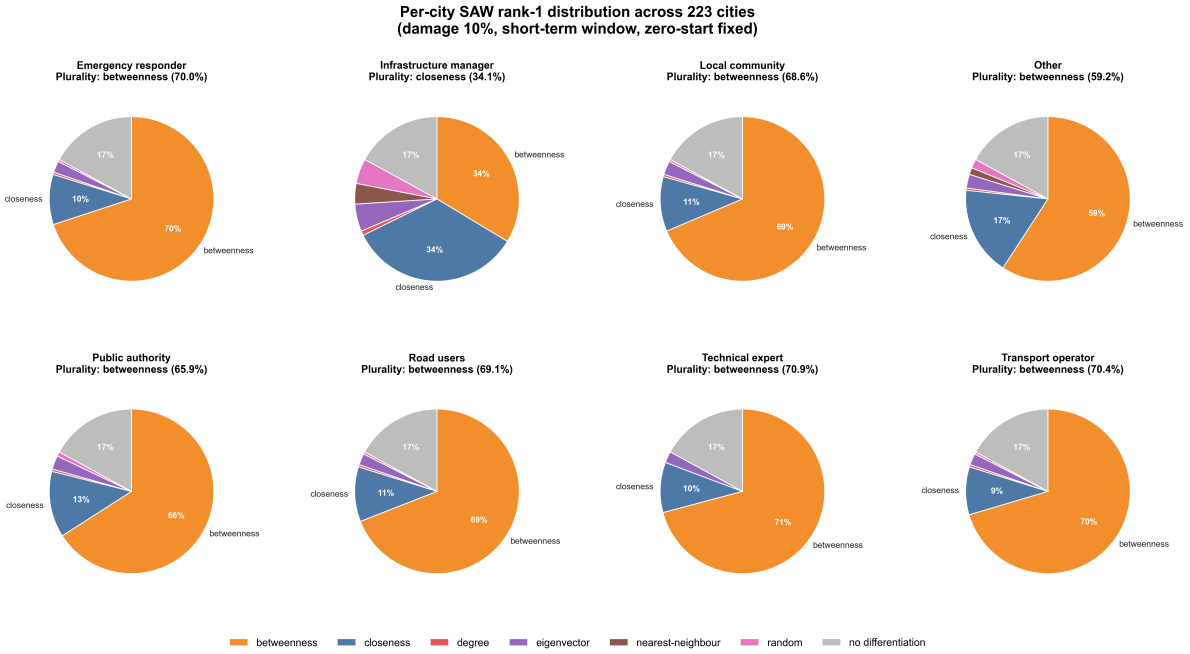


Figure F.1: Rank-one strategy distribution across cities under 10% damage in the short-term window.

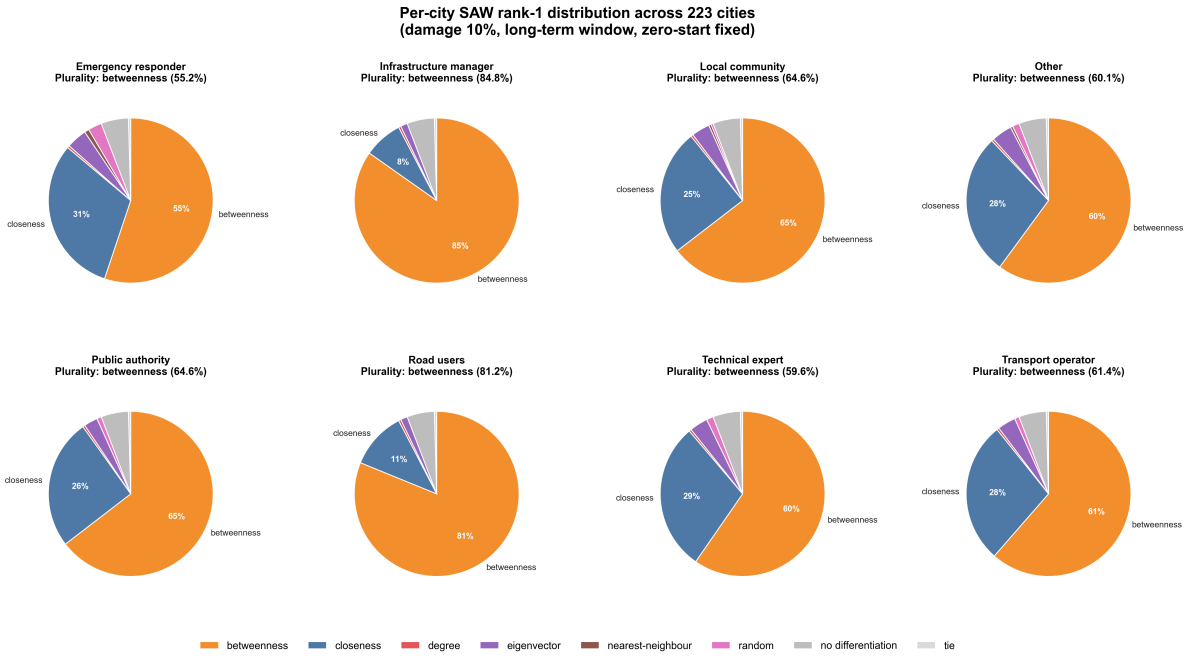


Figure F.2: Rank-one strategy distribution across cities under 10% damage in the long-term window.

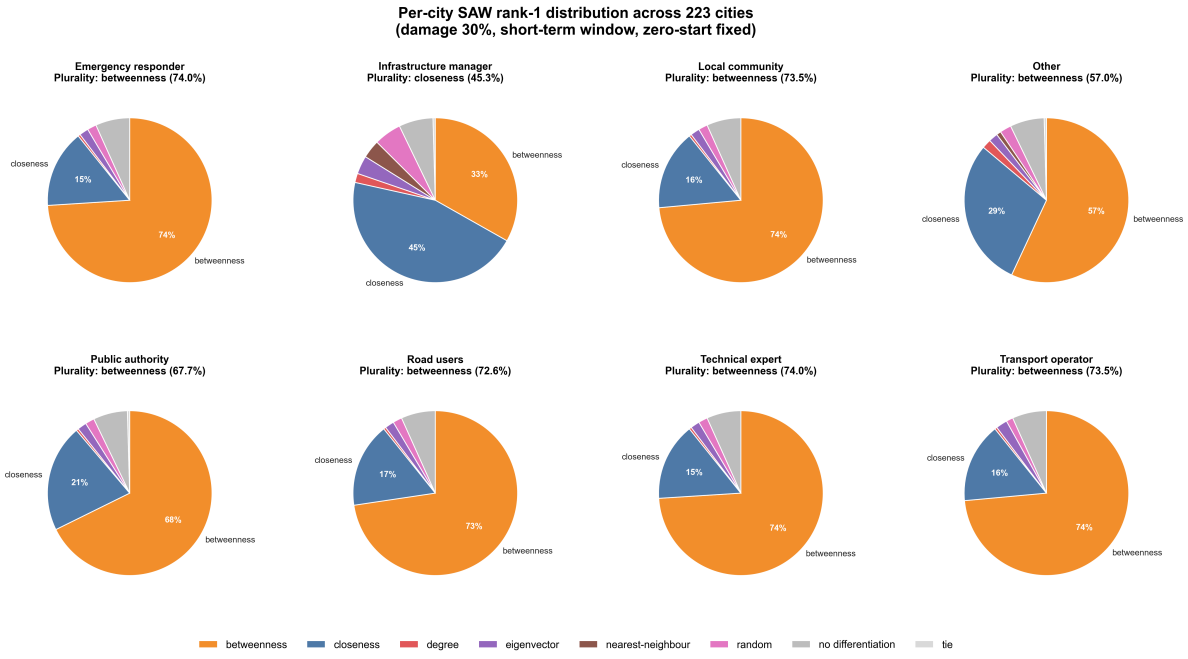


Figure F.3: Rank-one strategy distribution across cities under 30% damage in the short-term window.

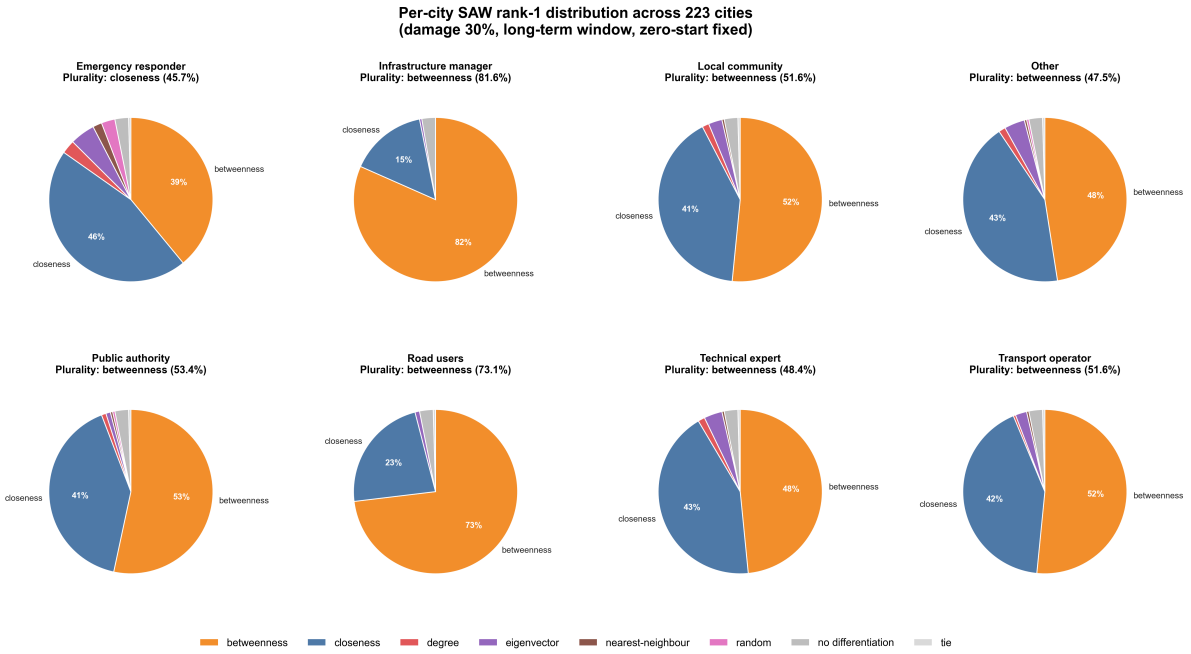


Figure F.4: Rank-one strategy distribution across cities under 30% damage in the long-term window.

F.2. Plurality rank-one shares with Wilson 95% confidence intervals

The rank-one shares were calculated across the 223 analysed city networks. For each damage level, recovery window, stakeholder group, and strategy, let x denote the number of networks in which that strategy was recorded as rank one, and let $n = 223$. The observed rank-one share is therefore $\hat{p} = x/n$.

Wilson score $95\,\%$ confidence intervals were calculated for the observed rank-one shares because they remain well behaved for binomial proportions, including proportions close to zero or one. For an observed proportion $\hat{p}$, the interval was calculated as

$$\frac{\hat{p} + z^2/(2n)}{1 + z^2/n} \pm \frac{z}{1 + z^2/n}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}},$$

where $z = 1.96$ for a $95\,\%$ confidence interval[82].

Tied and non-differentiated outcomes remained part of the denominator but were recorded as separate outcome categories when calculating the underlying rank-one distributions. The intervals describe the precision of the observed cross-city rank-one shares. They do not incorporate uncertainty in the BWM criterion weights or recovery trajectories and should not be interpreted as paired significance tests between strategies.

Table F.1 reports, for each damage level, recovery window, and stakeholder group, the strategy with the largest observed rank-one share and its Wilson $95\,\%$ confidence interval.

Table F.1: Plurality rank-one strategy shares with Wilson 95% confidence intervals.

Damage	Window	Stakeholder	Strategy	Count	Share [95% CI]
10%	short-term	Emergency responder	Betweenness	156	70.0% [63.6–75.6%]
10%	short-term	Infrastructure manager	Closeness	76	34.1% [28.2–40.5%]
10%	short-term	Local community	Betweenness	153	68.6% [62.2–74.3%]
10%	short-term	Public authority	Betweenness	147	65.9% [59.5–71.8%]
10%	short-term	Road users	Betweenness	154	69.1% [62.7–74.8%]

Damage	Window	Stakeholder	Strategy	Count	Share [95% CI]
10%	short-term	Technical expert	Betweenness	158	70.9% [64.6–76.4%]
10%	short-term	Transport operator	Betweenness	157	70.4% [64.1–76.0%]
10%	short-term	Other	Betweenness	132	59.2% [52.6–65.4%]
10%	long-term	Emergency responder	Betweenness	123	55.2% [48.6–61.5%]
10%	long-term	Infrastructure manager	Betweenness	189	84.8% [79.5–88.9%]
10%	long-term	Local community	Betweenness	144	64.6% [58.1–70.6%]
10%	long-term	Public authority	Betweenness	144	64.6% [58.1–70.6%]
10%	long-term	Road users	Betweenness	181	81.2% [75.5–85.8%]
10%	long-term	Technical expert	Betweenness	133	59.6% [53.1–65.9%]
10%	long-term	Transport operator	Betweenness	137	61.4% [54.9–67.6%]
10%	long-term	Other	Betweenness	134	60.1% [53.5–66.3%]
30%	short-term	Emergency responder	Betweenness	165	74.0% [67.9–79.3%]
30%	short-term	Infrastructure manager	Closeness	101	45.3% [38.9–51.8%]
30%	short-term	Local community	Betweenness	164	73.5% [67.4–78.9%]
30%	short-term	Public authority	Betweenness	151	67.7% [61.3–73.5%]
30%	short-term	Road users	Betweenness	162	72.6% [66.4–78.1%]
30%	short-term	Technical expert	Betweenness	165	74.0% [67.9–79.3%]
30%	short-term	Transport operator	Betweenness	164	73.5% [67.4–78.9%]
30%	short-term	Other	Betweenness	127	57.0% [50.4–63.3%]
30%	long-term	Emergency responder	Closeness	102	45.7% [39.3–52.3%]
30%	long-term	Infrastructure manager	Betweenness	182	81.6% [76.0–86.1%]
30%	long-term	Local community	Betweenness	115	51.6% [45.0–58.0%]
30%	long-term	Public authority	Betweenness	119	53.4% [46.8–59.8%]
30%	long-term	Road users	Betweenness	163	73.1% [66.9–78.5%]
30%	long-term	Technical expert	Betweenness	108	48.4% [42.0–55.0%]
30%	long-term	Transport operator	Betweenness	115	51.6% [45.0–58.0%]
30%	long-term	Other	Betweenness	106	47.5% [41.1–54.1%]

G

Sensitivity and robustness analyses

G.1. BWM configuration sensitivity

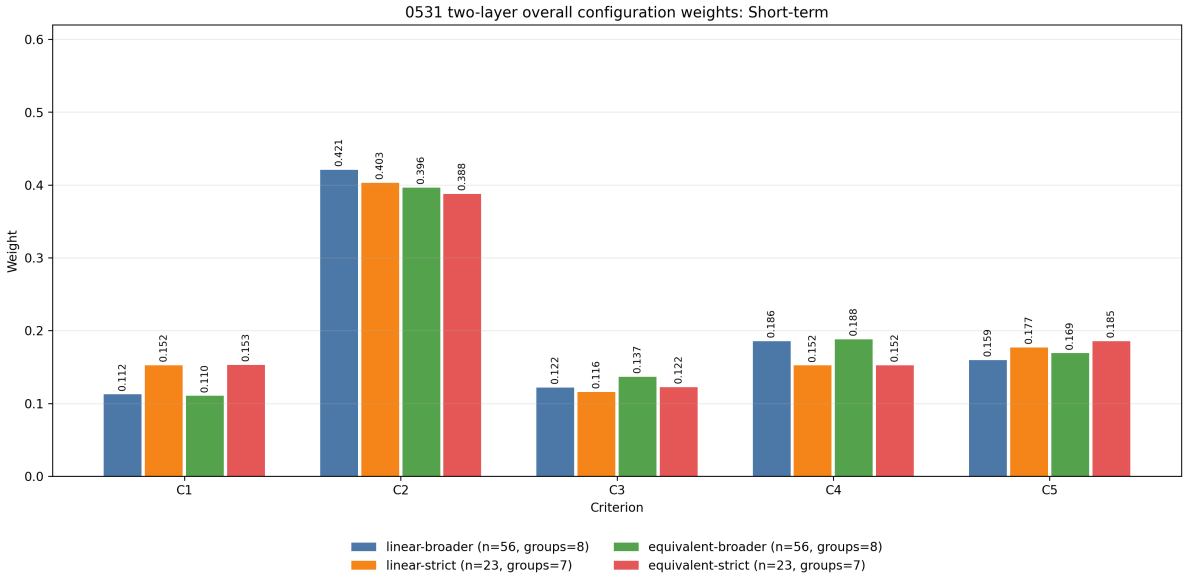


Figure G.1: Overall criterion weights under the four BWM configurations (linear-broader, linear-strict, equivalent-broader, equivalent-strict) in the short-term window. The weights are almost identical across configurations, with the largest difference in any single criterion below 0.02.

Table G.1: Sensitivity of plurality rank-one strategies to the BWM solver and respondent-sample configuration. Each cell reports the share of the 223 cities in which betweenness is the plurality top-ranked strategy; betweenness is the plurality winner in every configuration.

Configuration	10% short	10% long	30% short	30% long
Equivalent–broader	67.7%	71.3%	71.7%	59.6%
Equivalent–strict	67.7%	70.4%	71.3%	58.3%
Linear–broader	69.1%	71.3%	71.7%	59.6%
Linear–strict	68.2%	70.9%	71.7%	58.3%

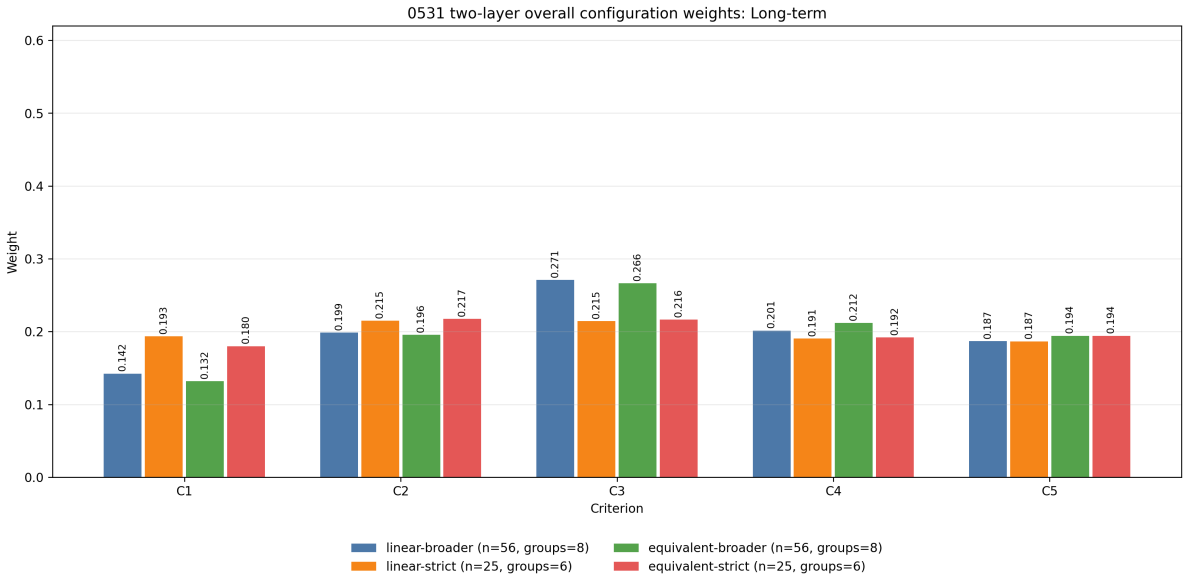


Figure G.2: Overall criterion weights under the four BWM configurations (linear-broader, linear-strict, equivalent-broader, equivalent-strict) in the long-term window. The weights are almost identical across configurations, with the largest difference in any single criterion below 0.02.

G.2. Recovery-stage boundary sensitivity

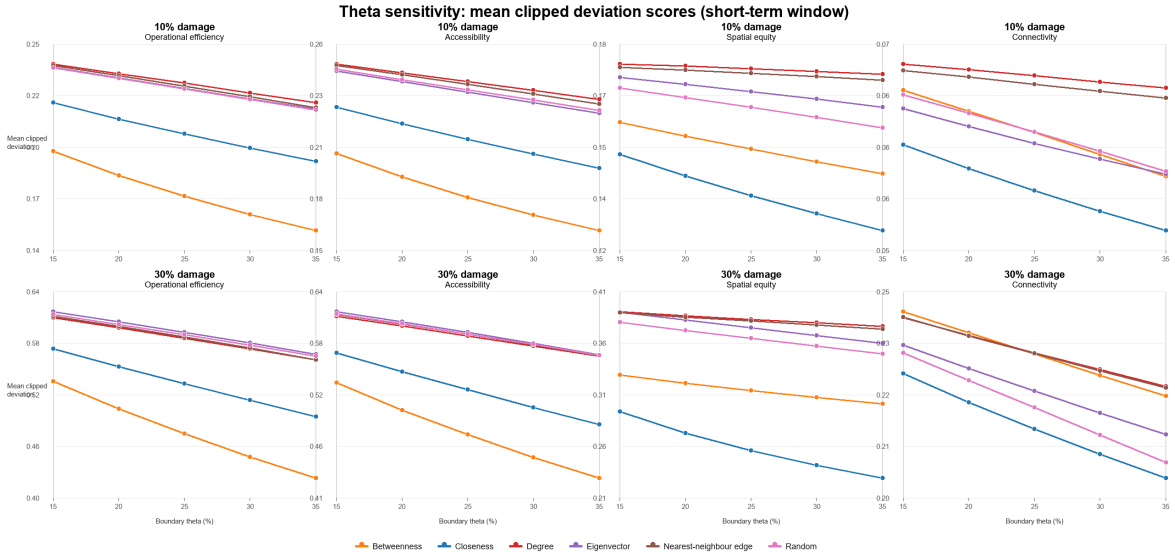


Figure G.3: Sensitivity of mean clipped deviation scores to the recovery-stage boundary θ in the short-term window. Lower values indicate better recovery performance.

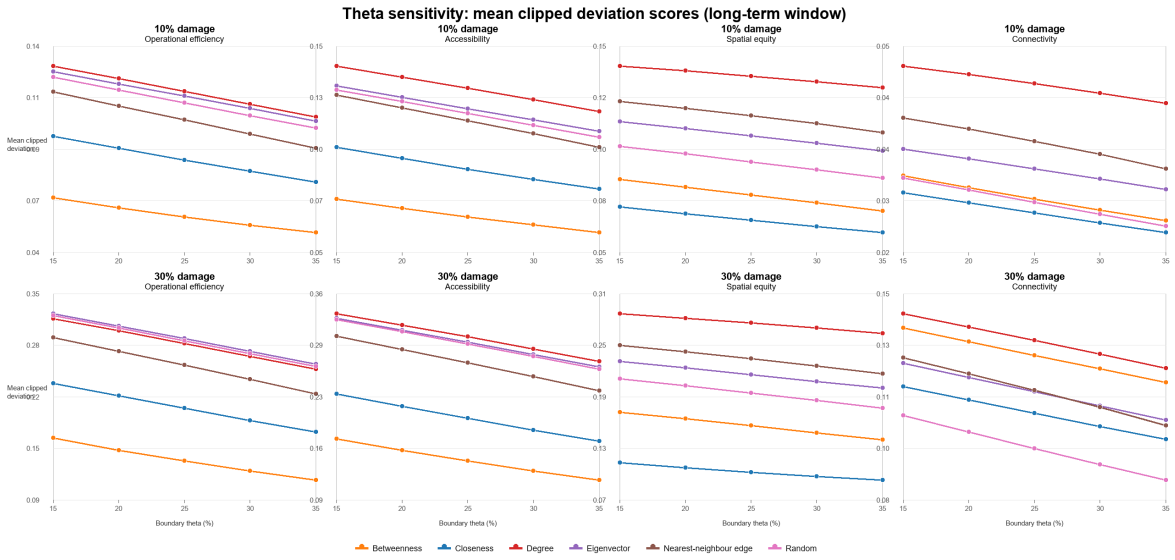


Figure G.4: Sensitivity of mean clipped deviation scores to the recovery-stage boundary θ in the long-term window. Lower values indicate better recovery performance.

G.3. Stakeholder group-weight sensitivity

Table G.2: Equal-group baseline plurality strategies. Each stakeholder group is assigned equal contribution, $1/8$, before aggregating the criterion weights.

Scenario	Plurality strategy	Share	Second strategy	Second share
10% short	Betweenness	80.0%	Closeness	15.1%
10% long	Betweenness	73.3%	Closeness	22.4%
30% short	Betweenness	74.4%	Closeness	21.3%
30% long	Betweenness	58.3%	Closeness	38.9%

Table G.3: Stakeholder group-weight sensitivity for 10% damage, short-term stage. One stakeholder group is assigned dominant weight p, with the remaining weight distributed equally across the other seven groups. Percentages use all 223 cities as the denominator; ambiguous or no-differentiation cases remain in the denominator but are not assigned to a strategy.

Dominant group	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$	$p = 0.6$
Emergency responder	B (66.4%)	B (67.3%)	B (68.6%)	B (68.6%)	B (68.6%)
Local community	B (66.4%)	B (66.8%)	B (67.7%)	B (68.2%)	B (68.6%)
Public authority	B (65.9%)	B (65.9%)	B (65.9%)	B (65.9%)	B (65.9%)
Road users	B (66.4%)	B (67.3%)	B (67.7%)	B (68.2%)	B (68.2%)
Technical expert	B (66.8%)	B (67.7%)	B (69.1%)	B (69.1%)	B (69.1%)
Transport operator	B (66.8%)	B (68.2%)	B (68.2%)	B (68.2%)	B (69.5%)

Table G.4: Stakeholder group-weight sensitivity for 10% damage, long-term stage. Percentages use all 223 cities as the denominator.

Dominant group	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$	$p = 0.6$
Emergency responder	B (68.2%)	B (64.1%)	B (63.2%)	B (62.3%)	B (62.3%)
Local community	B (68.6%)	B (68.2%)	B (68.6%)	B (67.7%)	B (65.9%)
Public authority	B (69.1%)	B (68.6%)	B (66.8%)	B (66.4%)	B (65.5%)
Road users	B (70.4%)	B (71.7%)	B (74.0%)	B (76.2%)	B (77.1%)
Technical expert	B (69.1%)	B (68.2%)	B (66.4%)	B (64.6%)	B (64.1%)
Transport operator	B (69.1%)	B (68.6%)	B (66.4%)	B (65.5%)	B (64.6%)

Table G.5: Stakeholder group-weight sensitivity for 30% damage, short-term stage. Percentages use all 223 cities as the denominator.

Dominant group	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$	$p = 0.6$
Emergency responder	B (69.5%)	B (70.0%)	B (72.2%)	B (72.2%)	B (72.6%)
Local community	B (69.5%)	B (69.5%)	B (70.4%)	B (71.3%)	B (71.7%)
Public authority	B (69.1%)	B (69.1%)	B (68.6%)	B (68.6%)	B (68.6%)
Road users	B (69.1%)	B (69.5%)	B (70.0%)	B (70.4%)	B (71.7%)
Technical expert	B (69.5%)	B (69.5%)	B (70.9%)	B (71.7%)	B (71.7%)
Transport operator	B (69.5%)	B (70.4%)	B (71.7%)	B (71.7%)	B (71.7%)

Table G.6: Stakeholder group-weight sensitivity for 30% damage, long-term stage. Percentages use all 223 cities as the denominator.

Dominant group	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$	$p = 0.6$
Emergency responder	B (55.2%)	B (51.6%)	B (50.2%)	B (48.0%)	B (46.6%)
Local community	B (56.5%)	B (55.6%)	B (55.2%)	B (54.3%)	B (52.9%)
Public authority	B (57.0%)	B (56.1%)	B (55.6%)	B (55.2%)	B (55.2%)
Road users	B (57.4%)	B (60.5%)	B (62.3%)	B (63.7%)	B (66.8%)
Technical expert	B (56.5%)	B (55.6%)	B (54.3%)	B (53.4%)	B (52.5%)
Transport operator	B (56.5%)	B (55.6%)	B (54.3%)	B (54.3%)	B (53.8%)

G.4. Delft network-boundary sensitivity

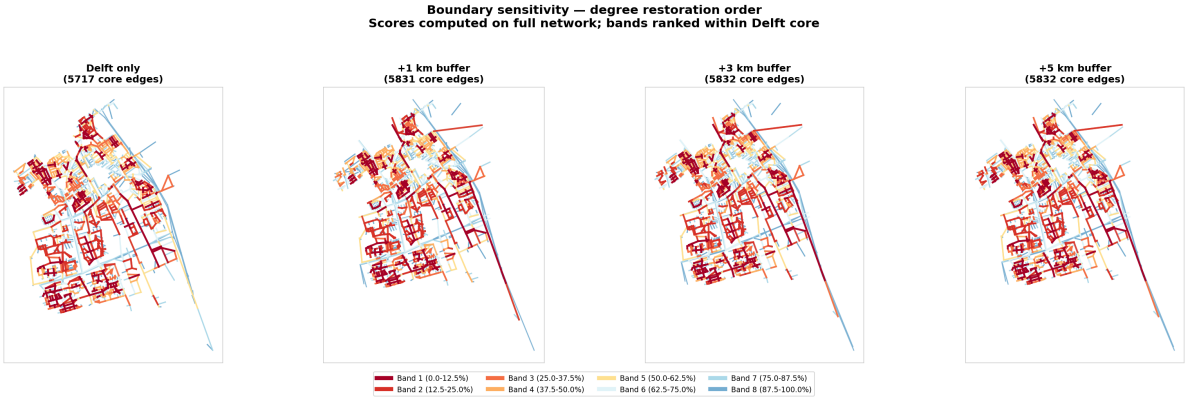


Figure G.5: Boundary sensitivity of the degree-based restoration order. Centrality is computed using each full network extent, while the visualised restoration bands are ranked only within the Delft core.

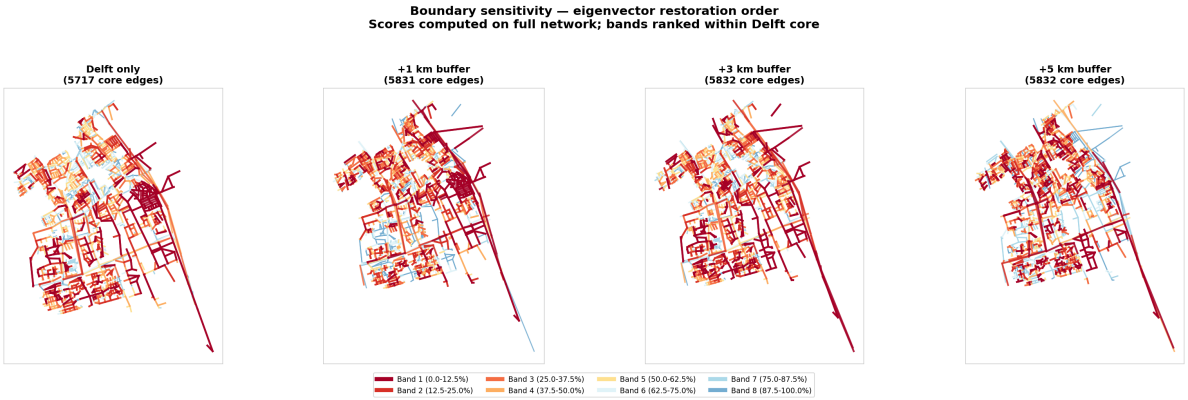


Figure G.6: Boundary sensitivity of the eigenvector-based restoration order. The maps compare the relative priority of Delft core roads under different buffer sizes.

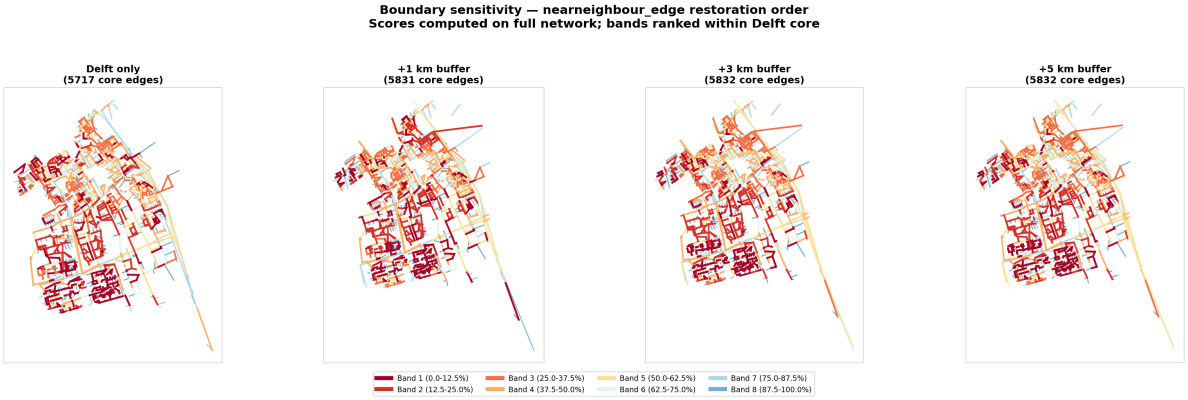


Figure G.7: Boundary sensitivity of the near-neighbour edge-based restoration order. Only Delft core edges are shown, while the buffer areas are used as computational context.

H

Scientific Paper

A Stage-Specific MCDA Framework for Comparing Post-Disaster Road Recovery Strategies

Yung-Ting Lu*, Patrick Stokkink†, Srijith Balakrishnan†, and Adam Pel* *Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
†Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands

***Abstract*—Post-disaster road recovery requires priorities to be set under limited resources. However, recovery performance depends on several criteria, and different stakeholder groups may have different priorities during short-term and long-term recovery. Existing approaches rarely combine stakeholder priorities, recovery stages, and the comparison of recovery strategies within one transparent evaluation. This study proposes a stage-specific BWM–SAW framework that combines stakeholder-derived criterion weights with simulation-based strategy performance to rank six predefined recovery strategies. The framework is demonstrated on 223 road networks, with Delft used as a complementary city-level application. The results show that short-term priorities focus strongly on access to critical facilities, while long-term priorities are more evenly distributed and differ more across stakeholder groups. Under the tested networks, damage scenarios, and modelling assumptions, betweenness ranks first most often. Closeness becomes more competitive under higher damage, during long-term recovery, and when greater importance is given to spatial equity or connectivity. The Delft application further shows that the strategy with the highest system-level ranking does not always perform best for every local accessibility outcome. The framework therefore supports the early-stage comparison of recovery strategies and gives stakeholders a structured basis for discussing recovery objectives and trade-offs, while system-level rankings should be interpreted alongside local checks suited to the decision at hand.**

***Index Terms*—Post-disaster recovery, road network resilience, multi-criteria decision analysis, Best-Worst Method, stakeholder priorities, recovery stages.**

I. Introduction

The effects of disasters on transport systems often extend beyond the damaged assets themselves. Transport systems support the movement of people, goods, repair teams, and emergency responders, so the speed at which they recover affects emergency response, access to essential services, and wider societal recovery [1], [2]. Road networks connect residents to employment, schools, healthcare, food, shelters, and other essential activities, while also supporting logistics and emergency response [3], [4]. Their wide spatial distribution exposes them to natural hazards. A global multi-hazard risk assessment estimated that approximately $27\,\%$ of road and railway infrastructure worldwide is exposed to at least one natural hazard [5]. When several roads are damaged simultaneously and repair crews, equipment, budgets, and reliable damage information are limited, infrastructure cannot be restored everywhere at once. Recovery planning must therefore determine which repairs should receive priority [2].

This prioritisation has no single obvious answer. Consider a simple recovery situation after a flood, in which a road authority must choose between reopening a major arterial road and repairing a smaller local connection. The arterial road may improve traffic flow and freight movement across the network. The local road may reconnect an isolated community to a hospital or emergency shelter. Emergency responders may prioritise the local road during short-term recovery, whereas infrastructure managers may prefer the arterial road because of its role in overall network functioning. The preferred recovery action therefore follows from the criterion being emphasised, the recovery stage being considered, and the stakeholder perspective being applied [6]–[8]. In practice, such trade-offs are often addressed through experience rather than formal decision-support tools [9].

Existing research addresses separate parts of this problem. Optimisation models can schedule the restoration of damaged roads or recovery projects under resource constraints, but they usually define recovery quality through one or a limited set of technical objectives [10], [11]. Multi-dimensional simulation can reveal trade-offs among operational efficiency, accessibility, spatial equity, and connectivity, but it does not directly produce a strategy ranking that reflects stakeholder preferences [6]. Stakeholder-oriented research recognises that different actors hold different priorities, yet these considerations are often represented through qualitative discussion or professional judgement. MCDA can integrate criterion weights with alternative performance, but road-recovery applications usually rank objects other than recovery strategies. Taken together, existing approaches rarely bring stakeholder priorities, recovery stages, and the comparison of recovery strategies together within a single transparent evaluation [12].

The objective of this study is to develop a stage-specific evaluation framework that combines stakeholder priorities with simulation-based recovery performance to compare post-disaster road recovery strategies. The framework treats short-term and long-term recovery as decision contexts with different priorities and incorporates the relative importance assigned to resilience criteria by different stakeholder groups.

This study addresses the following research questions:

1) How do stakeholder priorities for road-recovery resilience criteria differ across stakeholder groups and recovery stages?
2) How can stakeholder-derived criterion weights be used to rank predefined road recovery strategies based on simulated recovery performance?
3) How do the resulting strategy rankings vary across road networks, stakeholder groups, recovery stages, and

disruption levels in the cross-city, system-level analysis?
4) How do recovery strategies compare in a detailed city-level application, using Delft as a case study?

To answer these questions, this study proposes a stage-specific BWM–SAW framework. BWM converts stakeholder judgements into criterion weights specific to each recovery stage and stakeholder group. SAW then combines these weights with simulation-based recovery performance to generate composite scores and rankings for six predefined recovery strategies. Each combination of road network, damage level, recovery stage, and stakeholder profile forms a separate decision context.

The framework is demonstrated through two applications, one across many city networks and one within a single city. The cross-city, system-level application compares ranking patterns across multiple road networks, damage levels, recovery stages, and stakeholder profiles. Delft serves as a complementary city-level application and uses a more detailed facility-based representation of accessibility to compare recovery strategies within a single urban context. The main contribution of this study is the stage-specific BWM–SAW framework itself. It enables predefined recovery strategies to be compared systematically across stakeholder perspectives and recovery stages, while making trade-offs among recovery objectives more explicit within the decision process.

The remainder of this paper is organised as follows. Section II reviews related work and identifies the research gap. Section III presents the stage-specific BWM–SAW framework and the two applications. Section IV reports the results, covering stakeholder priorities across groups and recovery stages, the cross-city strategy rankings, and the city-level comparison in Delft. Section V discusses the implications, and Section VI concludes with recommendations for science and practice.

II. Related Work

A. Multi-Dimensional and Stage-Dependent Road Recovery

Post-disaster road recovery is commonly formulated as a restoration-sequencing or optimisation problem. Under constraints related to repair resources, workforce availability, damaged-road conditions, and travel demand, these models determine the order in which roads or repair activities should be restored [12]. Common objectives include reducing total recovery time, increasing recovery speed, lowering total travel time, meeting transport demand, and minimising unmet demand [13]–[16]. These models provide an important technical basis for repair scheduling and resource allocation. However, they usually rely on a small number of predefined objectives, which reduces the meaning of successful recovery to one or a limited set of performance indicators.

Transport resilience involves several related dimensions of performance. Existing indicators include travel time, service level, capacity, connectivity, cost, and qualitative measures, showing that the recovery state of a road network cannot be fully represented by one indicator [17]. Road-infrastructure resilience also includes technical, socioeconomic, and geospatial aspects [18]. A strategy that rapidly restores traffic flow may therefore leave weaknesses in service accessibility, the spatial distribution of recovery benefits, or network connectivity. Balakrishnan and Stokkink compared recovery strategies through simulation across operational efficiency, accessibility, spatial equity, and connectivity, and showed that no strategy performed best across all dimensions [6]. Their analysis makes the trade-offs between recovery logics visible, but it does not use stakeholder preferences to combine the multi-dimensional results into a strategy ranking that can directly support comparison and selection.

Post-disaster recovery is also stage-dependent because the tasks, constraints, and objectives may change over time [12]. Kaviani et al. distinguish response, short-term recovery, and long-term recovery, while Rey and Bar-Gera connect long-term recovery to extended reconstruction scheduling [10], [11]. Other studies use different indicators across recovery stages. Zhang et al., for example, evaluate mitigation, emergency response, and long-term recovery through robustness, emergency accessibility, and traffic-flow efficiency, respectively [7]. These studies establish the need for stage-specific evaluation. However, the indicators used in each stage and their relative importance are normally defined by researchers or embedded in the model, rather than being separately expressed by the stakeholders affected by recovery decisions.

B. Stakeholder-Informed Recovery Prioritisation

Road recovery involves stakeholder groups with different responsibilities and needs, including road users, local communities, technical experts, public authorities, and emergency actors. These groups may assign different importance to essential-service accessibility, network efficiency, spatial equity, and connectivity. Recovery prioritisation therefore involves both technical performance and value judgements. Disaster-recovery research has shown that differences in organisational objectives, weak coordination, and limited community engagement can affect resource use and recovery outcomes [8]. Responsibilities distributed across government levels and private organisations may also lead to fragmented recovery actions [19], [20]. The resulting priorities consequently depend on which stakeholders participate and which needs are represented.

The recovery-practice literature recognises the importance of stakeholder participation, local knowledge, and community needs. However, stakeholder priorities are usually represented through qualitative discussion, consultation, or professional judgement. Research involving Australian post-disaster road-reconstruction practitioners found that social, economic, and environmental considerations entered decisions largely through experience, tacit knowledge, and local understanding, without a systematic prioritisation procedure [9]. The literature therefore explains why stakeholder involvement matters, but rarely converts the priorities of different stakeholder groups into quantitative weights that can be directly used to rank recovery strategies. Studies that separately elicit, compare, and apply stakeholder weights for short-term and long-term recovery are even more limited.

C. MCDA Integration and Research Gap

Multi-criteria decision analysis provides a structure for combining multiple criteria, preference information, and alternative performance when objectives compete [21]. BWM can translate stakeholder judgements into criterion weights using a limited number of pairwise comparisons [22], while SAW combines weights and alternative performance through a weighted sum to produce comparable scores [23]. Performance inputs for MCDA may be obtained from quantitative data, models, or simulations [24]. Scholten et al., for example, combined modelled outcomes of water-network rehabilitation strategies with preference-based weights, while Bellsolà Olba et al. used simulation outputs to evaluate vessel-traffic scenarios [25], [26]. These applications show that simulation-based performance can support MCDA. However, MCDA studies in road recovery usually rank resilience indicators, infrastructure conditions, or reconstruction projects rather than alternative recovery strategies, and they generally apply one fixed set of weights throughout the recovery process.

The studies closest to the present research address separate parts of this problem. Balakrishnan and Stokkink compare the multi-dimensional performance of recovery strategies and reveal trade-offs between criteria, but do not use stakeholder-derived weights or distinguish recovery stages [6]. Zhang et al. apply different performance indicators across recovery stages, but the stage-specific priorities are defined by the researchers [7]. Mohammadnazari et al. use BWM and ranking methods to prioritise reconstruction projects, whereas Morshed et al., Xu et al., and Mukesh et al. use AHP, ANP, or related methods to evaluate resilience models, indicators, or infrastructure conditions [27]–[30]. Zamanifar and Hartmann develop decision attributes for transportation recovery, but focus on attribute selection rather than converting simulated strategy performance into stage-specific rankings [31]. These studies provide separate foundations for multi-dimensional evaluation, stage differentiation, stakeholder weighting, and MCDA aggregation. What remains missing is a method that combines stage-specific stakeholder-derived weights with simulation-based, multi-criteria strategy performance to rank recovery strategies across different decision contexts.

This study addresses this gap through a stage-specific BWM–SAW framework. The framework separately derives stakeholder criterion weights for short-term and long-term recovery and combines them with simulation-based strategy performance. A cross-city application examines ranking patterns across road networks, damage levels, recovery stages, and stakeholder profiles, while Delft provides a complementary city-level application with a more detailed facility-based accessibility representation.

III. Methodology

A. Framework Overview

This study develops a stage-specific multi-criteria decision analysis framework for comparing post-disaster road recovery strategies. As shown in Fig. 1, the framework links two analytical streams. The Best-Worst Method derives short-term and long-term criterion weights from stakeholder judgements, while road-network simulations evaluate the recovery performance of the candidate strategies. Simple Additive Weighting combines these inputs to rank the strategies for each stakeholder perspective and recovery stage.

The framework is demonstrated through cross-city and Delft applications. The cross-city application examines system-level patterns across multiple road networks, while the Delft application provides a more detailed city-level analysis using local facility data. Both applications follow the same weighting, simulation, and aggregation logic. The six recovery strategies, the 10% and 30% damage scenarios, and the two applications are study-specific settings used to demonstrate the framework, which can also be applied to other strategies, road networks, damage conditions, and accessibility representations.

B. Decision Problem Formulation

Recovery strategies. Six recovery strategies are compared: five centrality-based strategies and one random baseline, adopted from Balakrishnan and Stokkink [6]. For the centrality-based strategies, static priority scores are calculated for all road edges on the intact network and remain fixed throughout recovery. For each damage realisation, all strategies restore the same damaged-edge set but use different restoration orders. *Betweenness* prioritises edges used by many shortest paths; *weighted degree* prioritises edges connecting highly connected nodes; *nearest-neighbour centrality* combines local node strength with edge travel time; *closeness* prioritises edges connected to nodes with short distances to the rest of the network; and *eigenvector centrality* prioritises edges adjacent to other highly central edges. The *random baseline* restores the damaged edges in a random order. These structure-based rules allow the strategies to be compared consistently across different road networks.

Evaluation criteria. Recovery performance is evaluated against five criteria that extend the four recovery dimensions of Balakrishnan and Stokkink [6]. Operational efficiency (C1) is measured as graph-level network efficiency based on inverse shortest-path travel times [32]. Accessibility measures how easily destinations can be reached, and is separated in the stakeholder survey into access to critical facilities (C2) and access to routine facilities (C3). The cross-city simulation provides one aggregate accessibility metric, in which each destination is weighted by its pre-disruption closeness centrality and discounted by the shortest-path travel time from the origin, so C2 and C3 share this metric. The Delft application measures critical- and routine-facility accessibility separately using the same weighting. Spatial equity (C4) is measured as Moran's I of node-level accessibility, and connectivity (C5) as the average local directed clustering coefficient across all nodes [33]. The survey therefore uses five criteria while the cross-city simulation provides four metrics.

Recovery stages. Road recovery is divided into short-term emergency response and long-term repair and reconstruction because priorities and decision constraints change over time [7], [12]. In the stakeholder survey, short-term recovery was defined as temporary repair within approximately seven days [34], [35], while long-term recovery was defined as reconstruction over several weeks to months. Their implementation

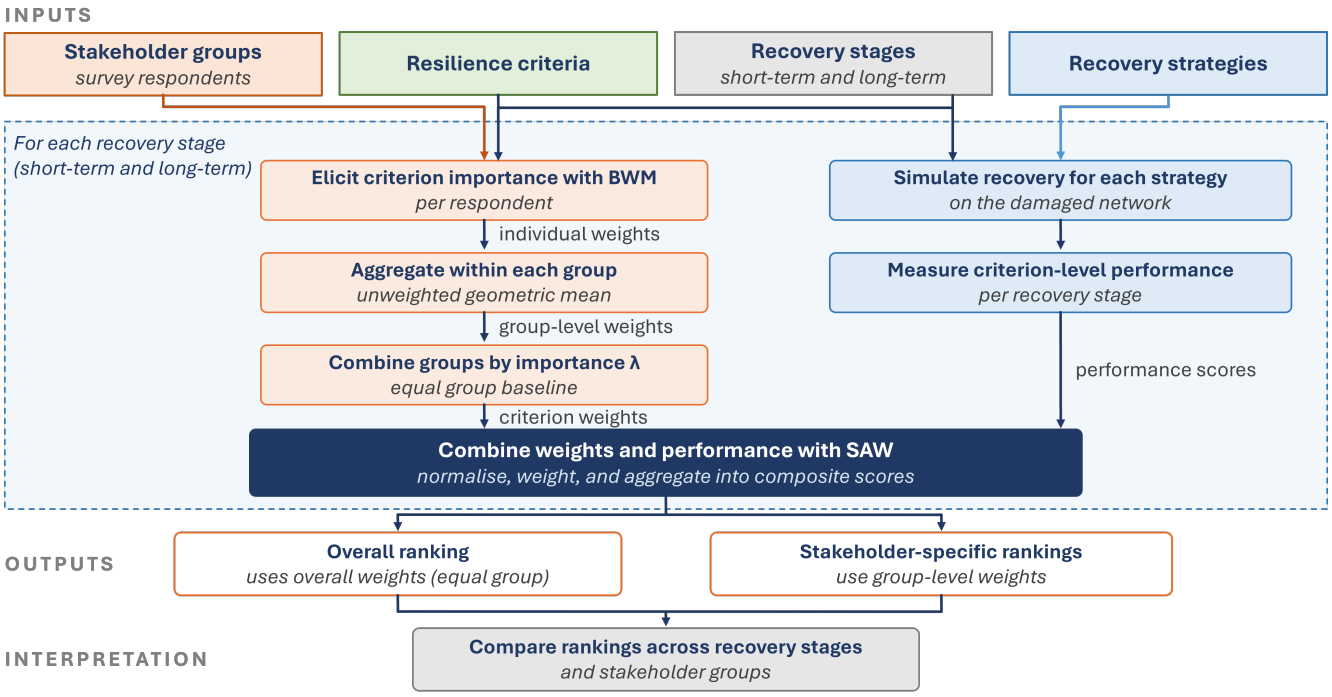


Fig. 1. Overview of the stage-specific MCDA framework.

as progress-based simulation windows is described in Section III-D.

Stakeholder groups. Eight stakeholder role types were derived by grouping stakeholders identified in the disaster-recovery and infrastructure-project literature: public authorities, infrastructure managers, technical experts, emergency responders, transport and logistics operators, contractors, local communities, and affected road users [10], [36]–[39]. These groups define the stakeholder perspectives used to derive group-specific criterion weights and recovery-strategy rankings.

C. Stakeholder Weighting with the Best-Worst Method

BWM formulation. The main analysis uses the linear Best-Worst Method (BWM) to estimate criterion weights separately for each respondent and recovery stage [40]. Respondents select the most important criterion B and the least important criterion W, then use a 1–9 scale to compare B with every criterion and every criterion with W. Let a_{Bj} and a_{jW} denote these comparisons. For the five criteria considered in this study, the normalised weights are obtained by solving

$$\begin{aligned} \min_{\mathbf{w},\,\xi_L} \quad & \xi_L \\ \text{s.t.} \quad & |w_B - a_{Bj} w_j| \le \xi_L, \qquad j = 1,\dots,5, \\ & |w_j - a_{jW} w_W| \le \xi_L, \qquad j = 1,\dots,5, \\ & \sum_{j=1}^{5} w_j = 1, \qquad w_j \ge 0, \quad j = 1,\dots,5. \end{aligned} \tag{1}$$

Here, w_j is the weight of criterion j, and ξ_L is the largest deviation between the stated comparisons and the estimated weights. Solving Eq. (1) produces one criterion-weight vector for each respondent and recovery stage.

Screening. Response validity was checked with the equivalent BWM of Wu *et al.* [41], which finds the smallest modification of a respondent's stated comparison vectors that satisfies the full BWM consistency condition. A response was excluded if its corrected comparisons still failed that condition, or if the selected best criterion was tied with another criterion in the Best-to-Others vector. All remaining responses were retained for the main analysis.

Weight aggregation. For each recovery stage, individual criterion weights were aggregated within each stakeholder group using the unweighted geometric mean, treating respondents within the same group as equally important [42]. For criterion i, stakeholder group g, and recovery stage s, the geometric mean is

$$\mathrm{GM}_{ig}^{s} = \left(\prod_{r \in R_g^s} w_{ir}^{s} \right)^{1/N_g^s}, \tag{2}$$

and the normalised group-level weight is

$$\bar{w}_{ig}^{s} = \frac{\mathrm{GM}_{ig}^{s}}{\sum_{k=1}^{5} \mathrm{GM}_{kg}^{s}}, \tag{3}$$

where R_g^s is the set of valid respondents in group g for stage s, N_g^s is the number of respondents in this set, and w_{ir}^s is respondent r's weight for criterion i. Overall weights were then calculated by equally aggregating the eight stakeholder groups represented in the valid sample:

$$w_{i,\mathrm{overall}}^{s} = \sum_{g=1}^{G} \lambda_g \bar{w}_{ig}^{s}, \qquad G = 8, \qquad \lambda_g = \frac{1}{G} = \frac{1}{8}. \tag{4}$$

Computing the overall weights from group-level values prevents groups with more respondents from automatically dominating the result. Equal group weights were used because the survey did not elicit the relative importance of the stakeholder groups. No respondent identified as a contractor and one respondent selected the "Other" category, so the eight groups in Eq. (4) are the eight roles reported in Table I.

Survey and coding. The questionnaire was distributed through community groups, professional networks, and professional and academic mailing lists. It elicited criterion judgements separately for the short-term and long-term recovery stages. Section A collected respondents' backgrounds, primary stakeholder roles, professional experience, and countries of practice, with each respondent assigned to one stakeholder group based on their primary role. Section B introduced the hypothetical disruption scenario, five evaluation criteria, and two recovery stages. Sections C and D collected the short-term and long-term BWM comparisons, respectively, together with open-ended explanations for the most- and least-important choices. The Mandarin explanations were translated into English and coded inductively to support interpretation of the weights; AI-assisted initial coding was reviewed and finalised by the author [43]–[45].

D. Simulation-Based Recovery Performance

Recovery trajectories. For each damage realisation, a fixed share of road edges is removed at random following Balakrishnan and Stokkink [6]. The same damaged-edge set is restored under each strategy according to its static edge-priority order, and the network performance metrics are recalculated at successive restoration steps. Let $M_i(x)$ denote the raw value of metric i when x percent of the damaged edges have been restored, and let $M_i(0^-)$ denote its pre-disruption value. The baseline performance ratio is

$$f_i(x) = \frac{M_i(x)}{M_i(0^-)}, \tag{5}$$

where $f_i(x) = 1$ represents the pre-disruption baseline. Tracking this ratio across restoration progress produces one recovery trajectory for each strategy, criterion, damage realisation, and damage level. For spatial equity, the ratio is interpreted as a baseline-deviation measure of Moran's I.

Stage-specific deviation scores. To align the progress-based simulations with the time-based stages used in the stakeholder survey, each trajectory is divided into a short-term window covering 0%–25% restoration and a long-term window covering 25%–100%. Aydin et al. report that one of their recovery strategies reached 25% recovery in 144.51 h, or about six days, close to the seven-day survey horizon [46]. The 25% boundary is therefore treated as a context-informed modelling assumption. For a recovery window $w = [a, b]$, performance is summarised using the area-above-curve (AAC) deviation score

$$\mathrm{AAC}_w = \frac{\int_a^b [1 - f(x)]\,\mathrm{d}x}{b - a}, \tag{6}$$

where lower values indicate better recovery performance. The integral is approximated using the trapezoidal rule, with linear interpolation when a window boundary falls between two restoration steps. The AAC score is calculated separately for each damage realisation. Any score outside $[0, 1]$ is clipped to this range before the scores are averaged across realisations to obtain the simulation input for SAW.

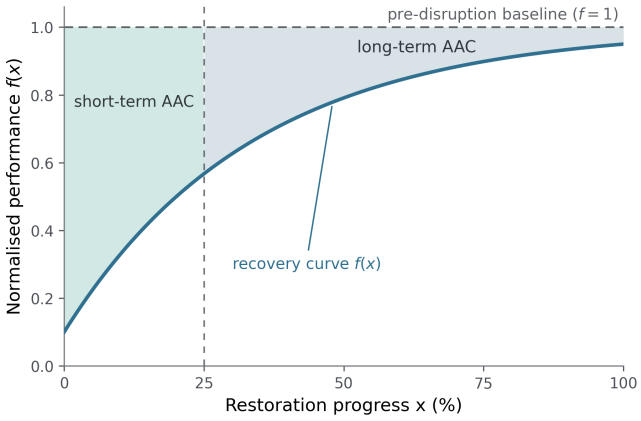


Fig. 2. The AAC deviation. Shaded areas between the pre-disruption baseline ($f = 1$) and the recovery curve give the short-term (0%–25%) and long-term (25%–100%) deviation scores. Lower values indicate better recovery.

E. SAW Aggregation and Strategy Ranking

Normalisation. Simple Additive Weighting (SAW) combines the BWM criterion weights with the mean AAC deviation scores obtained from the recovery simulations [23]. Because lower AAC values indicate better recovery, the deviation scores are converted into benefit-oriented scores using criterion-wise cost normalisation [47]:

$$r_{ij} = \frac{\max_k d_{ik} - d_{ij}}{\max_k d_{ik} - \min_k d_{ik}}, \tag{7}$$

where d_{ij} is the mean AAC deviation score of strategy j on criterion i. The maximum and minimum are calculated across the six strategies for the same city, damage level, recovery stage, and criterion. If all strategies have the same deviation score, they are assigned $r_{ij} = 1$.

Aggregation and ranking. Each strategy receives a composite score S_j, calculated as the weighted sum of its normalised criterion scores. In the cross-city application, critical-service and routine-activity accessibility share one simulation score, so their BWM weights are combined. In the Delft application, the two accessibility criteria are represented separately. The corresponding SAW scores are

$$\begin{aligned} S_j^{\mathrm{cross}} &= w_1 r_{\mathrm{eff},j} + (w_2 + w_3) r_{\mathrm{acc},j} + w_4 r_{\mathrm{equity},j} + w_5 r_{\mathrm{conn},j}, \\ S_j^{\mathrm{Delft}} &= w_1 r_{\mathrm{eff},j} + w_2 r_{\mathrm{crit},j} + w_3 r_{\mathrm{rout},j} + w_4 r_{\mathrm{equity},j} + w_5 r_{\mathrm{conn},j}. \end{aligned} \tag{8}$$

Because SAW is compensatory, stronger performance on one criterion can partly offset weaker performance on another [48]. Strategies are ranked in descending order of S_j, separately for each city, damage level, recovery stage, and stakeholder perspective. When two or more strategies receive the same highest score, the case is retained as a tied rank-one outcome

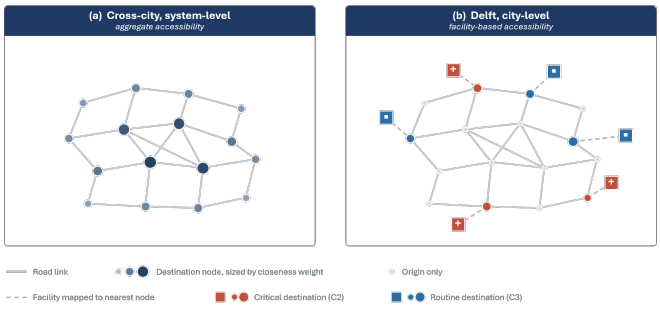


Fig. 3. Accessibility representations in the two applications. In the cross-city application (a), every road node is a destination, weighted by its pre-disruption closeness centrality, and critical (C2) and routine (C3) access share one aggregate score. In the Delft application (b), facilities are linked to their nearest road node and only those nodes are destinations, with critical and routine access scored separately using the same weighting. Every road node remains an origin in both applications.

but is not attributed to any single strategy in the rank-one shares. Tied cases are excluded from analyses that require a unique first-ranked strategy.

F. Cross-City and Delft Applications

Figure 3 summarises how accessibility is represented in the two applications.

Cross-city application. T cross-city analysis used the road-network recovery dataset of Balakrishnan and Stokkink [6], which covers cities and towns from 156 countries across six continents, extracted from OpenStreetMap and capped at 2500 road edges each. After removing networks without complete data for all required combinations of strategy, damage level, criterion, and recovery stage, 223 networks were retained. Each network was evaluated under $10\,\%$ and $30\,\%$ damage using 20 random damage iterations, with the same disrupted network state used for all six strategies within each iteration. The simulation provided operational-efficiency, aggregate-accessibility, spatial-equity, and connectivity scores. Aggregate accessibility was represented by closeness-weighted reachability, so critical-service and routine-activity access shared one performance score.

Delft application. The Delft road network was retrieved from OpenStreetMap on 13 June 2026 using OSMnx [49]. The drivable network was extracted with the `drive` filter, simplified, and restricted to its largest connected component. Edge weights were defined by free-flow travel time based on road length and speed; missing speed values were imputed by road type, with a fallback speed of $30\,\mathrm{km\,h^{-1}}$. The final directed network contained 2441 nodes and 5717 edges. Facility records were linked to their nearest road-network nodes, with duplicate destinations within each set counted once. The 58 critical-facility nodes represented emergency response, healthcare and medicine, food, and water, including police and fire stations, hospitals, clinics, pharmacies, supermarkets, marketplaces, drinking-water points, and wells [50]–[52]. The 364 routine-facility nodes represented education, workplaces, and general retail, with supermarkets excluded because they were already included in the critical-facility set. Every road-network node was treated as an origin, while the two facility sets were used as separate destinations with weights fixed by their pre-disruption closeness centrality. The Delft simulations used the same $10\,\%$ and $30\,\%$ damage levels, 20 realisations, and recovery procedure as the cross-city application.

Local accessibility check. A supplementary analysis compared the two highest-ranked Delft strategies, betweenness and closeness, using 30 spatially balanced residential origins. Unlike the SAW evaluation, which uses the Delft city network, this check uses a network extended by a $3\,\mathrm{km}$ buffer, so that services outside the municipal boundary that residents can realistically reach are included as destinations. For emergency, medical, food, and water services, travel time was calculated from each origin to the nearest reachable facility. Median travel times were calculated over reachable observations, while unreachable cases were reported separately. This analysis was not included in the SAW ranking.

Recovery-stage boundary sensitivity. The main analysis used 25% restoration progress to separate the short-term and long-term stages. To examine the influence of this assumption, AAC scores were recalculated for $\theta \in \{15\%, 20\%, 25\%, 30\%, 35\%\}$, where the short-term window was defined as 0% to θ and the long-term window as θ to 100%. Mean criterion-level deviation scores were then compared across damage levels, recovery stages, criteria, and recovery strategies.

Stakeholder-group aggregation sensitivity. The overall weights combine the eight stakeholder groups equally, which assumes that no group is more influential than another. To examine this assumption, the aggregation was repeated with one group assigned a dominance share $p \in \{0.2, 0.3, 0.4, 0.5, 0.6\}$ and the remaining weight $1 - p$ distributed equally among the other groups. Only groups with more than one valid response were given a dominant role, since increasing the weight of a single respondent would not provide a meaningful group-level check. For each value of p, the overall criterion weights were recomputed and the SAW ranking was repeated for both recovery-stage contexts and both damage levels.

IV. Results

A. Stage-Specific Stakeholder Priorities

A total of 115 survey responses were collected, of which 61 contained all required answers. Table I reports their distribution by stakeholder role. Of the 61 complete responses, 48 came through community groups, 10 through professional networks, and 3 through professional or academic mailing lists. This recruitment reached road users and local community members more readily than the professional stakeholder groups, and most respondents are based in Taiwan (40 of 61). After stage-specific BWM screening, 56 valid responses were retained for both the short-term and long-term analyses, although the respondents included in the two stages were not identical. Group-level interpretation focuses on road users, local communities, and technical experts because these groups had sufficient sample sizes. The five smaller groups were treated as exploratory in the group-level comparison

TABLE I
DISTRIBUTION OF COMPLETED SURVEY RESPONSES BY STAKEHOLDER ROLE

Stakeholder role	n	**Share**
Road users	22	36.1 %
Local community	19	31.1 %
Technical experts / researchers	9	14.8 %
Emergency responders	3	4.9 %
Transport / logistics operators	3	4.9 %
Public authorities	3	4.9 %
Infrastructure managers	1	1.6 %
Other	1	1.6 %
Total	**61**	100 %

but remained included in the equal-group aggregation used to calculate the overall weights. A total of 51 respondents had valid weights in both stages and were included in the paired transition analysis.

Figure 4 shows the criterion weights of the three main stakeholder groups. In the short-term stage, all three groups assigned the highest weight to accessibility to critical services (C2). Its weight was 0.442 for road users, 0.501 for local communities, and 0.374 for technical experts, clearly exceeding the weights of the other criteria. Technical experts also assigned a weight of 0.217 to operational efficiency (C1), which was higher than in the other two groups and slightly above the equal-weight reference of 0.20. The short-term results therefore show a shared focus on critical-service access, with technical experts placing additional importance on network efficiency.

The long-term weights were more evenly distributed, and the highest-weighted criterion differed across the three groups. Road users continued to assign the highest weight to C2, although its weight decreased to 0.279 and became closer to routine-activity accessibility (C3), at 0.232. Local communities assigned the highest weight to connectivity (C5), at 0.227, while their weights for C2, C3, and spatial equity (C4) were all close to 0.20. Technical experts assigned the highest weight to C4, at 0.244, followed by C2 at 0.234. From the short-term to the long-term stage, the weight of C2 decreased by 0.163 for road users, 0.294 for local communities, and 0.140 for technical experts. The largest increase occurred in C1 for road users and local communities and in C4 for technical experts. These results show that the shared short-term emphasis on critical services was redistributed across different long-term objectives, with the direction of this change varying by stakeholder group.

Respondent-level priority transitions. The paired transition analysis examined whether respondents retained a largely fixed priority list across the two recovery stages or followed a simple sequential shift, in which the second short-term priority became the first long-term priority. The results support neither explanation strongly. Among criteria ranked first in the short-term stage, only 15.7% remained first in the long-term stage, showing that the original top priority was usually not maintained. At the same time, only 17.2% of the long-term first-ranked cases came from criteria ranked second in the short-term stage, while 67.2% came from criteria previously ranked third, fourth, or fifth.

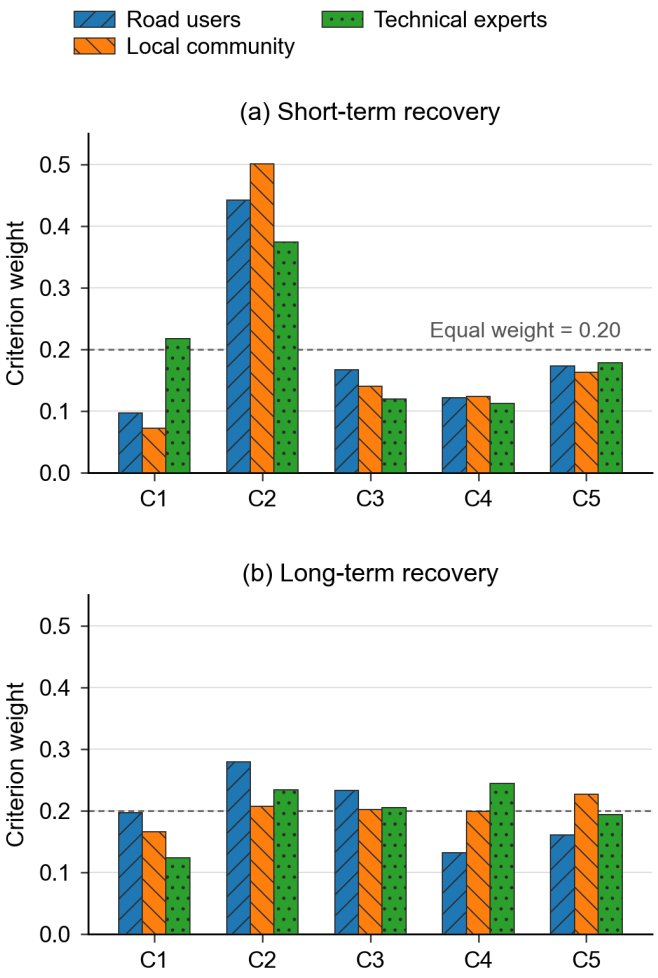


Fig. 4. Stage-specific criterion weights for the three main stakeholder groups. Panels (a) and (b) show the normalised group-level linear-BWM weights for short-term and long-term recovery, respectively. The dashed line indicates equal weighting across the five criteria, $w_i = 0.20$.

These patterns suggest that respondents did not simply keep the same ranking or move the remaining priorities up by one position. Instead, the relative importance of several criteria was reconsidered when the recovery stage changed. The respondent-level results therefore provide further evidence that recovery priorities are stage-dependent.

Interpretation of the open-ended responses. The open-ended explanations help interpret these weight patterns. Table II shows the themes identified in the reasons respondents gave for their most-important criterion. In the short-term stage, most explanations referred to life and basic needs, citing food, water, medicine, hospital access, and the minimum conditions required for survival. In the long-term stage, the dominant theme shifted to returning to normal daily life, and themes related to network efficiency, connectivity, equity, and longer-term resilience became more frequent. Some of these themes appeared only in the long-term explanations. This shift is consistent with the more even distribution of long-term

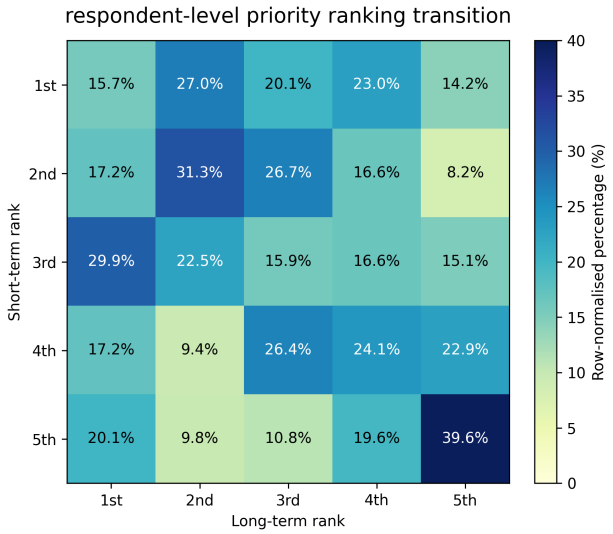


Fig. 5. Transition of respondent-level criterion rankings from short-term to long-term recovery. Each row shows the row-normalised distribution of criteria across long-term ranks, with darker shading indicating a higher percentage. Because each rank carries the same total fractional weight, the columns also sum to approximately 100%, apart from rounding differences.

TABLE II
THEMES IN THE OPEN-ENDED EXPLANATIONS OF THE MOST-IMPORTANT CRITERION ($n = 61$ IN EACH RECOVERY STAGE)

Theme	Short-term	Long-term
Life and basic needs	**45**	10
Emergency response and rescue access	7	1
Return to normal daily life	4	**21**
Connectivity and isolation prevention	3	8
Network efficiency and public benefit	1	10
Equity and inclusive recovery	1	6
Resilience and long-term preparedness	0	3
Ambiguous or unclear	0	2

criterion weights.

Overall, the results show that stakeholder priorities were both stage-specific and group-specific. The three main stakeholder groups shared a strong short-term emphasis on critical-service accessibility, whereas their long-term priorities were more evenly distributed and differed across groups.

B. Cross-City Strategy Rankings

Combining the stage-specific BWM weights with the normalised simulation results through SAW produced a strategy ranking for each road network, damage level, recovery stage, and stakeholder profile. Across the 223 road networks, most first-ranked results were concentrated in the betweenness and closeness strategies, with betweenness ranked first most often. Under 10% damage, betweenness was the most frequent first-ranked strategy for every stakeholder group in the long-term stage, and for all groups except the single-respondent infrastructure-manager group in the short-term stage. The same pattern held under 30% damage in the short-term stage, where betweenness ranked first in between 72.6% and 74.0% of the networks for the three main stakeholder groups. The clearest differences between stakeholder profiles appeared under 30% damage in the long-term stage and are therefore presented in Table III.

TABLE III
RANK-ONE STRATEGY SHARES ACROSS THE 223 ROAD NETWORKS UNDER 30% DAMAGE DURING LONG-TERM RECOVERY. THE THREE MAIN STAKEHOLDER GROUPS AND THE EXPLORATORY EMERGENCY-RESPONDER PROFILE ARE SHOWN. n IS THE NUMBER OF VALID LONG-TERM BWM RESPONSES IN EACH GROUP AFTER CONSISTENCY SCREENING, WHICH IS SMALLER THAN THE NUMBER OF COMPLETED RESPONSES IN TABLE I. COLUMNS DO NOT SUM TO 100% BECAUSE NETWORKS WITH A TIED FIRST-RANKED STRATEGY ARE NOT ATTRIBUTED TO A SINGLE STRATEGY.

Strategy	Road users ($n = 20$)	Local community ($n = 19$)	Technical experts ($n = 7$)	Emergency responders ($n = 2$)
Betweenness	**73.1%**	**51.6%**	**48.4%**	39.0%
Closeness	22.9%	40.8%	43.0%	**45.7%**
Degree	0.0%	1.3%	1.3%	2.7%
Eigenvector	0.9%	2.7%	3.6%	4.9%
Nearest-neighbour edge	0.0%	0.4%	0.4%	1.8%
Random	0.0%	0.0%	0.0%	2.7%

Under 30% damage in the long-term stage, betweenness remained the most frequent first-ranked strategy for all three main stakeholder groups, but closeness became more competitive. For the local community group, betweenness ranked first in 51.6% of the networks and closeness in 40.8%. For technical experts, the corresponding shares were 48.4% and 43.0%. Technical experts assigned their highest long-term weight to spatial equity (C4), for which closeness produced the lowest deviation scores. The increased closeness share for this group is therefore consistent with its criterion weights. Road users produced the largest betweenness share, with betweenness ranked first in 73.1% of the networks compared with 22.9% for closeness. Road users assigned 51.1% of their long-term weight to the two accessibility criteria, C2 and C3, combined, which is consistent with the continued advantage of betweenness for this group.

Among the five smaller stakeholder groups, emergency responders were the only group for which closeness became the most frequent first-ranked strategy under 30% damage in the long-term stage. Closeness ranked first in 45.7% of the networks, compared with 39.0% for betweenness. This pattern is consistent with the group's combined emphasis on spatial equity (C4) and connectivity (C5), which together accounted for 50.7% of its long-term weight. Closeness outperformed betweenness on both criteria, with mean deviation scores of 0.101 against 0.156 for spatial equity and 0.108 against 0.130 for connectivity. However, this result remains exploratory because the long-term emergency-responder weights were based on only two valid responses.

The effect of the recovery stage became more visible as the damage level increased. After excluding cases with a tied first-ranked strategy in either stage, the first-ranked strategy changed between the two recovery-stage contexts in 29.2% of the valid cases for technical experts, 25.4% for the local community group, and 14.6% for road users under

10% damage. Under 30% damage, these shares increased to 37.5%, 36.5%, and 19.2%, respectively. The increase across all three groups shows that differences between the short-term and long-term decision contexts became more visible under more severe network disruption. Road users had the lowest switching rate at both damage levels, which is consistent with their relatively stable emphasis on accessibility across the two recovery stages.

The strong results of betweenness and closeness did not arise because the two strategies restored the same links in a similar order. The average Spearman correlation between their intact-network edge-priority rankings was only $\rho = 0.35$. The two strategies therefore represent different restoration logics. The increased competitiveness of closeness in some stakeholder and network contexts reflects a meaningful change in the roads prioritised for restoration. As an illustration, Fig. 6 shows the two priority orders on the Delft network, where betweenness gives early priority to links connecting different parts of the network and closeness concentrates early priority in the central area.



Fig. 6. Static edge priority order on the Delft city network under the betweenness and closeness strategies. Band 1 contains the highest-priority eighth of the edge ranking and Band 8 the lowest. The bands summarise the priority order computed on the intact network, not the restoration steps used in the simulation. Black points mark the 58 mapped critical-facility nodes.

Recovery-stage boundary sensitivity. Changing the boundary between the short-term and long-term recovery stages affected the magnitude of the deviation scores but did not change the main criterion-level patterns. Across the tested boundaries, betweenness retained the lowest deviation scores for operational efficiency and accessibility at both damage levels and in both recovery stages. Closeness retained the lowest deviation scores for spatial equity. For connectivity, the best-performing strategy differed across damage levels and recovery stages but remained unchanged across the tested boundary values within each scenario. This sensitivity check examined criterion-level deviation scores and did not repeat the final SAW ranking.

Stakeholder-group aggregation sensitivity. The betweenness share varied with the dominant group. The largest variation occurred under 30 % damage in the long-term context, where it ranged from 46.6 % to 66.8 %. The choice of group weighting therefore affects the margin, while betweenness remains the most frequent first-ranked strategy in every tested scheme.

The BWM–SAW procedure translates stakeholder priorities into stage-specific strategy rankings. Betweenness ranked first in most decision contexts, while its margin over closeness narrowed under higher damage, in the long-term stage, and when more weight was placed on spatial equity and connectivity. The framework makes these differences in margin visible, which is the information needed when strategies are compared for a specific city and recovery stage.

C. Delft Application and Local Accessibility Outcomes

In the Delft application, access to critical facilities (C2) and access to routine facilities (C3) are scored against separate facility destination sets. Under this facility-based representation, betweenness ranked first and closeness ranked second across both damage levels, both recovery stages, and the three main stakeholder groups. The Delft SAW results therefore show that betweenness had the strongest overall performance when the five criteria were evaluated together using facility-based accessibility measures. At the criterion level, the two accessibility measures do not recover identically. Under 10% damage in the short-term stage, betweenness records an AAC deviation of 0.182 for routine access and 0.232 for critical access, a difference that a single aggregate metric cannot express.

The supplementary local accessibility check complements the system-level ranking by comparing the two highest-ranked strategies from a selected local road-user perspective. Under the 30% damage scenario and at 30% recovery progress, closeness produced shorter median travel times than betweenness across all four service groups, as shown in Figure 7(a). The largest difference occurred for emergency services, where the median travel time decreased from 3.55 to 3.02 minutes. Differences in unreachable shares were smaller, as shown in Figure 7(b). The two strategies produced similar results for emergency and food access, while closeness slightly reduced the unreachable share for medical services from 27.7% to 26.8% and for water services from 26.3% to 23.0%.

The Delft results show a difference between overall system performance and local user outcomes. Betweenness performed best in the five-criterion SAW evaluation, while closeness provided shorter travel times to nearby critical services for the selected residential origins. The local check therefore complements the city-level ranking by showing access outcomes that are not visible in the aggregated SAW score. At the city level, therefore, the multi-criteria ranking compares the recovery strategies, while local accessibility checks add information about their effects on residents.

V. Discussion

The results show that criterion priorities in post-disaster road recovery depend on the recovery stage. Short-term judgements concentrated on critical-service accessibility (C2). This

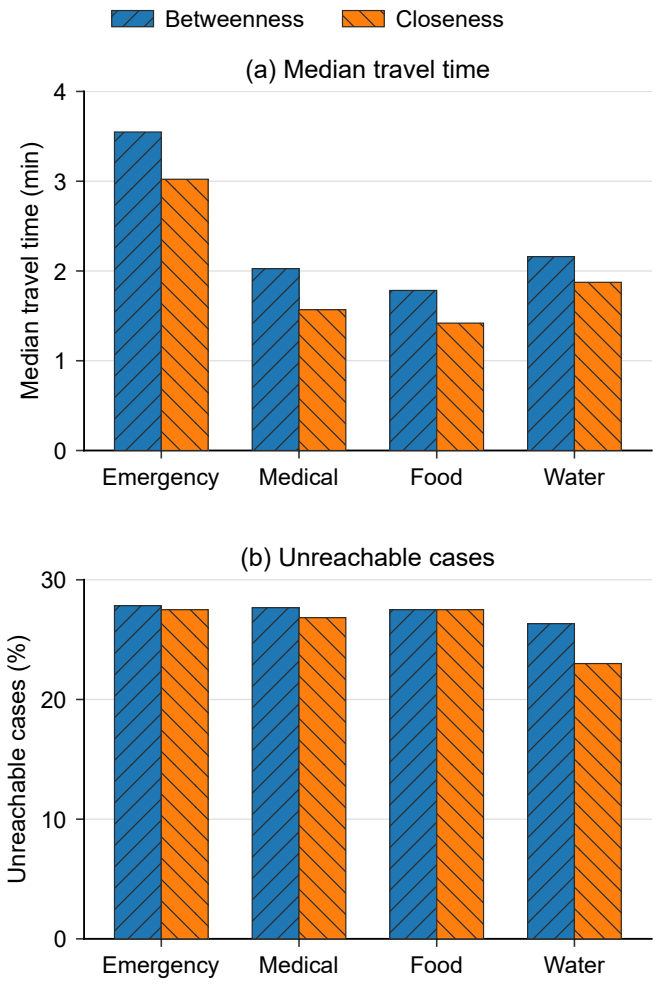


Fig. 7. Local accessibility outcomes for betweenness and closeness under 30 % damage at 30 % restoration progress: (a) median travel time to the nearest reachable facility and (b) unreachable cases by service group.

finding can be understood in light of empirical disaster research, as disrupted access to essential services directly affects residents' basic needs and medical care [53]. Long-term judgements were distributed more evenly across routine-activity accessibility, spatial equity, connectivity, and operational efficiency. The individual-level analysis further showed that respondents reassessed the criteria according to the recovery context instead of maintaining a fixed priority order. Previous studies have approached recovery stages mainly through technical performance measures, whereas this study identifies stage differences in stakeholder priorities. Zhang et al. distinguished emergency response from longer-term recovery using different network-performance indicators [7], while some restoration-ordering studies have evaluated longer-term recovery through total travel time or unmet demand [15], [16]. The present results show that long-term stakeholder judgements include a broader set of concerns than a single efficiency objective. Using one fixed set of weights throughout the recovery process may therefore overlook changes in decision priorities between stages.

Under the road networks, damage scenarios, and modelling assumptions tested in this study, betweenness was the most frequently top-ranked system-level strategy, mainly because it produced lower deviation scores for operational efficiency and accessibility. Its advantage nevertheless depended on the decision context. The two strategies order the same links differently and are strongest on different criteria, so they represent distinct restoration logics with contrasting criterion strengths. Closeness becomes a meaningful alternative when spatial equity or connectivity receives greater weight in long-term evaluation. These contrasting strengths may provide a basis for composite rules that combine the two orderings. Strategy rankings depend on both how each strategy performs across the criteria and how those criteria are weighted at each recovery stage. No single strategy should therefore be assumed to rank first across all recovery stages and stakeholder priorities.

The Delft application shows that the system-level SAW ranking and the local accessibility check may favour different strategies because they answer different decision questions. SAW uses five weighted criteria to compare the overall recovery performance of strategies across the road network, while the local check focuses on accessibility from selected residential origins to the nearest essential services. These perspectives provide complementary information because restoring roads that are important to the overall network and reconnecting residents to nearby essential services represent different recovery objectives. A strategy with stronger overall weighted performance may therefore provide less favourable outcomes for a specific local accessibility measure. The system-level ranking should be understood as a tool for comparing recovery strategies, rather than as a complete description of resident-level accessibility outcomes.

VI. Conclusion

This study developed a stage-specific BWM–SAW framework that combines stakeholder-derived criterion weights with simulation-based recovery performance to compare predefined road recovery strategies. Respondents evaluated the criteria differently in the short-term and long-term recovery contexts: access to critical facilities received the highest short-term weight across the three main stakeholder groups, while the long-term weights were more evenly distributed and the highest-weighted criterion differed across groups. Under the tested networks, damage scenarios, and modelling assumptions, betweenness was the most frequently first-ranked strategy, while closeness became more competitive in some higher-damage, long-term contexts. In Delft, betweenness ranked first in the facility-based SAW evaluation, while closeness produced shorter travel times for the selected origin-to-service trips in the local road-user checks.

The main contribution of this study is the stage-specific BWM–SAW framework, which brings stakeholder priorities, recovery-stage contexts, and simulation-based strategy performance into one evaluation process. It extends road-recovery and transport-resilience analysis by using stakeholder-derived weights to compare recovery logics across several criteria.

The separate elicitation of weights for the two recovery-stage contexts is methodologically valuable, because the long-term criterion rankings were generally not simple continuations of the short-term rankings. The study also shows that a system-level ranking does not describe location-specific outcomes. The cross-city application uses the road network of each city, but the destinations and outcomes it evaluates are generic, whereas a decision in a specific city depends on where people live, where the services are, and which outcomes matter locally. The Delft application shows how a local check can supply what the aggregate ranking does not.

In practical use, the framework does three things. It makes local and stage-specific stakeholder priorities visible, shows how these priorities affect strategy rankings and score margins, and links differences in the rankings to the underlying criterion weights and strategy performance. Together, these functions help stakeholders identify where recovery objectives align, where they conflict, and which trade-offs shape the evaluation. Discussion can therefore focus on the relative importance of recovery objectives in the specific recovery context, rather than on the final ranking alone.

When the framework is applied to a city, the system-level SAW ranking should be interpreted together with complementary analyses that reflect the specific decision context. These analyses may examine performance at a finer spatial scale, for particular services, or under additional local constraints. When the analyses favour different strategies, the difference reveals a trade-off between overall weighted performance and context-specific outcomes. By making the source of this difference visible, the framework gives stakeholders a concrete basis for discussing which recovery objectives should take priority.

Limitations. Several limitations shape the interpretation of the results. The BWM weights reflect stated preferences under hypothetical recovery contexts, while the stakeholder sample is uneven across groups and geographically concentrated. The results for the smaller groups should therefore be treated as exploratory. The strategy comparison also relies on controlled modelling assumptions. Random link failure does not capture the spatial concentration of real hazard damage, and the static restoration orders are not updated after the damaged roads become known. In addition, C2 and C3 share one aggregate accessibility metric in the cross-city analysis, meaning that only their combined weight influences the rankings. The interpretation of these rankings is further affected by the compensatory nature of SAW, since strong performance on one criterion can offset weak performance on another. The Delft local checks are also sensitive to the selected residential origins, facilities, and network boundary, so their findings remain specific to that analytical setup.

Future research. Future research should strengthen the empirical basis of the stakeholder weights and test the framework under more realistic recovery conditions. Larger and more geographically diverse samples would allow group-level differences to be examined more reliably, while elicited priorities could be compared with observed or historical recovery decisions. Hazard-specific and spatially correlated damage scenarios could then be used to test whether the ranking patterns remain stable when disruption is concentrated within particular areas. The candidate strategies could also be expanded to include adaptive, facility-targeted, and composite restoration rules, with the contrasting strengths of betweenness and closeness providing a starting point for their design. Finally, comparing SAW with non-compensatory or partially compensatory methods would show how strongly the rankings depend on the selected aggregation approach.

Taken together, the framework makes the trade-offs between stakeholder priorities, recovery-stage contexts, and recovery-performance dimensions visible. It gives a stage-specific way to compare post-disaster road recovery strategies, and it can support decision-makers by helping them see and weigh those trade-offs when recovery objectives compete.

REFERENCES

[1] United Nations Office for Disaster Risk Reduction and Centre for Research on the Epidemiology of Disasters, "The human cost of disasters: An overview of the last 20 years (2000–2019)," 2020, accessed 2026-07-09. [Online]. Available: https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019

[2] M. J. Benavidez and A. M. Mortlock, "Transport sector recovery: Opportunities to build resilience," World Bank Group and Global Facility for Disaster Reduction and Recovery, Washington, DC, Disaster Recovery Guidance Series, Oct. 2018. [Online]. Available: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/624471541689229858/transport-sector-recovery-opportunities-to-build-resilience

[3] K. Lucas, "Transport and social exclusion: Where are we now?" *Transport policy*, vol. 20, pp. 105–113, 2012.

[4] A. Rohr, P. Priesmeier, K. Tzavella, and A. Fekete, "System criticality of road network areas for emergency management services—spatial assessment using a tessellation approach," *Infrastructures*, vol. 5, no. 11, p. 99, 2020.

[5] 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*, vol. 10, no. 1, p. 2677, 2019.

[6] S. Balakrishnan and P. S. A. Stokkink, "Trade-offs in road network recovery: Analyzing resilience, efficiency, and connectivity from user perspective," *Transportation Research Procedia*, 2026, world Conference on Transport Research (WCTR 2026), Toulouse.

[7] W. Zhang, N. Wang, C. Nicholsonc, and M. H. Tehrani, "A stage-wise decision framework for transportation network resilience planning," *arXiv preprint arXiv:1808.03850*, 2018.

[8] A. Puri, M. Elkharboutly, and N. A. Ali, "Identifying major challenges in managing post-disaster reconstruction projects: A critical analysis," *International Journal of Disaster Risk Reduction*, vol. 107, p. 104491, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221242092400253X

[9] A. Gajanayake, T. Khan, and G. Zhang, "Post-Disaster Reconstruction of Road Infrastructure: Decision making processes in an Australian context," *European Journal of Transport and Infrastructure Research*, vol. 20, no. 1, pp. 1–16, Feb. 2020. [Online]. Available: https://journals.open.tudelft.nl/ejtir/article/view/3919

[10] A. Kaviani, R. G. Thompson, A. Rajabifard, and M. Sarvi, "A model for multi-class road network recovery scheduling of regional road networks," *Transportation*, vol. 47, no. 1, pp. 109–143, Feb. 2020. [Online]. Available: https://doi.org/10.1007/s11116-017-9852-5

[11] D. Rey and H. Bar-Gera, "Long-term scheduling for road network disaster recovery," *International journal of disaster risk reduction*, vol. 42, p. 101353, 2020.

[12] L. S. Almeida, F. Goerlandt, and R. Pelot, "Trends and gaps in the literature of road network repair and restoration in the context of disaster response operations," *Socio-Economic Planning Sciences*, vol. 84, p. 101398, 2022.

[13] W. Zhang, N. Wang, and C. Nicholson, "Resilience-based post-disaster recovery strategies for road-bridge networks," *Structure and Infrastructure Engineering*, vol. 13, no. 11, pp. 1404–1413, Nov. 2017, _eprint: https://doi.org/10.1080/15732479.2016.1271813. [Online]. Available: https://doi.org/10.1080/15732479.2016.1271813

[14] A. Kaviani, R. G. Thompson, and A. Rajabifard, "Improving regional road network resilience by optimised traffic guidance," *Transportmetrica A: Transport Science*, vol. 13, no. 9, pp. 794–828, 2017.

[15] C. Gokalp, P. N. Patil, and S. D. Boyles, "Post-disaster recovery sequencing strategy for road networks," *Transportation research part B: methodological*, vol. 153, pp. 228–245, 2021.

[16] T. Zhao and Y. Zhang, "Transportation infrastructure restoration optimization considering mobility and accessibility in resilience measures," *Transportation Research Part C: Emerging Technologies*, vol. 117, p. 102700, 2020.

[17] M. Z. Serdar, M. Koç, and S. G. Al-Ghamdi, "Urban transportation networks resilience: indicators, disturbances, and assessment methods," *Sustainable Cities and Society*, vol. 76, p. 103452, 2022.

[18] N. Nickdoost, M. J. Shooshtari, J. Choi, D. Smith, and Y. Abdelrazig, "A composite index framework for quantitative resilience assessment of road infrastructure systems," *Transportation research part D: transport and environment*, vol. 131, p. 104180, 2024.

[19] K. Tierney, "Disaster governance: Social, political, and economic dimensions," *Annual review of Environment and Resources*, vol. 37, pp. 341–363, 2012.

[20] A. Cho, "Post-tsunami recovery and reconstruction: governance issues and implications of the great east japan earthquake," *Disasters*, vol. 38, no. s2, pp. s157–s178, 2014.

[21] S. Greco, R. Słowiński, and J. Wallenius, "Fifty years of multiple criteria decision analysis: From classical methods to robust ordinal regression," *European Journal of Operational Research*, vol. 323, no. 2, pp. 351–377, 2025.

[22] J. Rezaei, "Best-worst multi-criteria decision-making method," *Omega*, vol. 53, pp. 49–57, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0305048314001480

[23] F. Ciardiello and A. Genovese, "A comparison between topsis and saw methods," *Annals of Operations Research*, vol. 325, no. 2, pp. 967–994, 2023.

[24] C. Macharis and A. Bernardini, "Reviewing the use of multi-criteria decision analysis for the evaluation of transport projects: Time for a multi-actor approach," *Transport policy*, vol. 37, pp. 177–186, 2015.

[25] L. Scholten, A. Scheidegger, P. Reichert, M. Mauer, and J. Lienert, "Strategic rehabilitation planning of piped water networks using multi-criteria decision analysis," *Water research*, vol. 49, pp. 124–143, 2014.

[26] X. B. Olba, W. Daamen, T. Vellinga, and S. P. Hoogendoorn, "Multi-criteria evaluation of vessel traffic for port assessment: A case study of the port of rotterdam," *Case Studies on Transport Policy*, vol. 7, no. 4, pp. 871–881, 2019.

[27] Z. Mohammadnazari, M. Mousapour Mamoudan, M. Alipour-Vaezi, A. Aghsami, F. Jolai, and M. Yazdani, "Prioritizing post-disaster reconstruction projects using an integrated multi-criteria decision-making approach: A case study," *Buildings*, vol. 12, no. 2, p. 136, 2022.

[28] S. A. Morshed, M. Arafat, S. Mokhtarimousavi, S. S. Khan, and K. Amine, "8R Resilience Model: A stakeholder-centered approach of disaster resilience for transportation infrastructure and network," *Transportation Engineering*, vol. 4, p. 100058, Jun. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666691X21000142

[29] W. Xu, Y. Zhang, D. Proverbs, and Z. Zhong, "Enhancing the resilience of road networks to flooding," *International Journal of Building Pathology and Adaptation*, vol. 42, no. 2, pp. 213–236, 2024, num Pages: 24. [Online]. Available: https://www.proquest.com/docview/3032844826/abstract/763E7719807140A3PQ/1

[30] M. M.S., Y. B. Katpatal, and D. S. Londhe, "Measurement of city road network resilience in hazardous flood events," *International Journal of Disaster Resilience in the Built Environment*, vol. 15, no. 2, pp. 274–288, Nov. 2022. [Online]. Available: https://doi.org/10.1108/IJDRBE-11-2021-0155

[31] M. Zamanifar and T. Hartmann, "Decision attributes for disaster recovery planning of transportation networks; a case study," *Transportation research part D: transport and environment*, vol. 93, p. 102771, 2021.

[32] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical review letters*, vol. 87, no. 19, p. 198701, 2001.

[33] G. Fagiolo, "Clustering in complex directed networks," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 76, no. 2, p. 026107, 2007.

[34] S. Yan and Y.-L. Shih, "A time-space network model for work team scheduling after a major disaster," *Journal of the chinese institute of engineers*, vol. 30, no. 1, pp. 63–75, 2007.

[35] S.-S. Liu, M. F. A. Arifin, W. T. Chen, and Y.-H. Huang, "Emergency repair scheduling model for road network integrating rescheduling feature," *Applied Sciences*, vol. 11, no. 4, p. 1447, 2021.

[36] M. Mojtahedi and B. L. Oo, "The impact of stakeholder attributes on performance of disaster recovery projects: The case of transport infrastructure," *International Journal of Project Management*, vol. 35, no. 5, pp. 841–852, 2017.

[37] R. J. Yang, Y. Wang, and X.-H. Jin, "Stakeholders' attributes, behaviors, and decision-making strategies in construction projects: importance and correlations in practice," *Project Management Journal*, vol. 45, no. 3, pp. 74–90, 2014.

[38] C. Macharis, L. Turcksin, and K. Lebeau, "Multi actor multi criteria analysis (MAMCA) as a tool to support sustainable decisions: State of use," *Decision Support Systems*, vol. 54, no. 1, pp. 610–620, Dec. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167923612002047

[39] Y. Chang, S. Wilkinson, E. Seville, and R. Potangaroa, "Resourcing for a resilient post-disaster reconstruction environment," *International Journal of Disaster Resilience in the Built Environment*, vol. 1, no. 1, pp. 65–83, 2010.

[40] J. Rezaei, "Best-worst multi-criteria decision-making method: Some properties and a linear model," *Omega*, vol. 64, pp. 126–130, 2016.

[41] Q. Wu, X. Liu, L. Zhou, J. Qin, and J. Rezaei, "An analytical framework for the best–worst method," *Omega*, vol. 123, p. 102974, 2024.

[42] B. Srdjevic, Z. Srdjevic, K. M. Reynolds, M. Lakicevic, and S. Zdero, "Using analytic hierarchy process and best–worst method in group evaluation of urban park quality," *Forests*, vol. 13, no. 2, p. 290, 2022.

[43] P. Mayring, "Qualitative content analysis," *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [On-line Journal], http://qualitative-research.net/fqs/fqs-e/2-00inhalt-e.htm*, vol. 1, 06 2000.

[44] S. De Paoli, "Performing an inductive thematic analysis of semi-structured interviews with a large language model: An exploration and provocation on the limits of the approach," *Social Science Computer Review*, vol. 42, no. 4, pp. 997–1019, 2024.

[45] J. Gao, Y. Guo, G. Lim, T. Zhang, Z. Zhang, T. J.-J. Li, and S. T. Perrault, "Collabcoder: a lower-barrier, rigorous workflow for inductive collaborative qualitative analysis with large language models," in *Proceedings of the 2024 CHI conference on human factors in computing systems*, 2024, pp. 1–29.

[46] 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*, vol. 31, pp. 832–843, 2018.

[47] N. Vafaei, R. A. Ribeiro, and L. M. Camarinha-Matos, "Assessing normalization techniques for simple additive weighting method," *Procedia Computer Science*, vol. 199, pp. 1229–1236, 2022.

[48] S. Schär, E. Pohl, and J. Geldermann, "Analysing the compensatory properties of the outranking approach promethee," *Journal of Multi-Criteria Decision Analysis*, vol. 32, no. 2, p. e70013, 2025.

[49] G. Boeing, "Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks," *Computers, environment and urban systems*, vol. 65, pp. 126–139, 2017.

[50] C. Bach, A. K. Gupta, S. S. Nair, and J. Birkmann, "Critical infrastructures and disaster risk reduction," *National Institute of Disaster Management and Deutsche Gesellschaft für internationale Zusammenarbeit GmbH (GIZ), New Delhi*, vol. 72, 2013.

[51] A. Kocatepe, M. B. Ulak, G. Kakareko, E. E. Ozguven, S. Jung, and R. Arghandeh, "Measuring the accessibility of critical facilities in the presence of hurricane-related roadway closures and an approach for predicting future roadway disruptions," *Natural Hazards*, vol. 95, no. 3, pp. 615–635, 2019.

[52] S. A. Enderami, E. Sutley, J. Helgeson, L. Dueñas-Osorio, M. Watson, and J. W. van de Lindt, "Measuring post-disaster accessibility to essential goods and services: proximity, availability, adequacy, and acceptability dimensions," *Journal of Infrastructure Preservation and Resilience*, vol. 5, no. 1, p. 12, 2024.

[53] F. Yuan, H. Farahmand, R. Blessing, S. Brody, and A. Mostafavi, "Unveiling dialysis centers' vulnerability and access inequality during urban flooding," *Transportation Research Part D: Transport and Environment*, vol. 125, p. 103920, 2023.