D3.3

Validated STT2 methodology & toolkit

Date
26.06.2026

Author: Ali Cheshomi, Arka Bhattacharyya, Martijn Warnier, Nazli Yonca Aydin

Organisation: TU Delft

Funded by the European Union

Project funded by the European Union's Horizon Europe and the UK Research and Innovation (UKRI) programme under the UK government's Horizon Europe funding guarantee grant agreement n°10062626

D3.3
Validated STT2 methodology & toolkit

Revision 5

Grant Agreement	101121356
UKRI numbers	10062626
Call identifier	HORIZON-CL3-2022-DRS-01
Project full name	AGnostic risk management for High Impact Low probability Events
Due Date	30.06.2026
Submission date	30.06.2026
Project start and end	01.10.2023 - 30.09.2027
Authors	Ali Cheshomi, Arka Bhattacharyya, Martijn Warnier, Nazli Yonca Aydin

Abstract

This deliverable presents the Stress Test Tier 2 (STT2) methodology and associated toolkit developed within the AGILE project. Building on the AGILE tiered stress testing framework, STT2 provides a structured, participatory, and semi-quantitative approach for assessing the resilience of critical infrastructure (CI) systems under High-Impact Low-Probability (HILP) event scenarios. The methodology focuses on identifying and analyzing physical, informational, organizational, and decision-related interdependencies that may contribute to cascading effects, systemic vulnerabilities, and coordination bottlenecks after disasters.

The deliverable describes the conceptual foundations, objectives, implementation process, analytical procedures, practical tools, resource requirements, and methodological boundaries of STT2. The methodology is organized around three main phases: preparation and scoping, participatory workshop implementation, and post-workshop analysis. Through stakeholder engagement, interdependency mapping, decision mapping, and structured analysis, STT2 supports the development of shared system understanding and provides evidence to inform resilience assessment and further modeling within the AGILE framework.

The deliverable also reports on the evaluation and refinement of the methodology, highlighting its applicability across different case-study contexts and spatial scales. Overall, STT2 offers a practical and scientifically grounded approach for exploring system behavior under stress and supporting resilience-oriented decision-making.

Document revision history

Issue	Date	Comment	Author
V1.0	08.04.2026	First draft	Ali Cheshomi
V2.0	20.05.2026	Second draft	Ali Cheshomi, Arka Bhattacharyya, Martijn Warnier, Nazli Yonca Aydin
V3.0	31.05.2026	First revision	Ali Cheshomi, Arka Bhattacharyya, Martijn Warnier, Nazli Yonca Aydin,

			Rabea Schulz, Gordana Čveljo, Beatriz Rosa, Jose Palma
V4.0	17.06.2026	Second revision	Ali Cheshomi, Arka Bhattacharyya, Martijn Warnier, Nazli Yonca Aydin, Rabea Schulz, Gordana Čveljo, Beatriz Rosa, Isabel Rodríguez
V5.0	26.06.2026	Third revision	Ali Cheshomi, Rabea Schulz

Acknowledgment

Project funded by the European Union's Horizon Europe under the grant agreement n°101121356 and the UK Research and Innovation (UKRI) programme. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Nature of the deliverable[1] **R**

Dissemination level

PU	Public, fully open. e.g., website	✔
SEN	Sensitive, limited under the conditions of the Grant Agreement	
CL	Classified information under the Commission Decision No2015/444	

[1] Deliverable types:
R: document, report (excluding periodic and final reports).
DEM: demonstrator, pilot, prototype, plan designs.
DEC: websites, patent filings, press and media actions, videos, etc.
OTHER: software, technical diagrams, etc.

Table of Contents

List of figures

List of tables

Abbreviations

STT	Stress Test Tier
HILP	High-Impact Low-Probability
Dx.x	Deliverable x.x
DSR	Design Science Research
DSRM	Design Science Research Methodology
TUD	Technische Universiteit Delft
GDPR	General Data Protection Regulation
CI	Critical Infrastructure
LoS	Loss of Service
AUC	Area Under the Curve
AAC	Area Above the Curve

Executive summary

Purpose of the ST methodology

The AGILE project aims to advance resilience assessment and improve understanding, anticipation, and management of High Impact Low Probability (HILP) events through the development of a threat-agnostic, system-level stress-testing framework. Building upon the conceptual foundations established in D3.1, this deliverable presents the validated Stress Test Tier 2 (STT2) methodology and its associated toolkit.

STT2 represents the intermediate level stress testing within AGILE's tiered stress testing framework. It operationalizes resilience stress testing through a participatory, semi-quantitative methodology designed to capture interdependencies across socio-technical systems. The methodology addresses key limitations of existing approaches, including their tendency to analyze infrastructure sectors in isolation, their reliance on hazard-specific scenarios, and their limited consideration of coordination and decision-making dynamics.

The primary purpose of STT2 is to assess how systems behave under extreme and uncertain stresses by deriving the performance of critical functions and the interactions within our socio-technical systems. The proposed methodology enables the identification of cascading effects, systemic vulnerabilities, and coordination bottlenecks that may arise during HILP events, particularly in the early response and recovery phases.

Validation/evaluation outcomes

The process of developing, evaluating, and validating STT2 is iterative and takes place through implementation across Europe, following a Design Science approach. The methodology has been implemented in four case studies so far. The completed implementation and evaluation in Rotterdam, together with the implementation of the methodology in three other case studies, provided insights into the methodological robustness and usability across complex multi-sectoral contexts. Further case studies are currently being implemented within the AGILE project in three more case studies, and the outcomes will be used to continue refining the methodology.

Readiness for operational use

The STT2 methodology has reached a level of development that allows for its application in real-world contexts, supported by a structured implementation process, facilitation guidelines, and analysis tools. Its design enables application across different spatial scales, infrastructure sectors, and governance settings, providing a practical basis for exploring system interdependencies and resilience under HILP events.

At the same time, the methodology is still being refined. Ongoing and upcoming case study applications within the AGILE project are being used to further test and improve its components, particularly in relation to facilitation practices, consistency of outputs, and integration with other tiers of the framework. As such, the current version should be understood as a robust and usable approach that continues to evolve through iterative validation and learning.

1. Introduction & Scope

Scope of the deliverable

This deliverable presents the design, development, and validation of the Stress Testing Tier 2 (STT2) methodology developed within the AGILE project. Building upon the conceptual framework and guiding principles defined in D3.1, it develops and operationalizes these principles into a methodology suitable for real-world implementation. This document provides a comprehensive description of the conceptual foundations and methodological design, implementation process, and evaluation and validation processes of the STT2 methodology. The aim is to provide both a scientifically grounded and practically applicable methodology that can support resilience assessment and decision-making in complex socio-technical systems.

Role of stress testing in AGILE, role of the STT2 within the ST framework

Within the AGILE framework, stress testing serves as a central mechanism for assessing the resilience of socio-technical systems under extreme and uncertain conditions. It complements other analytical and modeling approaches by providing a structured process for exploring how systems behave when exposed to stressors that exceed normal operating conditions.

The AGILE stress testing framework adopts a tiered structure, consisting of four levels. STT0 defines the system and its boundaries and stakeholders, and STT1 focuses on establishing a baseline qualitative understanding of the system, mainly by focusing on identifying the critical functions of the system. STT2, which is the focus of this deliverable, will go one step beyond that and introduce an approach that focuses on mapping and the exploration of interdependence. STT3 incorporates advanced modeling and quantitative analysis.

STT2 plays a central role by bridging qualitative insights and quantitative modeling, ensuring that the stress testing process is grounded in real-world system dynamics and stakeholder knowledge.

Scope of STT2

The scope of STT2 is defined by its focus on participatory stress testing of interdependent socio-technical systems under HILP conditions. The methodology is designed to capture system behavior across multiple dimensions, including physical infrastructure interdependencies, feedback loops, organizational coordination mechanisms, and the mapping of coordination decisions between actors and information flows. A key feature of STT2 is its emphasis on the early response and recovery phases of a crisis. These phases are critical for understanding how disruptions evolve over time and how decisions made under uncertainty influence system outcomes. By focusing on these dynamic aspects, the methodology provides insights that go beyond static assessments of system robustness. Furthermore, the methodology adopts a threat-agnostic perspective, allowing it to be applied to a wide range of HILP scenarios independent of specific hazard types. This enhances its flexibility and relevance in contexts characterized by deep uncertainty.

2. Conceptual Foundations

2.1. Theoretical foundations of the STT2 methodology

The development of the STT2 methodology is grounded in the principles of Design Science Research (DSR), which provides a structured approach for designing, implementing, and evaluating innovative solutions to complex problems. The methodology follows an iterative process that integrates theoretical knowledge with practical insights, ensuring both rigor and relevance. In particular, the approach builds on the framework proposed by Hevner et al. (2004), which conceptualizes DSR as a set of interacting cycles linking the problem environment, the knowledge base, and the design process. This is complemented by the Design Science Research Methodology (DSRM) introduced by Peffers et al. (2007), which provides a step-by-step process for developing and evaluating the methodological artifact. In line with this approach, the STT2 methodology is evaluated and refined through cycles of methodological development, theoretical reflection, and empirical testing.

2.2. Relevance and application of the theoretical foundations

The STT2 methodology operationalizes the key principles defined in the AGILE stress testing framework presented in D3.1: system focus, threat-agnosticism, resilience perspective, and the tiered approach. These principles are not only conceptual foundations but are actively embedded in the design and implementation of the methodology. Their relevance lies in addressing the limitations of traditional stress testing approaches, while their application within STT2 ensures that the methodology is both scientifically grounded and practically meaningful.

System focus

A central principle of STT2 is the adoption of a system-focused perspective, which conceptualizes critical infrastructures as interconnected socio-technical systems rather than isolated components. This perspective is essential for capturing the complex interdependencies that characterize modern infrastructures and for understanding how disruptions propagate across sectors.

Within STT2, this system focus is operationalized through the explicit mapping of interdependencies between infrastructure components, organizations, and decision-making processes. The system and its boundaries are defined during the preparation phase through close collaboration between the method developer (TUD) and the stress test host (case study partners), as well as by drawing outputs from STT0 and STT1 where available, and are presented to participants at the beginning of the workshop. Rather than analyzing individual sectors in siloes, the methodology examines how changes in one part of the system affect others, thereby enabling the identification of cascading effects and systemic vulnerabilities. This is particularly relevant in HILP contexts, where failures often emerge from interactions between system elements rather than from single-point disruptions.

The participatory nature of the methodology further reinforces this system perspective, as stakeholders collectively construct a shared representation of the system, including its boundaries, critical functions, and interdependencies. This shared understanding is crucial for capturing the complexity of real-world systems and for ensuring that the analysis reflects diverse perspectives and knowledge domains.

Threat-agnosticism

The STT2 methodology adopts a threat-agnostic approach, representing a shift from traditional hazard-specific stress testing toward an analysis of how systems behave under disruption. In the participatory workshop, a HILP scenario is still used as a practical entry point for discussion. Its purpose is not to limit the analysis to one specific hazard, but to

provide participants with a realistic and concrete stress context in which they can identify critical infrastructures, decisions, information needs, coordination requirements, and interdependencies. This scenario-based framing is important because stakeholders can more easily reason about system behavior when they are placed in a plausible disruption context that resembles practical emergency-management and training exercises.
The HILP scenarios are developed by the methodology developers in collaboration with the case-study host to ensure that they are sufficiently realistic for participants, and tailored to local conditions while remaining broad and flexible enough to support the mapping of interdependent systems. Once the physical, informational, organizational, and decision-related interdependencies have been mapped, the resulting networks can be stress-tested in a threat-agnostic way. Instead of modeling only the original scenario trigger, stress can be introduced through agnostic approaches, such as failing each part of the network one by one or randomizing failure locations across the network. This shifts the focus from the cause of disruption to the consequences of system degradation, cascading effects, and the system's ability to maintain or recover critical functions.
This approach is particularly relevant for HILP events, which are characterized by high uncertainty and limited historical data. By decoupling the analysis from specific hazard characteristics after the mapping phase, STT2 enables the identification of vulnerabilities that are inherent to the system's structure and functioning. In this way, the HILP scenario supports realistic dependency mapping, while the subsequent stress-testing analysis remains applicable beyond the specific hazard used in the workshop. This aligns with the AGILE framework's emphasis on resilience as an emergent property of system interactions.

Resilience perspective

The STT2 methodology is grounded in a resilience perspective, which focuses on the capacity of systems to absorb, adapt to, and recover from disruptions while maintaining critical functions. (Connelly et al., 2017). This perspective extends beyond traditional risk-based approaches, which typically emphasize the prevention of failures by considering how systems behave under conditions of stress and uncertainty. Within STT2, this perspective is operationalized by focusing on the dynamic evolution of systems during the early response and recovery phases of a crisis. These phases are critical for understanding how disruptions unfold over time and how decisions made under uncertainty influence system response outcomes. The methodology also emphasizes the identification of bottlenecks and vulnerabilities that affect the system's ability to maintain or restore its critical functions. This includes not only physical disruptions but also failures in coordination, delays in decision-making, and breakdowns in information exchange. By capturing these dimensions, STT2 provides a more comprehensive assessment of system resilience, reflecting the complex and adaptive nature of socio-technical systems.

Tiered approach and differentiation between the tiers

The STT2 methodology is embedded within AGILE's tiered stress testing framework, which provides a structured approach for conducting resilience assessments at different levels of complexity and analytical depth. The tiered approach enables flexibility in the application of stress testing, allowing users to select the appropriate level of analysis based on their objectives, available resources, and data.
Within this framework, STT2 occupies an intermediate position between Tier 1 (STT1) and Tier 3 (STT3), serving as a bridge between qualitative system understanding and advanced quantitative modeling, while whenever possible building on the system scoping conducted in Tier 0 (STT0). STT0 provides the initial definition of the system boundaries, relevant stakeholders, and scope of the stress testing exercise. STT1 focuses on the identification of system components, critical functions, and potential cascading effects through qualitative methods. It provides a high-level overview of the system and its vulnerabilities, but does not explicitly quantify interdependence. STT2 builds upon this foundation by

introducing a semi-quantitative and participatory approach that captures and structures interdependence across physical and organizational dimensions. It enables a more detailed and nuanced understanding of system dynamics, while remaining accessible and applicable in data-scarce environments by identifying and parameterizing interdependencies and feedback loops between systems and determining the potential for cascading failures across physical and organizational dimensions. STT3 further extends the analysis through advanced modeling techniques, such as network simulations and quantitative assessments of system performance. It provides a higher level of analytical precision but requires more extensive data and resources.
The differentiation between the tiers is not only methodological but also functional. While STT1 supports system scoping and initial understanding, and STT3 provides detailed modeling and simulation, STT2 plays a critical role in translating stakeholder knowledge into structured representations of interdependence and in generating insights that inform both decision-making and further and more advanced modeling efforts.

2.3. Assumptions and constraints

The application of the STT2 methodology is based on a set of underlying assumptions and is subject to practical, methodological, and ethical constraints. These considerations are essential for ensuring transparency in the interpretation of results and for supporting responsible implementation across different contexts.
A fundamental assumption of the methodology is the active participation of relevant stakeholders identified through the stakeholder mapping, representing different sectors, organizational roles, and decision-making levels. As STT2 relies on a participatory approach to capture interdependencies and system dynamics, the quality and completeness of the analysis are closely linked to the diversity, expertise, and engagement of participants. It is assumed that stakeholders possess sufficient knowledge of their respective systems, roles, and dependencies to meaningfully contribute to the mapping and analysis process.
Another key assumption concerns the availability of contextual information to support system definition and scenario development. While the methodology is designed to operate under conditions of uncertainty and limited data - characteristic of HILP events - it still requires a minimum level of background information, such as system boundaries, key actors, and critical functions. This information is typically provided through collaboration with case study hosts, supporting documentation, and insights from STT0 and STT1, when available.
The methodology also assumes that participants are able to engage in structured reflection and collaborative problem-solving, including the articulation of interdependencies, decision processes, and coordination mechanisms. This requires an environment that facilitates open dialogue and trust among participants, which in turn depends on effective facilitation.
In terms of constraints, the STT2 methodology is inherently semi-quantitative and context-dependent. While this allows flexibility and applicability in data-scarce environments, it also means that results are not fully generalizable or directly comparable across different case studies without careful interpretation. The reliance on stakeholders' input and their perceptions of the system introduces a degree of subjectivity, which influences the faithful identification and prioritization of interdependencies and vulnerabilities.
Practical constraints include the time and resource requirements associated with organizing and conducting participatory workshops, as well as the need for skilled facilitators capable of guiding discussions and ensuring that outputs are systematically captured. Variations in stakeholder availability, engagement levels, and institutional contexts also affect the consistency of implementation and the results obtained.

Ethical considerations are central to the application of the methodology. Given that discussions may involve sensitive information related to critical infrastructures, vulnerabilities, and organizational processes, it is essential to ensure confidentiality and responsible handling of information. Participants should be clearly informed about the purpose of the exercise, how the data will be used, and any limitations regarding data sharing. Where necessary, anonymization of results should be applied to prevent the identification of specific organizations or sensitive system weaknesses.
From a security perspective, particular care must be taken to avoid the unintended disclosure of critical vulnerabilities that could be exploited if improperly shared. The level of detail included in outputs and reports should therefore be carefully managed, balancing analytical usefulness with security considerations. This is especially relevant when working with operators of critical infrastructures and public authorities.
In relation to privacy, the methodology involves the collection of qualitative inputs from individuals, including opinions, experiences, and decision-making practices. As such, it is important to ensure that data collection complies with applicable data protection regulations, including the General Data Protection Regulation (GDPR). Personal data should be minimized, securely stored, and used only for the purposes of the stress testing exercise.
Finally, it is important to recognize that the STT2 methodology is implemented within an iterative learning framework. As such, assumptions and constraints are continuously revisited and refined based on insights from case study applications. This adaptive approach allows for the progressive improvement of the methodology while ensuring that ethical, security, and practical considerations remain integral to its development and application.

3. Objectives and Benefits

Objectives and scope of the methodology

The STT2 methodology has been developed to address key limitations in existing stress testing approaches. Its primary objective is to identify and analyze interdependence and feedback loops between systems to determine potential points of failure and cascading disruptions. This semi-quantitative methodology provides a structured and participatory approach for analyzing how critical infrastructure systems behave under stress, with particular emphasis on physical interdependencies and organizational interdependencies, system dynamics, and the evolution of disruptions under conditions of uncertainty.
In contrast to traditional methodologies that focus on hazard-specific risk assessment or component-level robustness, STT2 adopts a system-level perspective that prioritizes the performance of critical functions and the interactions between system components. The methodology is specifically designed to identify vulnerabilities that arise not only from physical disruptions but also from failures in coordination, decision-making, and information exchange. This reflects the understanding that, in the context of HILP events, systemic failures are often driven by complex interactions across multiple domains rather than by isolated technical breakdowns.
A key objective of the methodology is therefore to capture and analyze interdependencies across physical and organizational layers, enabling the identification of cascading effects that may emerge during extreme and uncertain disruption scenarios. In doing so, the methodology contributes to a more comprehensive understanding of systemic risk and resilience, going beyond the limitations of siloed infrastructure assessments.
Another central objective is to support stakeholder learning and shared understanding. By engaging diverse actors in a participatory process, STT2 creates a structured environment for collaborative problem exploration, enabling participants to reflect on their roles, dependencies, and coordination mechanisms. This process not only generates analytical

outputs but also fosters awareness of system complexity and interconnectivity, which is essential for improving preparedness and response capabilities.
The scope of the methodology encompasses the analysis of system behavior during the early response and early recovery phases of a crisis. These phases are critical for understanding how disruptions propagate and how decisions made under uncertainty influence system outcomes. By focusing on these dynamic stages, STT2 complements existing approaches that tend to emphasize preparedness or immediate response while overlooking the transitional phases where cascading failures often unfold and long-term resilience.
Furthermore, the methodology is designed to be threat-agnostic, allowing it to be applied across a wide range of scenarios without being constrained by specific hazards. This enhances its applicability in contexts characterized by deep uncertainty, where future threats may not be known or predictable.

Systems, sectors, and decision-making levels targeted

The STT2 methodology is designed for application to complex socio-technical systems, with a particular focus on critical infrastructures that support essential societal functions. These systems are characterized by high levels of interdependency, both within and across sectors, and by the involvement of multiple organizational actors operating at different levels of governance.
The methodology is applicable across a wide range of infrastructure sectors, including but not limited to energy, transport, water, digital infrastructure, and emergency services, in line with the definition of critical infrastructures provided in D3.1. Importantly, the approach is not limited to individual sectors but is explicitly designed to capture cross-sectoral interdependencies, recognizing that disruptions in one sector can have cascading impacts on others.
In addition to physical infrastructures, the methodology targets the organizational and institutional structures that govern system operation and response. This includes public authorities, infrastructure operators, emergency management agencies, and other relevant stakeholders. By incorporating these actors, STT2 acknowledges that system resilience is shaped not only by technical components but also by decision-making processes, coordination mechanisms, and information flows. The methodology is applicable across multiple decision-making levels, including:

- Operational level, where real-time decisions are made during crisis response
- Tactical level, involving coordination between organizations and sectors
- Strategic level, where policies, planning, and resource allocation decisions are defined

By engaging stakeholders across these levels, STT2 enables a comprehensive analysis of how decisions and actions at different levels interact and influence system performance. This multi-level perspective is particularly important in HILP contexts, where coordination across organizational boundaries is critical for effective response and recovery.
The flexibility of the methodology allows it to be applied at different spatial scales, from local and urban systems to regional and national contexts. This adaptability ensures that it can be tailored to the specific characteristics and needs of each case study.

Expected outputs and added value

The STT2 methodology generates a range of outputs that support both analytical understanding and practical decision-making.
A primary output is the development of interdependency maps, which capture the relationships between infrastructures, organizations, and decision-making processes. These maps provide a structured representation of the system, highlighting key

dependencies and potential pathways for cascading effects. In addition, the methodology enables the identification of critical vulnerabilities and bottlenecks, including those related to coordination failures, information gaps, and delays in decision-making. These insights are particularly valuable in HILP contexts, where such factors can significantly amplify the impact of disruptions.

Another important output is the enhanced understanding of system dynamics, including how disruptions evolve over time and how interactions between system components influence overall performance. This dynamic perspective supports more informed decision-making and helps stakeholders anticipate potential challenges during crisis situations.

Beyond these analytical outputs, the methodology provides significant added value through its participatory and learning-oriented approach. By involving stakeholders from different sectors and levels of governance, STT2 fosters a shared understanding of system interdependence and promotes cross-sector collaboration. This can contribute to improved coordination and more effective resource allocation in real-world contexts.

Furthermore, the methodology serves as a bridge between qualitative system understanding and quantitative modeling. The structured outputs generated through STT2 can inform subsequent analysis in STT3, ensuring continuity within the AGILE stress testing framework.

Overall, the added value of STT2 lies in its ability to integrate technical and organizational perspectives, operate under conditions of uncertainty, and provide actionable insights that support resilience enhancement in complex socio-technical systems. The method also supports the development of resilience strategies tailored for HILP events by revealing possible system bottlenecks and highlighting priority areas for intervention across both infrastructure and organizational networks.

4. STT2 methodology

The STT2 methodology was developed through an iterative design process that combined insights from research and practice. The starting point was a literature review conducted within the AGILE project to identify what had already been done in stress testing and where important gaps remained. This review highlighted the need for approaches that go beyond sector-specific and hazard-specific stress testing, address physical and organizational interdependencies, and capture cascading effects under HILP conditions (Cheshomi et al., 2026). In parallel, the methodology development was informed by participation in and observation of first-responder training exercises. These practical experiences helped the method developers understand how scenarios are used in emergency-management practice, how stakeholders interact during exercises, and where existing training formats could be extended to support systematic data collection and interdependency analysis.

Inspired by this research and practice inputs, the first version of STT2 was designed as a two-day participatory workshop, which was presented to project partners for their feedback. The two-day format was intended to provide sufficient time for problem structuring, physical interdependency mapping, decision and coordination mapping, resource discussion, and reflection across the early warning, early response, response, and recovery phases. However, during preparation and implementation planning, it became clear that a two-day format was difficult to realize in practice. Case-study hosts faced limitations in securing stakeholder availability for two full days, and there was a risk that the participant composition would change between days, reducing continuity in group discussions and affecting the comparability of outputs.

For this reason, the methodology was refined into a one-day format. To make this feasible while preserving the core methodological logic, several practical adjustments were introduced. The focus of the workshop was narrowed to the early response and early

recovery phases. This focus was selected because these phases are particularly important for understanding how disruptions evolve, how cascading effects emerge, and how decisions made under uncertainty influence system outcomes. The early response phase captures the immediate consequences of system stress, including urgent operational dependencies, coordination needs, and time-critical decisions. The early recovery phase captures the transition from immediate response to the restoration of critical functions, where resource prioritization, inter-organizational coordination, and infrastructure dependencies remain highly relevant. Together, these two phases provide a manageable but analytically meaningful timeframe for examining both the immediate and evolving effects of HILP events. Plenary reporting and cross-group discussions were reduced to prioritize group work and data collection. The number of critical infrastructures selected by participants was limited to a more manageable set, allowing groups to complete pairwise interdependency assessments within the available time. The dedicated resource-token allocation exercise was removed from the core agenda and retained only as an optional extension when time and group capacity allow.

Overall, the development journey of STT2 was one of practical refinement rather than fundamental redesign. The core methodological aim remained the same: to use stakeholder knowledge to map physical, informational, organizational, and decision-related interdependencies and to transform these outputs into structured evidence for stress testing. The main changes were introduced to address needs and gaps identified in the literature and practice, while improving feasibility, continuity of participation, clarity for facilitators and participants, and the analytical usability of the outputs.

4.1. STT2 Methodology description

Through a structured, participatory, and semi-quantitative approach, the STT2 methodology is designed to capture and analyze interdependencies within socio-technical systems and to explore how disruptions propagate across physical and organizational layers under HILP conditions. As illustrated in Figure 1, the methodology is structured into three main phases:

1) **Preparation and scoping**,
2) **Participatory implementation (workshop)**,
3) **Post-workshop analysis.**

These phases ensure a systematic process for defining the system and co-designing the event scenarios, engaging stakeholders, collecting data, and generating structured outputs that support resilience assessment. In the following sections, each of these phases will be described in more detail.

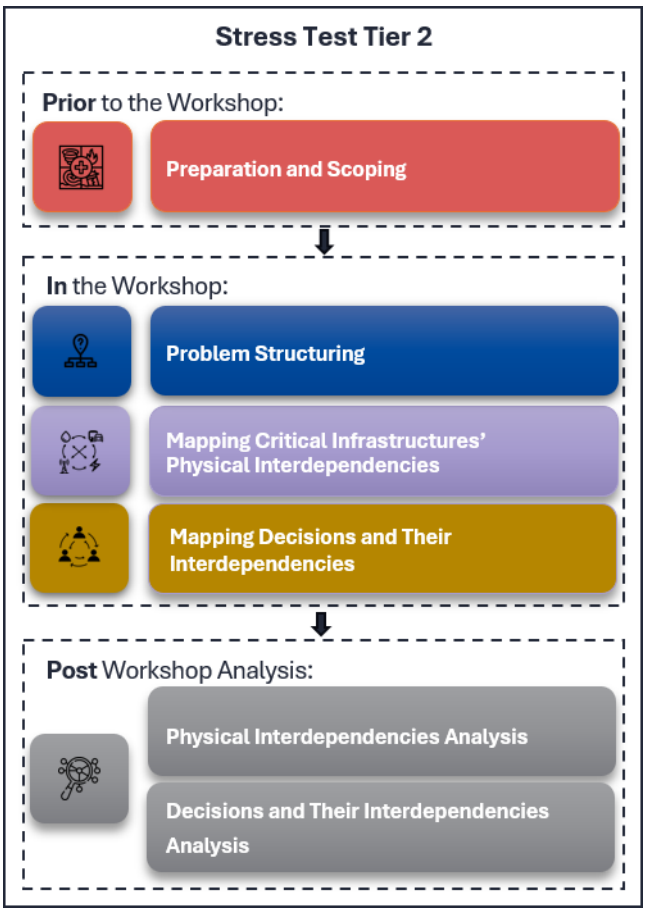

Figure 1: STT2 Methodology

4.1.1. Preparation, scoping, inputs, system definition, practicalities

Preparatory steps between the stress test host and the method developer

The preparation phase is conducted collaboratively between the stress test host and the methodology developers. This phase is critical for ensuring that the stress test is contextually relevant while maintaining methodological consistency.

Key preparatory activities include:

- Definition of the scope of the analysis (based on STT0 and STT1, when available).
- Definition of the system under analysis (based on STT0 and STT1, when available).
- Collection and review of relevant background information (based on STT0 and STT1, when available).
- Co-design of the stress scenario.
- Identification of key stakeholders and participants to be involved (based on STT0 and STT1, when available).
- Preparation of the workshop material.

Each of these activities is described in detail below.

Defining the scope of the analysis

This step involves establishing the boundaries, objectives, and level of detail of the stress testing exercise. It includes defining the level of the analysis (e.g., city, region, or national level) – previously defined in STT0 – as well as selecting the infrastructure sectors and critical functions to be considered. The scope is co-designed in close collaboration with the stress test host, ensuring alignment with the specific priorities, concerns, and context of the case study. In addition, this step also defines the temporal focus of the analysis, with

STT2 typically concentrating on the early response and early recovery phases of a crisis, where cascading effects and decision-making dynamics are most critical. The scope further determines the level of granularity of the analysis, including whether the focus is on high-level sectoral interactions or more detailed subsystem components. Inputs from STT0, such as system descriptions and stakeholder mapping, should be used when available to ensure that the selected scope captures the most relevant components and stakeholders of the system. In parallel, outputs from STT1 provide additional insights into key system elements and initially identified interdependencies that may require further analysis. These inputs support informed decisions regarding the geographical and temporal boundaries, as well as the appropriate level of granularity of the analysis. Importantly, the scope is defined in a way that remains sufficiently broad to enable the exploration of interdependencies and cascading effects, in line with the threat-agnostic approach, while also ensuring that the exercise remains feasible within the time and resource constraints of the workshop.

Defining the system boundary under analysis and stakeholders mapping

This step can be built directly on the preparatory activities conducted in STT0 and the initial system characterization carried out in STT1 to better ensure continuity within the AGILE tiered stress testing framework. It aims to establish a clear, shared understanding of the system to be analysed, along with a structured identification of the stakeholders who operate within, influence, or depend on it. The system is defined as a socio-technical network composed of interconnected infrastructure components, organizations, and decision-making processes. Drawing on the outputs of STT0, this step can use existing system descriptions, stakeholder lists, and contextual information to identify key system elements, including critical infrastructures, essential services, and critical functions. In addition, insights from STT1, such as the identification of key components, initial interdependencies, and critical functions, can be used to further refine and structure the system definition at a higher level of detail. This process also involves clearly defining system boundaries, including what is considered within the scope of the stress test and what remains external but may still influence system behaviour. Particular attention is given to interdependencies between system components, as these are central to understanding cascading effects in the context of HILP events.

Closely linked to system definition is the stakeholder mapping process, which can be built on STT0 and further develops the stakeholder identification initiated there, if necessary. This step aims to ensure that all relevant actors are represented, including infrastructure operators, public authorities, emergency management organizations, and other entities involved in maintaining critical functions and decision-making. Building on the engagement initiated in STT1, the mapping process also considers different decision-making levels, operational, tactical, and strategic, to capture how decisions are made and coordinated across the system. This is particularly important for analyzing organizational and informational interdependencies, which are key drivers of system performance during disruption. This step provides a structured and consistent foundation for the subsequent phases of STT2. It ensures that the stress-testing exercise reflects the complexity of real-world systems while remaining aligned with the overall AGILE framework.

Collection and review of relevant background information

This step involves collecting and reviewing available information to support the design and implementation of the stress testing exercise. Relevant inputs may include contingency and emergency response plans, risk and resilience assessments, infrastructure descriptions, policy documents, and publicly available data. In addition, expert knowledge

from the stress test host and stakeholders is used to complement formal sources. Given the HILP context and participatory nature of the methodology, the aim is not to achieve exhaustive data collection, but rather to ensure that sufficient contextual knowledge is available to support the scenario design and a better understanding of the system for stress test developers.

Co-design of the stress scenario

The stress scenario is developed collaboratively between the methodology developers and the stress test host. It is designed to represent a plausible yet challenging situation capable of triggering cascading effects across the system. The co-design process ensures that the scenario is relevant to the specific case study context, while remaining sufficiently flexible to allow the exploration of different system responses, interdependencies, and decision-making processes.

The main objective of the scenario is to provide a structured setting that reflects the defined scope, objectives, and system boundaries. It should involve a wide range of stakeholders within these boundaries and be formulated in a way that allows participants to realistically engage with the situation. This supports a more effective mapping of interdependencies, coordination mechanisms, and decision-making processes. At the same time, the scenario should be sufficiently engaging and challenging, going beyond expected scenarios, routine exercises, and training settings, while remaining generic enough to allow participants to explore multiple potential cascading effects. It should strike a balance between being detailed enough to support realistic interpretation and open enough to encourage exploration and discussion.

The development of the scenario builds on previous preparatory steps, including the definition of scope, system boundaries, and the collection of background information. The methodology developers work closely with the host expert team to define key scenario characteristics, such as the type of disruption or failure, timing (e.g., season, time of day, weekday or weekend), depending on the types of events, location of the triggering event, if necessary, duration, and sequence of events. Based on these inputs, an initial draft of the scenario is prepared by the methodology developer team and shared with the host experts for feedback. The scenario is then refined iteratively until a final version is agreed upon.

When the scenario includes multiple temporal stages, such as early response and early recovery, additional coordination with the host team is required to define the progression of events. This includes clarifying how much time has passed since the initial disruption, what actions have already been taken, and what challenges remain to be addressed in subsequent phases. This ensures that each stage of the scenario is clearly framed and supports meaningful engagement during the workshop.

Where relevant, the scenario can build on and further refine the scenario developed in Tier 1, ensuring continuity within the AGILE framework while enhancing its level of detail and applicability for the STT2 exercise.

Identification of key stakeholders and participants to be involved

This step involves identifying and selecting the relevant stakeholders and infrastructure sectors that should participate in the stress testing exercise. The objective is to ensure a comprehensive representation of the system under analysis. This includes actors from critical infrastructure sectors such as energy, transport, water, and digital infrastructure, as well as public authorities, emergency management organizations, and other entities involved in maintaining or restoring critical functions.

This step, where available, builds on the stakeholder mapping conducted in STT0 and the stakeholders identified and engaged in STT1. These earlier tiers can provide an important basis for selecting participants for STT2, ensuring continuity and consistency across the AGILE stress testing process. Using the outputs of STT0 and STT1 also helps avoid starting stakeholder identification from scratch and supports a more coherent understanding of the system boundaries, critical functions, and relevant actors.

Building on this basis, the case-study host identifies and selects a variety of relevant stakeholders to reflect different roles and responsibilities across the system, including operational, tactical, and strategic decision-making levels. Particular attention is given to capturing interdependencies across sectors and organizations, ensuring that both technical operators and coordination actors are included. This diversity of participants is essential for enabling a realistic mapping of interdependencies, information flows, and decision-making processes during the workshop, and for ensuring that the stress testing exercise reflects the complexity of real-world systems.

Preparation material and the team for the workshop

The final preparatory activity consists of developing and adapting the materials required for the workshop and training of the facilitation team. This includes facilitation tools such as templates, critical infrastructure lists, maps, and guidance documents, as well as scenario descriptions for participants. Materials are tailored to the specific context of the case study while maintaining alignment with the overall methodology. In addition, the methodology developer briefs and trains the facilitators and note takers before the workshop to ensure that they understand the objectives of each session, the expected outputs, the use of the templates, and the logic of the exercises.

Proper preparation of these materials in collaboration with the case-study host is essential for ensuring a structured implementation. The host contributes local knowledge, maps, terminology, and contextual information, while the methodology developer ensures methodological consistency and analytical usability. Together, these preparatory activities support facilitators and note takers in guiding the workshop process, capturing outputs consistently, and producing data that can be used for post-workshop analysis.

4.1.2. Participatory Implementation

The STT2 methodology is implemented through a structured participatory workshop designed to translate stakeholder knowledge into systematic stress testing outputs. The participatory approach is central to the methodology because many of the interdependencies that shape the resilience of socio-technical systems are not fully captured in existing datasets, plans, or technical models. Instead, they are often embedded in the operational experience, institutional knowledge, coordination routines, and decision-making practices of the actors who manage, operate, regulate, or depend on these systems. This design responds to the gaps identified in the AGILE literature review, which highlighted the need for stress testing approaches that move beyond sector-specific and hazard-specific analysis and better capture cascading effects, physical dependencies, coordination needs, information flows, and decision-making processes under HILP conditions (Cheshomi et al., 2026).

The workshop brings together stakeholders from different sectors and organizations, including critical infrastructure operators, emergency management actors, public authorities, service providers, and other relevant actors involved in maintaining or restoring critical functions. Participants work in small (4-7 participants per group), facilitated groups around a common HILP scenario. The scenario provides a realistic stress

context that helps participants reason about system behavior under disruption, while the structured exercises ensure that their knowledge is captured in a comparable and analyzable form through templates, maps, interdependency matrices, and decision tables. A key aim of the participatory implementation is to capture different types of interdependencies that influence system resilience. The methodology distinguishes between physical interdependencies and decision-related interdependencies. Physical interdependencies refer to the functional relationships between critical infrastructure components, services, or sectors, for example, how the disruption of electricity, transport, water, digital infrastructure, or healthcare may affect the performance of other parts of the system. Mapping these relationships is essential for identifying possible cascading effects and critical nodes in the infrastructure network. Decision-related interdependencies capture the organizational and governance dimensions of crisis response and recovery. These include dependencies related to information needs, information sources, resource availability, authority, prioritization, and coordination requirements. They may occur within the same organization, for example, between operational units, departments, or decision levels, but also across organizations, where decisions depend on information exchange, joint action, or coordination between infrastructure operators, public authorities, emergency services, and other stakeholders. Distinguishing physical and decision-related interdependencies is important because system vulnerability during HILP events may result not only from infrastructure disruption but also from delayed decisions, missing information, unclear responsibilities, or coordination bottlenecks.

The time frame of the workshop is also methodologically important. STT2 distinguishes between the early response and early recovery phases because interdependencies are dynamic rather than static. The infrastructures, actors, information flows, resources, and decisions that are critical immediately after a disaster may differ from those that become critical during the transition to recovery. By mapping physical and decision-related interdependencies separately for these two phases, the methodology captures how system priorities, dependencies, and vulnerabilities evolve over time. This supports a more dynamic understanding of resilience than a single static mapping exercise.

The implementation process follows a one-day workshop format. However, the duration and level of detail can be adapted depending on the case study context, the number of participants, the complexity of the system, and the availability of stakeholders. A typical implementation consists of an introductory plenary session, five main working sessions, and a final wrap-up and feedback session. The five working sessions are:

1) Problem structuring;
2) Physical interdependency mapping for early response;
3) Decision and organizational interdependency mapping for early response;
4) Physical interdependency mapping for early recovery;
5) Decision and organizational interdependency mapping for early recovery.

The figure and table below provide an overview of the workshop day. The table also summarizes the main objectives of each session and its key outputs.

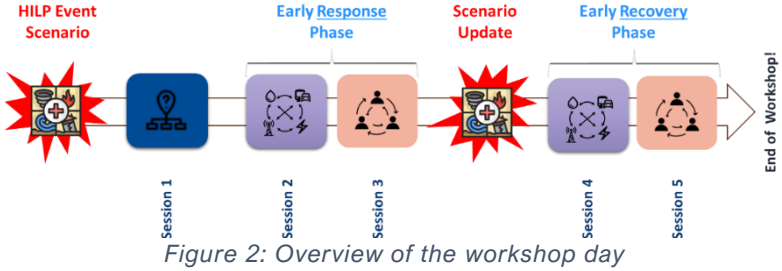

Figure 2: Overview of the workshop day

Table 1: Example agenda for STT2 participatory implementation

Workshop phase	Session	Main objective	Main output
Introduction	Arrival, welcome, workshop objectives, system boundaries, and HILP scenario	Introduce the AGILE project, STT2 methodology, case study scope, scenario, ground rules, and group assignments.	Shared understanding of the workshop purpose, scenario, system boundaries, and expected outputs
Problem structuring	Session 1: Problem Structuring	Identify stakeholder concerns, objectives, and possible intervention measures.	Clustered overview of concerns, objectives, and intervention measures
Early response	Session 2: Physical Interdependency Mapping	Identify critical infrastructures and assess their physical interdependencies during early response.	Network map and interdependency matrix for early response
Early response	Session 3: Decisions and Their Interdependency Mapping	Identify priority decisions, required resources, information needs, and coordination requirements during early response	Decision table for early response
Early recovery	Session 4: Physical Interdependency Mapping	Reassess critical infrastructures and physical interdependencies during early recovery.	Network map and interdependency matrix for early recovery
Early recovery	Session 5: Decisions and Their Interdependency Mapping	Identify priority decisions, required resources, information needs, and coordination requirements during early recovery	Decision table for early recovery
Wrap-up	Feedback and closing reflection	Reflect on the process, collect feedback, and identify key learning points	Participant feedback and final reflections

Introduction: workshop framing and scenario presentation

The workshop starts with an introductory plenary session. The purpose of this session is to create a common starting point for all participants. The host welcomes participants and introduces the local case study context. The plenary lead (usually the developer of STT2) then presents the AGILE project, the purpose of stress testing, and the role of STT2 within the wider tiered stress-testing framework. This introduction explains that STT2 focuses on the participatory mapping and analysis of interdependencies in socio-technical systems under HILP conditions.

The system boundaries and the HILP scenario are then introduced. This step is important because the quality of the workshop outputs depends on all participants working with the same understanding of the scenario, the geographical or organizational scope, and the

time phases under consideration. The scenario is presented in a concise but concrete way, explaining the triggering event, affected areas, disrupted services, and expected uncertainty. The aim is not to predict a specific future event, but to create a sufficiently challenging stress context that allows stakeholders to explore how the system behaves under extreme conditions.

Participants are then informed about the workshop structure, the five sessions, the expected outputs, and their role in the process. Group assignments are also explained. Groups are ideally prepared in advance to ensure a balanced mix of sectors, organizations, expertise, and decision-making roles. Ground rules are established to support constructive participation. These include respect for different perspectives, equal opportunity to contribute, confidentiality of sensitive information, and recognition that HILP events do not have a single correct interpretation or solution.

Session 1: Problem Structuring

The first working session is the problem structuring exercise. Its objective is to elicit how participants understand the problem space before the more detailed interdependency mapping begins. This session helps participants move from a broad crisis scenario to a more structured understanding of what is at stake, what needs to be protected, and what actions may be possible.

Participants are asked to identify three types of input:

1) Concerns,
2) Objectives
3) and Intervention measures.

Concerns refer to the main vulnerabilities, fears, or potential negative consequences that participants associate with the scenario, considering the services their organization is providing and the services that they are receiving in order to be able to function. **Objectives** refer to the priorities that participants want to protect or achieve during the crisis for their organizations, such as maintaining lifesaving services, ensuring public safety, protecting vulnerable populations, restoring essential services, or preserving access to critical locations. **Intervention measures** refer to possible actions that could reduce the impacts of the scenario or support response and recovery for their organizations.

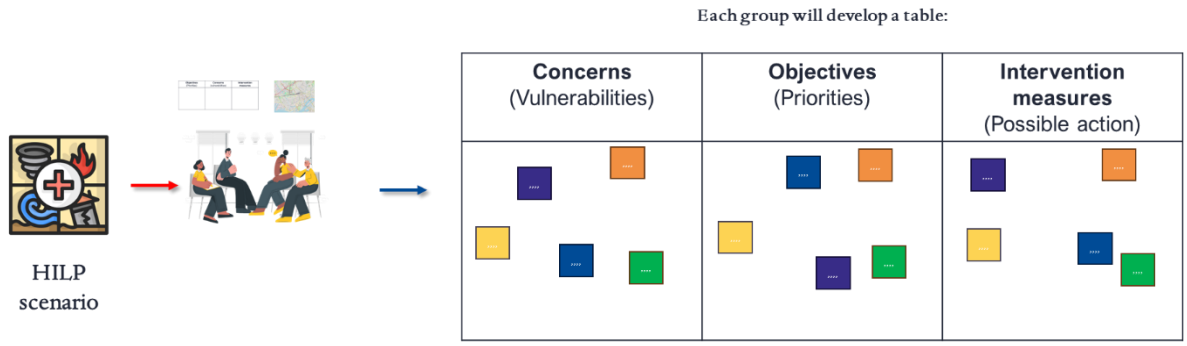

Figure 3: Example of problem structuring session

The exercise begins with individual brainstorming. Each participant writes down their own concerns, objectives, and possible interventions on sticky notes. This individual step is important because it allows different perspectives to emerge before the group discussion starts. After this, participants place their inputs on a shared A0 sheet or board divided into three columns: objectives, concerns, and intervention measures. The facilitator guides the group through each column, asking participants to briefly explain their inputs where needed. Similar inputs are then clustered together. The main output of this session is a

structured overview of the group's understanding of the problem, including the priorities, vulnerabilities, and possible actions identified by participants. The exercise also helps reveal competing or potentially contradictory priorities within the group, such as tensions between safety, continuity of services, public communication, and long-term recovery objectives.

Session 2: Physical Interdependency Mapping for the Early Response Phase

The second session focuses on physical interdependency mapping during the early response phase. The objective is to identify which CIs are most relevant immediately after the HILP event and to map how disruptions in one infrastructure may affect others. This session captures the physical and functional relationships between infrastructure components and supports the identification of cascading effects.

The session begins with a plenary explanation of the task and the meaning of physical interdependencies. Participants, after going back to their groups, are then asked to individually identify the critical infrastructures they consider most important for the early response phase for their own organizations, considering the service that they are providing and the services that they are receiving that are necessary at the early response stage of the scenario. These may include, for example, electricity supply, telecommunications, roads, hospitals, drinking water, wastewater, emergency shelters, ports, fuel supply, or other locally relevant infrastructure systems. The selection should be based on the scenario, the early response time frame, and the importance of the infrastructure for the participant's organization.

Color code	Sector	Sub-Sector	Component	Code
	Energy	Electricity	Powerplants	E1
			Transmission grids	E2
			Distribution networks	E3
		Oil	Importing oil, refining, treatment and storage, including pipelines	E4
			Transmission	E5
			Distribution	E6
		Gas	Production, refining, treatment and storage, including pipelines	E7
			Transmission	E8
			Distribution	E9
		Hydrogen	Production, storage	E10
			Transmission	E11
			Distribution	E12
	Transport	Aviation	Airports	T1
			Air traffic control	T2
		Rail	Railways	T3
			Major stations	T4
			control systems	T5
		Maritime & Ports	passenger water transport	T6
			freight water shipping	T7
			ports and their related facilities	T8
			cargo terminals	T9
		Road	Major road and highways	T10
			Bridges	T11
			Tunnels	T12
			Major stations	T13
		Drinking water	Reservoirs, groundwater wells	W1
			Treatment plants	W2
			Distribution Pipes, pumping stations	W3
			Collection: Sewer systems	W4

Figure 4: Example of a part of the list of critical infrastructures

After the individual selection, the group discusses the proposed infrastructures and agrees on a smaller set of priority infrastructures, usually between six and nine. This number is intentionally limited to keep the mapping exercise feasible within the available time while still capturing the most relevant interdependencies. The selected infrastructures are then placed on a map if spatial information is relevant. When an infrastructure is spatially distributed, such as a road network or electricity grid, the group may select a representative node or location for the purpose of the exercise.

The core of the session is the pairwise assessment of interdependencies. For each selected infrastructure, the group discusses how its severe disruption would affect each of the other selected infrastructures. The question guiding the assessment is: if this infrastructure fails or is severely degraded in the early response phase, how much does it affect the other infrastructure? The relationship is directional, meaning that the impact from infrastructure A to infrastructure B may be different from the impact from infrastructure B to infrastructure A. For this reason, reciprocal dependencies are assessed separately.

The group assigns an impact level to each dependency using a simple qualitative scale: **no effect, low impact (L), medium impact (M), or high impact (H)**. A low impact indicates a minor or manageable effect. A medium impact indicates a significant disruption that affects performance but does not necessarily stop the receiving infrastructure from functioning. A high impact indicates a severe dependency, where the receiving infrastructure may be unable to function properly if the source infrastructure fails. The facilitator ensures that the direction of the dependency and the impact level are recorded consistently on the map and in the interdependency matrix.

Dependencies are recorded in a **cause–and–effect matrix**. Each **row** represents the disrupted/source node, and each **column** represents the affected node. For every directed arrow in the network, the strength shown on the arrow is entered in the corresponding matrix cell. If there is **no arrow**, the value is **0**. The diagonal is marked with **–** because a node is not assessed against itself.

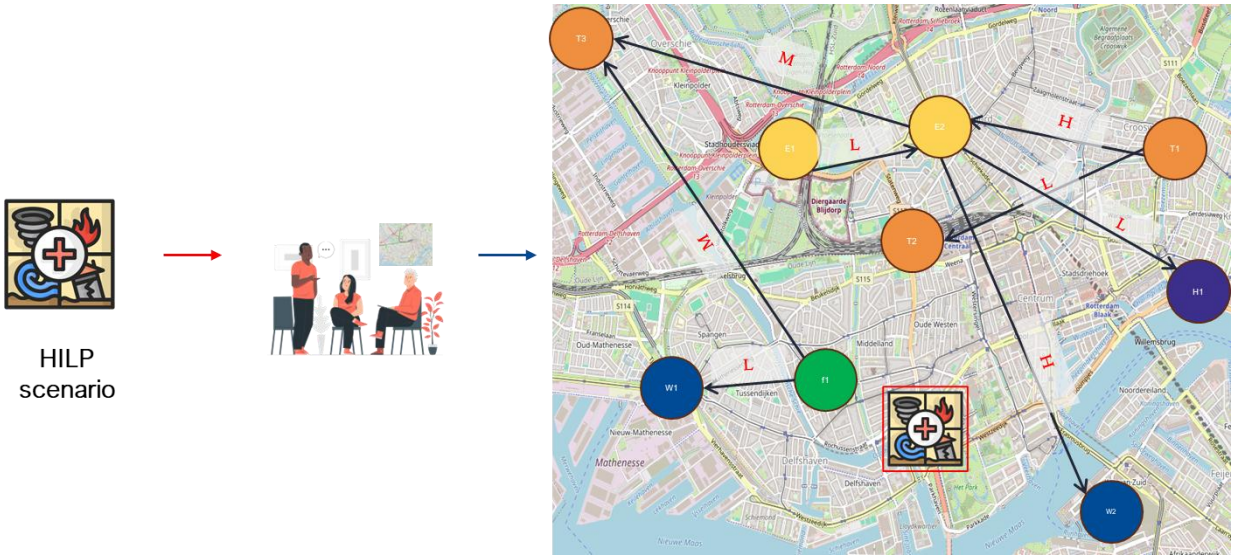

Figure 5: Physical interdependency map

Cause ↓ / Effect →	T3	E1	E2	T1	T2	H1	W1	F1	W2
T3	–	0	0	0	0	0	0	M	0
E1	0	–	L	0	0	0	0	0	0
E2	0	0	–	0	0	L	0	0	H
T1	0	0	H	–	L	0	0	0	0
T2	0	0	0	0	–	0	0	0	0
H1	0	0	H	0	0	–	0	0	0
W1	0	0	0	0	0	0	–	0	0
F1	M	0	0	0	0	0	L	–	0
W2	0	0	0	0	0	0	0	0	–

Figure 6: Example of Cause–Effect matrix

After the network is developed, and if there is time remaining, the group reflects on the emerging structure. Participants are asked to identify critical nodes, highly dependent infrastructures, possible cascading effects, and surprising dependencies. This reflection is important because it moves the exercise beyond simply drawing connections. It helps participants interpret what the network means for system resilience during the early response phase. The main outputs of this session are a physical interdependency network. map, a completed interdependency matrix, and a set of qualitative insights on critical components and cascading effects documented by the note taker.

Session 3: Decision and Organizational Interdependency Mapping for the Early Response Phase

The third session focuses on decision-making, information needs, resource dependencies, and coordination requirements during the early response phase. While the previous session maps physical dependencies between infrastructures, this session captures organizational and decision interdependencies that influence the ability of actors to respond effectively.

The session begins with an explanation that decisions during HILP events are rarely isolated. They often depend on information from other organizations, access to scarce resources, coordination with multiple actors, legal or institutional mandates, and the status of physical infrastructures. For this reason, the decision-mapping session aims to make these dependencies explicit.

Participants first individually identify the most important decisions that their organization, or the wider response system, would need to make during the early response phase. Examples may include decisions about evacuation, prioritization of emergency access routes, allocation of backup power, public warning, closure of infrastructure, deployment of emergency services, protection of vulnerable populations, or prioritization of repairs. Each participant writes their proposed decisions (2 per person) on sticky notes.

The group then discusses the proposed decisions and agrees on the top five to seven priority decisions. These decisions are placed in order of priority in a pre-structured decision table. For each decision, the group records the responsible organization for the decision, the resources required, the information needed, the source of that information, the coordination required, and the reason why the decision is considered a priority. This

structure makes it possible to identify dependencies between actors, resources, information flows, and decision-making processes.

Decision Made	Responsible Decision-maker	Resource Allocated	Information Needed	From Whom?	Coordination Required With	Rationale / Priority Reason
Send search and rescue teams to collapsed buildings in city center	Fire Department Incident Commander	1 rescue units, medical staff, heavy equipment	Exact locations of collapsed structures, population density, accessibility of roads	Local police, fire department	Hospitals, Red Cross, Fire Department, Police	To save lives within the golden 72-hour window after collapse
.....	.....	.....	.....	.....	.....	.....

Figure 7: Example of decision table

A key element of this session is the discussion of information and coordination needs. Participants are encouraged to specify what information is needed, from whom it should be obtained, and how delays, uncertainty, or missing information could affect the decision. They are also asked to identify which organizations need to coordinate before or during the decision. This allows the methodology to capture interdependencies that are not visible in physical infrastructure maps, such as dependence on crisis communication channels, situational awareness, authorization procedures, or coordination between public and private actors.

The main output of this session is a completed decision table for the early response phase. This table provides structured information on priority decisions, resource requirements, information dependencies, coordination needs, and reasons for prioritization. It can later be analyzed to identify critical bottlenecks in information and coordination dependencies.

Session 4: Physical Interdependency Mapping for the Early Recovery Phase

The fourth session repeats the physical interdependency mapping exercise but **shifts the focus from early response to early recovery**. This phase is treated separately because the relevance of infrastructures and the nature of dependencies may change as the crisis evolves. In the early response phase, the focus may be on lifesaving operations, emergency access, immediate communication, rescue capacity, and stabilization. In the early recovery phase, attention may shift toward restoring essential services, implementing temporary solutions, maintaining public service continuity, debris removal, repair logistics, supply chains, and longer-term functionality.

The session begins with a transition briefing that explains the early recovery phase and clarifies how it differs from early response. After receiving an updated scenario, participants are asked to reconsider which infrastructures are most important in this later phase of the scenario. Some infrastructures selected during early response may remain important, while others may become less central. New infrastructures or services may also become relevant as the system moves from immediate crisis management toward stabilization and restoration.

The group again selects a limited number (six to nine) of priority infrastructures and assesses the dependencies between them. The same pairwise logic is applied: if one infrastructure is severely disrupted in the early recovery phase, how much does it affect another infrastructure? The use of the same mapping logic across both phases allows comparison between early response and early recovery. This comparison is important

because it shows whether critical nodes, dependencies, and cascading effects remain stable or change over time.
The output of this session is a physical interdependency network map and interdependency matrix for the early recovery phase. These outputs help identify which infrastructures become more or less critical during recovery and where restoration efforts may be most constrained by dependencies on other systems. The session also supports reflection on temporal dynamics, showing that system vulnerability is not static but changes as operational priorities, resource needs, and service requirements evolve.

Session 5: Decision and Organizational Interdependency Mapping for the Early Recovery Phase

The fifth session focuses on decision and organizational interdependencies during the early recovery phase. As in the early response session, participants identify key decisions, required resources, information needs, and coordination requirements. However, the content of the decisions is expected to change because the system has moved from immediate response toward restoration and recovery.
Participants begin by identifying the most important early recovery decisions from their organizational perspective. These may include decisions on restoration priorities, reopening transport routes, temporary service provision, allocation of repair crews, public communication about service restoration, support for displaced populations, reopening schools or public facilities, waste and debris management, or coordination of reconstruction activities. The group then agrees on the top five to seven priority decisions and records them in the decision table.
For each decision, the group identifies the resources needed, the information required, the source of that information, the organizations involved in coordination, and the reason for prioritizing the decision. The facilitator encourages participants to reflect on how dependencies have changed compared to the early response phase. For example, a decision that was initially dependent on emergency services may later become dependent on infrastructure operators, municipal authorities, contractors, logistics providers, or social support organizations.
The main output of this session is a completed decision table for the early recovery phase. Together with the early response decision table, it enables analysis of how decision priorities, information dependencies, coordination needs, and resource bottlenecks evolve over time. This temporal comparison is one of the key strengths of STT2, as it captures not only what the system depends on, but also when and how these dependencies become critical.

Wrap-up and feedback

The workshop concludes with a plenary wrap-up. The plenary lead briefly summarizes the work completed during the day, outlines the next steps for analyzing the collected data, and invites participants to reflect on the workshop by completing an online survey.
The feedback survey supports the evaluation and refinement of the STT2 methodology. Participants are invited to reflect on the clarity of the process, the usefulness and the usability of the methodology for them, the realism of the scenario, and the relevance of the methodology for their own organization or context. This feedback is important because STT2 is designed as an iterative methodology that can be adapted and improved through case study applications.

4.1.2.1. Role of facilitators

Facilitators play a central role in the participatory implementation of STT2. Each working group is assigned a facilitator, ideally supported by a note taker. The facilitator acts as a neutral process guide rather than as a content expert or decision-maker. Their main responsibility is to ensure that the group follows the methodological steps, that all

participants have the opportunity to contribute, and that the required outputs are completed in a consistent and usable form.
At the beginning of each session, facilitators explain the task, expected output, and available time. They help participants understand the purpose of the exercise and keep the group focused on the scenario and time phase under discussion. Facilitators also manage group dynamics by encouraging quieter participants to contribute, preventing dominant voices from taking over, and ensuring that different organizational and sectoral perspectives are considered.
During the problem structuring session, facilitators guide participants through the identification, explanation, and clustering of objectives, concerns, and intervention measures. They encourage participants to clarify the reasoning behind their inputs, especially when disagreements, trade-offs, or constraints emerge. This helps ensure that the outputs capture not only the final clustered items, but also the reasoning behind them.
During the physical interdependency mapping sessions, facilitators guide the systematic selection and assessment of critical infrastructures. They ensure that the group selects a manageable number of infrastructures, places them clearly on the map where relevant, and assesses dependencies in a consistent way. A key facilitation task is to ask clear pairwise questions, such as: if this infrastructure is severely disrupted in this phase, what is the impact on the other infrastructure? Facilitators ensure that the direction of each dependency is clear, that the impact level is agreed by the group, and that the map and matrix are consistent with each other.
During the decision and organizational interdependency mapping sessions, facilitators help participants move from broad discussion to structured documentation. They guide the group in selecting priority decisions and filling in the decision table. For each decision, they ask what resources are needed, what information is required, who provides that information, which organizations must coordinate, and why the decision is important. They also encourage discussion of resource constraints, information gaps, coordination bottlenecks, and changes in decision priorities.
Facilitators are also responsible for time management and process discipline. They set or follow time limits for each step, remind the group of remaining time, and ensure that the session reaches the required output. They coordinate closely with note takers to make sure that important discussions, disagreements, assumptions, justifications, and insights are captured. This is essential because the later analysis depends not only on completed maps and tables, but also on understanding the reasoning behind the recorded dependencies and decisions.
Overall, the facilitator ensures the balance between open stakeholder discussion and methodological structure. The participatory setting allows stakeholders to contribute contextual knowledge and practical experience, while the facilitator ensures that this knowledge is translated into comparable and analyzable outputs. In this way, facilitators directly support the reliability, consistency, and usefulness of the STT2 methodology across different case-study contexts.

4.1.2.2. Tools

The STT2 methodology is supported by a set of practical tools that guide the preparation, implementation, analysis, and reporting of the participatory stress test. These tools are designed to ensure methodological consistency across case studies while allowing adaptation to different contexts, hazards, spatial scales, stakeholder groups, and workshop settings. The toolkit does not function as a rigid protocol; rather, it provides a structured set of templates, guidance documents, and analytical procedures that help hosts, facilitators, note takers, and analysts implement the methodology in a comparable and transparent way.
The tools are grouped into five main categories that represent the STT2 toolkit: 1. agenda templates, 2. facilitation guides, 3. practical implementation guides, 4. analysis guidance,

and 5. report templates. Together, these tools support the full STT2 process, from workshop preparation to the communication of results.
Table 2 provides an overview of the main categories of tools currently used to support the STT2 methodology. The table summarises their purpose, limitations, and adaptability, while recognizing that the detailed templates and guidance documents are still being refined through ongoing case-study implementations. In line with the iterative development approach of STT2, the current deliverable therefore describes the tool categories rather than including the full toolkit as appendices. The updated and consolidated versions of the tools will be included in a later version of the methodology deliverable once further validation cycles have been completed.

Table 2: Practical tools that guide the preparation, implementation, analysis, and reporting of the STT2

Tool	Purpose	Limitations	Adaptability
Agenda templates	Structure the workshop day and define the sequence, duration, setting, and responsibility for each session.	Timing may need adjustment depending on participant numbers, translation needs, local constraints, or the complexity of the case.	Can be adapted to half-day, two-day, or extended formats; sessions can be shortened, extended, or split across multiple days.
Facilitation guides	Provide detailed instructions for facilitators on how to run each workshop session and guide participant discussions.	The quality of outputs depends on facilitator experience, group dynamics, and the ability to manage time and discussion.	Can be adapted to local language, case-specific scenario, stakeholder profile, and in-person or online implementation
Practical implementation guides	Support hosts in preparing logistics, participant recruitment, materials, room setup, registration, equipment, and follow-up	Cannot fully anticipate local administrative, cultural, venue, or stakeholder availability constraints	Can be tailored to the host organization, venue, available resources, participant profile, and national or local requirements
Note taker and documentation templates.	Support systematic data capture during group work, including key discussions, disagreements, assumptions,	Outputs may still be incomplete if discussions move too quickly, if handwriting is unclear, or if note takers are not	Can be adapted to the number of groups, workshop language, digital or paper-based formats, and the

	matrices, and decision tables.	sufficiently briefed.	required level of analytical detail
Analysis guidance	Provides procedures for transforming workshop outputs into structured analytical outputs, such as interdependency matrices, network representations, decision-dependency tables, and bottleneck analysis	Analysis depends on the quality, completeness, and consistency of workshop data; subjective interpretation may be required when outputs are ambiguous.	Can be adapted to different levels of analytical depth
Report templates	Provide a common structure for reporting case study results, key findings, methodological reflections, and recommendations.	Templates may not capture all case-specific insights or locally sensitive issues; results may need to be anonymized or aggregated.	Can be adapted to different audiences, including project partners, case study hosts, policy stakeholders, infrastructure operators, and scientific reporting

Agenda templates

The agenda template structures the participatory implementation of STT2. It defines the sequence of sessions, the indicative duration of each session, the workshop setting, and the responsible actors. A typical agenda includes an introduction, a problem structuring session, two physical interdependency mapping sessions, two decision and organizational interdependency mapping sessions, and a final wrap-up and feedback session.

The main purpose of the agenda template is to ensure that all key components of STT2 are covered in a logical order. It also helps the host and methodology team plan the workshop realistically, including the time needed for plenary explanations, group work, coffee breaks, lunch, transition between sessions, and feedback collection. By providing a predefined structure, the agenda template supports consistency across case studies and helps ensure that comparable outputs are generated.

However, the agenda template has limitations. The proposed timing may not fit all contexts. Workshops with more participants, more groups, interpretation or translation requirements, complex scenarios, or highly detailed discussions may require more time. Conversely, smaller or more focused workshops may be implemented in a shorter format. In addition, unforeseen issues such as late arrivals, technical problems, or extended discussions may require adjustments during implementation.

The agenda template is therefore adaptable. Sessions can be shortened, extended, merged, or divided across more than one day. For example, in a highly complex case

study, the early response and early recovery phases may be implemented on separate days. In a more focused case, the number of critical infrastructures or decisions considered by each group may be reduced to keep the exercise manageable. The agenda can also be adapted for online or hybrid settings, although this requires additional attention to digital facilitation and data capture.

Facilitation guides

The facilitation guides provide detailed instructions for the facilitators who lead the group exercises. They explain the objectives of each session, the required materials, the expected output, and the steps to follow. The facilitation guides are particularly important because STT2 relies on structured participation. Stakeholders are encouraged to bring their own knowledge and experience, but this knowledge must be captured systematically to be useful for analysis.

The purpose of the facilitation guides is to support consistency between groups and case studies. They help facilitators ask comparable questions, guide participants through the same methodological logic, and ensure that outputs such as maps, matrices, and decision tables are completed. For example, during physical interdependency mapping, facilitators are guided to ask how the disruption of one infrastructure affects another to clarify the direction of dependency and support the group in assigning an impact level. During decision mapping, facilitators guide participants to identify priority decisions, information needs, resource requirements, coordination needs, and reasons for prioritization.

The main limitation of facilitation guides is that they cannot replace facilitator skills. The quality of a session depends on the facilitator's ability to manage group dynamics, encourage balanced participation, clarify misunderstandings, and keep the group focused. Some groups may require more support than others, especially when participants come from very different sectors or have different levels of familiarity with the scenario. There is also a risk that facilitators apply the guide too rigidly, limiting discussion and reducing the richness of stakeholder input.

The facilitation guides are adaptable to each workshop's specific context. They can be translated into the local language, shortened into quick-reference sheets, or expanded with case-specific examples. The examples used in the guide can be adapted to the dominant hazards, infrastructures, governance arrangements, or stakeholder groups in the case study. The facilitation logic can also be adapted for online implementation, although online facilitation requires clearer time management, digital templates, and more explicit instructions for participant interaction.

Note-taker and documentation templates

The note taker and documentation templates support the systematic recording of workshop discussions and outputs. They are used to capture key insights, disagreements, assumptions, justifications, critical nodes, cascading pathways, information needs, coordination requirements, and resource conflicts. These templates are essential because the maps and tables produced during the workshop do not always capture the reasoning behind participants' choices.

The purpose of these templates is to improve data quality. Note takers help bridge the gap between discussion and analysis by ensuring that outputs are complete, consistent, readable, and interpretable after the workshop. In the physical interdependency mapping sessions, note takers support the completion of the interdependency matrix and record key discussions about critical nodes and cascading effects. In the decision mapping sessions, they document decision logic, information needs, coordination challenges, resource conflicts, and changes in priorities.

The main limitation is that note-taking is selective. It is neither possible nor desirable to record every statement word-for-word. Important insights may be missed if the discussion moves quickly, if participants speak at the same time, or if note takers are not familiar with the methodology. This makes note taker preparation very important, establishing beforehand what kinds of information should or should not be noted down. The quality of documentation may also be affected by unclear handwriting, incomplete tables, inconsistent terminology, or ambiguous references to organizations and infrastructures.
The templates can be adapted to each case study's needs. They can be simplified for smaller workshops or expanded for more detailed analytical needs. They can be translated into the workshop language and used in either paper-based or digital formats. They can also be adjusted to align with the intended post-workshop analysis.

Practical implementation guides

The practical implementation guides support the host organization and the methodology team in preparing and running the workshop. These guides cover logistical and organizational aspects such as venue requirements, room setup, catering, participant invitations, registration, equipment, printed materials, maps, coordination tables, interdependency matrices, group composition, audio recorders for documenting group discussions where consent has been obtained, and follow-up activities.
The purpose of these guides is to make the methodology operationally feasible. STT2 workshops require careful preparation because the quality of the outputs depends not only on the methodology itself, but also on whether the right participants are present, whether the room setup supports group discussion, whether materials are available, and whether facilitators and note takers understand their roles. The implementation guide, therefore, translates the methodological design into practical requirements.
The limitations of practical implementation guides are linked to local constraints. The guide can specify ideal conditions, such as sufficient room separation, one facilitator and one note taker per group, printed maps, and complete workshop materials, but these conditions may not always be possible. Hosts may face constraints related to budget, staff availability, venue layout, participant availability, language, security restrictions, or access to sensitive information. These constraints can affect the depth and consistency of the workshop outputs.
The implementation guides provide the preferred structure for preparing and conducting the STT2 workshop and should be followed as closely as possible to ensure methodological consistency and data quality. At the same time, some practical adaptation may be necessary depending on the resources and constraints of each case-study host. For example, if separate rooms are not available, groups may be placed far apart in a larger room, with careful attention to noise levels and the ability of facilitators and note takers to capture the discussion. Any adaptation should therefore aim to remain as close as possible to the guidance while preserving the core logic of STT2 and the completeness of the expected outputs.

Analysis guidance

The analysis guidance supports the transformation of workshop outputs into structured results. The workshop produces several types of raw data, such as physical interdependency maps, interdependency matrices, and decision tables. The analysis guidance explains how these materials can be cleaned, interpreted, coded, and converted into analytical outputs.
The purpose of the analysis guidance is to ensure transparency and consistency in the post-workshop analysis. For physical interdependencies, the analysis includes cleaning

the matrices, standardizing infrastructure names, checking the direction and strength of dependencies, and identifying influential or vulnerable nodes. For decision and organizational interdependencies, the analysis includes extracting decisions, resources, information dependencies, coordination dependencies, involved actors, and potential bottlenecks. The analysis also compares the early response and early recovery phases to understand how dependencies and priorities change over time.

The main limitation of the analysis guidance is that workshop data can be incomplete, ambiguous, or inconsistent. Participants may use different names for the same actor or infrastructure, some dependencies may be discussed but not written down, and some relationships may require interpretation during data cleaning. The analysis is therefore not purely mechanical. It requires judgment from the analysts, especially when translating qualitative workshop material into structured matrices or networks. To address this, assumptions and uncertainties should be documented clearly.

The analysis guidance is adaptable to different levels of analytical depth. The level of analysis can therefore be adapted to the available data, the needs of the host, and the objectives of the case study.

Report templates

The report templates provide a common structure for documenting and communicating STT2 results. A report may include the case study context, workshop scope, participant profile, scenario description, methodology, session outputs, interdependency analysis, decision and coordination analysis, key bottlenecks, lessons learned, limitations, and recommendations. The report template helps ensure that findings from different case studies are presented in a consistent format.

The purpose of the report template is to support communication and comparability. It allows the methodology team to present results in a way that is understandable for case study hosts, project partners, stakeholders, and, where appropriate, wider policy or practitioner audiences. It also supports internal learning across the AGILE project by making it easier to compare findings between case studies, identify recurring bottlenecks, and assess how the methodology performs in different contexts.

The main limitation of report templates is that they may not capture all case-specific issues. Some findings may be sensitive, especially when they relate to infrastructure vulnerabilities, organizational weaknesses, or coordination gaps. In such cases, results may need to be anonymized, aggregated, or communicated only to specific audiences. Another limitation is that a standard report structure may not fully reflect the priorities of every host organization. Some hosts may be more interested in operational recommendations, while others may prioritize methodological learning or inputs for further modeling.

The report template is therefore adaptable to the audience and purpose. A technical report can provide detailed matrices, network results, and methodological reflections. A host-oriented report can focus on practical findings, bottlenecks, and recommendations. A project-level report can focus on cross-case comparison and methodological validation. A public-facing version can communicate high-level insights while excluding sensitive details. This adaptability is important for ensuring that STT2 outputs are useful while respecting confidentiality, ethics, and security considerations.

Overall role of the toolkit

Overall, the STT2 toolkit supports the implementation of a structured but flexible participatory stress-testing methodology. The tools help ensure that the process is replicable across case studies, while still allowing adaptation to local conditions. This

balance between consistency and flexibility is essential for the AGILE project because STT2 must be applicable across different geographical scales, governance settings, infrastructure systems, stakeholder groups, and HILP scenarios.
The tools should be understood as supporting instruments rather than standalone solutions. Their effectiveness depends on appropriate preparation, careful facilitation, meaningful stakeholder participation, good documentation, and transparent analysis. When these conditions are met, the toolkit enables the systematic capture of physical, informational, organizational, and decision-related interdependencies, thereby supporting the broader objective of STT2: to identify vulnerabilities, cascading pathways, coordination bottlenecks, and resilience-enhancing opportunities in complex socio-technical systems.

4.1.2.3. Resources and practical requirements

The implementation of STT2 requires a combination of local host resources and methodological support from the STT2 method developer. The host plays a central role in enabling the workshop in the local case study context, while the method developer ensures that the workshop is implemented consistently with the STT2 methodology and that the outputs can be used for subsequent analysis. This division of responsibilities is important because the methodology depends on both local contextual knowledge and specialized methodological knowledge related to participatory stress testing, interdependency mapping, and post-workshop analysis.
The practical requirements can be grouped into two main categories:

1) resources required from the host,
2) and resources or knowledge that need to be provided by the method developer.

The host is primarily responsible for the local organization of the workshop, including stakeholder mobilization, venue arrangements, local contextual input, practical materials, and providing facilitators and note takers. The method developer is primarily responsible for the methodological design, facilitation structure, analytical templates, facilitator briefing, quality assurance, and post-workshop data analysis.

Resources required from the host

The host organization provides the local foundation for the STT2 implementation. This includes the practical arrangements needed to run the workshop and the contextual knowledge needed to ensure that the stress test is relevant to the case study setting. The host is expected to support the workshop before, during, and after the event.

Stakeholder mobilization

The first requirement is **stakeholder mobilization**. The host is usually best placed to identify and invite relevant participants because it has knowledge of the local governance structure, infrastructure operators, emergency management actors, public authorities, service providers, and other organizations that should be represented. The quality of the STT2 outputs depends strongly on the diversity and relevance of participants. Therefore, the host should aim to involve stakeholders from different sectors, operational roles, and decision-making levels. The host may also support the preparation of invitation materials, registration information, reminders, and follow-up communication.
The host also provides **contextual knowledge**. This includes information on the local system, relevant critical infrastructures, governance arrangements, existing emergency management structures, previous incidents, local vulnerabilities, and ongoing resilience or risk management initiatives. This knowledge is needed to define the system boundaries – through STT0, when available – adapt the HILP scenario to the case study context, and ensure that the workshop discussions remain realistic and useful for participants. The host

may also provide or validate maps, infrastructure lists, local terminology, and relevant background information for participants.

Staff

A second requirement is finding the **staff** needed to support the workshop implementation, which is the host's responsibility. This includes staff for registration and welcoming participants, a host representative who can introduce the local case study context, and one facilitator and ideally one note taker for each working group. The facilitators and note takers are therefore provided by the host organization or arranged by the host with support from relevant local or consortium partners when needed.

Facilitators are responsible for guiding the group's work, keeping the discussion focused, ensuring that all participants have the opportunity to contribute, and helping the group complete the required outputs. Note takers support the facilitator by documenting key discussions, agreements, disagreements, assumptions, justifications, and insights that are not fully captured in the maps, matrices, or decision tables. They also help ensure that the outputs are complete, readable, and suitable for later analysis.

Although facilitators and note takers are provided by the host, a methodological briefing by the STT2 method developer before the workshop has proven to be highly valuable. This briefing helps address questions and concerns from the facilitation team, clarifies the purpose and expected outputs of each session, and ensures that facilitators and note takers understand how to apply the STT2 guidance consistently during the workshop. This is important because STT2 has a specific facilitation logic, including how to structure the problem-framing exercise, how to guide pairwise interdependency assessment, how to distinguish physical, information, coordination, and resource dependencies, and how to document decision-making processes in a consistent way.

In addition to the developer's briefing, the host should organize a mock workshop or dry run with all facilitators and note takers before the actual workshop. The purpose of the mock workshop is to ensure that facilitators and note takers understand their roles, are comfortable with the session flow, and are able to use the workshop materials correctly. During the mock workshop, the team should test the agenda, timing, room setup, maps, interdependency matrices, decision tables, sticky notes, recording procedures, and documentation templates. This rehearsal also allows the host and method developer to identify practical problems before the actual event, such as unclear instructions, missing materials, insufficient time allocation, translation issues, or uncertainty about how to record specific types of dependencies. The mock workshop is therefore a very important quality-control step that increases the likelihood of consistent facilitation, complete outputs, and usable data.

The host also needs to provide practical support during the event. This includes helping with room setup, participant registration, distribution of materials, time coordination, catering, and troubleshooting. In the context of AGILE methodology development, additional language support may be required when workshops are conducted in a local language. In such cases, the host should ensure that facilitators and note takers can communicate effectively with participants and that outputs can be translated into the English language after the workshop. This is also important for clarifying context-specific terminology, institutional roles, and infrastructure names. However, when STT2 is applied independently in a single local context, translation may not be necessary if the workshop and subsequent analysis are conducted in the same language.

Venue and room setup

A third requirement is the **venue and room setup**. The workshop requires a plenary space for introductions, explanations, and wrap-up, as well as sufficient space for group work.

Ideally, groups should work in separate rooms or in a large room with enough distance between groups to minimize noise. This is especially important when discussions are recorded or when several groups are working simultaneously. Each group requires enough table and wall space to work with printed maps, A0 sheets, sticky notes, matrices, and decision tables. The room should also have suitable lighting, temperature, ventilation, seating, and projection equipment.
The host is also expected to provide or arrange the necessary materials and equipment. These typically include name tags, sign-in sheets, printed agendas, pre-workshop briefing materials, printed maps, printed critical infrastructure lists, interdependency matrices, decision tables, pens, markers, sticky notes, dot stickers, flip charts, and pin boards or walls. Depending on the workshop format, laptops, tablets, audio-recording devices, projectors, screens, adapters, extension cables, and internet access may also be required. Catering, including coffee breaks, water, and lunch where relevant, should also be arranged by the host.

Post-workshop data handling

A fourth requirement is **post-workshop data handling**. After the workshop, the facilitators and note takers are responsible for digitizing the collected data from their groups. This includes transferring the content of physical maps, interdependency matrices, decision tables, sticky notes, handwritten notes, and additional observation sheets into the agreed digital templates. Where the workshop is conducted in a local language, in the context of the AGILE project, facilitators and note takers are also responsible for translating the collected data into English. This step is essential because the method developer can only conduct consistent cross-group and cross-case analysis if the data are readable, complete, digitized, and available in a common language.
The digitization and translation process should preserve the original meaning of participant inputs. Facilitators and note takers should avoid over-interpreting, simplifying, or adding unsupported information. If a statement, dependency, actor, infrastructure name, or decision is unclear, this should be flagged as an uncertainty rather than silently corrected. Original photographs of maps, matrices, decision tables, and sticky-note outputs should also be stored and shared with the method developer, so that the digitized data can be checked against the original workshop outputs if needed.
In addition to logistical resources, the host should provide several competencies. These include knowledge of the local institutional setting, the ability to identify relevant stakeholders, familiarity with the critical infrastructure and emergency management context, and the capacity to coordinate practical arrangements. The host should also be able to judge whether certain information is sensitive and whether workshop outputs need to be restricted or communicated only to selected audiences. This is particularly important because the workshop may reveal vulnerabilities, dependencies, or coordination bottlenecks that should not be shared publicly without review.

Resources and knowledge required from the method developer

While the host provides the local, logistical, facilitation, documentation, digitization, and translation resources, several aspects of STT2 cannot be easily implemented by the host alone without significant support or preparation. These aspects relate mainly to the methodological design, facilitation logic, analytical consistency, and transformation of workshop outputs into stress-testing results.

Methodological framework

The method developer provides the overall **methodological framework** for the STT2 implementation. This includes defining the workshop logic, explaining the role of STT2

within the AGILE tiered stress-testing framework, and ensuring that the workshop captures the intended dimensions of the methodology: physical interdependencies, decision-making dependencies, information flows, coordination needs, and temporal differences between early response and early recovery. Without this methodological guidance, the workshop could become a general discussion or training exercise rather than a structured stress-testing process.

Local adaptation of the workshop

The method developer also supports the **adaptation of the workshop** to the case study context. This includes working with the host to define the system boundaries based on STT0, refining the HILP scenario, identifying relevant time phases, selecting suitable workshop materials, and adjusting the agenda to the available time and participants. The method developer ensures that adaptations do not compromise the core logic of the methodology. For example, the number of infrastructures or decisions may be adjusted, but the structured mapping of dependencies and the distinction between response and recovery phases should be preserved.

Workshop templates and guidance

A key responsibility of the method developer is the **provision of workshop templates and guidance documents**. These include agenda templates, facilitation guides, practical implementation guidance, note taker templates, physical interdependency matrices, decision tables, infrastructure lists, analysis templates, and reporting templates. These tools are necessary to ensure that each group produces outputs in a format that can be compared, cleaned, and analyzed after the workshop. The host may be able to provide local content, but the design of these templates requires methodological knowledge of how the outputs will be used analytically.

The method developer does not necessarily provide the facilitators and note takers, as these roles are primarily assigned by the host. However, the method developer provides the guidance and preparation needed for these roles to be performed consistently with the STT2 methodology. This includes facilitator guides, note taker instructions, session templates, briefing materials, and pre-workshop training or briefing sessions.

Support for the mock workshop or dry run

The method developer also **supports the mock workshop or dry run**. During this rehearsal, the method developer helps the host test whether facilitators and note takers understand the methodology, whether the timing is realistic, whether the materials are complete, and whether the instructions are clear. The method developer can use the mock workshop to clarify methodological questions, correct misunderstandings, and identify risks to data quality before the actual workshop takes place. This support is especially important when facilitators and note takers are not familiar with stress testing, interdependency mapping, or the AGILE methodology.

The method developer explains the objective of each session, the expected outputs, the meaning of dependency direction and impact levels, the distinction between different types of dependencies, and the minimum data quality requirements for later analysis. This support is essential because facilitators and note takers may be familiar with the local context but not with the methodological logic of STT2. Without this preparation, the workshop may still generate useful discussion, but the outputs may be incomplete, inconsistent, or difficult to analyze across groups and case studies.

Methodological supervision

During the workshop, the method developer may also provide **methodological supervision**. This means supporting facilitators when questions arise, checking whether

maps, matrices, and decision tables are being completed correctly, and helping resolve uncertainties about how to record specific dependencies or decisions. In this way, the host provides human resources for facilitation and documentation, while the method developer ensures that these resources are prepared and supported to implement the methodology correctly.

Quality assurance

Another important responsibility of the method developer is **quality assurance**. The method developer checks whether the required outputs are complete, whether the maps and matrices are consistent, whether the decision tables contain enough information for analysis, and whether the notes are sufficiently clear. This quality control should take place during the workshop, where possible, because missing or unclear information is much harder to reconstruct afterward. For example, if an arrow is drawn on a map but not recorded in the matrix, or if a decision table lists an information need but not its source, the method developer or trained facilitators should identify and resolve the issue before the session ends.

Post-workshop guidance

The method developer also provides **guidance for post-workshop digitization and translation**. Although facilitators and note takers are responsible for digitizing and translating the collected data, the method developer should provide clear templates, naming conventions, coding instructions, and data-quality requirements. This ensures that the digitized material can be integrated into the analysis without losing the meaning of the original workshop discussions. The method developer may also review the digitized outputs, compare them with photographs of the original materials, and ask the host for clarification where necessary.

Post-workshop analysis

The method developer is responsible for the post-workshop analysis. This includes collecting and organizing photographs, maps, matrices, decision tables, notes, translated digital templates, and feedback forms; cleaning and standardizing the data; resolving terminology inconsistencies; coding physical, informational, and coordination dependencies; and transforming workshop outputs into analytical formats. Depending on the level of analysis, this may include creating interdependency matrices, network representations, decision-dependency tables, bottleneck summaries, and cross-phase comparisons between early response and early recovery.

This analytical work requires methodological expertise because workshop outputs are often incomplete, ambiguous, or expressed in context-specific language. The analyst should maintain analytical integrity and data usability when processing the workshop data. Dependency relationships need to be established based on documented evidence or expert-validated assumptions. Analysts may need to distinguish between a physical dependency, an information dependency, a coordination dependency, and a resource conflict. They may also need to document assumptions, flag uncertainties, and decide how to represent reciprocal dependencies or multi-actor relationships. These tasks require familiarity with the STT2 analytical framework and cannot be easily delegated to the host without prior training.

Reporting

Finally, the method developer supports **reporting and methodological learning**. This includes preparing case study outputs, summarizing key findings, identifying critical infrastructures and coordination bottlenecks, comparing outputs across groups and phases, and reflecting on the strengths and limitations of the implementation. The method

developer also uses feedback from participants, hosts, facilitators, and note takers to refine the STT2 methodology and toolkit for future applications.

4.1.3. Post-workshop analysis process and output

4.1.3.1. Physical interdependency ST analysis and Output

This part introduces the physical interdependency stress-testing analysis used to transform the workshop data into interpretable outputs. As shown in Figure 2, the process begins with the data collected from each workshop group, which is used to develop a physical interdependency network. Stress is then introduced into selected parts of the network, and the resulting propagation of disruption is calculated to examine how failures may cascade across interconnected critical infrastructure subsectors. The purpose of this analysis is to explore how disruptions in specific components or subsectors can affect the wider system, including both direct and indirect impacts. By simulating different stress conditions, the analysis helps identify systemic vulnerabilities, influential subsectors, critical components, and potential drivers of network-wide service loss. The outputs support a more detailed interpretation of how failures in each subsector of interest may affect other subsectors, providing useful insights for infrastructure operators, emergency managers, and policy-makers in prioritizing resilience measures and targeted interventions. The following subsections describe each step of this analytical process in more detail.

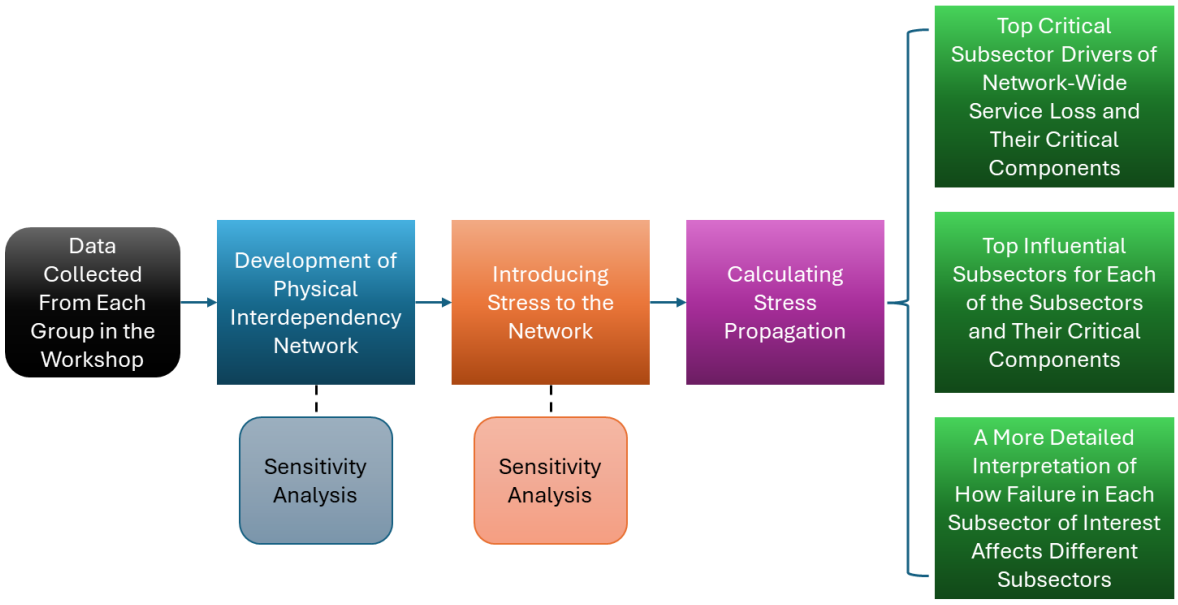

Figure 8: Post-workshop physical interdependency ST analysis process and output

Input data used in the analysis

The simulation utilizes two primary datasets provided:

- **The CIs list:** The CI component list functions (e.g., Figure 4) as a reference dictionary for interpreting the codes used in the interdependency matrices. It maps each specific component code to its corresponding infrastructure subsector. For example, a component such as E1, representing electricity power plants, is linked to the broader Electricity subsector. This mapping enables the simulation results to be interpreted not only at the component level, but also at the subsector level.
- **The physical interdependency matrices:** In the sessions dedicated to physical interdependency mapping, each group developed a CI network map and a corresponding interdependency matrix. The matrix (e.g., Figure 6) provides an analytical representation of the interdependencies identified in the map, allowing the analyst to further process and analyze in the simulation. Each matrix was produced by a different group. The rows and columns represent specific CI

component codes, such as E1 or T3, while the matrix cells indicate the assessed level of dependency between components. These dependency levels are expressed using categorical ratings: H for high dependency, M for medium dependency, L for low dependency, and empty or zero values where no relevant dependency was identified.

Data analysis steps

Development of physical interdependency network

Data aggregation from the component to the subsector level

The level of analysis in STT2 is defined during the preparation and scoping phase by the methodology developer together with the case-study host. This decision depends on the system boundaries, geographical scale, available data, workshop objectives, and the level at which participants can meaningfully discuss dependencies. Since STT2 can be applied to different types of systems, ranging from neighborhood level systems to urban, regional, and national systems, the appropriate level of analysis cannot be fixed. In some cases, component analysis may be possible and useful, while in other cases it is impractical because the system is too large or complex.

In the STT2 implementations conducted so far, participants have been asked to identify infrastructures and dependencies at a relatively detailed level (component level), where this was meaningful for the case-study context. This was a deliberate methodological choice. Asking participants to start from concrete examples helps move the discussion beyond generic statements such as “energy depends on transport” and encourages them to reflect on how dependencies materialize in practice, which elements are involved, and where bottlenecks may emerge. Although this level of detail can be challenging for participants, it supports richer data collection and makes assumptions, uncertainties, and differences in perspective more explicit.

For the purpose of STT2 analysis, however, these detailed inputs are often aggregated to a higher analytical level, such as the subsector or sector level. This aggregation is particularly useful when the objective is to evaluate system-level effects of HILP events rather than to reproduce detailed technical system behavior. For example, components such as electricity grids, power plants, or substations may be grouped under the electricity subsector, while major roads, bridges, or tunnels may be grouped under road transport. This allows the analysis to preserve the richness of stakeholder input while producing outputs that are feasible, interpretable, and useful for identifying broader patterns of dependency, vulnerability, and cascading effects.

A subsector- or sector-level representation supports high-level policy-making, cross-sectoral comparison, strategic resource allocation, and crisis-management planning. It helps reveal whether one subsector acts as a bottleneck for another and supports communication between departments and organizations that need to understand the main dependencies between infrastructure domains. Aggregation should therefore be understood as a methodological choice aligned with the scale and purpose of the stress test, rather than as a loss of detail.

After the workshop, the analysis can aggregate these detailed inputs to a higher level when this is more appropriate for system-level interpretation. For example, components (e.g., electricity grids, oil pipelines, major roads) identified by participants can be grouped into subsectors (e.g., electricity, oil, road) or even sectors (e.g., energy, transport) to reveal broader patterns of dependency, vulnerability, and cascading effects. This approach combines the advantages of detailed stakeholder elicitation with the clarity of higher-level analysis.

To conduct the analysis, the component-level data is first transformed into subsector-level data. Before aggregation, the categorical dependency ratings are converted into numerical values as follows:

- No dependency = 0,
- Low = 1,
- Medium = 2,
- and High = 3.

The rows and columns of each physical interdependency matrix are then mapped from specific component codes, such as E1: electricity power plants or E2: electricity transmission grids, to their corresponding subsectors, such as Electricity (see Figure 4).
This mapping creates duplicate subsector names in the matrix. For example, multiple electricity-related components may all be mapped to the same Electricity subsector. To address this, the values are grouped by subsector and aggregated using the arithmetic mean. In other words, all component-level dependencies between two subsectors are averaged to produce a **single subsector-to-subsector dependency score**.
For example, if components within the Drinking Water subsector depend on two different Electricity components, with one dependency rated as High and another as Low, the corresponding numerical values are 3 and 1. The aggregated subsector-level dependency score is therefore (3+1)/2=2. This results in an overall Medium dependency between the Drinking Water and Electricity subsectors. The arithmetic mean is used because it provides a balanced representation of the typical dependency intensity between two subsectors within a single group's matrix. It reflects the general degree of interconnection without overemphasizing isolated extreme values. In contrast, using the maximum value at this stage would mean that one highly dependent component could make an entire subsector appear strongly dependent, even if most components within that subsector have weak or no dependency. The mean, therefore, provides a more proportionate and interpretable basis for subsector-level stress testing.

Cross-group data aggregation

After a subsector-level matrix has been calculated for each individual group, the group-level matrices are combined into a **single master subsector interdependency matrix**. This step integrates the physical interdependencies identified by different groups into one consolidated representation of the system.
To do this, the subsector matrices generated by the different groups are layered and compared link by link. For each subsector-to-subsector relationship, the highest dependency score reported by any group is selected. The resulting value is then rounded to the nearest whole number, where necessary, and converted back into the categorical dependency scale: 0 = no dependency, 1 = Low, 2 = Medium, and 3 = High.
For example, if Group 1 assesses the dependency from Electricity to Water as 2, corresponding to a Medium dependency, while Group 2 assesses the same dependency as 3, corresponding to a High dependency, the master matrix assigns a value of 3, or High, to this link.
The maximum value is used at this stage to reflect a conservative and risk-averse approach to stress testing. Workshop-based assessments are based on participant knowledge, perception, and expertise, and different groups may identify different aspects of the system. As a result, it is natural that groups will identify and value dependencies differently; some dependencies may be underestimated, overlooked, or not discussed in each of the groups. If one group identifies a severe dependency while another group assigns a lower value or does not identify the link, averaging across groups could dilute a potentially critical vulnerability.
Using the maximum, therefore, ensures that any dependency identified as critical by at least one group is retained in the final system representation. This approach is appropriate for stress testing because the aim is not to estimate an average perception of dependency,

but to identify and preserve potentially important pathways through which disruptions may cascade across subsectors.

Missing links treatment after aggregating group networks

In the workshop-based critical CI analysis, interdependency networks are developed through participant elicitation. Each group identifies and assesses a limited number of CI components (5-9). This limitation is necessary for both methodological and practical reasons. First, participants are asked to prioritize the infrastructures they consider most important from the perspective of their organizations, roles, responsibilities, and service dependencies. This reflects the reality of extreme events, where decision-makers and policymakers often need to prioritize under time pressure and resource constraints and cannot focus on all infrastructures simultaneously. Second, limiting the number of selected components reduces participants' cognitive burden and allows sufficient time for discussion of the strength and importance of the dependencies among the selected components, which will lead to a better quality and more in-depth group assessment.

As a result, the networks developed by different groups may vary in two main ways. First, groups may select different CI components based on their organizational perspectives, priorities, responsibilities, and perceived stakes. Second, even when similar components are selected, the assessed strength of dependencies may differ across groups, as discussed above. When groups select different sets of CI components, aggregating their interdependency matrices into a single master network may generate potential links for which no direct participant input is available. In this context, a missing link refers to a possible dependency relationship for which no rating was provided by participants during the workshop. These missing links may correspond to dependencies that were not discussed, were overlooked, or were excluded because of the limited number of components selected by each group. Nevertheless, some of these links may still be relevant for understanding cascading effects across the wider CI system.

Aggregating the data at the subsector level helps to reduce this problem to some extent. By moving from component-level analysis to a higher subsector-level representation, the density of the data increases, and some missing component-level relationships are indirectly captured through broader subsector-to-subsector relationships. In this sense, subsector-level aggregation acts as a form of data-density improvement. However, it does not fully eliminate the issue of missing links. In other words, even after the aggregation, there might still be missing links.

To reduce this limitation, a sensitivity analysis is conducted by assigning alternative dependency ratings to the missing links and examining how these assumptions influence the network structure and the resulting cascading effects. This approach allows the robustness of the network results to be explored under different assumptions. After the master network is generated as a result of aggregation, the missing links are identified, and alternative scenarios are created by assigning different dependency ratings to these links. Specifically, the missing links are tested under four assumptions: 0 = no dependency, L = low dependency, M = medium dependency, and H = high dependency. This makes it possible to examine how sensitive the stress testing results are to different interpretations of the missing relationships.

This step is important because treating all missing links as zero may artificially fragment the network. In network-based cascading failure analysis, a missing link that is assumed to represent no dependency can act as an artificial break in the cascade pathway. As a result, the simulated disruption may stop at that point, even though a real-world dependency could exist. This could give decision-makers a false sense of security regarding some relationships between CIs by underestimating the potential for cascading effects.

By explicitly testing different assumptions for missing links, the method follows precautionary logic. In the context of disaster risk management and stress testing, it is

often more useful to examine the consequences of possible hidden dependencies than to assume that unobserved dependencies do not exist. This does not mean that all missing links are treated as real or strong dependencies in the final interpretation. Rather, the sensitivity analysis helps assess whether the overall system's behavior changes significantly when these uncertain links are assigned different levels of dependency. If the results remain robust across scenarios, confidence in the network findings increases. If the results change substantially, the missing links should be treated as important sources of uncertainty and may require further expert validation.

Introducing stress to the network

To simulate a stress profile, an initial stress is introduced into the network by defining which subsector or subsectors experience the initial damage, such as Electricity, Water, or Transport. This initial stress can be introduced in different ways depending on the purpose of the analysis.

One option is to apply **randomized failures**, where the initially affected subsectors are selected randomly. This allows the system to be tested across a broad range of possible disruption conditions without assuming a specific hazard scenario in advance. Another option is to apply **targeted stress scenarios**, where specific subsectors are selected based on user interest, expert judgment, known vulnerabilities, or scenario-specific concerns. A further option is to test the failure of each node one by one, allowing the role and criticality of every subsector in the network to be systematically examined.

For this stress testing method, the preferable option is combining randomized failures with targeted stress scenarios, or alternatively, testing all subsectors individually together with targeted stress scenarios. These approaches support a threat-agnostic analysis by exploring how the system behaves under a wide range of stress conditions, while still allowing users to investigate specific scenarios of interest. As a result, the method can identify both general systemic vulnerabilities and scenario-specific cascading effects. To introduce stress into the network, the user can define a base functionality percentage for the selected node, namely the subsector or subsectors to which the initial stress is applied. This value represents the remaining level of service of the affected subsector after the stress has been introduced.

The intensity of the stress can be adjusted depending on the goal of the analysis. In some cases, the stress may represent the complete failure of one or more subsectors to provide their services, corresponding to 0% functionality. In other cases, it may represent a partial reduction in service level, ranging from minor degradation to a considerable loss of functionality. This flexibility allows the method to capture different degrees of disruption and assess how varying levels of service reduction may propagate through the CI network. For example, if the Electricity subsector is set to operate at 30% functionality, this represents a 70% reduction in service level. Such partial service loss can represent situations such as partial blackouts, reduced output flows, degraded asset operations, or limited service availability. This allows the simulation to move beyond a simple binary distinction between "functioning" and "failed" and instead capture more realistic degradation patterns. This is particularly relevant for stress testing, as many real-world disruptions do not immediately lead to complete infrastructure collapse. Instead, infrastructures may continue to operate at reduced capacity.

To account for uncertainty in the initial stress level, the user can define sensitivity bounds around the base functionality percentage, for example, ±10%. Based on this input, the simulation evaluates three alternative stress conditions:

1) a lower-bound case,
2) the base target case,
3) and an upper-bound case.

For example, if the selected subsector is assumed to operate at 30% functionality with a sensitivity bound of ±10%, the simulation can assess the cascading effects under 20%,

30%, and 40% functionality. This allows the analysis to explore how changes in the initial level of service influence the wider network response.
This step is important because decision-making under disaster conditions involves considerable uncertainty. A single deterministic estimate of functionality may be misleading, especially when the actual level of service degradation is difficult to measure or predict. Sensitivity analysis helps assess whether the system is mostly robust to minor changes in stress levels and degrades gradually as stress increases, or whether cascading effects intensify abruptly once a critical threshold is reached.
In this way, the sensitivity bounds help evaluate how responsive the network is to changes in the level of service of the initially stressed subsector. If small changes in the initial level of service lead to large differences in cascading effects, the system can be considered highly sensitive to that subsector. This provides useful insight into potential tipping points, hidden vulnerabilities, and priority areas for resilience planning.

Calculating stress propagation

Formulas for calculating the CIs' level of services

A fundamental assumption in CI cascade analysis is that a reduction in the level of service of a dependency-providing subsector can constrain the level of service of other subsectors that depend on the services it provides. However, there is uncertainty regarding how this service loss should be propagated through the network, particularly when the data are collected through a participatory process, and the analysis is conducted at a higher level with less focus on details.
To address this uncertainty and avoid relying on a single deterministic propagation rule, the stress testing method applies **three alternative mathematical formulas**. They are methodological assumptions introduced by the developers to explore different plausible ways in which service losses may propagate through the interdependency network. Each formula represents a different interpretation of how losses in dependency-providing subsectors may affect dependent subsectors. This allows the analysis to compare different cascade behaviors and provides a more comprehensive view of potential systemic vulnerabilities.
The example below is theoretical and illustrative. It is not based on the results of any specific STT2 case study or workshop. It is included only to demonstrate how the proposed formulas operate.
Example Scenario:
Assume that a Hospital depends on five upstream infrastructures:

Table 3: Example hospital dependency inputs for physical interdependency stress test analysis.

Dependency provider infrastructure	Dependency level	Weight	Functionality	Loss of Service
Electricity	High	0.5	40%	60%
Water	High	0.5	80%	20%
Internet	Medium	0.3	50%	50%
Gas	Medium	0.3	70%	30%
Roads	Low	0.2	100%	0%

The dependency levels in the table represent the qualitative dependency ratings identified through the participatory mapping process. In the workshop, participants assess whether the dependency of one infrastructure on another is low, medium, or high. For the purpose of the post-workshop cascade analysis, these qualitative levels need to be translated into numerical values. The weights shown in the example are therefore developer-defined analytical coefficients used to operationalize the qualitative dependency levels in the service-loss propagation formulas.

In this illustrative example, the weight values low = 0.2, medium = 0.3, and high = 0.5 are used to preserve the ordinal relationship between dependency levels and to give stronger influence to dependencies assessed as high. These values should not be interpreted as empirically measured parameters or universal constants. Rather, they are modeling assumptions introduced to enable exploratory stress testing of the mapped interdependency network. Where more detailed technical data or validated expert estimates are available, these weights can be adjusted or calibrated.
In this example, the Hospital depends strongly on Electricity and Water, moderately on the Internet and Gas, and weakly on Roads. The loss of functionality of each upstream infrastructure is calculated as the difference between full functionality and its remaining functionality.

For example, if Electricity operates at 40% functionality, its loss of functionality is:

$$\mathbf{1 - 0.40 = 0.60}$$

or 60%.

The MAX Formula: Largest Service Losses Determine the Impact

The MAX formula provides a bottleneck-oriented estimate of cascading impacts. It assumes that, within each dependency category (High, Medium, or Low), service-loss impacts largely overlap rather than accumulate. As a result, the dependent subsector is constrained by the most severe service loss in that category, which is treated as the dominant bottleneck.
For example, if both Electricity and Water are classified as high dependencies for a Hospital, but Electricity experiences a much larger service reduction, the Electricity loss determines the impact for the high-dependency category. The smaller Water-related loss is not treated as an additional independent penalty.
Therefore, the MAX formula can be considered bottleneck-sensitive but not very conservative. It is useful for identifying dominant dependency bottlenecks, but it may underestimate cascading impacts when multiple service losses affect different aspects of the dependent subsector's operation.
This formula is useful for identifying strict bottlenecks in the network. Subsectors that are highly impacted under the MAX formula can be interpreted as particularly fragile because their functionality is strongly constrained by the weakest upstream dependency.

$$F_i = 1 - \sum_{k \in K} w_k \cdot max_{j \in S_{ik}}(L_j)$$

Where:

- F_i: The final Level of Service of subsector i after accounting for cascading failures from its upstream dependencies.
- $k \in K$: The dependency category K = (High, Medium, Low)
- w_k: The weight assigned to that category.
- j: An individual upstream dependency
- S_{ik}: The set of all upstream dependencies to the subsector i within the dependency category k
- L_j: The loss of functionality of the upstream dependency j

Example Calculation:

High Category (w_H = 0.5):
Compares Electricity (60% loss) and Water (20% loss).
The algorithm takes the MAX loss: max (60%, 20%) = 0.60

Medium Category (w_M= 0.3):
Compares Internet (50% loss) and Gas (30% loss).

The algorithm takes the MAX loss: max (50%, 30%) = 0.50

Low Category (w_L= 0.2):
Road is completely functional (0% loss)
The algorithm takes the MAX loss: max (0%) = 0.00

Result:
The Hospital suffers a (0.5 × 0.60) + (0.3 × 0.50) + (0.2 × 0.0) = 0.45 total penalty.
Final Level of Service of Hospital: 100% − 45% = 55%

The weighted SUM Formula: Combined Service Losses Determine the Impact

The SUM formula provides a conservative estimate of cascading impacts. It assumes that service losses from all supporting subsectors accumulate and jointly reduce the level of service of the dependent subsector. In other words, each service loss is treated as an additional source of pressure on the dependent subsector, rather than being absorbed within a single dominant bottleneck.
For example, if a Hospital experiences a partial loss of both Electricity and Water, the SUM formula assumes that both losses contribute to the overall reduction in the Hospital's Level of Service (LoS). This reflects a cumulative degradation logic, where multiple partial disruptions can collectively produce a severe impact, even if no single supporting subsector fails completely.
This makes the SUM formula particularly useful for identifying subsectors that are sensitive to compound stress conditions. It can reveal vulnerabilities that may not be visible when only the largest individual dependency loss is considered. Therefore, the SUM formula is especially relevant for assessing subsectors that rely on several supporting services simultaneously and may require broader redundancy across multiple dependencies.
Because it assumes that service losses accumulate, the weighted SUM formula can be considered a highly conservative propagation assumption. It is appropriate for stress testing because it helps explore severe but plausible cascading effects under conditions of uncertainty.

$$F_i = 1 - \sum_{k \in K} w_k \cdot \sum_{j \in S_{ik}} (L_j)$$

Where:
F_i: The final Level of Service of subsector i after accounting for cascading failures from its upstream dependencies.
$k \in K$: The dependency category K = {High, Medium, Low}
w_k: The weight assigned to that category.
j: An individual upstream dependency
S_{ik}: The set of all upstream dependencies to the subsector i within the dependency category k
L_j: The loss of functionality of the upstream dependency j

Note: Because the weighted sum can theoretically produce a total loss greater than 100% when several upstream subsectors experience severe losses, the calculated loss of service (LoS) is capped at 100%. In other words, once the dependent subsector reaches complete loss of service, the value remains at 100% and does not increase further.

Example Calculation:
High Category (w_H= 0.5):
Sums Electricity (60% loss) and Water (20% loss).
Total category loss: 60% + 20% = 80%.

Medium Category (w_M= 0.3):

Sums Internet (50% loss) and Gas (30% loss).

Total category loss: 50% + 30% = 80%.

Low Category (w_L= 0.2):

The road is fully functional (0% loss).

Total category loss: 0%

Result:

The Hospital suffers a (0.5 × 0.80) + (0.3 × 0.80) + (0.2 × 0.0) = 0.40 + 0.24 = 0.64 total penalty.

Final level of service of Hospital: 100% - 64% = 36%

Therefore, even if no single dependency fails completely, the "death by a thousand cuts" from multiple shortages can grind the facility to a halt much faster (dropping it to 36% instead of MAX's 55% Los).

The "G*P" Formula: Critical Services Drive Impact, Non-Critical Services Adjust Performance

The G*P formula provides a hybrid estimate of cascading impacts. It sits between the two more extreme assumptions represented by the MAX and SUM formulas. While the MAX formula assumes that service losses within a dependency category largely overlap, and the SUM formula assumes that all service losses accumulate, the G*P formula introduces a more differentiated logic.

The formula distinguishes between critical dependencies and non-critical dependencies. Critical dependencies are those without which the dependent subsector may experience a substantial reduction in its ability to provide basic services. Non-critical dependencies, by contrast, do not necessarily determine whether the subsector can continue operating, but they influence the efficiency, quality, or capacity of service provision.

This distinction makes the G*P formula closer to real-world infrastructure behavior. In practice, the loss of some services may severely constrain basic operations, while the loss of others may create inefficiencies, delays, or reduced performance without causing immediate collapse. The proposed G*P formula below reflects this behavior.

$$F_i = F_{\{min,i\}} + \left(1 - F_{\{min,i\}}\right) * G * P$$

$$G = \prod_{j \in C(i)} F_j^{w_j}$$

$$P = \prod_{\{k \in N(i)\}} (\epsilon_k + (1 - \epsilon_k) F_k)^{v_k}$$

Where:

F_i: Level of service of the subsector i

$F_{min,i}$: The baseline level of service of subsector i that can be maintained even under severe disruption

j: Index of critical dependencies

k: Index of non-critical dependencies

$C(i)$: Set of critical dependencies of the subsector i

$N(i)$: Set of non-critical dependencies of the subsector i

F_j: Level of service of critical dependency j

F_k: Level of service of non-critical dependency k

w_j: Weight associated with critical dependency j

v_k: Weight associated with non-critical dependency k

ϵ_k: Elasticity parameter for non-critical dependency k, with $0 < \epsilon_k < 1$

The G*P topology splits the network edges. It distinguishes between "Critical dependencies" (without which the downstream system can be heavily impacted) and "Non-Critical dependencies" (without which efficiency suffers, but a minimum operational threshold is maintained).
At its core, the formula decomposes system functionality into two multiplicative components:

- A **geometric component (G)** representing *critical dependencies without which the critical service cannot function.*
- A **penalty/performance component (P)** representing *non-critical dependencies. Losing these services does not lead to a complete shutdown of the critical service, but reduces its functionality.*

The geometric component (G) aggregates critical inputs using a multiplicative (geometric) structure, meaning that the system is highly sensitive to failures in **essential and critical services**. If any critical dependency experiences severe degradation or reaches zero, the entire system's functionality is strongly reduced or even collapses. This reflects real-world conditions where certain infrastructures are indispensable for most basic operations. However, this strict multiplicative behavior can lead to an overly rigid representation of system collapse, as many real-world systems retain a limited level of functionality even under extreme disruption (e.g., emergency protocols, backup generators, or manual operations). For this reason, the formulation introduces a minimum functionality level, $F_{min,i}$, which represents the baseline operational capacity that can be sustained independently of external dependencies. This term ensures that the model captures not only the vulnerability of systems to critical failures but also their inherent ability to maintain a minimal level of service, thereby preventing unrealistic total collapse in situations where residual functionality is expected.
In contrast, the penalty component (P) introduces an elastic degradation mechanism for **non-critical dependencies**. These inputs do not directly determine system survival but instead affect operational efficiency. The inclusion of an **elasticity floor (ϵ)** ensures that even if non-critical services fail completely, the level of service of the subsector does not drastically reduce, but instead operates at a reduced performance level. This reflects the reality that systems can often adapt to the loss of secondary services, albeit with reduced effectiveness. To set the value of this parameter, ask this question: If this supplier is completely lost, what percentage of the level of service (from 0 to 1) can the receiver maintain on its own? Another possibility is that, to reduce complexity and establish a consistent baseline for system-wide behavior, a uniform elasticity parameter (e.g., ε = 0.2) can be assumed for all non-critical dependencies, representing a conservative estimate of inherent resilience across the network.
This structure ensures that failures for critical dependencies dominate system viability (through G) and non-critical failures influence efficiency but not total collapse (through P). For example, if the electricity supply to a hospital is cut, the hospital cannot perform critical operations. This severe loss is captured through G, which drives the overall functionality down substantially. If internet connectivity is lost, the hospital can still treat patients, but administrative systems, remote diagnostics, and coordination suffer. This reduced performance is captured through P, which applies a moderate penalty on top. Because G and P are multiplied together, a catastrophic critical failure dominates the outcome, while non-critical losses adjust the result without causing collapse on their own. As a result, the G*P formulation provides a balanced representation between the extremes of the MAX and SUM approaches. It avoids the overly optimistic assumption of single bottlenecks (MAX) and the overly pessimistic accumulation of all losses (SUM), instead offering a **layered understanding of dependency importance**.

$$F_i = F_{\{min,i\}} + \left(1 - F_{\{min,i\}}\right) \cdot \left(\prod_{j \in C(i)} F_j^{w_j} \right) \cdot \left(\prod_{\{k \in N(i)\}} (\epsilon_k + (1 - \epsilon_k) F_k)^{v_k} \right)$$

Example Calculation:
(Assuming $\mathrm{F_{min}} = 0$, and an elasticity floor $\epsilon = 0.2$)
We decided on the critical and non-critical services based on the input from the workshop. If the dependency rating is H, the service is critical; if it is M or L, it is non-critical.

Weight Normalization in G*P: Unlike MAX and SUM, the geometric products require the exponents (weights) of each group (Critical and Non-Critical) to sum perfectly to `1.0`. Otherwise, the penalty would artificially inflate.
Critical Group (G) Normalization: Electricity (H=0.5) + Water (H=0.5) = 1.0. Since they already sum to 1.0, the normalized formula weights (w_j) remain **0.5** for both.

Non-Critical Group (P) Normalization: Internet (M=0.3) + Gas (M=0.3) + Roads (L=0.2) = 0.8. We divide each by the total (0.8) to get normalized weights (v_k): Internet (0.3/0.8 = 0.375), Gas (0.3/0.8 = 0.375), Roads (0.2/0.8 = 0.25).

Step A: Calculate G, meaning the influence of critical services on the hospital
Critical services: Electricity and Water are critical services because the rating for these dependencies in the workshop is “H”, which means high dependency.

Normalized weights for G:
w_{Elc}= 0.5, w_{Elc}= 0.5

Calculation:
Electricity (60% loss), Water (20% loss)

$$G = (0.4^{0.5}) \times (0.8^{0.5}) = \sqrt{0.32} \approx 0.566$$

Step B: Calculate P, meaning the influence of non-critical services on the hospital
Non-Critical services: Internet, Gas, and Roads, because the rating for dependency on Internet and Gas was M, and for Roads was L.

Normalized weights for P:
$v_{Int} = 0.375$, $v_{Gas} = 0.375$, $v_{Roads} = 0.25$

Calculation:
Internet loss (50%), Gas loss (30%), Road loss (0%)

- Internet penalty factor:

$$0.2 + (0.8 \times 0.5) = 0.60$$

- Gas penalty factor:

$$0.2 + (0.8 \times 0.7) = 0.76$$

- Road penalty factor:

$$0.2 + (0.8 \times 1.0) = 1.00$$

- Calculation of P:

$$P = (0.6^{0.375}) \times (0.76^{0.375}) \times (1.0^{0.25}) \approx 0.825 \times 0.902 \times 1.0 \approx 0.744$$

Final Functionality: , $G \times P = 0.566 \times 0.744 \approx 42.1\%$

The G*P output (42.1%) is widely considered the most realistic balance, finding the grounded reality between the forgiving bottleneck nature of the MAX approach (55%) and the highly punishing cumulative nature of the SUM approach (36%).

Physical interdependency analysis outputs and interpretation of cascading service losses

After developing the physical interdependency network, represented by the aggregated master matrix, and defining the initial stress scenarios, the stress testing simulation can be applied to the network. The simulation **uses the three service dependency formulas** described in the following parts to estimate how reductions in the level of service of one or more subsectors propagate through the CI network.

Using three formulas rather than a single propagation rule is important because the analysis involves several sources of uncertainty. The interdependency data are collected through a participatory process, the analysis is conducted at a subsector level, and the actual behavior of CI systems under extreme stress is difficult to predict with certainty. Comparing the outputs of the three formulas, therefore, provides a broader and more robust understanding of potential cascading effects and system vulnerabilities.

In addition to comparing the results of the three formulas, **two types of sensitivity analysis are conducted:**

1) The first addresses uncertainty related to missing links in the aggregated network.
2) The second examines the sensitivity of the results to changes in the assumed intensity of the initial stress scenario.

Visualizing Level of Service During Cascading Effects

After the initial stress is introduced into the system, the simulation tracks how the LoS of each subsector changes over successive cascade steps. These results can be visualized through line charts showing the degradation of each subsector over time or cascade iterations.

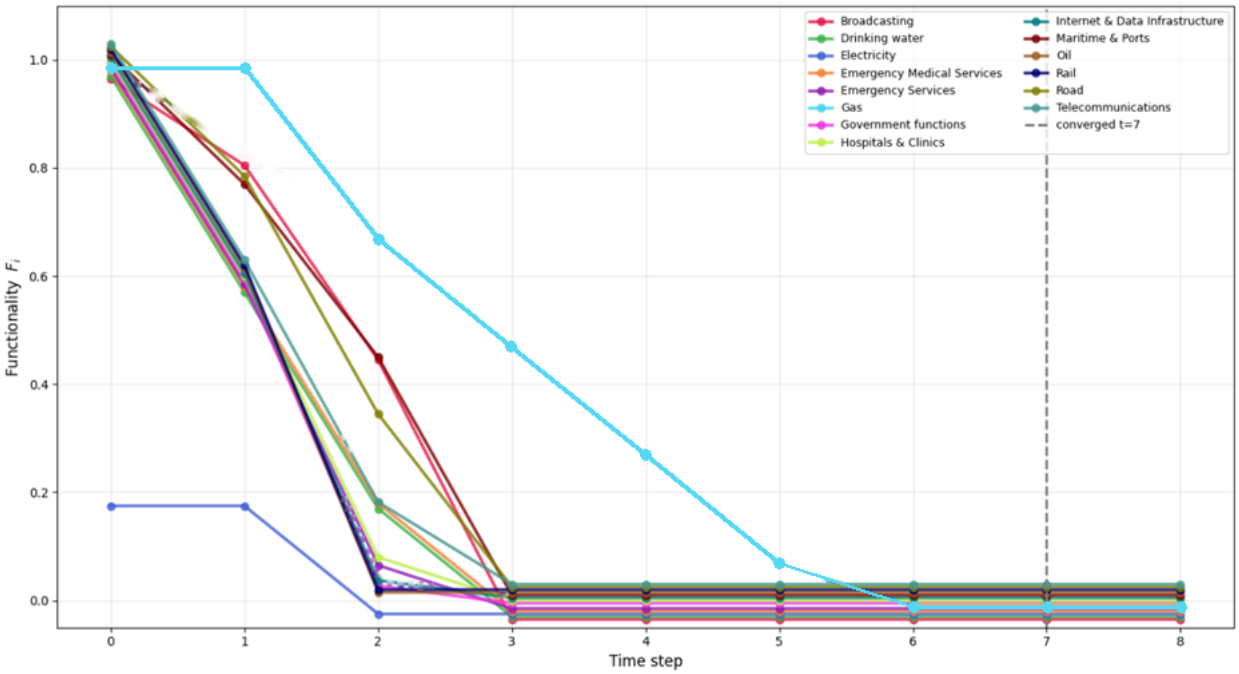

Figure 9: Example of charts showing the degradation of each subsector over time (cascade iterations) because of an initial major loss (stress) in the level of service of electricity

Figure 9 illustrates a synthetic example of the cascade analysis output, a network with 14 subsectors (nodes), showing how the level of service of different subsectors degrades over successive cascade iterations. In this illustrative example, an initial major stress is introduced by reducing the level of service of the electricity subsector by 80%. The chart then shows how this initial loss propagates through the interdependency network over time, affecting other subsectors according to their dependency relationships and the selected propagation formula. The figure is intended to demonstrate the type of dynamic degradation pattern that can be produced by the STT2 analysis, rather than to represent a specific empirical workshop result.

In these charts, the x-axis represents the cascade steps. Step 0 corresponds to the initial stress. Step 1 represents the immediate impact of the initial 80% loss of services introduced to the electricity subsector, directly on dependent subsectors. Subsequent steps represent secondary, tertiary, and further cascading effects until the system reaches stability. The y-axis represents the LoS of each subsector, ranging from 1.0, meaning 100% service provision, to 0.0, meaning complete service failure, or to the predefined minimum LoS if a residual service level is assumed.

These visualizations support the interpretation of cascade dynamics in three main ways:

- **Direct knock-on**: A sharp decline at Step 1 indicates that the subsector is strongly and directly affected by the initially stressed subsector.
- **Late onset failure**: A subsector whose LoS remains stable during the first steps but declines later can be interpreted as being affected through indirect cascading pathways. This indicates that the subsector may not be directly dependent on the initially stressed subsector, but becomes vulnerable as other supporting subsectors lose service capacity.
- **Convergence**: The point at which the LoS line stabilizes represents the post-cascade condition of the subsector under the given scenario. This can be interpreted as the remaining service capacity if no additional emergency interventions are introduced.

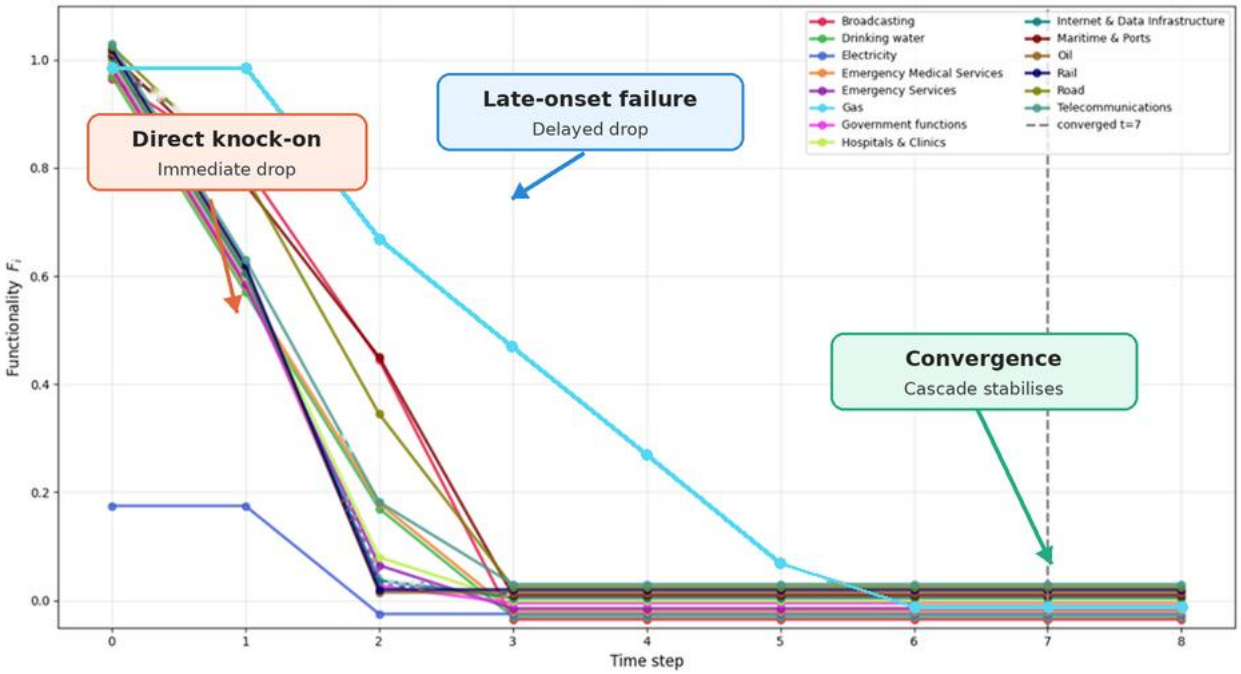

Figure 10: Example of cascade-pattern categories in STT2 service-level degradation analysis.

Quantitative Impact Metrics: AUC and AAC

In addition to visual interpretation, the simulation quantifies the impact of cascading service losses using two complementary metrics: the Area Under the Curve (AUC) and the Area Above the Curve (AAC).

The **Area Under the Curve** (AUC) measures the total service capacity retained by a subsector during the cascade. A higher AUC indicates that the subsector maintained a higher LoS for a longer period. Therefore, a higher AUC can be interpreted as greater resilience to the triggering stress scenario. (Bruneau et al., 2003).

The **Area Above the Curve** (AAC) measures the cumulative service loss experienced by a subsector during the cascade. It represents the area between the full-service baseline and the subsector's LoS curve. A larger AAC indicates that service was lost earlier, more severely, or for a longer period. Therefore, a higher AAC indicates a more severe impact from the triggering stress scenario.

Since AUC and AAC are perfectly complementary, with AAC being the total possible level of service (100%) minus the AUC, reporting both would be redundant. By focusing exclusively on AAC, we consistently frame the results in terms of vulnerability and loss, which directly aligns with the primary objective of identifying the most critical and fragile points within the interdependent infrastructure network.

In the simulation, the LoS of the subsector i is tracked over cascade steps t, where $F_i(t)$ represents the LoS of the subsector i at a time or cascade step t. The value 1.0 represents **full service provision**, while 0.0 represents **complete service failure**. For a given simulation horizon T, until the cascade converges, the AUC can be calculated using numerical integration, such as the trapezoidal rule.

The cumulative service loss, or AAC, is defined as:

$$AAC_i = (1.0 \times T) - \int_0^T F_i\,(t)\,dt$$

where:

AAC_i: Cumulative service loss experienced by subsector i

$F_i(t)$: Level of Service of subsector $i$at cascade step t

T: Simulation horizon until cascade convergence

$1.0 \times T$: Maximum possible service capacity over the simulation horizon

The AAC is particularly useful for ranking subsectors according to the severity of the cascading impact they experience. Subsections with higher AAC values are more severely affected by the initial stress scenario.

Main Analytical Outcome

The simulation produces three main types of analytical outputs:

1. **Critical subsector drivers of network-wide service loss**
2. **Most influential stressors for each subsector**
3. **Detailed scenario-specific interpretation across the three formulas**

Each output provides a different perspective on vulnerability, criticality, and cascading behavior in the CI network.

Critical subsector drivers of network-wide service loss

The first outcome identifies **subsectors that act as major drivers of cascading service loss** across the wider network. These can be interpreted as system-critical subsectors or cascade-driving subsectors. If these subsectors are stressed, they generate the most severe ripple effects across other parts of the CI system. The objective of this analysis is to identify subsectors whose disruption would create the highest collateral impact on the rest of the network.

Calculation Method

First, an initial service loss, such as a 50% reduction in LoS, is applied to one selected source subsector. The cascading effect is then simulated across the entire network until the system reaches convergence. For each affected target subsector, the AAC is calculated. To isolate the wider network effect, the service loss experienced by the source subsector itself is excluded from the calculation. This means that the criticality score focuses only on the collateral consequences imposed on other subsectors.

The system-wide criticality score of a source subsector is calculated as:

$$Criticality(S_{source}) = \sum_{j \neq source} AA\,C_{S_{source} \rightarrow S_j}$$

where:

S_{source} The subsector where the initial stress is introduced

S_j Any other subsector affected by the cascade

$AAC_{S_{source} \to S_j}$ Cumulative service loss experienced by subsector j due to stress in the source subsector.

Subsectors with the highest criticality scores can be interpreted as key drivers of network-wide service disruption. From a resilience planning perspective, strengthening these subsectors can generate wider system benefits because their protection reduces cascading impacts across multiple other subsectors.

Example:
This output identifies which subsectors generate the largest cascading impacts across the wider CI network when they are initially stressed. The score represents the total collateral service loss, measured through the sum of AAC values induced in all other subsectors. The initially stressed subsector itself is excluded from this calculation.

Table 4: Ranking of subsectors by overall stress impact under the MAX, SUM, and G×P stress propagation formulas, based on AAC values.

Rank	MAX Formula	AAC	SUM Formula	AAC	G*P Formula	AAC
1	Road	438.2	Electricity	463.4	Electricity	440.1
2	Electricity	432.8	Road	458.6	Road	435.7
3	Internet & Data Infrastructure	426.6	Hospitals & Clinics	451.2	Internet & Data Infrastructure	431.3
4	Drinking water	418.9	Drinking water	444.8	Drinking water	425.6
5	Rail	414.7	Internet & Data Infrastructure	442.1	Hospitals & Clinics	417.4
6	Hospitals & Clinics	409.3	Emergency Services	433.9	Rail	411.2
7	Maritime & Ports	401.4	Telecommunications	429.3	Emergency Services	405.8
8	Emergency Services	397.5	Rail	423.6	Telecommunications	399.4
9	Telecommunications	391.8	Maritime & Ports	416.9	Maritime & Ports	397.1
10	Gas	386.1	Gas	407.2	Gas	389.5
11	Oil	378.6	Oil	401.4	Oil	380.7
12	Broadcasting	371.9	Emergency Medical Services	394.8	Emergency Medical Services	374.2
13	Emergency Medical Services	366.4	Broadcasting	387.5	Broadcasting	368.9
14	Government functions	358.7	Government functions	375.3	Government functions	360.4

The illustrative results show that Electricity, Road, and Internet & Data Infrastructure appear among the highest-ranked cascade-driving subsectors across the three formulas. This suggests that, under this example of stress test, disruptions in these subsectors would generate substantial collateral service loss across the wider CI network.
The comparison across formulas also provides additional insight. Under the MAX formula, Road is ranked first, indicating that it acts as a dominant bottleneck for several dependent

subsectors. Under the SUM formula, Electricity becomes the highest-ranked driver, suggesting that its disruption contributes to cumulative service degradation across multiple subsectors. Under the G*P formula, Electricity remains the highest-ranked driver, indicating that it is important not only as a general dependency but also as a critical service for maintaining the basic viability of other subsectors.
Subsectors that remain highly ranked across all formulas can be interpreted as robust candidates for resilience prioritization, because their importance is not dependent on one specific propagation assumption. In contrast, subsectors whose ranking changes substantially across formulas may require more detailed interpretation to understand whether their influence is driven by bottleneck effects, cumulative service losses, or critical-service dependencies.

Most influential stressors for each subsector

The second outcome identifies, **for each subsector, which external subsector poses the greatest threat to its service provision**. While the previous metric focuses on which subsectors cause the most damage across the whole network, this metric focuses on the vulnerability of each individual subsector. The objective is to answer the following question: For each subsector, which external subsector failure would cause the highest cumulative service loss?

Calculation Method

The model evaluates the full cross-impact matrix, which contains the AAC caused by every possible source subsector on every possible target subsector. For each target subsector, the model identifies the source subsector that produces the highest AAC.

This can be expressed as:

Table 5:Comparison of the most influential stressors affecting each target subsector across the MAX, SUM, and G×P stress propagation formulas, based on AAC values.

Target subsector	MAX Formula: most influential stressor	AAC	SUM Formula: most influential stressor	AAC	G*P Formula: most influential stressor	AAC
Broadcasting	Internet & Data Infrastructure	34.8	Telecommunications	36.4	Telecommunications	35.9
Drinking water	Electricity	35.2	Electricity	37.1	Electricity	36.6
Electricity	Road	34.6	Road	37.8	Internet & Data Infrastructure	36.2
Emergency Medical Services	Road	35.5	Hospitals & Clinics	38.3	Telecommunications	37.4
Emergency Services	Telecommunications	35.1	Telecommunications	38.0	Internet & Data Infrastructure	37.1
Gas	Electricity	33.7	Electricity	35.6	Electricity	34.2
Government functions	Telecommunications	34.4	Internet & Data Infrastructure	37.9	Telecommunications	36.8
Hospitals & Clinics	Electricity	36.1	Electricity	38.5	Electricity	38.1
Internet & Data Infrastructure	Electricity	35.8	Electricity	37.6	Electricity	36.9

Maritime & Ports	Road	34.9	Road	36.7	Road	36.1
Oil	Maritime & Ports	33.9	Road	36.2	Road	35.4
Rail	Electricity	35.4	Electricity	38.2	Electricity	37.7
Road	Gas	34.2	Oil	36.8	Oil	35.9
Telecommunications	Electricity	35.7	Electricity	37.4	Electricity	37.0

$$VulnerabilityThreat(S_{target}) = \arg \max_{i \neq target} \left(AAC_{S_i \rightarrow S_{target}} \right)$$

where:

S_{target} The subsector whose vulnerability is being assessed

S_i A possible source subsector where stress is introduced

$AAC_{S_i \rightarrow S_{target}}$ Cumulative service loss experienced by the target subsector due to stress in source subsector i

This output is useful for local contingency planning. It helps infrastructure operators identify which external subsector failures they should prioritize when designing backup systems, redundancies, mutual aid arrangements, or emergency procedures. For example, the results may show that the Water subsector is most vulnerable to a disruption in the Electricity subsector.

Example:
This output identifies, for each target subsector, which external subsector causes the highest cumulative service loss when stressed. This is useful for contingency planning because it shows which external dependency represents the greatest threat to each subsector's Level of Service.

The illustrative results show that several subsectors are most strongly affected by stress in the Electricity sector, particularly Drinking water, Gas, Hospitals & Clinics, Internet & Data Infrastructure, Rail, and Telecommunications. This indicates that Electricity acts as a key supporting service for several target subsectors in this example network.
The results also show that Road is an influential stressor for subsectors such as Electricity, Maritime & Ports, and Oil, suggesting that transport accessibility plays an important role in maintaining service continuity for these domains. Similarly, Telecommunications and Internet & Data Infrastructure appear as major stressors for communication-dependent and coordination-dependent subsectors, including Broadcasting, Emergency Services, and Government functions.
This type of output is useful because it provides a target-specific vulnerability perspective. While the system-wide criticality ranking identifies which subsectors cause the largest total impact across the network, the subsector vulnerability analysis identifies the most important external threat for each individual subsector. For example, the results may show that Hospitals & Clinics are most vulnerable to stress in Electricity, while Broadcasting is most vulnerable to stress in Telecommunications or Internet & Data Infrastructure. This distinction is important for designing targeted contingency plans, backup arrangements, and cross-sector coordination protocols.

Detailed scenario-specific interpretation across the three formulas

For specific stress scenarios of interest, the results can be interpreted in more detail by comparing the outputs of the three formulas: MAX, SUM, and G*P. This comparison helps explain not only which subsectors are affected, but also why they are affected under different assumptions about cascading service loss.
Each formula emphasizes a different mechanism of service-loss propagation.

1) The MAX formula mainly reflects the effect of dominant bottlenecks, where the largest service loss within each dependency category determines the impact.
2) The SUM formula reflects cumulative stress, where multiple service losses add up and jointly reduce the level of service of a subsector.
3) The G*P formula reflects a more differentiated service logic by distinguishing between critical services that affect basic viability and non-critical services that mainly influence performance.

Therefore, comparing subsector rankings across the three formulas is useful because it **helps identify the underlying type of vulnerability**. A subsector that ranks highly under one formula but not another may be vulnerable for a specific reason, such as dependence on a single bottleneck, exposure to multiple simultaneous losses, or sensitivity to the disruption of critical services.
The comparison is based primarily on AAC rankings. For each formula, subsectors are ranked according to their AAC values, with Rank 1 representing the subsector with the highest cumulative service loss. It is important to note that rankings are relative. A subsector may drop in rank from one formula to another even if its own AAC remains similar. This can happen because other subsectors experience larger increases in AAC and therefore move above it in the ranking. For example, the Roads subsector may have a similar AAC under both the MAX and SUM formulas, but its rank may shift from 2nd under MAX to 6th under SUM because other subsectors become more severely affected under the cumulative logic of the SUM formula.
When interpreting the rankings, the following points should be considered:

- If the same weights for High, Medium, and Low dependencies are used in both the MAX and SUM formulas, the AAC under SUM will always be equal to or greater than the AAC under MAX. This is because the SUM formula adds all service losses within each dependency category, while the MAX formula only takes the largest service loss in each category.

- The G*P formula follows a different mathematical logic and should not be directly compared with MAX and SUM using absolute AAC values alone. Instead, it is more appropriate to interpret the changes in its rankings of CIs compared to the SUM and MAX formulas. The G*P formula is **multiplicative**, meaning it compounds the effects of failures, whereas the MAX and SUM formulas are **additive**, calculating loss through simple summation. Because of this fundamental difference, the absolute AAC values they produce are not on the same scale and cannot be directly compared. Instead, the value lies in comparing the **relative rankings** of infrastructure impacts, as this reveals how the different modeling assumptions (additive vs. multiplicative) change our understanding of which components are most critical.

Interpreting Cross-Formula Patterns:

Comparing the ranking of subsectors across the three formulas can reveal different types of vulnerability. The following patterns are the most important patterns that can be used to support interpretation:

- Single-Bottleneck Victim
- Compound-Stress Victim
- Elasticity Survivor
- Consistently High-Impact Subsector

- **Single-Bottleneck Victim:**

  Pattern: High rank under MAX, but lower rank under SUM and G*P.

Interpretation: This indicates that the **subsector is mainly affected by one dominant service loss**. Under the MAX formula, this dominant loss strongly influences the result. However, under the SUM formula, other subsectors may experience more compound service losses and therefore move higher in the ranking.

If the AAC under MAX and SUM is approximately equal, this suggests a relatively pure single-bottleneck case. The subsector is mainly vulnerable to one key supporting service within each dependency category. If the AAC under SUM is higher than under MAX, but the subsector still drops in rank, this suggests a dominant-bottleneck case with some additional minor service losses. The main vulnerability still comes from one major dependency, while other disruptions play a secondary role.

Planning implication: The priority should be to identify and strengthen the dominant supporting service that creates the main bottleneck.

Example: In the example plot, Telecommunications ranks highly under the MAX formula but becomes less highly ranked under the SUM and G*P formulas as a result of the introduced stress to a specific subsector in the network. This suggests that its impact is mainly driven by one dominant dependency bottleneck. When cumulative losses or the distinction between critical and non-critical dependencies are considered, other subsectors become more severely affected.

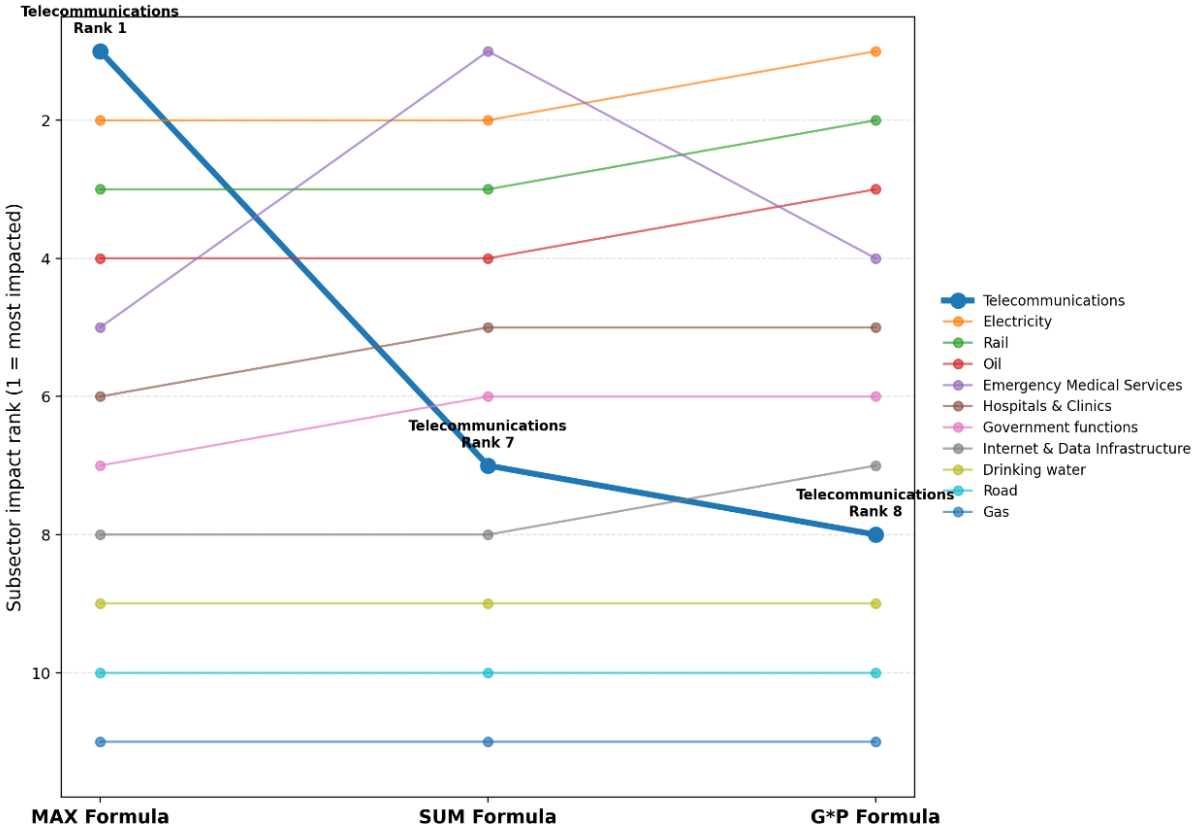

Figure 11: Example of a single-bottleneck victim pattern.

Telecommunications is vulnerable to a specific supporting service. The main resilience strategy should focus on identifying and strengthening that dominant dependency, for example, through redundancy, backup systems, or priority restoration arrangements.

- **Compound-Stress Victim:**

Pattern: Low or moderate rank under MAX, but high rank under SUM and often also high under G*P.

Interpretation: This indicates that the subsector is not primarily affected by one single severe service loss. Instead, it is **vulnerable because several supporting services experience partial degradation at the same time**. These losses accumulate and produce a severe total impact. This type of subsector may appear relatively stable under a bottleneck-based interpretation, but highly vulnerable under a cumulative-stress interpretation.

Planning implication: The subsector requires broad contingency planning across multiple supporting services, rather than protection against only one dominant dependency.

Example: In the example plot below, Hospitals & Clinics is not among the most impacted subsectors under the MAX formula, but becomes highly ranked under the SUM and G*P

formulas. This means that several supporting services may degrade simultaneously, and their combined effect can be severe on the hospitals and clinics.

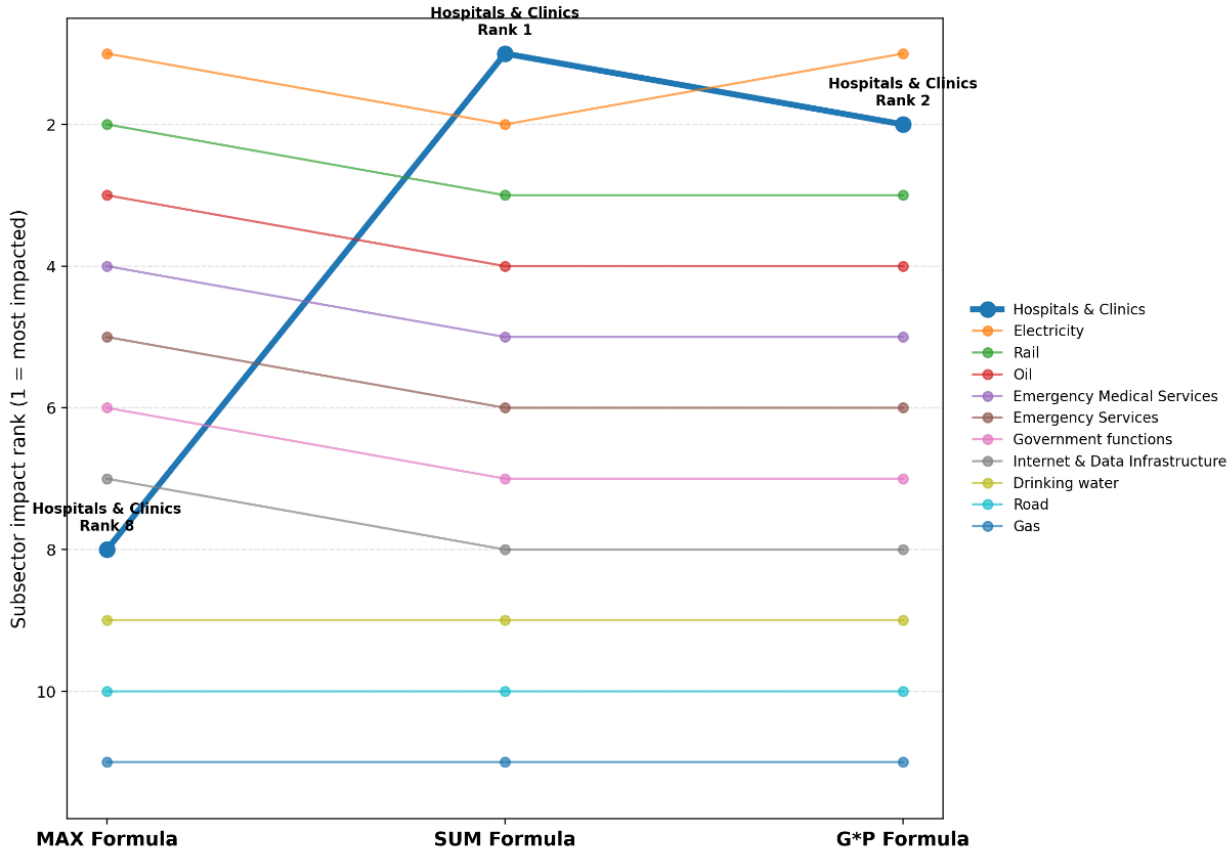

Figure 12:Example of a compound-stress victim pattern.

Hospitals & Clinics may require broad contingency planning across multiple supporting services, for example, electricity, water, telecommunications, road access, and emergency medical services. The priority is not only to protect against one dominant dependency, but to improve redundancy across several service relationships.

- **Elasticity Survivor:**

  Pattern: High rank under MAX or SUM, but substantially lower rank under G*P.

Interpretation: This indicates that the **subsector appears vulnerable when all losses are treated as direct constraints or cumulative penalties**. However, under the G*P formula, the most affected dependencies are likely to be non-critical services. As a result, the subsector may experience reduced efficiency or degraded performance, but not necessarily a severe loss of basic service viability. This pattern suggests that the subsector has some capacity to absorb or adapt to the loss of non-critical services.

Planning implication: The subsector may require operational adjustments and efficiency support during disruptions, but it may not be at immediate risk of complete service collapse.

Example: In the example plot below, Road ranks highly under MAX and SUM but drops under G*P. This suggests that many of the affected dependencies may be non-critical. The

subsector may experience reduced performance, delays, or operational inefficiencies, but it may still maintain a minimum level of service.

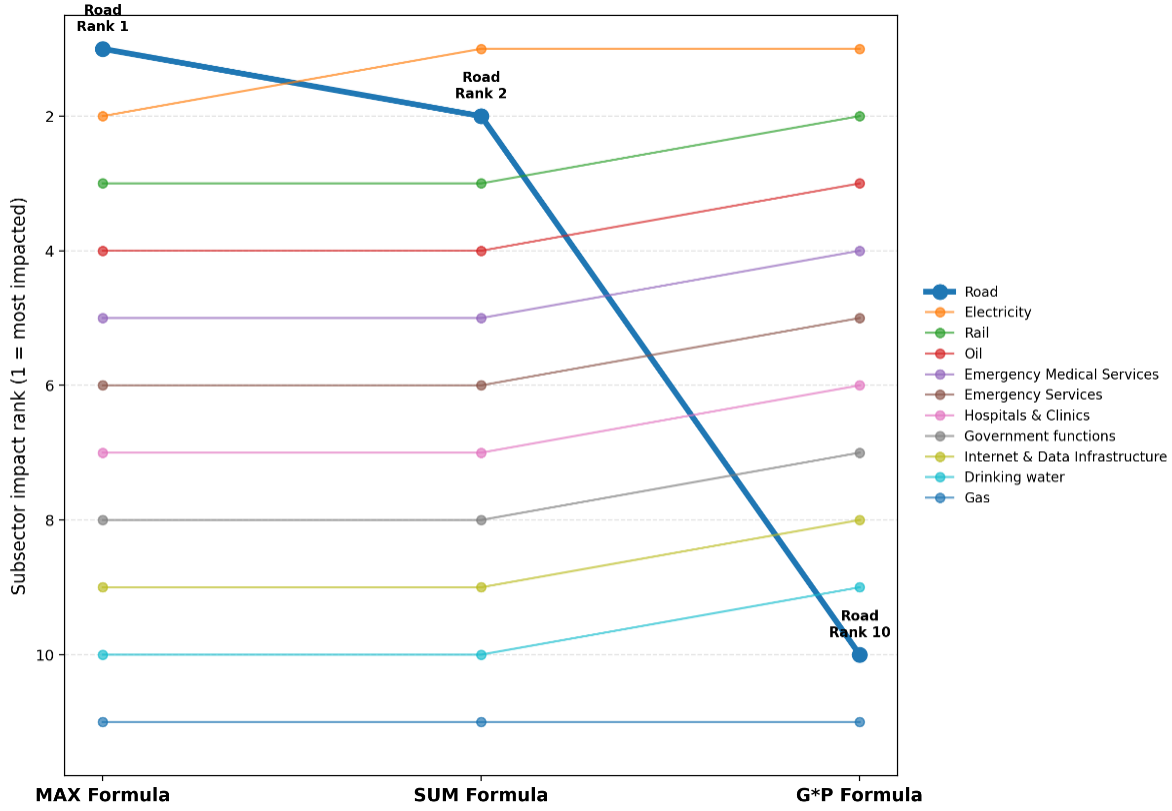

Figure 13: Example of an elasticity survivor pattern.

Road infrastructure may be operationally disrupted but not necessarily at risk of immediate service collapse. Planning should therefore focus on degraded-operation management, traffic rerouting, emergency access prioritization, and adaptive operational measures.

- **Consistently High-Impact Subsector**

  Pattern: High rank across MAX, SUM, and G*P.

Interpretation: This indicates a robust vulnerability finding. The **subsector is affected under multiple assumptions about service-loss propagation**. It may be exposed to a dominant bottleneck, cumulative service losses, and disruption to critical dependencies. If the AAC values under MAX and SUM are both high and close to the maximum possible area, this suggests that the subsector experiences severe and persistent service loss under the stress scenario.

Planning implication: This subsector should be treated as a high-priority candidate for resilience investment, emergency planning, and cross-sector coordination.

Example: In the example plot below, Emergency Medical Services remains highly ranked across MAX, SUM, and G*P. This suggests that it is affected by a dominant bottleneck, cumulative service losses, and critical dependency disruptions.

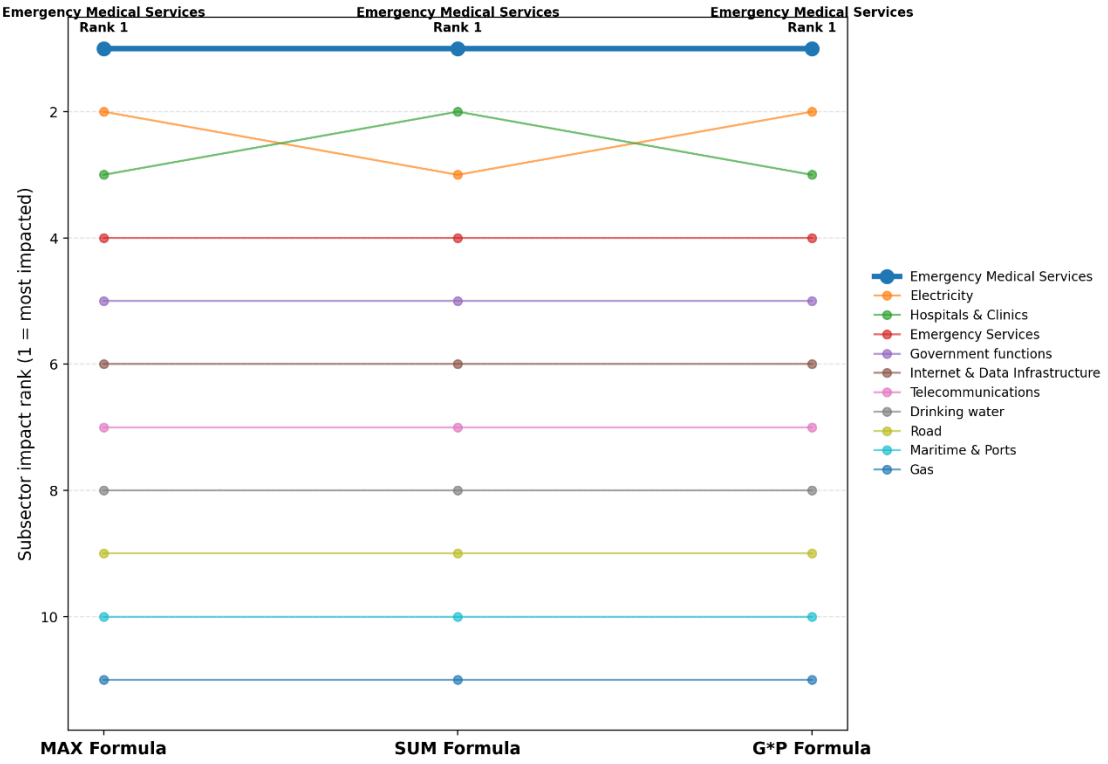

Figure 14: Example of a consistently high-impact subsector pattern.

Emergency Medical Services should be treated as a high-priority candidate for resilience investment and emergency planning. Its vulnerability is not dependent on one specific modeling assumption, which makes the finding more robust.

Sensitivity Analysis Outputs Interpretation

Sensitivity analysis is used to assess the robustness of the simulation results under uncertain assumptions. In this method, two types of sensitivity analysis are applied:

1. **Sensitivity analysis for missing links in the network**
2. **Sensitivity analysis for the intensity of the initial stress scenario.**

In the next parts, we will look into how these sensitivity analyses should be interpreted:

1. Sensitivity analysis for missing links in the network

After aggregating the group-level interdependency matrices into one master network, some potential links may remain missing. These missing links represent relationships for which no direct rating was provided by workshop participants. However, some of these links may still be relevant for cascading effects. To **assess the influence of these missing links**, the simulation is repeated under **different assumptions about their dependency ratings**. Specifically, missing links are assigned alternative ratings of:

- 0 = no dependency
- L = low dependency
- M = medium dependency
- H = high dependency

The same propagation simulations are then conducted for each rating scenario and for each of the three formulas. The results should be analyzed separately for MAX, SUM, and

G*P, with particular attention to changes in subsector rankings and AAC values. Extreme comparisons are especially informative. For example, comparing the results when missing links are assigned **0** or **L** against the results when they are assigned **H** can show whether the network is highly sensitive to assumptions about unobserved dependencies.
If the rankings remain relatively stable across missing-link scenarios, this suggests that the results are robust. In this case, assigning a moderate value, such as **M**, to missing links can be considered a reasonable and cautious modeling choice. However, if the rankings change substantially across scenarios, this indicates that the results are highly sensitive to missing-link assumptions. In such cases, two options can be considered. The first is to adopt a conservative modeling approach by assigning **H** to missing links. The second is to conduct an additional expert validation survey to obtain more reliable estimates of the importance of these uncertain links.

2. Sensitivity analysis for the intensity of the initial stress scenario

A second sensitivity analysis is conducted by **varying the intensity of the initial stress scenario**. For example, the introduced loss of LoS can be varied by ±10% around the baseline stress assumption. This is important for three reasons:
First, the **initial stress level is often uncertain**. It may be based on expert judgment, assumptions, scenario narratives, or approximate estimates. Varying the stress level helps assess whether the model results remain stable when the assumed stress intensity changes moderately. Second, **CI systems may respond non-linearly to changes in service levels**. A 10% increase or decrease in initial stress may have limited effects in some cases but may trigger disproportionate cascading effects in others. Sensitivity analysis can therefore help identify possible tipping points or thresholds. Third, sensitivity analysis provides a **basic model validation check**. If a small and plausible change in the initial stress level produces unrealistic or highly unstable results, this may indicate a problem in the model structure, mathematical logic, or implementation.
A ±10% variation is used as a practical rule of thumb because it represents a plausible but meaningful deviation from the baseline assumption. It is large enough to reveal potential sensitivity, but not so large that it represents a completely different scenario. After comparing the results for the baseline stress scenario, the +10% scenario, and the −10% scenario across the three propagation formulas, the stability of subsector rankings can be assessed. If no significant changes are observed, this suggests that the model results are robust to moderate variations in the initial stress level. If the rankings or AAC values change substantially, the results should be interpreted with caution, and the subsectors affected by these changes should be examined in more detail.
This sensitivity check has already been applied in the first STT2 case study. In that application, the subsector rankings and AAC-based interpretation remained stable under the ±10% variation of the initial stress level, suggesting that the model outputs were robust to moderate changes in the assumed initial disruption. The results of this sensitivity analysis will be reported in the case-study-specific deliverables.

4.1.3.2. Decisions and Their Interdependency ST analysis and Output

Input data used in the analysis

The input data for the decision interdependency analysis are collected during the workshop session on Mapping Decisions and their Interdependencies. As mentioned before, in this session, participants first identify important decisions from the perspective of their organization, the scenario, and the relevant time phase, such as early response or early recovery. The group then selects the most relevant decisions (4-6) and completes a structured decision table. For each decision, the table captures the responsible organization and location, allocated resources, information needed, information sources, coordination requirements, and the reason for prioritizing the decision. The workshop

protocol specifies that the main output of this exercise was a **completed decision table with at least four decisions**.
The raw workshop tables are not immediately suitable for network analysis because they contain free-text descriptions, combined actors in single cells, inconsistent naming of organizations, and sometimes ambiguous distinctions between information needs and coordination requirements. Therefore, the raw data is cleaned and transformed into a structured dataset of socio-technical interdependencies. The original decision table format included the following columns:

- Decision Made,
- Organization/ Infrastructure / Location,
- Resource Allocated,
- Information Needed,
- From Whom?
- Coordination Required With
- and Rationale / Priority Reason.

The cleaning process follows several steps. First, **each decision is reviewed and given a distinct decision entry**. Where multiple decisions or actions are written in one cell, they are separated only when the original meaning clearly supports this. Second, **resources are extracted and preserved as contextual attributes of the decision**, rather than treated as network nodes unless they clearly refer to an actor or organization. Third, **actors mentioned** in the *From Whom?* and *Coordination Required With* columns **are split into separate entries** when multiple organizations were listed together. For example, if one cell mentioned police, fire department, and municipalities, each actor was represented separately in the cleaned dataset.
Next, **actor names are standardized** to reduce duplication. This step ensures that variations such as “municipality,” “Municipality Rotterdam,” or “municipal services” are treated consistently when they refer to the same actor. **Each actor is also assigned a node type**, such as operational actor, coordination actor, information provider, decision-making actor, or critical infrastructure/service actor. This classification supports later analysis of which types of actors provide information, coordinate action, or depend on others.
The cleaned data distinguishes between two main dependency types:

1) information dependencies
2) coordination dependencies.

Information dependencies are derived mainly from the *Information Needed* and *From Whom?* columns. These represent cases where a decision-making actor depends on another actor for situational information, capacity information, risk information, or operational status updates. **Coordination dependencies** are derived from the *Coordination Required With* column. These represent cases where actors need to align actions, request support, allocate resources, or implement operational measures together.
A key cleaning rule is that coordination dependencies are recorded as reciprocal relations when mutual coordination was implied. For example, if the raw data indicates that an actor needed to coordinate with another organization, the cleaned edge list includes both directions where appropriate. This reflects the mutual nature of coordination, while information dependencies were generally kept directional, from the information provider to the actor using the information for a decision.
The cleaning process also aims to avoid over-interpreting the workshop data. Dependencies are only added when they were directly supported by the original table entries. When an actor, direction, or dependency type is unclear, the entry was either left conservative or flagged as an assumption. This is important because the purpose of the cleaning is not to create an idealized crisis-management model, but to preserve the meaning of participants’ responses in a form suitable for analysis.

The resulting cleaned dataset contains two analytical outputs. The first output is a **decision-context dependency table**, which keeps each decision linked to its resources, information needs, coordination requirements, actor types, decision category, and phase. The second output is a **network edge list**, which can be used for network analysis by representing actors as nodes and information or coordination dependencies as edges. Together, these outputs make it possible to analyze how crisis decisions depend on information flows, coordination relations, and resource-related interactions between organizations.

Developing the Network of Interdependent Organizations

As explained above, data is collected on the communication patterns between organizations involved in, or affected by, the HILP scenario across two distinct operational stages: early response and early recovery. For network analysis, the cleaned data were transformed into an edge-list structure. Each dependency was represented as a directed relation between a source and a target, with additional attributes describing the edge content, dependency type, decision category, related decision, resources, and phase. Based on this data, an initial network model was developed for each stage to demonstrate structural interdependencies among the participating organizations.

The methodology translates the collected workshop data into a quantifiable mathematical network structure. This process requires mapping discrete crisis organizations, such as the Municipality of Rotterdam or the regional safety authority, as individual nodes within a graph. Interactions between these entities are modeled as directed edges to capture the highly asymmetrical nature of crisis communication, acknowledging that a command issued in one direction does not necessarily imply a reciprocal flow of information. Crucially, the resulting network model is strictly unweighted, i.e., all operational edges carry an identical, uniform weight to ensure the analysis evaluates the pure topological architecture of the response system.

To ensure the analysis measures the true structural integrity of the network rather than simply the volume of communication, the dataset undergoes a data cleaning process. This step filters out redundant edges between identical pairs of organizations concerning the same crisis phase. It prevents high-frequency routine updates from artificially inflating the perceived strength of a connection.

Within the collected data, communication flows are categorized into two primary typologies:

1) information
2) coordination.

Information flow primarily consists of situational updates and intelligence gathering regarding the HILP incident. Conversely, ***coordination flow*** governs operational execution, including crisis response directives, resource management, and external crisis communication. In addition to creating a comprehensive, interdependent network for each stage (early response and early recovery), these overarching networks were further stratified by flow type. Consequently, for each stage, one integrated macro-network was generated, followed by two distinct subnetworks representing the isolated information and coordination flows.

Centrality Analysis

To understand the structural hierarchy and operational influence of individual organizations within the HILP scenario, a comprehensive centrality analysis is conducted. Crucially, this analysis is not limited to the overarching network of each operational stage. Instead, the centrality algorithms are independently executed across the main network as well as the isolated information and coordination subnetworks. This multi-layered approach is essential for identifying functional disparities within the crisis architecture. By calculating metrics across distinct subnetworks, the methodology reveals shifting operational profiles,

demonstrating whether an organization acts as a critical hub for gathering situational awareness but remains peripheral when issuing tactical coordination commands.
The **structural influence of each organization is quantified** using four distinct network metrics, beginning with In-Degree and Out-Degree centrality to measure immediate, localized connectivity.
In-Degree centrality calculates the total **number of incoming communication flows** directed toward a specific node. High scores in this metric highlight organizations that function as primary sinks for field intelligence or major recipients of operational directives. Conversely, **Out-Degree centrality** enumerates the **outgoing flows**, identifying the primary sources of situational updates and the central authorities responsible for dictating crisis response commands.
By comparing these directional metrics across the information and coordination subnetworks, the analysis effectively categorizes organizations into distinct operational archetypes, distinguishing pure intelligence providers from command-and-control authorities.
Beyond immediate local connections, the methodology evaluates the **broader topological positioning of each organization** through Betweenness and Eigenvector centrality.
Betweenness centrality quantifies the frequency with which a specific organization sits on the shortest communication path between other pairs of nodes in the network. Organizations scoring high in this metric act as vital structural bridges and, consequently, represent potential communication bottlenecks. Their presence is required to connect disparate operational clusters, meaning their removal or failure would severely disrupt cross-agency coordination.
Eigenvector centrality introduces the concept of relative influence, measuring not just the sheer volume of an organization's connections, but the strategic importance of those connections. An organization achieves a high Eigenvector score if it is directly linked to other highly influential, well-connected command hubs, indicating a position of deeply embedded systemic power.

Ultimately, the centrality analysis output tracks how institutional power and communication bottlenecks occur in different stages and provides targeted insights into the resilience of the administrative hierarchy.

Stress Testing the Network

To rigorously evaluate the resilience and structural vulnerability of the crisis response architecture, a stress-testing simulation is applied to the network models. This analytical phase moves beyond static observation to dynamically model how the administrative governance structure withstands operational shocks during the HILP scenario. The simulation introduces stress through two distinct degradation mechanisms: targeted edge removal and targeted node removal. Edge removal simulates the failure of specific, isolated communication channels, representing realistic crisis complications such as severed radio frequencies, interoperability failures, or procedural breakdowns between liaison officers. Conversely, node removal models a more catastrophic systemic shock, representing the complete incapacitation of an entire organization due to physical destruction, overwhelming resource depletion, or communication blackout. By systematically removing these elements both iteratively in a Leave-One-Out isolation approach and sequentially based on their previously calculated centrality rankings, the methodology exposes hidden administrative dependencies and single points of failure.
To quantify the propagation of this structural stress and evaluate the shifting effectiveness of the network, the methodology leverages a suite of network-level metrics standard to disaster response evaluation, drawing upon the framework articulated by Nowell et al. (2018). Following each simulated failure of a node or edge, the network's compromised state is immediately assessed against its baseline using five critical indicators.

1) First, changes in overall network Density are measured to track the rapid decay of systemic communication cohesion.
2) Second, shifts in structural Centralization are calculated to determine whether the loss of a critical hub forces the network into a highly decentralized but functional web, or if it simply paralyzes the remaining command structure.
3) Third, the simulation closely monitors the Average Path Length across the remaining operational structure. An increase in this specific metric serves as a direct proxy for operational delay, mathematically quantifying the additional administrative hurdles and detours that time-sensitive information must clear when primary communication routes are severed.
4) Fourth, the network's Clustering Coefficient is recalculated to observe whether localized, tight-knit collaborative sub-teams, such as specific medical or public safety operational groups, remain structurally intact or dissolve under systemic pressure.
5) Finally, the simulation tracks the total Number of Components, which measures holistic connectivity. An increase in this final metric flags the most severe consequence of a structural failure, i.e., network fragmentation. When the number of components rises, it signals that the network has fractured into isolated islands, leaving specific organizations entirely cut off from central command and broader situational awareness.

Through this continuous, iterative process of removal, measurement, and restoration, the analysis generates a comprehensive vulnerability profile of the crisis response network. By correlating the loss of specific edges and nodes with severe degradations in Density, Centralization, Path Length, Clustering, and Connectivity, the methodology identifies the network's critical load-bearing infrastructure. Ultimately, these outputs allow decision-makers to anticipate how localized communication breakdowns cascade into systemic service losses and provide the empirical foundation necessary to design strategic redundancies and more robust governance protocols.

4.1.4. Method boundaries

The STT2 methodology has been designed as an intermediate, participatory, and semi-quantitative stress-testing approach within the broader AGILE tiered stress-testing framework. Its boundaries are defined by its intended purpose: to support the structured identification and analysis of interdependencies, cascading effects, decision dependencies, information needs, coordination requirements, and resource bottlenecks in critical infrastructure systems under HILP conditions. The methodology is not intended to replace performance-based engineering analysis, operational modelling, or quantitative simulation. Rather, it provides a structured bridge between qualitative system understanding and more advanced analytical or modelling approaches.

A first boundary concerns the level of analysis. STT2 focuses on system-level and functional interdependencies rather than detailed technical failure mechanisms of individual assets. It is designed to explore how disruption to one infrastructure, organization, decision, or information flow may affect other parts of the system. It does not aim to calculate precise probabilities of failure, technical fragility curves, repair times, or asset-level performance thresholds. Where such data are available, they may support the preparation or analysis of the stress test, but they are not a prerequisite for applying the methodology.

A second boundary concerns the type of evidence generated. The methodology produces semi-quantitative and qualitative evidence based on structured stakeholder elicitation. Outputs include problem-structuring results, physical interdependency maps, interdependency matrices, decision tables, documentation of information and coordination dependencies, and qualitative insights on bottlenecks and cascading effects. These

outputs provide a structured representation of stakeholder knowledge and perceived system dependencies. However, they should not be interpreted as statistically validated measurements of system performance. The strength of STT2 lies in making hidden interdependencies visible for decision-makers instead of producing deterministic predictions.

A third boundary relates to threat-agnosticism. STT2 is designed to be adaptable across different hazards and stress contexts. However, in practice, the workshop still requires a scenario to provide participants with a concrete basis for discussion. This means that the methodology is threat-agnostic in its analytical logic, but scenario-informed in its implementation. The scenario is used to create a shared stress context, not to limit the analysis to one specific hazard. Therefore, care must be taken to design scenarios that are specific enough to support meaningful discussion, but broad enough to allow exploration of systemic vulnerabilities and cascading effects beyond the initial triggering event.

A fourth boundary concerns temporal scope. The current STT2 design focuses mainly on the early response and early recovery phases of a crisis. These phases were selected because they are critical for understanding how disruptions evolve, how decisions are made under uncertainty, and how coordination and information dependencies influence system outcomes. The methodology can be adapted to other phases, such as preparedness, long-term recovery, or reconstruction, but such adaptations require changes to the workshop design, scenario framing, and analytical templates.

A fifth boundary concerns stakeholder participation. STT2 depends on the presence and active engagement of relevant stakeholders. The quality of the outputs is strongly influenced by who participates, the diversity of sectors and organizations represented, the level of operational and strategic knowledge in the room, and the willingness of participants to share information. If key organizations are absent, certain dependencies may remain invisible. If participants have limited knowledge of operational interdependencies, the outputs may be less detailed. For this reason, stakeholder selection and preparation are important methodological boundary conditions.

A sixth boundary concerns geographic scale and complexity. STT2 can be applied at different geographical and organizational scales, including urban, regional, national, or cross-border contexts. However, the larger and more complex the system, the more difficult it becomes to represent all relevant infrastructures, actors, decisions, and dependencies within a one-day workshop. The methodology, therefore, requires scoping choices. These may include limiting the number of infrastructures mapped, focusing on selected critical functions, dividing participants into thematic groups, or conducting multiple workshops for different sub-systems or phases.

A seventh boundary concerns facilitation and documentation capacity. The methodology requires trained facilitators and note takers who understand both the local context and the methodological logic of STT2. Poor facilitation may lead to incomplete outputs, inconsistent scoring, unclear dependency directions, or domination of the discussion by a few participants. Similarly, insufficient note-taking or incomplete documentation may reduce the analytical value of the workshop outputs. The mock workshop, facilitator guidance, note taker guidance, and post-workshop digitization process are therefore important safeguards, but they do not fully remove the dependency on human judgment and process quality.

An eighth boundary concerns confidentiality and security. STT2 may reveal sensitive information about critical infrastructure vulnerabilities, dependencies, organizational bottlenecks, and coordination weaknesses. The level of detail included in workshop outputs and reports must therefore be managed carefully. The methodology is intended to support learning and resilience improvement, not public exposure of sensitive vulnerabilities. Anonymization, aggregation, restricted reporting, and host review may be required depending on the sensitivity of the findings.

Finally, STT2 should be understood as an iterative learning methodology. It provides structured evidence for identifying vulnerabilities and bottlenecks, but it does not

automatically prescribe solutions. Recommendations and resilience-enhancing measures require further interpretation, validation with stakeholders, and, where appropriate, integration with other AGILE stress test tiers or complementary modelling approaches.

5. Evaluation

The evaluation of the STT2 methodology assesses whether the methodology is scientifically grounded, practically usable, analytically valuable, and suitable for application across different case study contexts. In line with the design science approach underpinning the methodology, evaluation is not treated as a one-off activity but as an iterative process of testing, reflection, adjustment, and refinement. The purpose of the evaluation is therefore twofold: first, to assess the current performance of the methodology, and second, to identify improvements that can increase its robustness, usability, and transferability in future applications.

5.1. Criteria

The evaluation focuses on three main criteria:
1) methodological soundness,
2) usability and applicability,
3) and analytical value.

These criteria reflect the dual objective of STT2: to provide a theoretically grounded stress-testing approach and to ensure it can be implemented in real-world stakeholder settings.

5.1.1. Methodological soundness

Methodological soundness refers to the extent to which STT2 is coherent, theoretically grounded, internally consistent, and aligned with the conceptual principles of the AGILE stress-testing framework. This includes whether the methodology operationalizes the core principles of system focus, threat-agnosticism, resilience orientation, and tiered stress testing, which is adaptable to different contexts.

A methodologically sound STT2 implementation should enable stakeholders to move beyond single-sector or single-asset perspectives and instead consider critical infrastructures as interconnected socio-technical systems. It should support the identification of physical interdependencies, information dependencies, coordination dependencies, and decision-making bottlenecks. It should also maintain a clear distinction between the scenario as a discussion frame and the methodology's broader threat-agnostic purpose.

Methodological soundness is also assessed in terms of traceability. The outputs of each session should be clearly linked to the methodology's objectives. For example, problem structuring should support the identification of priorities and concerns; physical interdependency mapping should support the identification of cascading effects; and decision mapping should support the identification of information, coordination, and resource dependencies. In addition, the methodology should generate outputs that can be documented, digitized, analyzed, and compared across groups and case studies.

5.1.2. Usability, usefulness, and applicability

Usability refers to whether the STT2 methodology can be understood and applied in practice by the different actors involved in the implementation, including hosts, facilitators, note takers, participants, and analysts. This includes the clarity of the workshop instructions, the feasibility of the agenda, the adequacy of the facilitation and note-taking

guides, the ease of working with templates, and the practicality of the overall workshop format.
Usefulness refers to whether the methodology provides value for the host organization and participating stakeholders. This value may include improved understanding of physical, informational, organizational, and decision-related interdependencies; identification of vulnerabilities and bottlenecks; recognition of coordination challenges; and support for shared learning among stakeholders.
Applicability refers to whether STT2 can be implemented across different case study contexts, geographical scales, governance settings, stakeholder groups, and HILP scenarios. Since AGILE case studies differ in terms of system complexity, stakeholder experience, institutional arrangements, available data, and local risk context, the methodology needs to be structured enough to ensure consistency but flexible enough to be adapted to different implementation environments.
To assess usability, usefulness, and applicability, feedback was collected through a survey (see Appendix) from participants involved in the implementation. Feedback meetings were held with the host, facilitators, and relevant stakeholders to reflect on the workshop process, the clarity of the sessions, the feasibility of the facilitation tasks, the suitability of the materials, and the usefulness of the outputs. In addition, a participant survey was shared with workshop participants to collect their input on the relevance, clarity, and added value of the methodology from their perspective.
The feedback collected through these meetings and surveys was used to identify practical challenges and improvement needs. These inputs informed refinements to the methodology, including clearer facilitation instructions, improved note-taking guidance, more structured templates, better preparation requirements, and the inclusion of a mock workshop before the actual implementation. This iterative feedback process is important for ensuring that STT2 remains applicable across different contexts while maintaining a coherent methodological core.
Overall, usability, usefulness, and applicability are assessed not only by whether the methodology can be implemented once, but by whether it can be replicated and adapted across diverse contexts. The evaluation, therefore, considers whether the methodology is understandable for users, useful for stakeholders, feasible for hosts, and sufficiently adaptable to support stress testing in different systems, locations, and institutional settings.

5.1.3. Analytical value

Analytical value refers to the extent to which the methodology produces outputs that are useful for identifying vulnerabilities, bottlenecks, cascading pathways, critical nodes, information dependencies, coordination gaps, and resource constraints. The evaluation considers whether the workshop outputs, such as interdependency matrices, network representations, decision-dependency tables, bottleneck summaries, and cross-phase comparisons, can be transformed into meaningful analytical products.
The analytical value of STT2 also depends on insights that the methodology provides that would not easily emerge from document review, expert interviews, or single-sector analysis alone. In particular, the methodology is expected to reveal hidden or under-discussed dependencies between infrastructures, organizations, information flows, and decisions. It should also support comparison between early response and early recovery, thereby showing how system vulnerabilities and priorities change over time.

5.2. Methods, sources of data, and evidence

The evaluation of STT2 is based on multiple sources of data and evidence. This multi-source approach is necessary because the methodology has both scientific and practical

objectives. It must be assessed not only by reviewing whether the design is conceptually coherent, but also by examining how it performs in real workshop conditions.

The first source of evidence is the **review of workshop outputs**. This includes problem structuring sheets, physical interdependency maps, interdependency matrices, decision tables, facilitator notes, note taker documentation, photographs of group outputs, and digitized post-workshop datasets. These materials are reviewed to assess whether the expected outputs were produced, whether they are complete, whether dependencies are recorded consistently, and whether the data can be used for further analysis.

The second source of evidence is **feedback from participants**. Participant feedback can be collected through feedback forms, surveys, plenary reflections, or informal comments during and after the workshop. This evidence helps assess whether participants understood the purpose of the methodology, found the sessions relevant, considered the scenario realistic and useful, and perceived the process as valuable for understanding interdependencies and resilience challenges.

The third source of evidence is **feedback from facilitators and note takers**. Facilitators and note takers provide important insights into the practical feasibility of the methodology. Their feedback helps identify whether the session instructions were clear, whether the timing was realistic, whether the materials were sufficient, whether participants understood the tasks, and where additional guidance or simplification may be needed.

The fourth source of evidence is **feedback from the case study host**. The host can assess whether the methodology was relevant to the local context, whether the right stakeholders were involved, whether the scenario and system boundaries were appropriate, and whether the results are meaningful for local decision-making. Host feedback is also important for identifying sensitivities, contextual constraints, and opportunities for further use of the outputs.

The fifth source of evidence is **direct observation by the method developer or methodological support team**. During the workshop, the method developer can observe how groups interact with the methodology, whether facilitators apply the instructions consistently, whether participants engage with the tasks, and whether outputs are completed as intended. These observations provide evidence on both methodological soundness and practical usability.

The sixth source of evidence is **post-workshop analysis**. The process of cleaning, digitizing, translating, coding, and analyzing the data provides important evidence on the analytical value of the methodology. If the workshop outputs can be transformed into structured interdependency datasets, network representations, and bottleneck analyses, this indicates that the methodology produces analytically useful outputs. If substantial interpretation, correction, or clarification is required, this indicates areas where the methodology or templates need refinement.

The seventh source of evidence is **validation with the host** and, where appropriate, with participants. After analysis, results can be shared with the host or stakeholders to check whether the identified dependencies, vulnerabilities, and bottlenecks are contextually accurate and meaningful. This validation step is important because STT2 outputs are based on stakeholder elicitation and interpretation; therefore, results should be reviewed.

The available methods and data also have limitations. Participant feedback may reflect individual perceptions and may be influenced by workshop dynamics, facilitation, expectations, or the participant's organizational role. Facilitator and note taker feedback may focus more on process issues than analytical outcomes. Workshop outputs may be incomplete or inconsistent if time was limited, if discussions were complex, or if some groups interpreted instructions differently. The data are also context-specific and cannot be treated as statistically representative of all possible stakeholders or system conditions.

Another limitation is that the methodology is evaluated in real-world settings, where controlled experimental comparison is not possible. Different case studies vary in scale, hazard context, stakeholder composition, available data, and institutional culture. As a result, differences in outputs may reflect both the methodology and the context in which it

was applied. This limitation is addressed by using multiple evidence sources, documenting assumptions, and comparing findings across case studies where possible.

5.3. Findings

Evidence of and challenges regarding methodological soundness

The evaluation indicates that the STT2 methodology provides a coherent structure for operationalizing a system-focused, resilience-oriented, and threat-agnostic stress-testing approach. The sequence of problem structuring, physical interdependency mapping, decision interdependency mapping, and cross-phase comparison enables participants to examine critical infrastructure systems as interconnected socio-technical systems rather than as isolated sectors.

The system focus is supported by the explicit mapping of relationships between infrastructures, organizations, information flows, resources, and decisions. The physical interdependency sessions help participants identify how the disruption of one infrastructure may affect others, while the decision-mapping sessions extend this perspective by capturing dependencies related to information and coordination. This combination strengthens the methodological soundness of STT2 by addressing both the technical and organizational dimensions of system resilience.

The methodology also demonstrates a practical form of threat-agnosticism. STT2 does not require detailed hazard probabilities, fragility curves, or asset-level failure models. Instead, it uses a HILP scenario as a stress context to help participants reason about how their system may behave under disruption. This scenario-based framing is important because stress scenarios are already commonly used in training exercises for first responders and other crisis-management stakeholders. As a result, the scenario provides participants with a familiar and realistic basis for discussion, helping them imagine system behavior in conditions that are closer to practical emergency and disaster-management exercises. At the same time, the scenario is not the final object of analysis. Its main purpose is to support the identification and mapping of interdependencies. Once these interdependencies have been mapped, the resulting network can be stress-tested beyond the specific hazard used in the workshop. For example, failures can be introduced randomly, applied to selected critical nodes, or tested across different parts of the system. In this way, the methodology remains adaptable to different hazards and contexts while still providing participants with a concrete and realistic entry point for discussion. However, the observations of the methodology developer and the feedback from facilitators also show that threat-agnosticism requires careful facilitation. Participants may naturally focus on the specific triggering event presented in the scenario, which can narrow the discussion if it is not managed properly. Facilitators, therefore, need to emphasize that the scenario is a vehicle for exploring system vulnerabilities and interdependencies, rather than the only event of interest.

The resilience perspective is reflected in the methodology's focus on early response and early recovery. By separating these phases, STT2 captures how dependencies and priorities change over time. This is methodologically important because critical nodes during the immediate response may differ from those during recovery. Similarly, coordination and information needs may shift as the crisis evolves. The distinction between phases, therefore, supports a more dynamic understanding of system resilience.

A key methodological challenge is maintaining consistency across groups. Even with common templates and instructions, different facilitators and participant groups may interpret dependency strength, infrastructure categories, decision priorities, or coordination needs differently. This does not undermine the methodology, but it highlights the importance of facilitator briefing, mock workshops, clear scoring instructions, and post-workshop quality control.

Evidence of and challenges regarding usability and applicability

The evaluation indicates that STT2 is suitable for stakeholder workshops because it builds on familiar participatory formats, including group discussion, mapping, prioritization, and table-based documentation. Based on the results of the feedback survey, the participants can engage well with the methodology and found it useful.

Evidence from the participant feedback survey provides positive indications regarding the usability and usefulness of the STT2 methodology. The survey responses show that participants were generally able to follow and engage with the workshop process. All respondents indicated that they would participate again in this type of workshop, and the average score for this item was very high. Participants also reported that the workshop was easy to follow and that the different sessions were well integrated. At the same time, most respondents disagreed with statements suggesting that the workshop was unnecessarily complex or excessively demanding. This suggests that, although the methodology requires active engagement and structured thinking, the participatory format was understandable and manageable for participants.

The feedback also indicates that the workshop generated useful insights for participants. The physical interdependency mapping session for the early response phase was most frequently identified as useful, followed by the problem-structuring session and the physical interdependency mapping session for early recovery. The decision-making sessions were also considered useful by several respondents, particularly because they helped participants reflect on information needs, coordination requirements, and decision priorities. Moreover, a majority of respondents indicated that none of the sessions were less useful, suggesting that the overall structure of the workshop was perceived as relevant.

The survey further shows that participants remained actively engaged during the workshop. Respondents reported high levels of participation, focus on critical infrastructures and decisions, and sufficient attention to think about possible actions and adaptations. This is important for STT2 because the quality of the methodology depends on participants being able to contribute their operational and institutional knowledge. The feedback therefore supports the conclusion that the workshop format is suitable for eliciting stakeholder knowledge and translating it into structured outputs for stress-testing analysis. At the same time, the feedback points to several practical considerations for future implementations.

From the modeling perspective, the modeling approach used in STT2 does not require the same level of data, computational resources, etc. as data-intensive approaches such as agent-based models, digital twins, or detailed engineering simulations. Instead, it relies on participatory mapping and simplified cascade analysis to represent how disruptions may propagate across infrastructures and decision-making processes. While the post-workshop analysis still requires methodological and analytical expertise, the overall approach remains less data-intensive and more accessible than advanced simulation-based modeling, and it serves a different purpose compared to those approaches. This is an important strength, particularly in data-scarce contexts or in settings where stakeholders hold practical knowledge that is not captured in formal datasets.

The workshop structure is also useful for creating a shared understanding among stakeholders. By bringing different organizations together, the methodology helps participants identify dependencies that may not be visible from their own organizational perspective. This learning function is an important practical benefit of STT2. The process supports not only data collection but also reflection, dialogue, and mutual awareness among actors who may need to coordinate during HILP events. Some participants across the case studies indicated that the insights generated during this exercise would be brought to their governing bodies and incorporated into their strategic planning processes.

At the same time, the evaluation highlights several usability challenges. The workshop is intensive, especially when implemented in a one-day format. Participants need to

understand the scenario, contribute to problem structuring, select critical infrastructures, assess pairwise dependencies, identify key decisions, and document information and coordination needs. This requires careful time management and strong facilitation.
The physical interdependency mapping can be demanding, particularly when many infrastructures are selected. A group that selects too many nodes may struggle to complete all pairwise assessments within the available time. For this reason, the methodology works best when the number of infrastructures is limited to a manageable set, as defined in the facilitation instructions, while still representing the most relevant system components. The use of the interdependency matrix and a strict row-by-row approach helps prevent missing links, but it requires discipline from facilitators and note takers.
The decision-mapping sessions are generally useful but can also be challenging. Participants may initially propose broad actions rather than specific decisions. Facilitators, therefore, need to help groups formulate decisions clearly and identify the related resources, information needs, information sources, coordination requirements, and reasons for prioritization. This requires more methodological guidance than a standard discussion-based workshop.
Applicability across contexts is a strength of STT2, but it requires adaptation. Different case studies vary in geographical scale, governance arrangements, infrastructure complexity, stakeholder experience, and available data. In smaller or more familiar contexts, stakeholders may be able to identify dependencies relatively easily. In larger or more complex contexts, the system may need to be divided into geographical zones, or the level of detail may need to be adjusted. For example, instead of mapping individual infrastructure assets, the workshop may focus on subsector- or sector-level critical infrastructures. Similarly, decisions may need to be framed at a more strategic level when the workshop is applied to regional or national levels. Applying the methodology at higher governance levels also creates additional practical challenges. For example, in national-level applications, some relevant decisions may be made at ministerial or senior policy levels. In such cases, involving the most appropriate participants can be difficult, either because of availability constraints or because the required decision-making authority is distributed across multiple institutions. This does not limit the overall applicability of STT2, but it shows that careful stakeholder selection, scoping, and adaptation of the level of analysis are necessary for each implementation context.

Evidence of and challenges regarding analytical value and effectiveness

The methodology produces analytical value by generating structured outputs that can be used to identify critical nodes, cascading effects, decision bottlenecks, information dependencies, coordination requirements, and resource conflicts. The physical interdependency maps and matrices provide a basis for analyzing how disruptions may propagate across infrastructure systems. The decision tables provide a basis for analyzing how decisions depend on resources, information, coordination, and organizational responsibilities.
One important analytical strength of STT2 is that it captures both physical and organizational dimensions of interdependency. Traditional infrastructure analysis may identify technical dependencies, such as electricity supporting telecommunications or transport supporting emergency services. STT2 also captures dependencies such as who needs information from whom, which actors must coordinate before a decision can be implemented, and where limited resources may create conflicts between priorities. This broader perspective increases the analytical value of the methodology for resilience assessment.
Another analytical strength is the comparison between early response and early recovery. This comparison can show how vulnerabilities evolve over time. Some infrastructures may

be immediately critical in the response phase, while others become more important during recovery. Similarly, some decisions may be urgent immediately after the event, while others become central once the focus shifts to restoration and continuity. This temporal comparison supports a more nuanced understanding of bottlenecks and resilience capacities.

However, the analytical value of STT2 depends heavily on data quality. If maps are incomplete, if dependency directions are unclear, if matrices do not match the maps, or if decision tables lack information sources or coordination actors, the post-workshop analysis becomes more difficult. This confirms the importance of note taker guidance, real-time quality checks, photographs of outputs, digitization templates, and post-workshop clarification with facilitators and hosts.

A further analytical limitation is that the scoring of dependency strength is based on stakeholder judgment. This is appropriate for a participatory semi-quantitative methodology, especially in data-scarce contexts, but the scores should be interpreted as elicited expert assessments rather than objective measurements. Where possible, the results should be triangulated with documents, operational data, technical models, or validation meetings.

Context-dependent effects relevant for the method

The evaluation shows that the performance of STT2 is influenced by several context-dependent factors.

One important factor is **stakeholder composition**. Groups with diverse sectoral representation tend to identify a wider range of interdependencies, while groups dominated by one sector may produce narrower outputs. The presence of both operational and strategic actors can also enrich the discussion, as operational actors often understand practical dependencies, while strategic actors understand governance and decision-making constraints.

A second factor is **prior collaboration among participants**. In contexts where organizations already know each other and have experience working together, discussions may move more quickly, and coordination dependencies may be easier to identify. In contexts where stakeholders are less familiar with each other, the workshop may first serve as a trust-building and sense-making exercise before detailed interdependencies can be mapped.

A third factor is **geographical and system scale**. Urban-scale applications may allow more concrete mapping of specific infrastructures, locations, and operational dependencies. Regional, national, or cross-border applications may require a higher level of abstraction because the number of relevant infrastructures, actors, decision levels, and governance arrangements increases. The methodology can be applied at these scales, but scoping becomes more important.

A fourth factor is **language and terminology**. Based on the implementation experience so far, holding the workshop in the local language is encouraged where possible. Even when participants are fluent in English, discussions often flow more smoothly, faster, and more naturally when they can use their usual working language. This reduces communication pressure and allows participants to explain operational details, local practices, and institutional arrangements more easily and comfortably. However, when workshops are conducted in local languages, facilitators and note takers play a crucial role in ensuring that outputs are digitized and translated accurately after the workshop. Differences in terminology for infrastructures, organizations, decisions, or crisis-management structures can affect the analysis if they are not standardized during the post-workshop processing.

A fifth factor is the **sensitivity of information**. In some contexts, participants may be reluctant to discuss vulnerabilities, dependencies, or coordination weaknesses openly. This may limit the level of detail in the outputs. Clear communication about confidentiality,

anonymization, and the purpose of the workshop is therefore important for encouraging meaningful participation.

5.4. Improvements

Adjustments already implemented

Several adjustments have already been implemented to improve the clarity, usability, and analytical quality of STT2.

First, the **role of facilitators has been clarified in greater detail**. The facilitator guidance now provides more explicit instructions on how to introduce each task, manage time, support balanced participation, and guide groups through the mapping exercises. This adjustment was needed because the quality of participatory outputs depends strongly on facilitation. Clearer facilitator instructions make the process more consistent across groups and reduce the risk that the exercise becomes an unstructured discussion.

Second, the **role of note takers has been strengthened**. The note taker guidance now emphasizes three core functions: capturing key insights, supporting the process, and ensuring data quality. Note takers are instructed to document not only the final outputs but also the reasoning behind them, including disagreements, bottlenecks, critical nodes, cascading pathways, information needs, and coordination challenges. This adjustment was introduced because stronger note-taking improves the analytical value of the workshop data and helps preserve the reasoning behind the recorded outputs.

Third, the **list of critical infrastructures** used in the physical interdependency mapping sessions **has been adapted to each workshop context**. This ensures that the terminology used is familiar to participants and that the list reflects the local system under analysis. Critical infrastructure that the host and their expert team identify as not applicable to the case study context is removed in advance. This makes the selection process easier for participants and helps them focus on the infrastructures most relevant to their own context.

Fourth, the **decision-mapping templates have been refined**. The decision table now asks explicitly for the responsible decision-maker, rather than for responsible organizations and locations. This adjustment was introduced to reduce confusion and to help facilitators guide participants more easily through the exercise.

Fifth, the **resource-allocation exercise using limited resource cards has been removed** from the decision-mapping sessions. In the initial design, this exercise was included at the end of the decision-mapping sessions. Each group received a set of limited resource cards and was asked to allocate them to the priority decisions identified during the session. The purpose was to introduce resource scarcity and observe whether decision priorities changed under constrained conditions. However, after testing this in the first case-study implementation, it became clear that the exercise was difficult to implement within the available workshop time. In addition, designing meaningful resource cards required substantial case-specific preparation, as the resources had to be adapted to the local context and scenario. Considering the time required and the relatively limited added analytical value of the output, the exercise was not continued as a standard component of the methodology.

Six, the **importance of a mock workshop or dry run has been added as a practical requirement**. The host is expected to organize a rehearsal with facilitators and note takers before the actual workshop. This adjustment was introduced because facilitators and note takers need to be familiar with the session flow, templates, timing, and expected outputs before working with participants. Although some facilitators may have experience with emergency-management training exercises, the STT2 workshop is different. While it is inspired by practice, it requires a structured approach to data collection using specific materials, templates, and analytical categories. The mock workshop helps identify missing materials, unclear instructions, timing problems, and data-quality risks before the actual event.

Seventh, **post-workshop digitization and translation have been clarified as responsibilities of facilitators and note takers**, with guidance from the method developer. This adjustment was introduced because raw workshop outputs are often produced in local languages and physical formats. Digitization and translation are necessary to make the data usable for analysis and comparison across groups and case studies. Clarifying this responsibility improves the continuity between workshop implementation and post-workshop analysis. This process also helps ensure consistent terminology among groups, while accurately preserving local terms, contextual meanings, and case-specific details. Clarifying this responsibility improves the continuity between workshop implementation and post-workshop analysis.

Planned refinements

At the current stage, **no major structural refinements to the STT2 methodology are planned**. The methodology has already gone through several iterative design, implementation, evaluation, and refinement cycles. These cycles allowed the method development team to identify practical challenges, adjust the workshop structure, improve the facilitation and note-taking guidance, refine the templates, clarify the responsibilities of the host and method developer, and strengthen the post-workshop data handling process.

The refinements already implemented have been tested in subsequent workshop applications and have led to positive implementation outcomes. In particular, the clearer facilitator instructions, strengthened note-taker role, adapted critical infrastructure lists, revised decision-mapping table, mock workshop requirement, and clarification of digitization and translation responsibilities have improved the usability, consistency, and analytical quality of the methodology. Based on the feedback received from hosts, facilitators, note takers, stakeholders, and participants, these adjustments have addressed the main implementation challenges identified so far.

Future refinements are expected to be minor and context-specific rather than fundamental. These may include adapting the agenda to local time constraints (e.g., usual time for lunch break), tailoring the critical infrastructure list to the case study context, translating materials into the local language, adjusting the level of detail to the geographical scale of the workshop, and refining examples used in facilitator briefings. Such adaptations are part of the normal implementation process and do not require changes to the core methodological design.

Future applications will continue to be monitored to ensure that the methodology remains applicable across different contexts, scales, stakeholder settings, and HILP scenarios. If new challenges emerge in later implementations, they will be documented. However, at this stage, the evaluation indicates that the existing refinement cycles have produced a workable and robust version of the methodology. Therefore, the focus of future work will be on consistent application, quality assurance, and learning across case studies, rather than on further redesign of the STT2 methodology.

6. Conclusions

The STT2 methodology provides a structured participatory approach for stress testing interdependent critical infrastructure systems under HILP conditions. It contributes to the AGILE stress-testing framework by bridging qualitative system understanding and more advanced analytical or modeling approaches. Its main strength lies in its ability to capture physical, informational, organizational, and decision-related interdependencies through stakeholder participation.

The methodology has several important strengths. It supports a system-level perspective by moving beyond isolated infrastructure sectors and encouraging stakeholders to map the critical parts of the system and examine how disruptions may propagate across this

interconnected system. It operationalizes a threat-agnostic logic by focusing on system behavior under stress rather than on the probability of a specific hazard. It adopts a resilience perspective by examining early response and early recovery, thereby capturing how vulnerabilities and dependencies evolve over time. It is also practical in data-scarce contexts because it does not require extensive quantitative datasets as a prerequisite for implementation. Another strength is its learning value. The workshop format creates a space for stakeholders to share perspectives, identify hidden dependencies, discuss coordination challenges, and develop a common understanding of systemic vulnerabilities. This makes STT2 not only an analytical methodology but also a process for stakeholder learning and reflection.

At the same time, the methodology has limitations. It relies on diverse stakeholder participation, facilitator quality, and careful post-workshop processing. It cannot cover all infrastructures, actors, decisions, or dependencies in complex systems within a limited workshop format. Its outputs should therefore be interpreted as structured stakeholder-based evidence, not as exhaustive representations of the system.

Replication of STT2 requires several conditions. A suitable host organization must be able to mobilize relevant stakeholders, provide contextual knowledge, arrange the workshop logistics, assign facilitators and note takers, organize a mock workshop, and support digitization and translation of outputs. The method developer must provide the methodological structure, templates, facilitation guidance, note-taker guidance, analytical procedures guide, and quality assurance. A clear scope, a well-designed HILP scenario, appropriate materials, and sufficient time for preparation and analysis are also necessary.

Future implementations will further test and refine the methodology across different AGILE case-study contexts. These applications will provide additional evidence on methodological soundness, usability, applicability, analytical value, and transferability. Through this iterative process, STT2 will continue to evolve as a practical and scientifically grounded methodology for identifying vulnerabilities, cascading effects, and coordination bottlenecks in complex socio-technical systems exposed to HILP events.

References

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., & Von Winterfeldt, D. (2003). A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities. *Earthquake Spectra*, *19*(4), 733–752. https://doi.org/10.1193/1.1623497

Cheshomi, A., Aydin, N. Y., & Comes, T. (2026). A scoping review and stress testing framework for High Impact Low Probability (HILP) events. *International Journal of Disaster Risk Reduction*, *142*, 106194. https://doi.org/10.1016/j.ijdrr.2026.106194

Connelly, E. B., Allen, C. R., Hatfield, K., Palma-Oliveira, J. M., Woods, D. D., & Linkov, I. (2017). Features of resilience. *Environment Systems and Decisions*, *37*(1), 46–50. https://doi.org/10.1007/s10669-017-9634-9

Nowell, B., Steelman, T., Velez, A.-L. K., & Yang, Z. (2018). The structure of effective governance of disaster response networks: Insights from the field. The American Review of Public Administration, 48(7), 699-715.

Appendix

Post workshop participants feedback survey

1. Validation Survey – Tier 2 Stress Test Workshop

Thank you for participating in the Tier 2 Stress Test Workshop.

This survey is designed using validated academic instruments to evaluate the workshop's usability, usefulness, and situational awareness. Your feedback will help us improve future workshops.

Please answer honestly. Your responses will remain anonymous.

2. Background & Preparation

1. What is your sector/role? (e.g., Energy, Water, Transport, Emergency Services, Government, Academia, Other): .....

2. How clear were the objectives of the workshop to you?

☐ 1 – Not clear at all ☐ 2 ☐ 3– Somewhat ☐ 4 ☐ 5 – Very clear

3. How well were you informed about what to expect from the workshop?

☐ 1 – Not at all ☐ 2 ☐ 3– Somewhat ☐ 4 ☐ 5 – Very well

4. How well were you informed about practical details (e.g. agenda, location, registration and consent form, etc.)?

☐ 1 – Not at all ☐ 2 ☐ 3 – Somewhat ☐ 4 ☐ 5 – Very well

3. Prerequisites

Below are some statements regarding your previous knowledge and experiences. Please mark what characterizes you best:

1. In your daily work, how much are you involved in managing or preparing for disruptions in critical infrastructures (e.g., disaster management, emergency planning, or decision-making under crisis conditions)?

☐ 1 – Not at all ☐ 2 – Limited ☐ 3 – Fairly ☐ 4 – Strongly ☐ 5 – Very strongly

2. In your daily work, how much do interdependencies between infrastructures (e.g., power, water, transport, ICT) influence your activities or decisions?

☐ 1 – Not at all ☐ 2 – Limited ☐ 3 – Fairly ☐ 4 – Strongly ☐ 5 – Very strongly

3. In your daily work, how often do you collaborate with colleagues from other organizations or sectors (e.g., utilities, emergency services, government) to address complex problems?

☐ 1 – Not at all ☐ 2 – Rarely ☐ 3 – Sometimes ☐ 4 – Often ☐ 5 – Very often

4. Usability & Usefulness

For the following statements, please indicate your level of agreement on a 5-point scale:
1 = Strongly disagree … 5 = Strongly agree

5. Part A - System Usability Scale (SUS)

1. I think I would like to participate in this type of workshop again.

☐ 1 – Strongly disagree ☐ 2 ☐ 3 ☐ 4 ☐ 5 – Strongly agree

2. I found the workshop unnecessarily complex.

☐ 1 – Strongly disagree ☐ 2 ☐ 3 ☐ 4 ☐ 5 – Strongly agree

3. I think the workshop was easy to follow.

☐ 1 – Strongly disagree ☐ 2 ☐ 3 ☐ 4 ☐ 5 – Strongly agree

4. I think I would need support to fully benefit from the workshop.

☐ 1 – Strongly disagree ☐ 2 ☐ 3 ☐ 4 ☐ 5 – Strongly agree

5. I found the different sessions well integrated.

☐ 1 – Strongly disagree ☐ 2 ☐ 3 ☐ 4 ☐ 5 – Strongly agree

6. I thought there was too much inconsistency in the workshop.

☐ 1 – Strongly disagree ☐ 2 ☐ 3 ☐ 4 ☐ 5 – Strongly agree

7. I imagine most people would learn to benefit quickly from this workshop.

☐ 1 – Strongly disagree ☐ 2 ☐ 3 ☐ 4 ☐ 5 – Strongly agree

8. I found the workshop overly demanding.

☐ 1 – Strongly disagree ☐ 2 ☐ 3 ☐ 4 ☐ 5 – Strongly agree

9. I felt very confident participating in the workshop activities.

☐ 1 – Strongly disagree ☐ 2 ☐ 3 ☐ 4 ☐ 5 – Strongly agree

10. I needed to learn a lot before I could get the most out of the workshop.

☐ 1 – Strongly disagree ☐ 2 ☐ 3 ☐ 4 ☐ 5 – Strongly agree

6. Part B – Workshop-Specific Add-Ons

1. Which part(s) of the workshop did you find most useful?

(You can select one or more)

☐ Session 1: Problem structuring (objectives, concerns and action points)

☐ Session 2: Mapping of physical interdependencies in early response

☐ Session 3: Decision-making exercise in early response

☐ Session 4: Mapping of physical interdependencies in early recovery

☐ Session 5: Decision-making exercise in early recovery

2. Which part(s) of the workshop did you find least useful or feel could be improved?

(You can select one or more)

☐ Session 1: Problem structuring (objectives, concerns and action points)

☐ Session 2: Mapping of physical interdependencies in early response

☐ Session 3: Decision-making exercise in early response

☐ Session 4: Mapping of physical interdependencies in early recovery

☐ Session 5: Decision-making exercise in early recovery

☐ None of the sessions

7. Situational Awareness

Please rate each statement from 1 (Low) to 7 (High).

1. How unstable was the situation during the scenario (e.g., rapid escalation, sudden cascading failures)?

☐ 1 – Very stable / predictable ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 – Very unstable / highly unpredictable

2. How complex was the scenario in terms of interdependencies between critical infrastructures and organizations and actors?

☐ 1 – Simple / few interdependencies ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 – Very complex / many interdependencies

3. How many different factors were changing at the same time (e.g., infrastructure status, resource needs, coordination requirements)?

☐ 1 – Very few changes ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 – Many changes happening simultaneously

4. During the workshop, to what extent did you participate actively and stay engaged?

☐ 1 – Very passive / low energy ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 – Very active / highly engaged

5. How well were you able to concentrate on the most critical infrastructures and decisions, and complex interdependencies

☐ 1 – Very distracted / lost priorities ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 – Fully focused on what mattered most

6. During each session, was your attention mostly on one main issue, or was it spread across many different issues at the same time?

☐ 1 – Mostly focused on one issue ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 – Spread across many issues

7. During the workshop, did you feel you had enough attention left to think about possible cascading effects or new disruptions, or were you fully occupied with the immediate situation?

☐ 1 – Fully occupied with the immediate situation, nothing left ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 – Plenty of attention left to think ahead and adapt

8. To what extent did you feel you had enough information to understand interdependencies between infrastructures and make decisions during the HILP event scenario?

☐ 1 – Very little / incomplete information ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 – Sufficient and clear information

9. How familiar was this HILP event scenario compared to situations you have experienced in your professional work (e.g., smaller disruptions, drills, or past emergencies)?

☐ 1 – Completely new / unfamiliar ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 – Very familiar / similar to real challenges

8. Open Feedback

1. Do you have any suggestions for improving aspects such as the preparation, the scenario, the exercises, or the facilitation?