Planning for the Unknown:

Using Exploratory Modelling and Analysis to Create Dynamic Adaptive Policy Pathways in Infrastructure Construction Scheduling under Deep Uncertainty

Daan Nienhuis

July 7, 2025

Image source: Rijkswaterstaat, 2024

First Supervisor:	Prof.dr.ir Alexander Verbraeck
Second Supervisor:	Prof.dr.ir Lóránt Tavasszy
Advisor:	dr.ir. Patrick Steinmann
Company supervisor:	drs. Dirk van Uffelen
External supervisor:	dr.ir Ruud Binnekamp
Project Duration:	February 2025 – July 2025
Faculty:	Faculty of Technology, Policy and Management

Abstract

Megaprojects frequently face cost overruns and schedule delays, a pattern known as the *iron law of megaprojects*, which persists partly because traditional deterministic or probabilistic planning methods are inadequate for managing deep uncertainty. This form of uncertainty arises when the probability, timing, or impact of key events cannot be reliably estimated, often leading to unrealistic schedules and ineffective risk responses. This study investigates the question: *How can Exploratory Modelling and Analysis and Dynamic Adaptive Policy Pathways be applied to improve schedule robustness in infrastructure construction projects?* The research focuses on the Schiphol bridge reconstruction, which is part of the Veenix A9 BaHo project and is conducted in collaboration with Count & Cooper. A Discrete Event Simulation (DES) model is built in SimPy and structured using a task dependency graph derived from the project's original schedule via NetworkX. The model is sampled 10,000 times under baseline conditions using Latin Hypercube Sampling. In scenario discovery, Patient Rule Induction Method (PRIM) is used in combination with a scaling function identifying six high-impact scenarios, which serve both as inputs for robust policy search and as Adaptation Tipping Points (ATP) for the Dynamic Adaptive Policy Pathways (DAPP) schedule. Robust mitigation strategies are derived using a Multi-Objective Evolutionary Algorithm under the Multi-Objective Robust Decision Making framework, with a second PRIM experiment selecting four final robust policies. These policies correspond to at least one of the high-impact scenarios and form the backbone of a conditional DAPP schedule. The DAPP schedule is evaluated against a static baseline using 5,000 DES simulations with identical uncertainty sampling. In 20 comparative runs, it reduced project duration by an average of 67 days and cost by approximately €97.5 million on the entire project schedule. All three robust policies include the measures *new design*, *overtime labour*, and *electric machinery*, suggesting that a focused subset of actions can improve resilience even when the future is highly uncertain. Unlike prior Decision Making under Deep Uncertainty (DMDU) applications, which often focus on long-term or high-level strategic planning, this study embeds adaptive logic within a highly granular, task-level construction schedule based on real project data. This approach raises methodological challenges in how adaptation tipping points are defined, triggered, and monitored within network-based simulation. The findings demonstrate not only the feasibility of combining Exploratory Modelling and Analysis (EMA) and DAPP in operational construction settings but also the need for further research into real-time scenario recognition and policy switching mechanisms under uncertainty.

Contents

List of Acronyms

Glossary of Terms

Adaptation Tipping Point A condition under which a predefined pathway no longer meets its objectives, triggering a switch to an alternative policy. In this study, one ATP refers to the set of risk and uncertainty values connected to one high-impact scenario. i

Decision Making under Deep Uncertainty Decision making under deep uncertainty is a decision science practice and analytical framework that evaluates potential solutions across multiple plausible future scenarios rather than attempting to predict a single future outcome. i

Discrete Event Simulation A simulation technique where the operation of a system is represented as a chronological sequence of events. i

Dynamic Adaptive Policy Pathways A planning approach that sequences policy actions over time and incorporates switching logic based on adaptation tipping points. i

Exploratory Modelling and Analysis A modelling approach that explores a wide range of plausible futures to support decision-making under deep uncertainty. i

high-impact scenario A subset of simulated futures in which the project experiences severe outcomes on cost and duration. These are used as adaptation tipping points in the DAPP schedule. 31, 49

Latin Hypercube Sampling A statistical method used to efficiently sample multidimensional spaces for simulation experiments. i

measure An individual intervention or modification to mitigate the consequences of risk and uncertainty. 23, 45, 70, 76

Multi-Objective Evolutionary Algorithm An optimisation algorithm that evolves a population of solutions toward a Pareto front under multiple objectives. i

Multi-Objective Robust Decision Making An extension of EMA that uses multi-objective evolutionary algorithms to identify robust policies across many scenarios. i

Patient Rule Induction Method A greedy algorithm that finds a minimum spanning tree for a weighted undirected graph. In this study, PRIM identifies regions of the uncertainty space associated with poor performance or policy failure. i

policy A policy is a predefined combination of measures intended to mitigate the impact of evolving scenarios under uncertainty. 9

risk A discrete event that may or may not occur, often deeply uncertain due to unknown or contested odds of occurrence. 43

robust policy A set of measures with Pareto-optimal performance across all identified high-impact scenarios. 32, 54

scenario A coherent representation of a possible future state of the world, used to explore the effects of uncertainty. 2

uncertainty A continuous variable with unknown or contested distribution that influences project outcomes over time. 43

1 Introduction

1.1 Problem Statement

Over budget, over time, over and over again is the iron law of megaprojects (Flyvbjerg, 2014). Revealing more than 70 years of persistent budget and schedule overruns across different sectors and regions (Flyvbjerg et al., 2002, 2005). Infrastructure is one of the key domains in which these large-scale, complex undertakings are especially common (Chen et al., 2022). Within this sector, Larsson et al. (2016) and Ahmad et al. (2020) identify a variety of factors that influence cost, schedule, and quality in construction projects. Among these, Flyvbjerg (2014) argues that delays are particularly consequential, as they tend to escalate costs and reduce overall project benefits. As infrastructure projects become more complex and interdependent, schedules are exposed to increasingly dynamic risk interactions (Wang and Yuan, 2017), making timely project delivery a growing concern—not only for project teams, but also for the public systems, services, and economic activities that depend on them.

One key reason for persistent overruns is the way uncertainty is understood and managed (Aven, 2016). A risk can be distinguished from an uncertainty. According to Knight (1921), risk can be quantified while uncertainty arises from a lack of knowledge. This distinction explains why conventional risk management often fails in large construction projects, where uncertainty is highly nuanced (Aven, 2012, 2016). Aleatoric uncertainty is consistent with Knightian risk because it assumes that probabilities can be assigned, whereas Knightian uncertainty is consistent with epistemic uncertainty, where probabilities are unknown or cannot be defined (Aven, 2013). Traditional risk management typically focuses on aleatoric uncertainty and overlooks the epistemic type that is common in megaprojects (Fang et al., 2013). Furthermore, unpredictable events can appear more predictable in hindsight (Taleb, 2020). This can lead project planners to treat genuine uncertainty as if it were manageable risk, especially in megaprojects with unique features (Sadeghi et al., 2015). Relying on rare past events to guide forecasts can lead to unjustified confidence in predictive models, resulting in further schedule and cost overruns.

Maier et al. (2016) describe three ways of modelling the future, highlighting how the aforementioned misconceptions about risk and uncertainty are embedded. The first approach treats the future as deterministic, refining models only as new data emerge (Bankes, 1993). Flyvbjerg (2014) shows that such static forecasting often leads to overruns because it focuses on a single *most likely* scenario. The second approach accepts that the future is quantitatively uncertain and uses probability distributions to model unknowns (Aven, 2012). However, deep uncertainty arises when these probabilities themselves are unclear or impossible to define (Lempert et al., 2003). Popular techniques such as Monte Carlo simulation address aleatoric uncertainty, but fail when key probabilities cannot be fixed (Tegeltija et al., 2016; Feng et al., 2022). The third approach explores multiple possible futures, recognising that epistemic uncertainty can become so severe that it becomes deep uncertainty (Aven, 2013). Under these conditions, standard risk management falls short, highlighting the need for methods designed to deal with unpredictable and unquantifiable outcomes.

Deep uncertainty is especially relevant in the early stages of construction projects, where decisions made with incomplete knowledge have long-term consequences. Lau et al. (2018) observe that uncertainty is most prominent during planning. Williams and Samset (2010) and Mohd Nasir et al. (2016) build on this by emphasising the outsized impact of early decisions on final outcomes. Stanton and Roelich (2021) argue for flexibility and adaptability as key responses to this challenge. Despite these calls, current scheduling practices often fail to incorporate deep uncertainty in a meaningful way, increasing the risk of disruptions to infrastructure delivery with implications for both usage and public expenditure.

This gap highlights the need for new planning methods capable of supporting decisions under conditions of uncertainty that cannot be quantified. Under the umbrella of Decision Making under Deep Uncertainty (DMDU), Exploratory Modelling and Analysis (EMA) and Dynamic Adaptive Policy Pathways (DAPP) have emerged as promising alternatives. EMA enables the exploration of diverse futures by testing a large ensemble of models and conditions rather than relying on a single validated forecast (Bankes et al., 2001, 2013). DAPP offers a structured way to embed flexibility into plans, using empirically defined tipping points to guide policy adaptation (Marchau et al., 2019a).

These methods have shown success in long-term strategic water management (Haasnoot et al., 2013). Michas et al. (2020) used DAPP for solar photovoltaic planning, suggesting broader relevance. Feng et al. (2022) propose a data-driven approach for construction projects under deep uncertainty, acknowledging the relevance in the construction domain, but emphasise the need for more tailored methods. While such studies mark an important shift in recognising the limitations of probabilistic forecasting, the application of decision-making frameworks like EMA and DAPP remains largely unexplored in this context.

This study investigates whether EMA and DAPP can improve schedule robustness in infrastructure construction projects facing deep uncertainty. While the relevance of deep uncertainty in construction has been increasingly acknowledged, the application of structured DMDU methods such as EMA and DAPP remains limited. By integrating scenario discovery, robust policy evaluation, and adaptive scheduling into a single modelling framework, this research provides a structured alternative to static risk-based planning. In doing so, it contributes to both the theoretical understanding of deep uncertainty in project management and the practical development of more resilient infrastructure scheduling tools under deep uncertainty, ultimately taking on the long-standing *iron law of megaprojects*.

This study aims to contribute by:

- Creating a clear argument based on scientific literature as to why construction schedulers should consider deep uncertainty in the planning phase of large infrastructure projects.

- Introducing the Exploratory Modelling and Analysis framework to the infrastructure construction sector.

- Providing a novel approach to exploring high-impact scenarios under deep uncertainty in construction scheduling that goes beyond traditional risk analysis.

- Demonstrating how robust mitigation strategies can be identified and, when combined with the identified scenarios, serve as the foundation for a DAPP-based infrastructure construction schedule.

- Providing a quantitative method to compare a DAPP-based construction schedule to a traditional deterministic construction schedule on multiple project objectives.

1.2 Research Questions

As the literature reveals, infrastructure construction projects face significant challenges when confronted with deep uncertainty. Conventional planning methods frequently depend on fixed assumptions and probabilistic risk modelling, which may prove inadequate when future conditions cannot be reliably predicted or agreed upon. In response, a new generation of methods has emerged. These include EMA and DAPP, which aim to better account for a wide range of plausible futures and preserve flexibility over time. While these methodologies have been successfully applied in fields such as climate policy and water management, their integration into infrastructure construction scheduling remains limited. This gap is particularly relevant given the scale, complexity, and public importance of infrastructure delivery. This study explores the potential of EMA and DAPP to enhance schedule robustness in infrastructure construction, providing guidance on decision-making under uncertainty and informing adaptive strategies.

1.2.1 Main Research Question

How can Exploratory Modelling and Analysis and Dynamic Adaptive Policy Pathways be applied to improve schedule robustness in infrastructure construction projects under deep uncertainty?

1.2.2 Sub-Questions

The four sub-questions establish the theoretical and methodological foundation needed to answer the primary research question and bridge existing knowledge gaps.

sub-question 1: Deep Uncertainty

While the concept of deep uncertainty is discussed extensively in the literature, there remains debate about which forms are most prevalent in infrastructure construction. Projects are exposed to both discrete risk events and continuous stressors, but their manifestation and impact on scheduling differ significantly. This raises the question:

Which deep uncertainties commonly appear in infrastructure construction?

Sub-Question 2: Scenario Discovery

Given the range of risks and uncertainties that can affect construction schedules, identifying which combinations lead to failure is critical. Scenario discovery techniques are designed to uncover these high-impact conditions, but their application in construction planning remains limited. This prompts the question:

How can scenario discovery be used to find high-impact scenarios and adaptation tipping points in construction scheduling under deep uncertainty?

Sub-Question 3: Directed Search

Once high-impact scenarios are identified, the next challenge is to design mitigation strategies that remain effective across these deeply uncertain futures. Traditional optimisation approaches focus on finding the best solution for a specific future, but in uncertain environments, robust strategies are preferred because they perform reliably across a range of multiple futures. This leads to the question:

Which robust mitigation strategies can be identified that address high-impact scenarios affecting cost and duration?

Sub-Question 4: Pathway Development

Even when robust strategies are available, their effectiveness depends on the ability to switch between them as project conditions evolve. DAPP addressed this, however, it remains unclear how such pathways can be practically constructed in the context of infrastructure scheduling. This introduces sub-question 4:

How can dynamic adaptive policy pathways be constructed using adaptation tipping points and mitigation strategies for infrastructure construction scheduling?

2 Related work

2.1 Definition of Deep Uncertainty

The concept of deep uncertainty is increasingly recognised in fields where decision-making must occur under conditions of limited knowledge. In project management, the classical definition of risk, which describes it as "an uncertain event or condition whose occurrence affects at least one of the project objectives, such as scope, schedule, cost, or quality" (Project Management Institute, 2021), fails to clearly distinguish between quantifiable risks and broader categories of uncertainty. This interchanged use has long been criticised, dating back to Knight (1921), and is particularly problematic in complex domains such as infrastructure planning. Uncertainty can take different forms: *Aleatoric uncertainty* arises from natural variability—such as minor differences in task durations—and is typically captured using probabilistic techniques such as Monte Carlo simulation (Fox and Ülkümen, 2011). In contrast, *epistemic uncertainty* reflects a lack of knowledge—such as unknown subsurface conditions—and aligns more closely with Bayesian reasoning. While aleatoric uncertainty remains, even when a system is well understood, epistemic uncertainty may be reduced through further investigation (McCann, 2020; Chung et al., 2004). In reality, both types of uncertainty often coexist, and researchers have proposed hybrid models to accommodate both perspectives (Ahmadu et al., 2020; Okmen and Oztas, 2014; Sadeghi et al., 2010; Zamani et al., 2024). Yet even such models often assume an underlying probabilistic structure. When the future cannot be meaningfully represented with probabilities, or when stakeholders cannot agree on models or value judgments, the situation is said to involve *deep uncertainty* (Lempert et al., 2003; Aven, 2013). Lempert et al. (2003) defines deep uncertainty as:

> the condition in which analysts do not know or the parties to a decision cannot agree upon (1) the appropriate models to describe interactions among a system's variables, (2) the probability distributions to represent uncertainty about key parameters in the models, and/or (3) how to value the desirability of alternative outcomes.

Building on this, Walker et al. (2013) proposed five levels of uncertainty to distinguish ordinary uncertainty from deep uncertainty. Levels 1 to 3 describe situations where models, probabilities, or rankings are still applicable. Levels 4 and 5 describe cases where futures are either unranked or entirely unknown, thus exceeding the capacity of conventional probabilistic tools.

Table 2.1: Five levels of uncertainty

Level	Description	Classification
1	A clear future with sensitivity: a single system model having point estimation and sensitivity outcomes	Ordinary uncertainty
2	Alternate futures with probabilities: a single system model having a probabilistic parametrisation	Ordinary uncertainty
3	Alternate futures with rankings: several system models having point estimates ranked by perceived likelihood	Ordinary uncertainty
4	Multiple of plausible futures without rankings: several system models without ranking outcomes	Deep uncertainty
5	Unknown future: unknown system and unknown outcomes	Deep uncertainty

Source: Data from Walker et al. (2013)

This study adopts the classification of uncertainty as a foundation for distinguishing between ordinary and deep uncertainty in construction projects. By focusing on conditions where model structures or probability distributions cannot be reliably defined, it positions deep uncertainty as a central consideration in schedule design and evaluation. One example in this study is the inclusion of near-critical tasks with a small slack margin from the critical path, to reflect how future uncertainties may shift project bottlenecks in unforeseen ways.

2.2 Risk and Uncertainty in Relation to Current Modelling Solutions

Traditional stochastic methods such as Monte Carlo simulation remain widely used in construction scheduling under uncertainty. While useful for capturing natural variability, these methods often assume fixed distributions and do not reflect how disruptions evolve during project execution. Moret and Einstein (2016) designed a model that quantified risk events into one simulation framework but did not account for deep uncertainty in its uncertainty model. Furthermore, Event Chain Methodology (ECM) provides another realistic alternative by explicitly modelling the timing and impact of discrete risk events on continuous project parameters (Virine and Trumper, 2013). In ECM, tasks can transition into *excited* states when triggered by events, temporarily altering their duration, cost, or resource demands. This temporal sensitivity improves the simulation of knock-on effects, but ECM still operates within a probabilistic framework and does not fully capture the structural ambiguity associated with deep uncertainty.

Feng et al. (2022) presented a data-driven approach to support decision making in construction planning under deep uncertainty. Their method modified existing probability distributions using a scaling coefficient to generate a set of plausible outcomes. This technique mirrors the use of fuzzy logic to introduce epistemic uncertainty into a frequentist modelling context. By combining Discrete Event Simulation (DES) with Percent Deviation Index (PDI), they constructed robust construction schedules capable of performing across a range of uncertain futures. To further strengthen their approach, Feng et al. (2022) incorporated Patient Rule Induction Method (PRIM) to identify conditions under which even robust strategies would fail. These zones of fragility provided project planners with actionable early warning signs, enabling them to anticipate and mitigate failures before they occur. Their findings demonstrate that while deep uncertainty makes precise prediction impossible, identifying vulnerable system states can still inform meaningful planning decisions. The structure of the research by Feng et al. (2022) is similar to that of this study, particularly in its integration of simulation, robustness analysis and scenario discovery.

Within the broader EMA framework, two main stages are commonly distinguished: an initial open exploration of the uncertainty space, and a subsequent directed search for optimised strategies. These stages are typically connected through the use of scenario discovery and multi-objective optimisation. However, the precise sequence in which these steps are applied is not fixed. Some studies prioritise the selection of robust policies first, and then apply scenario discovery to uncover the specific conditions under which these strategies fail. This sequencing ensures that scenario discovery remains directly relevant to the solutions being considered, which can save time and enhance communication with decision makers (Feng et al., 2022; Giudici et al., 2020). Alternatively, Lempert et al. (2003) and Kwakkel and Pruyt (2013) have demonstrated the benefits of performing scenario discovery at the outset, using a baseline policy to identify the most critical uncertainties that threaten project success. This ordering helps clarify which vulnerabilities matter most before resources are allocated to strategy development, and places greater emphasis on understanding the structure of uncertainty itself.

Various techniques are available for the directed search phase, depending on the structure of the policy space. Multi-Objective Robust Decision Making (MORDM) and its recent extension, Multi-Scenario Multi-Objective Problem (MSMOP) are commonly used to search for policies that balance performance and robustness across many futures (Shavazipour et al., 2021; Kasprzyk et al., 2013). While MORDM can accommodate mixed-variable decision spaces, the MSMOP extension is particularly effective in continuous domains. In scenarios involving both binary and continuous decision levers—such as those found in infrastructure planning—lighter tools like PRIM can be used to uncover structural patterns in robust policy performance (Bryant and Lempert, 2010). While PRIM is often associated with scenario discovery, its logic can also be repurposed to support policy evaluation by identifying consistent input–output relationships within specific scenario clusters. This study opts to use PRIM in the solution space, following an experimental setup for the constraints on project outcomes.

While prior studies have developed various methods to incorporate uncertainty into construction scheduling—such as stochastic simulation, fuzzy logic, or event-based models—many continue to rely on fixed probabilistic assumptions. This study takes a different approach by using the EMA framework to explore structural scenario diversity rather than estimating predefined distributions. PRIM is applied not only for scenario discovery but also to inform the design of policies that perform well across a wide range of conditions. This helps clarify how deep uncertainty can be addressed in infrastructure scheduling beyond traditional modelling techniques, by integrating EMA, MORDM, and DAPP into a single simulation framework. The combined approach offers a novel contribution by enabling dynamic adaptation and robust policy design under uncertainty in infrastructure construction.

2.3 Exploratory Modelling and Analysis in Relation to Dynamic Adaptive Policy Pathways

To navigate deep uncertainty, EMA offers a fundamentally different perspective. EMA is based on the principle that a single computational experiment or model cannot capture all possible futures; instead, it samples across many *a priori* models, each representing a conditional future state (Bankes et al., 2001, 2013). EMA is particularly effective in exploring a wide range of plausible scenarios and model structures, making it particularly suited for situations involving deep uncertainty. By not committing to a single predictive model, EMA allows for the integration of various analytical approaches, including Bayesian methods that incorporate prior knowledge and Frequentist methods that rely on observed data frequencies. This exploratory approach differs from point-estimate approaches by focusing on multiple plausible trajectories rather than relying on one overarching model. As Kwakkel and Pruyt (2013) emphasise, EMA's strength lies in enabling decision makers to develop adaptive plans rather than seeking to optimise a single outcome—an especially useful approach in construction, where project conditions can shift dramatically over time. While uncertainty can influence every stage of a construction project, the early planning phase is particularly critical, as choices made here significantly affect final performance outcomes (Lau et al., 2018; Williams and Samset, 2010; Mohd Nasir et al., 2016).

DAPP combines Dynamic Adaptation Planning (DAP) with Adaptation Pathways (AP) to account for evolving conditions and irresolvable uncertainties during a project's life cycle (Marchau et al., 2019a). Although this approach might seem to conflict with the need for a thorough front-end plan, as noted by Williams and Samset (2010), it actually complements it by preventing lock-in to any single *best* solution and keeping the schedule flexible. This is achieved through ongoing monitoring of Adaptation Tipping Points (ATPs) —key indicators that trigger a shift to alternate pathways when thresholds are exceeded (Kwadijk et al., 2010). Haasnoot et al. (2013) applied DAPP to the long-term water management of the Rhine Delta in the Netherlands. More recently, Michas et al. (2020) developed a DAPP-based modelling toolbox for solar photovoltaic construction planning, demonstrating that the approach is also applicable to shorter-term infrastructure projects. This suggests that DAPP is also a suitable framework for the context of this research. Haasnoot et al. (2013) acknowledge the DAPP methodology can be extended with embedding it into a simulation framework. Michas et al. (2020) introduced three quantitative modelling tools, including scenario discovery in combination with PRIM, to find stressing scenarios. Like this study, Michas et al. (2020) also incorporate a mechanism for monitoring ATPs during simulation. Their approach however goes a step further by dynamically updating the ATP conditions by periodically reapplying PRIM within the simulation, allowing for adaptive thresholds as system states evolve.

Whereas Michas et al. (2020) embed PRIM directly within the simulation loop to adapt ATP values dynamically, other approaches—including the one explored in this study—position PRIM outside the simulation environment. In this structure, a set of ATPs is defined prior to simulation using an iterative scenario discovery process inspired by Guivarch et al. (2016). Rather than relying on a single PRIM run, this method builds a sequence of scenario families, each representing distinct high-impact conditions. These ATPs then serve as fixed monitoring thresholds that trigger predefined robust policies during the simulation. In contrast, the dynamic mechanism developed by Michas et al. (2020) recalculates ATPs during execution: whenever the success rate of the currently active policy falls below a critical level, the simulation is paused and PRIM is re-applied to identify the conditions associated with past success. These new clusters of input variables are used to redefine ATP triggers on the fly. In doing so, PRIM is no longer just a scenario discovery tool but becomes an embedded monitoring function, allowing the DAPP framework to adapt in real time to unfolding uncertainty and shifting policy effectiveness.

Once high-impact scenarios have been identified through scenario discovery, the next step is to design policies that remain effective across these divergent futures. To achieve this, this study adopts MORDM, a method that extends EMA by incorporating multi-objective evolutionary algorithms into the policy design process (Kasprzyk et al., 2013). MORDM enables the exploration of trade-offs between conflicting objectives—in this case, project duration and cost—across a wide range of deeply uncertain scenarios. Rather than converging on a single optimal strategy, MORDM searches for Pareto-efficient solutions that remain robust across the sampled high-impact scenarios. This makes the method particularly suited to dynamic infrastructure projects where planners must balance performance criteria under conditions of uncertainty. The robust policies generated by MORDM then serve as the candidate pathways within the DAPP framework, ready to be activated when a corresponding ATP is triggered.

While (Michas et al., 2020) present a robust example of dynamic ATP monitoring, their model input consists of high-level investment decisions in solar PV capacity over multi-year time steps. In contrast, this study applies DAPP logic to an operational infrastructure schedule with over 300 discrete activities, each defined by task-specific durations, dependencies, and sequencing constraints. The ATPs developed here are not only condition-based but also anchored to specific tasks within this networked schedule, making them time-sensitive by design. This added granularity increases the realism of adaptive logic but also introduces significant challenges for real-time monitoring. To the best of the author's knowledge, ATPs have not previously been constructed at such a detailed operational level. By embedding DAPP, EMA, and MORDM into a full construction simulation with granular schedule inputs, this study extends the methodological scope of adaptive planning to the level of operational infrastructure delivery.

3 Research design

3.1 Introduction

Sub-question 1 is addressed through a literature review. Sub-questions 2 through 4 will be answered by the methods outlined in this research design. This study uses a real-world deterministic project schedule from the Schiphol bridge as modelling input with the objective to create a DAPP-based conceptual schedule. The design of this research integrates multiple methodologies drawn from the framework of DMDU. In the introduction, the structure of the study and its alignment with these methodologies are first clarified. secondly the case description is given. In Section 3.2 a flow chart of the research design is presented, after which the methods are discussed in the last Section of the research design.

3.1.1 Integration of Methodological Frameworks

This chapter presents the methodological design of this study by showing how the different frameworks introduced in the previous chapters are combined into a single modelling process. While EMA, Multi-Objective Robust Decision MORDM, and DAPP have each been discussed in terms of their theoretical foundations and application in past studies, this section clarifies how they interact within the structure of this specific research. The conceptual model in Figure 3.1 illustrates this integration.

At the highest level, EMA is a research approach for analysing complex and uncertain systems through computational experimentation (Bankes, 1993). Originally introduced as a stand-alone method, EMA has since become a foundational element within the broader framework of DMDU, which encompasses several techniques for supporting decision-making under deep uncertainty (Marchau et al., 2019b). Two of these techniques structure the modelling process of this study in two key phases: open exploration and directed search.

Open exploration begins with sampling a wide range of plausible future states of the world, followed by scenario discovery to identify key vulnerabilities in the project schedule. These steps define the uncertainty space and reveal conditions under which the schedule may fail. The results are then used in the directed search that builds on the MORDM framework, which identifies robust strategies by balancing trade-offs between competing project objectives.

The robust policies emerging from MORDM are subsequently integrated into DAPP. While EMA and MORDM support the exploration and evaluation of strategies under uncertainty, DAPP provides a decision structure for implementing these strategies over time. In this study, DAPP is used to map out adaptive schedules, where policy switches are triggered by predefined ATPs. These tipping points are derived from the scenario discovery phase and embedded into the simulation as monitoring conditions.

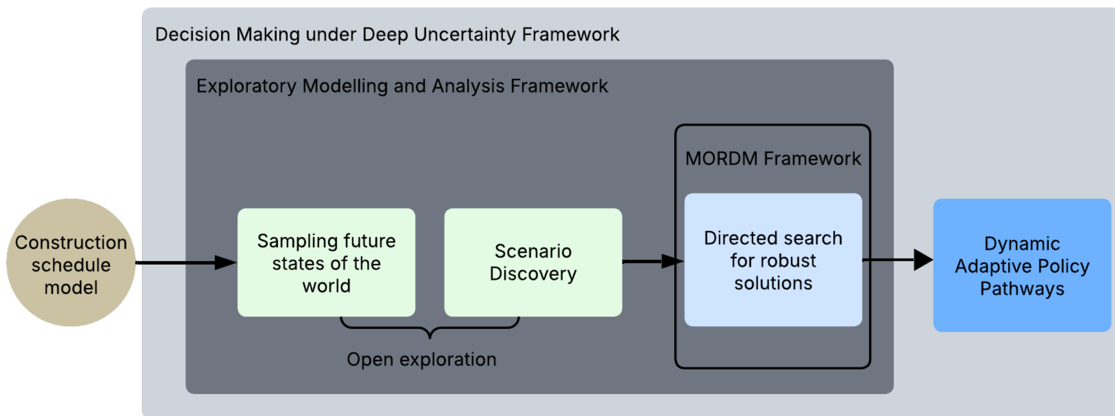

Figure 3.1: Conceptual model of used frameworks in relation to the structure of this study

3.1.2 Case Description and Data

For this research, the studied case is the Schiphol bridge in the Netherlands. The reconstruction of the bridge is part of the Veenix A9 BaHo project which is the reconstruction of the A9 highway between Badhoevedorp and Amsterdam-Holendrecht. Data is provided in cooperation with project management company Count & Cooper. Veenix is currently the largest project undertaken by Count & Cooper and forms part of the Schiphol–Amsterdam–Almere (SAA) program, the most extensive road expansion initiative in the Netherlands to date. The project is executed by Count & Cooper, FCC Construction, Macquarie, and Rijkswaterstaat.

The Schiphol Bridge is one of seventeen civil structures scheduled for construction or renovation within the project and is the first to be completed. It will be widened through reconstruction, Figure 3.2 shows a digitally constructed image of the appearance of the to be finished structure. Selecting the Schiphol Bridge as the focus of this research aims to generate insights that can support the company throughout the remainder of the project, which will continue until 2027. Moreover, the construction process did not proceed as originally planned, making it a relevant and compelling case for testing a novel scheduling approach.

The reconstruction of the Schiphol Bridge is divided into three distinct phases. All scheduling activities examined in this research can be traced back to work carried out within one of these phases. The schedule is designed to maintain optimal flow of traffic during construction as the A9 is an important connection in the ring of Amsterdam. Figure 3.3 shows the traffic flow on the Schiphol bridge during each of the building phases.

- Phase 1: Expansion Southern deck
  - Traffic can continue on both ends of the highway.
- Phase 2: Replacing northern flap
  - Redirecting traffic to southern deck expansion. No traffic on northern deck during construction.
- Phase 3: Civil works and replacing southern flap.
  - Traffic is temporarily redirected to the newly installed northern deck. The southern deck remains unused until the project is completed, at which

point it will support increased traffic flow resulting from the lane expansion.

Figure 3.2: Aerial view of the the finished Schiphol bridge. This is a computer adjusted image. Reproduced with permission from Count & Cooper (2023).

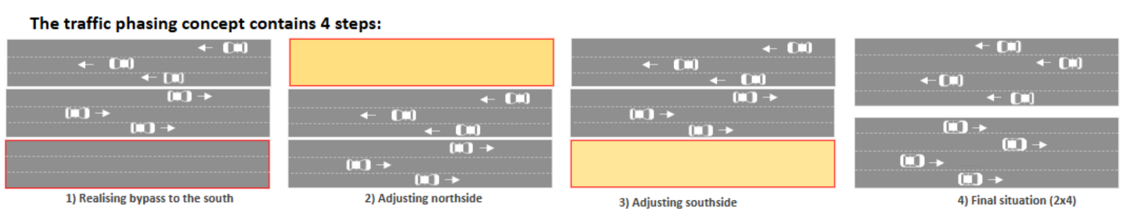

Figure 3.3: The impact of the construction phasing on the traffic on the Schiphol bridge. Reproduced with permission from Count & Cooper (2023)

3.2 Flow Chart of Research Design

The flowchart in Figure 3.4 outlines the ten main steps undertaken in this research. The first three steps, shown in grey, represent the construction of the model and are discussed in section 3.3.1. Steps 4 through 10 correspond to the analytical methods used to derive the results. Each rectangle is colour-coded according to the section of the research design it belongs to. All steps produce outputs that serve as inputs for subsequent techniques. A more detailed version of this flow chart is presented in Appendix Section A.1.

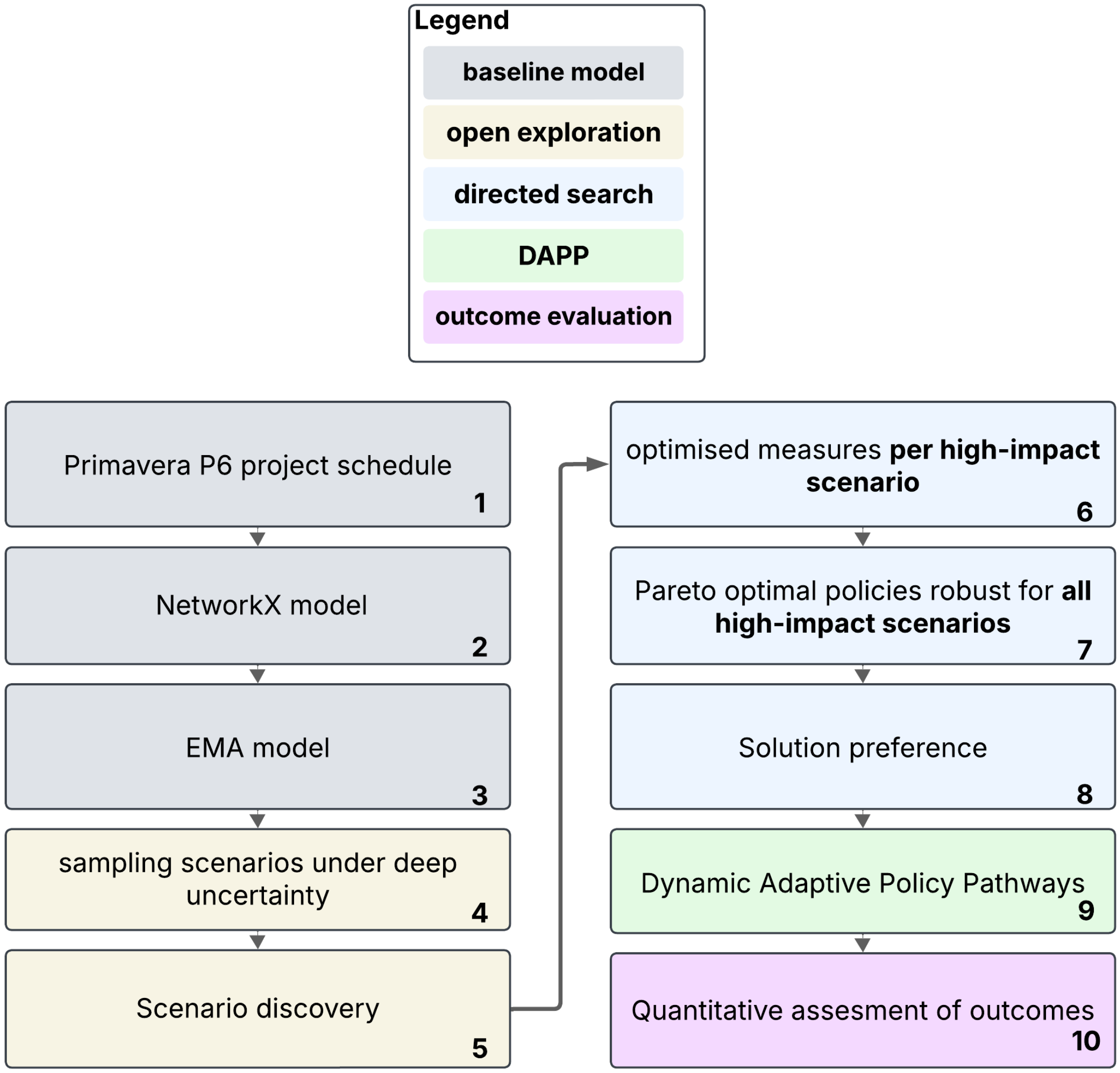

Figure 3.4: 10 steps are executed to go from a deterministic project schedule to a DAPP-based schedule concept. Each step can be seen as an intermediary product that is input for the next step. The colour coded rectangles each belong to a specific section of the research design which is provided in the legend.

3.3 Methods

3.3.1 Model Description

3.3.1.1 NetworkX Model

The project schedule for the construction of the Schiphol bridge was created in Primavera P6 planning software by a project team at Veenix. In Primavera, a schedule follows the structure of a Gantt chart. Figure 3.5 shows a representation of the scheduled activities in Primavera. Construction activities have relationships and dependencies and start either at a planned time or when the preceding activity is completed.

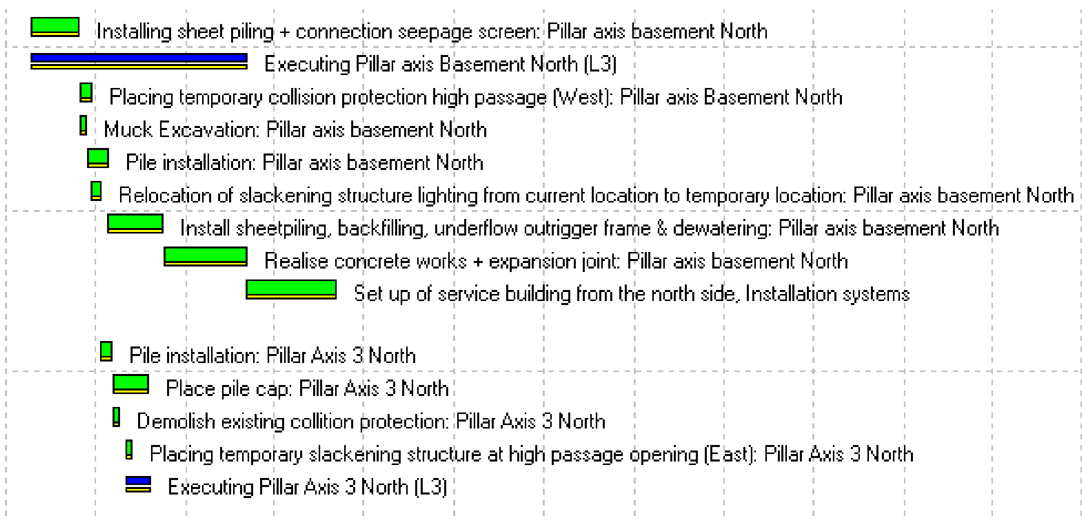

Figure 3.5: Gantt chart from Primavera that shows a small part of the pillar installation of the north side of the Schiphol bridge. Each green bar represents an activity as currently scheduled whereas a blue line represents a baseline. Because this is a zoomed-in image the axis are lost. On the X-axis the time should be displayed while on the Y-axis the tasks are presented.

The schedule in Primavera P6 is already a model, but its functionality is largely limited to deterministic planning and static visualisation. To enable more advanced analysis of uncertainty, adaptation, and system behaviour, the schedule must be converted into a format suitable for computational experimentation. This is done by translating the schedule into a NetworkX graph. NetworkX is a Python library used for the creation, manipulation, and study of complex networks and graphs (Hagberg et al., 2008).

At this stage, the graph model stores the information as presented in Table 3.1. Whereas the duration of tasks is based on the schedule, the costs are added. Although Primavera supports the inclusion of cost data, the baseline schedule used in this study does not contain cost estimates. However, cost is a crucial outcome in the analysis. To address this, daily rates were derived from project risk reports. All costs are scaled to a consistent order of magnitude, ensuring comparability across tasks. To approximate task-specific costs, six categories of activities were identified based on keyword matching in the activity names. Each category was assigned a daily rate derived from project risk reports. The following daily rates were applied:

- `pile installation` — €150,000 per day
- `place pile cap` — €80,000 per day
- `wall` — €200,000 per day
- `pour strip decks` — €60,000 per day
- `applying basement` — €250,000 per day
- `sheet piles` — €170,000 per day

For all other tasks that did not fall into one of these categories, a default daily rate of €50,000 was used. The final cost is calculated as the daily rate times the duration of the task in days.

Table 3.1: Variable descriptions for activity data. The table describes each column used in the project schedule.

Variable	Explanation	Example
Name_activity	Name of the activity	Pillar installation
Activity_ID	Unique activity ID assigned by Primavera	NM.23890
StartDate	Planned start date of the activity	20-02-2022
FinishDate	Planned end date of the activity	27-02-2022
ObjectId_activity	ID linking multiple similar activities to one working package	46174246
WBSObjectId	ID of the Work Breakdown Structure (WBS) element that groups related activities	1296928
Name_wbs	Name of the working package the activity belongs to	WP-005.2.3.1 Installing Pillars
Predecessors	List of activity IDs that must be completed before this activity starts	[4617306, 4617254]
Successors	List of activity IDs that follow this activity	[4617304, 4617253]
Duration	Total duration of the activity in days	7
DailyRate	Estimated daily cost of performing the activity	1000
Cost	Total cost calculated as DailyRate × Duration	7000

3.3.1.2 Critical Path and Slack Time

The project schedule consists of 614 tasks, of which 25 form the critical path. The critical path is the sequence of tasks that cannot be delayed without delaying the overall project. Its total duration, scheduled for 1290 days, represents the shortest possible time in which the project can be completed. In this study, all task durations and project timelines are measured in working days, excluding weekends and holidays, unless scheduled otherwise. This reflects standard industry scheduling practice and ensures that the modelled durations align with how construction time is managed in reality.

Slack time (or float) is the amount of time a task can be delayed without affecting the project's end date. Tasks on the critical path have zero slack, meaning any delay directly impacts the entire schedule. Under conditions of deep uncertainty, task durations can vary significantly between scenarios due to unpredictable or poorly understood risks. As these durations change, the critical path can shift — tasks that were previously non-critical may become critical, and vice versa. Acknowledging that the critical path can shift under deep uncertainty, this research focuses not only on the original critical path, but also on tasks with a limited amount of slack time. Tasks with low slack are those most likely to become critical in alternative scenarios.

To determine an acceptable slack threshold, this study performs a sensitivity analysis. For slack levels ranging from 1% to 20%, the number of additional activities included beyond the critical path is calculated. This is done by applying the Critical Path Method (CPM) to the NetworkX graph. For each task, the slack is calculated as the difference between its latest and earliest start times, using a forward and backward pass through the schedule.

$$\text{Slack} = \text{Latest activity start} - \text{Earliest activity start}$$

The activities on the critical path plus the activities falling within the amount of slack time will be used as the baseline model. The non critical tasks are filtered out of the analysis as they have less impact on the schedule.

3.3.1.3 EMA Workbench Model

This study uses the EMA Workbench, an open-source Python library designed to support the full range of EMA methods. Developed by Kwakkel (2017), the workbench enables users to design experiments and analyse results through integrated tools for sampling, simulation, scenario discovery, and robust decision-making under deep uncertainty.

While the NetworkX graph defines the system structure by capturing project activities and dependencies, it lacks the explicit definition of uncertainty and decision space required for exploratory modelling. Therefore, the schedule needs to be transformed into an EMA workbench model object with *uncertainties, levers and outcomes*. The model is conceptualised in Figure 3.6. The new model acts as a container that can tell the EMA workbench how the system works, what uncertainties affect it, what decision levers are available, and what outcomes are important.

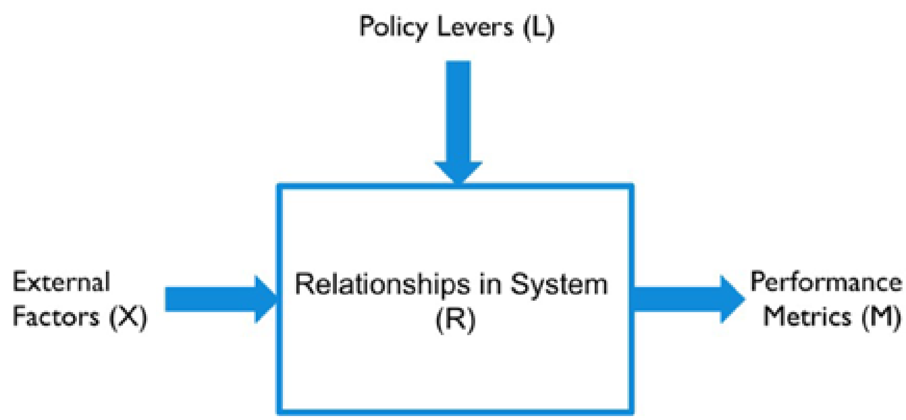

Figure 3.6: The XLRM-Framework conceptualises the relationships (R) between the modelled system, the affecting uncertainties (X), the levers (L), and its outcomes (M). *Source:* Data from Kwakkel (2017)

The external factors i.e. uncertainties and risk events influence the system beyond the control of the decision maker. The levers can be seen as solutions to mitigate the influence of said uncertainties. In this study, levers are referred to as measures to align with infrastructure industry standard. The outcomes are the metrics for performance of the system. The uncertainties, measures and outcomes that are added to the model are presented in Table 3.2.

While the EMA Workbench model is structured around uncertainties, levers, and outcomes, its internal execution in this study follows the principles of DES. The model is implemented in SimPy, a discrete-event simulation library in Python (Meurer et al., 2017). The Simpy model uses the task dependency graph from the NetworkX-schedule to structure event-based execution. Each activity is simulated as a discrete process that begins once its predecessors are completed, and time advances based on task durations influenced by uncertainty. This design qualifies the EMA model not only as a scenario-exploration framework, but also as a fully operational DES-engine, allowing for time-dependent modelling of cascading effects under deep uncertainty.

Table 3.2: Overview of uncertainties, measures, and outcomes used in the model. The type, name, description, and value range are shown.

Type	**Name**	**Description**	**Value Range**
Uncertainty	p_hard_layer_found	Probability of encountering a hard layer during muck excavation	0.10 to 0.25
Uncertainty	p_size_foundation_pillars	Probability that foundation pillar sizing is incorrect	0.05 to 0.10
Uncertainty	p_heavy_wind	Probability of heavy wind disrupting a critical lift	0.10 to 0.25
Uncertainty	p_uxo_found	Probability of discovering unexploded ordnance (UXO)	0.00 to 0.05
Uncertainty	p_critical_material_failure	Probability of a critical equipment or material failure	0.00 to 0.05
Uncertainty	p_influenza_wave	Probability of labour capacity drop due to an influenza wave	RAI becomes 0.6
Uncertainty	fuel_multiplier	fuel cost sampling value (triangular distribution)	min: 0.9 mode: 1.0 max: 1.5
Uncertainty	RAI	resource availability index (triangular distribution)	min: 0.7 mode: 0.9 max: 1.0
Uncertainty	delay_factor	General task delay factor (aleatoric uncertainty)	0.9 to 1.25
Lever	predrilling	Use predrilling to mitigate hard layer delay (adds time and cost)	True or False
Lever	extra_careful_installation	Mitigate pillar issues by stricter installation (partial success)	True or False
Lever	new_design	Redesign to address known risks (higher cost/time, full mitigation)	True or False
Lever	extended_search	Search for UXO early (lower consequences if found)	True or False
Lever	option_spare_part	Have option on backup parts to mitigate equipment failure (lower consequence but higher price)	True or False
Lever	electric_machinery	Invest in electric equipment to reduce fuel uncertainty (adds base cost)	True or False
Lever	overtime_labor	Use overtime to reduce delay (adds cost)	True or False
Lever	schedule_padding	Add time buffer to all task durations	1.0 to 1.2
Lever	budget_buffer	Add cost buffer to absorb risk-driven cost increases	1.0 to 1.1
Lever	num_backup_weekends	Pre-authorised backup weekends for critical lifts	0 to 4
Outcome	project_duration	Total time of the project	ScalarOutcome
Outcome	total_cost	Total cost of the project	ScalarOutcome

3.3.1.4 Risks, Uncertainties, and Measures

The discrete risks in this study are selected from a risk register used for the construction of the Schiphol bridge. Each risk has a given range of probability of occurrence which was established in risk management sessions over the course of the project. The risks were defined using uniform distributions, reflecting a lack of precise knowledge about their true shape and ensuring equal plausibility across the specified ranges.

The continuous uncertainties used in this research differ from the discrete risks found in the project's risk register. Three novel uncertainties—fuel multiplier, delay factor, and the proxy Resource Availability Index (RAI)—were validated by professionals at Count & Cooper as useful and plausible representations of systemic project variability. The delay factor accounts for aleatoric uncertainty inherent in project execution: even when activities are repeated under similar conditions, slight variations in duration can occur. Since the distribution of such randomness is task-specific and not empirically known, a general uniform distribution was chosen to reflect a transparent, non-assumptive approach within bounded ranges—aligned with the treatment of discrete risks.

Fuel multiplier represents global fuel price uncertainty and is modelled using a triangular distribution. The mode is set to 1.0, reflecting current fuel prices as the most likely condition during project execution. The maximum value of 1.5 accounts for historically observed spikes in fuel costs over the past 50 years, while the minimum of 0.9 captures the possibility of modest, bounded price drops. This distribution captures asymmetrical uncertainty, where extreme increases are more impactful than small decreases.

The RAI serves as a proxy for resource constraints and it is also modelled using a triangular distribution with a minimum of 0.7, a mode of 0.9, and a maximum of 1.0, where 1 indicates full availability of labour, equipment, and materials. In the absence of real-time resource data, the RAI offers a simplified yet effective way to reflect varying levels of resource pressure. The RAI acknowledges the importance of resource availability in scheduling while remaining minimal in implementation. The values of its distribution reflect the expectation that most construction phases only operate under light constraints due to professional planning, while still allowing for the possibility of severe shortages. This approach aligns with the resource-constrained scheduling challenges discussed by (Gómez Sánchez et al., 2023).

As for the measures, a consistent approach was applied. The risk register typically provided one or two predefined measures per risk with a cost and likelihood of reoccurrence. The measures associated with the continuous uncertainties were discussed and validated through professional judgment.

3.3.1.5 Model Logic

Each task in the schedule follows a set of conditional rules that determine how its duration and cost are affected by the presence of risk events and continuous uncertainties. While the effects of the risk events are defined at the task level, the continuous uncertainties are sampled at the start of each simulation and applied globally as fixed scenario parameters. This simplification was introduced to prevent short-lived fluctuations—such as a brief spike in fuel prices—from disproportionately influencing long-term planning decisions. By stabilising continuous inputs at the scenario level, the model remains focused on structural performance under broader uncertainty patterns rather than reactive shifts to transient conditions. The following examples illustrate how task behaviour changes in relation to the application of connecting measures.

For muck excavation tasks, if a hard layer is encountered and the predrilling mitigation is not applied, a delay of 60 to 120 days is introduced, along with an additional cost between €200,000 and €500,000. If predrilling is used, both task duration and cost increase by 20%, but the risk is avoided.

For foundation pillar installation tasks, if incorrectly sized pillars are discovered, a new design may be applied (adding 20% to duration), or the project may rely on extra careful installation. This latter option adds 5% to the duration and only succeeds with a 60% probability. If it fails, the task is penalised with a 30–60 day delay and up to €500,000 in additional cost.

For tasks depending on traffic diversion, such as flap or pillar installation, backup weekends may be scheduled in advance. Each additional weekend adds 7 days (a work week of delay plus the additional weekend of work) and incurs a planning cost between €100,000 and €300,000. If heavy wind occurs, each weekend has a 10–25% chance of being unusable. If no weekend is usable, a 6–8 month delay and a penalty cost between €2–10 million are applied. If a backup weekend is usable, the delay is avoided but a smaller cost penalty (up to €800,000) is still applied.

After the task-level logic, the simulation evaluates project-wide risks. If unexploded ordnance (UXO) is found, the outcome depends on whether extended search was applied. With extended search, the delay is 3–4 months and cost increases by €100,000–250,000. Without the measure, delay extends to 6–8 months and costs range from €200,000–500,000.

If the spare part policy is selected, a fixed cost of €400,000 is added. If a critical material failure occurs, the delay and cost depend on whether the spare part was available. With the option, the penalty is reduced to 36 days and €1.2–6 million. Without it, the project is delayed by 6 months and penalised by €2–10 million.

Electric machinery reduces fuel cost volatility. If applied, the fuel cost multiplier is reduced by 20%, but a fixed investment of €2 million is added. This adjusted multiplier is applied to all project cost at the end of the simulation.

In the event that an influenza wave occurs during the project timeline, the RAI is reduced to 0.6, simulating a 40% reduction in the availability of labour, equipment, and materials. This simplification applies the core logic of ECM, where a discrete disruption triggers a temporary shift in continuous project parameters. In this implementation, the RAI is adjusted globally across the schedule to reflect a systemic drop in capacity following a triggering event. While this uniform application simplifies the localised dynamics typically modelled in ECM, it preserves the method's essential mechanism for capturing cascading disruption.

If the overtime labour policy is applied, all task durations are reduced by 10% to reflect increased work capacity, while project costs are increased by 20% to account for the added expense of extended labour hours.

3.3.2 Open Exploration

This chapter is split into two parts and provides an answer to sub-question 2: *"How can scenario discovery be used to find high-impact scenarios and adaptation tipping points in construction scheduling under deep uncertainty?"*

Firstly, the process of generating samples of plausible future states of the world is explained, followed by the discovery of high-impact scenarios. A high-impact scenario represents a plausible combination of uncertainties that, if left unaddressed, leads to significant schedule and cost overruns. The sequence in which scenario discovery is performed in combination with the optimisation of robust policies is not definitive. This study adopts the scenario discovery-first approach in order to understand which uncertainties are most critical before mitigation strategies are explored (Lempert et al., 2003; Kwakkel and Pruyt, 2013).

3.3.2.1 Sampling

The EMA Workbench uses Latin Hypercube Sampling (LHS) by default. LHS divides the range of each uncertainty variable into intervals based on the number of experiments, ensuring that the entire uncertainty space is systematically explored (Saltelli et al., 2000). This sampling method is well suited for open exploration, as it stratifies uniform distributions to minimise variance and reduce sampling bias, thereby lowering the risk of overlooking critical combinations of uncertainties (Dutta and Gandomi, 2020).

After manually testing how many samples would provide a varied enough uncertainty space, this study opted for a simulation of 10,000 samples against a *baseline policy*. This phase examines which combinations of uncertainties could compromise the achievement of the scheduled project duration and cost objectives. At this stage, the measures presented in Table 3.2 are fixed at their base value and therefore have no effect on the outcomes. The measure values for the baseline policy are displayed in Table 3.3:

Table 3.3: Measure settings of the baseline policy.

Lever	**Value**
`predrilling`	0.0
`extra_careful_installation`	0.0
`new_design`	0.0
`extended_search`	0.0
`electric_machinery`	0.0
`overtime_labour`	0.0
`budget_buffer`	1.0
`schedule_padding`	1.0
`num_backup_weekends`	0.0

3.3.2.2 Scenario Discovery

Once scenarios have been generated, scenario discovery methods are used to isolate a smaller, more impactful subset. Traditional model-based analyses often overlook abrupt changes or surprising outcomes. In contrast, scenario discovery explicitly examines how input parameters interact and identifies threshold combinations that strongly predict policy-relevant outcomes (Lempert et al., 2003; Bryant and Lempert, 2010; Saltelli et al., 2000). By actively searching for disruptive scenarios, this approach ensures that unexpected but plausible outcomes are systematically considered in quantitative analysis. In the context of construction scheduling, this allows for the identification of uncertainty conditions that lead to significant project delays or cost overruns, enabling the development of more resilient planning strategies.

The PRIM algorithm identifies combinations of uncertain conditions that lead to extreme outcomes, making it well-suited for scenario discovery under deep uncertainty (Friedman and Fisher, 1999). In this study, PRIM is used to filter down high-impact scenarios from the 10,000 sampled futures. PRIM leverages predefined thresholds to identify which sampled conditions are most likely to exceed acceptable limits, thereby providing valuable interpretability to facilitate understanding of why certain parameter combinations might derail the plan (Bryant and Lempert, 2010). These extracted conditions effectively become ATPs: In a dynamic schedule, once an ATP is reached, a modification to the planning strategy may be necessary to ensure the project remains on schedule.

3.3.2.3 Iterative PRIM Cycles

This study adopts the modified PRIM procedure proposed by Guivarch et al. (2016) for iterative scenario discovery. Unlike the standard PRIM algorithm, which discards previously covered regions, the proposed approach retains these scenarios but reclassifies them as *not of interest*. This enables multiple iterations on the same dataset, allowing the discovery of distinct scenario families rather than converging prematurely on a single dominant box. A peeling threshold of 0.5 was selected to balance generality and explanatory power. Each PRIM box must therefore contain at least half of the remaining *bad cases*, consistent with the minimum density recommended by Bryant and Lempert (2010), which helps avoid overfitting to highly specific scenarios. After each iteration, the highest-density box is flagged as *not interesting*, and the process continues with the remaining cases until no box exceeds the 50% density threshold. Each box is constructed based on the predefined thresholds for cost or duration and is intended to isolate combinations of uncertainty values and risk events that are systematically associated with unfavourable outcomes. PRIM uses two key metrics to guide box selection: *coverage*, which indicates the share of all bad cases that the box captures, and *density*, which measures the proportion of cases inside the box that are truly bad. Each resulting box reflects a unique set of conditions within the uncertainty space that systematically leads to poor outcomes.

A combined threshold at the 60th percentile for both project duration and total cost is selected, targeting the worst-performing 40% of scenarios with respect to both objectives. This means that PRIM only looks at scenarios that exceed 1553 days in duration and €154,574,107 in cost simultaneously. This decision aims at finding scenarios where cost and time risks reinforce stress on the project schedule. This emphasises trade-off navigation between the project objectives in the search for solutions later on. The result of this iterative procedure is a subset of scenario families, where each family groups together scenarios with similar uncertainty values that lead to poor project outcomes. To assess whether these scenario families are also structurally distinct in terms of their underlying uncertainty configurations, Principal Component Analysis (PCA) is applied in the results section. This dimensionality reduction technique allows visual confirmation that the PRIM boxes are located in separate regions of the input space, as intended by the iterative approach.

3.3.2.4 High-Impact Scenarios

The scenario families present subsets of the initial 10,000 samples, offering a more targeted view of high-impact cases. However, further filtering is required to isolate a final set of representative scenarios. To find the high-impact scenarios a scaling function is applied on the normalised project outcomes. Within each scenario family the following scaling function is applied:

$$\text{ScalarScore}_i = 0.5 \cdot \text{Cost}_{i,\text{norm}} + 0.5 \cdot \text{Duration}_{i,\text{norm}}$$

A fixed 50/50 scalar score is used to rank scenarios by joint outcome severity. This approach supports an objective assessment of system stress by treating cost and duration equally, without embedding normative assumptions. Variable weightings are avoided to preserve the descriptive nature of scenario discovery, which aims to isolate structurally disruptive conditions rather than reflect stakeholder preferences.

From each scenario family, the scenario with the highest scalar score is selected as the most severe case. A second scenario is then identified from within the top 15 highest-scoring cases in that family, based on the maximum Euclidean distance in the uncertainty and risk event space. This method ensures that each pair includes one scenario that is both extreme and another that is maximally distinct, while still representative of high-impact conditions. The result of this final step is a set of high-impact scenarios that will be used in the directed search phase.

3.3.3 Directed search

In this section, sub-question 3: *"Which robust mitigation strategies can be identified that address high-impact scenarios affecting cost and duration?"* - is addressed. This section aims to identify measures that minimise project duration and total cost across the high-impact scenarios found through scenario discovery. Following the MORDM framework, a Multi-Objective Evolutionary Algorithm (MOEA) is used to search for robust, Pareto-efficient solutions. Specifically, the Nondominated Sorted Genetic Algorithm-II (NSGA-II) is applied, a MOEA known for effectively approximating the Pareto front in complex optimisation landscapes.

The optimisation is executed with 8,000 evaluations per run, a number determined through convergence metrics. As shown in Figure A.4 in Appendix A.2.4, these metrics demonstrate that the samples stabilises well before the 8000^{th} evaluation, confirming this threshold as sufficient for convergence. To account for the stochastic nature of evolutionary algorithms, five random seeds are used per scenario. The results are then aggregated by reporting the mean and standard deviation across the five seeds.

3.3.3.1 Selection of Robust Policies

NSGA-II produces solutions tailored to each individual high-impact scenario. While effective within their specific context, these scenario-specific solutions are not inherently robust, as they are not optimised to perform well under alternative future conditions. A solution that performs well in one scenario may fail completely in another, highlighting the need for a robustness-based evaluation across scenarios. In this study, a policy is considered robust if it demonstrates Pareto optimal performance when evaluated across all identified high-impact scenarios. This means that no other policy outperforms it on both cost and duration simultaneously across the same scenario set. Such robustness ensures that the selected strategies remain effective even under deep uncertainty This form of robustness is essential in real-world infrastructure planning, where the future is unpredictable and decisions must hold up under a variety of stress conditions.

This study initially explored the use of the MSMOP approach to identify robust policies across high-impact scenarios (Shavazipour et al., 2021). However, because MSMOP is designed for continuous decision variables, it cannot contain the boolean measures present in this study without significant adaptation. While MSMOP would have been the preferred approach for integrated robustness, these structural limitations call for an alternative. The analysis therefore applies PRIM to identify robust policy configurations in a post-optimisation step. In Appendix A.2.4.1, an extended discussion on the consideration is written.

Although PRIM is most commonly applied in the uncertainty space for scenario discovery, its algorithm is equally applicable to the solution space for policy evaluation. While the underlying mechanism remains unchanged, the focus of analysis differs: in the earlier stage, PRIM was used to isolate combinations of uncertainties associated with adverse outcomes. Here, it is repurposed to explore the policy space, with the goal of identifying combinations of measures that consistently produce favourable results across all high-impact scenarios. This approach builds on precedents in the EMA and Robust Decision Making literature, where PRIM has been applied to characterise decision spaces and robust design configurations (Hamarat et al., 2013; Kwakkel and Pruyt, 2013).

To assess the sensitivity of the robustness classification, PRIM is executed at three threshold levels: the 70$^{\text{th}}$, 80$^{\text{th}}$, and 90$^{\text{th}}$ percentiles of the observed cost and duration distributions. Since PRIM is sensitive to threshold selection, and because applying it to the solution space is less common, this sensitivity analysis adds transparency and exploratory power. It ensures that the identification of robust combinations of measures is not overly dependent on arbitrary percentile cut-offs, and that the discovered patterns are stable under varying definitions of outcome acceptability.

The input for PRIM consists of a full factorial design of scenario–policy experiments. These combinations are created by evaluating all optimised policies across the identified high-impact scenarios. This setup allows for a post-optimisation assessment of robustness: instead of selecting policies purely based on their performance under the scenario they were optimised for, PRIM is used to identify configurations of measures that perform well across the entire scenario set.

To explore how robust policies can be best identified using PRIM, two alternative approaches are compared. The distinction lies in the structure of the conditional input: one approach uses a joint threshold for cost and duration, while the other applies separate thresholds for each objective. Given the uncertain correlation between cost and duration, treating the outcomes independently may provide more flexibility and granularity in characterising conditionally robust solutions. These are policies that meet the robustness criterion in either cost or duration, but not necessarily both. However, a joint threshold offers a better combined perspective on overall policy performance, ensuring that selected solutions perform acceptably on both objectives simultaneously. Figure 3.7 illustrates both approaches.

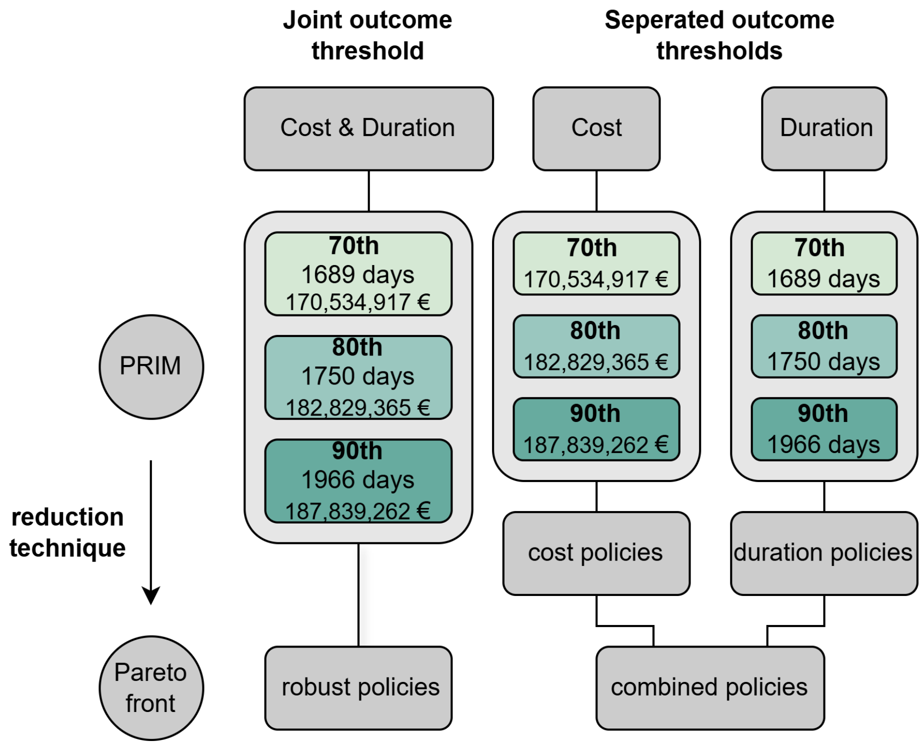

Figure 3.7: Proposed design experiment for PRIM in the directed search stage. Policies are found through a combined condition threshold on cost and outcome as well as a separated threshold. For the final reduction the Pareto front is calculated. For the separated thresholds the non-dominated policies are merged. The policies from the joint threshold are deliberately named robust as they are selected on both cost and duration conditions. The combined policies from the separated thresholds are merely conditionally robust for their respective single objective.

Contrary to scenario discovery, a peeling threshold of 0.8 was used for PRIM. This ensured that only high-density boxes were accepted, exposing structural differences across the robustness thresholds and making it easier to identify whether robust policies clustered differently under looser or stricter performance criteria. The use of percentile thresholds for outcome-based classification—such as the 70^{th}, 80^{th}, or 90^{th} percentiles—serves not to exclude deep uncertainty, but to distinguish between degrees of robustness within it. The full design space already represents deeply uncertain conditions, as all sampled scenarios are drawn from wide distributions that capture plausible but highly variable futures.

The duration outcome exhibits strong nonlinearity: while the difference in project duration between the 70^{th} and 80^{th} percentiles is 61 days, the jump from the 80^{th} to the 90^{th} percentile exceeds 216 days. This indicates a heavy-tailed distribution, where the most extreme cases may dominate statistical patterns despite being rare. In this context, the 80^{th} percentile highlights strong-performing policies under deep uncertainty without overemphasising rare tail cases that could dominate pattern discovery.

Because of this non-linear behaviour and the way the peeling trajectory developed the 80^{th} percentile was selected for the joint threshold. The separated thresholds provide more flexibility, making it possible to select the 90^{th} percentile for duration and the 70^{th} percentile for cost. This distinction is justified by the fact that cost and duration have different distributional characteristics and decision relevance; applying outcome-specific thresholds allows PRIM to identify robust regions more fairly and precisely within each outcome space.

The approach by Guivarch et al. (2016) used in the scenario discovery allows for high coverage of the uncertainty space (as explained in section 3.3.2.3). Because PRIM is used again in the directed search phase, it could be argued that the same method would ensure varied coverage of the solution space as well. For both the joint threshold as the separated thresholds it is not possible to perform more than one PRIM iteration suggesting that the solution space is structurally narrow. The robust policies appear to be concentrated within a single region of the solution space. Instead, the Pareto front is calculated on the three final PRIM boxes: one box from the joint threshold experiments as well as two for the separated experiment. Following the experiment, two sets of robust policies remain, one found by each threshold.

3.3.3.2 Solution Preference per High-Impact Scenario

With the high-impact scenarios defined and both sets of robust policies identified—one from the joint threshold and one from the separated thresholds—one final step remains to complete the modelling setup. While both policy sets are deemed robust across the high-impact scenarios, the relative appeal of a given policy may depend on which scenario ultimately unfolds.

In practice, decision makers may prioritise certain outcomes over others for various strategic or contextual reasons. However, this study limits the preference ranking to the two key outcomes in this study: total cost and project duration. The measures associated with each policy are presented in the results section and may be considered in future applications for broader decision support.

To rank near-equally robust policies within each high-impact scenario, a scalar scoring function is applied that combines project duration and total cost. Since the correlation between cost and duration cannot be directly validated in the current modelling setup, a sensitivity analysis is performed to assess how different weightings influence policy selection. Because of their varying values both cost and duration are normalised using *MinMaxScaler* from the SKlearn Python library (Pedregosa et al., 2011).

To evaluate the sensitivity of policy selection to scalarised weightings, a range of duration-to-cost ratios was tested. The following combinations were tested:

Duration	**Cost**
0.2	0.8
0.3	0.7
0.4	0.6
0.5	0.5
0.6	0.4
0.7	0.3
0.8	0.2

Table 3.4: Tested weight ratios for duration and cost in the scalar function.

Each combination was applied to assess how the scalar score influences policy selection across high-impact scenarios. The lowest scalar score for each policy is selected per high impact scenario. The applied scalar score is:

$$\text{ScalarScore}_{i,j} = w_{\text{cost}} \cdot \Delta C^*_{i,j} + w_{\text{duration}} \cdot \Delta D^*_{i,j} \tag{3.1}$$

- $\text{ScalarScore}_{i,j}$: Combined score of policy i in scenario j
- w_{cost}: Weight assigned to cost
- w_{duration}: Weight assigned to duration
- $\Delta C^*_{i,j}$: Normalised cost improvement of policy i in scenario j
- $\Delta D^*_{i,j}$: Normalised duration improvement of policy i in scenario j

While all solutions could perform under all high-impact scenarios, the scalar function enables a structured selection of the most preferable subset of strategies to be integrated into DAPP. The application of this will be discussed in the next Section.

3.3.4 Dynamic Adaptive Policy Pathways

This section provides the design to answer the fourth and final sub-question : *"How can dynamic adaptive policy pathways be constructed using adaptation tipping points and mitigation strategies for infrastructure construction planning?"*. To answer this question, the approach introduced by Haasnoot et al. (2013) is followed to construct the DAPP. In the previous two chapters, two fundamental attributes are found through scenario discovery and directed search.

Scenario discovery provides scenarios based on specific values for uncertainties and risk events. Each combination of values becomes an ATP. ATPs represent the conditions under which the construction schedule would result in cost and schedule overruns. The robust mitigation strategies founded through MORDM can act as contingency plans to avoid reaching the ATPs. Following the idea of dynamic planning, each high-impact scenario represents a pathway. Each pathway has its own set of resilient mitigation strategies. Due to the high level of uncertainty, it is not feasible to determine in advance which pathway will unfold, making baseline planning a necessary starting point. Each pathway allows for dynamic contingency planning, as the ATP values will determine whether or not a high-impact scenario develops that threatens the achievement of the project objectives.

Multiple pathways are often illustrated as a metro map or a decision tree, where time or changing conditions form one axis of the diagram (Marchau et al., 2019a). This is similar to the initial planning pulled from Primavera P6. Figure 3.8 shows one way to illustrate the pathways.

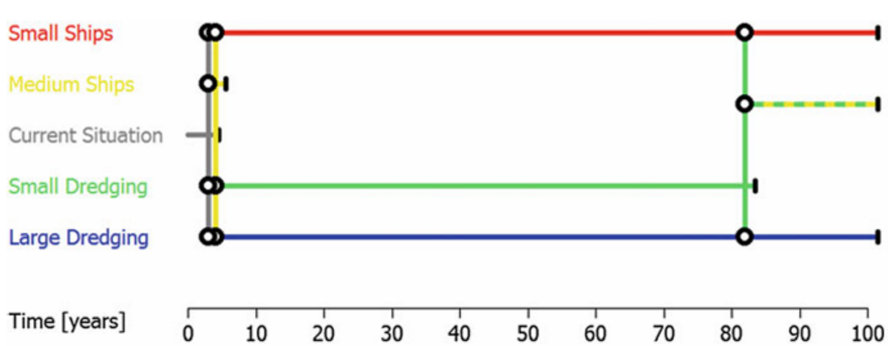

Figure 3.8: An adaptation pathways map for flood management reproduced from (Marchau et al., 2019a): Four pathways are displayed over time. Each dot represents an Adaptation Tipping Point.

To translate this conceptual framework into a working simulation model, the adaptive EMA structure is extended with pathway-specific logic. Each high-impact scenario identified during scenario discovery corresponds to a unique DAPP pathway, with its own robust mitigation strategy derived from directed search. These solutions are embedded in the model and conditionally activated through ATP detection: when a scenario's characteristic uncertainty pattern is observed, the model dynamically switches to the corresponding robust policy. This results in a DAPP-enabled simulation where adaptation occurs in real time based on how uncertainty unfolds. A baseline DES model was constructed in parallel using the same schedule and sampling structure but without any adaptive switching logic. Together, these two models form the foundation of a quantitative analysis, enabling a controlled comparison between adaptive and static scheduling. This setup isolates the effect of dynamic adaptation under deep uncertainty.

3.3.4.1 Quantitative Assessment of Outcomes

The quantitative analysis compares two DES models built on the exact same schedule structure, event logic, and uncertainty inputs. Both models use SimPy to simulate the project as a directed task network, structured to prevent loops, where each activity is executed as a discrete-time process under uncertainty in duration and cost. The baseline DES applies a static configuration in which all measures remain neutral, representing a fixed schedule that cannot respond to evolving conditions. In contrast, the adaptive DES, based on DAPP principles, extends this structure by incorporating a real-time adaptation mechanism. It continuously monitors the unfolding uncertainty state and, upon detection of a matching high-impact scenario profile, activates a predefined robust policy associated with that scenario. Figure 3.9 presents a conceptualisation of the design, illustrating how the effect of adaptive planning is isolated. All other elements—such as the schedule, task logic, and uncertainty sampling—are held constant, allowing any differences in outcomes to be attributed solely to the presence or absence of dynamic adaptation.

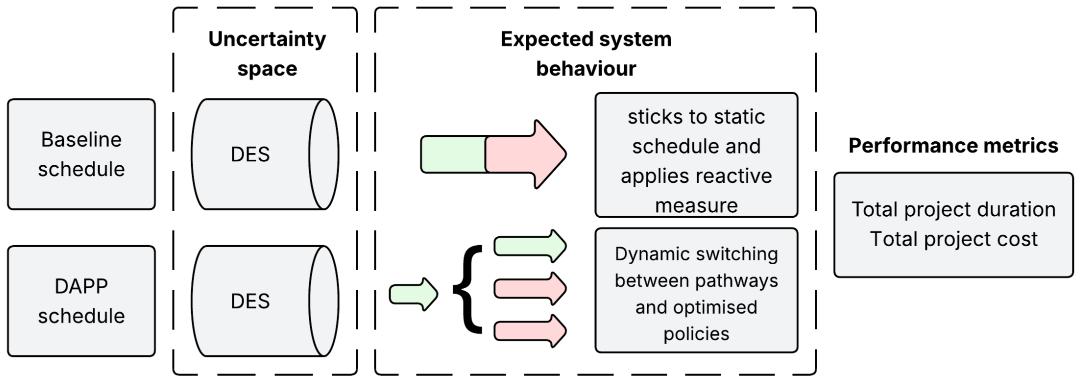

Figure 3.9: conceptualisation of the quantitative analysis. Two discrete event simulations are run. One representing the original baseline schedule and the other representing the DAPP-based schedule. For both simulation total project cost and project duration are monitored for comparison.

To evaluate both models under uncertainty, the shared DES-engine was executed 5,000 times across five seeds using LHS across the defined uncertainty space. Each simulation run begins with the baseline plan and tracks the unfolding values of key uncertainties at each timestamp. If the realised values fall within 20% of the ATPs, the corresponding robust policy is initiated. This buffer was introduced to reflect the anticipatory nature of adaptive planning—it ensures that policy switches occur early enough to be effective, rather than reacting too late when the tipping point has already been crossed. This logic aligns with the concept of a *sell-by date* proposed by Haasnoot et al. (2013), which recognises that some adaptive actions must be taken before the ATP is reached in order to remain feasible and effective. In this model, ATPs are identified based on specific scenario patterns rather than on a single measurable threshold within the system. As such, triggering a policy slightly before the full pattern is realised improves detection reliability in a discrete-event simulation context. Finally, the model assumes that only one pathway switch can occur per run, simplifying the complexity of real-world dynamics while retaining the core DAPP principle: structured flexibility in response to evolving conditions.

Although mitigation measures are often applied reactively in practice, their context-specific nature makes them unsuitable for generalisation within a simulation framework. Designing a baseline model that fully captures real-world scheduling behaviour would exceed the scope of this study. Instead, the focus is on demonstrating the value of incorporating deep uncertainty into planning from the outset. In this sense, the comparison functions as a structured proof of value for the DAPP approach.

The DES experiment is performed on the full project schedule, rather than the 76-activity graph discussed in Section 3.3.1.2. This approach reflects a deliberate modelling choice: while robust policies are optimised using a reduced graph focused on critical and near-critical activities, their effectiveness is evaluated on the complete schedule to ensure external validity. By applying adaptive logic across the entire network, the analysis captures not only local improvements but also broader systemic effects. This separation ensures that the robustness of the adaptive strategies is not evaluated on the same reduced model used for optimisation, but instead tested on the full schedule to assess generalisability under realistic project conditions.

Both policy sets from the joint PRIM threshold as well as the separated PRIM thresholds are quantitatively assessed following the DAPP-switching logic. Each schedule follows the same uncertainty logic in the DES framework and is identically sampled using LHS. By evaluating their performance against the deterministic baseline, the analysis aims to determine which approach yields more effective policies in terms of project duration and total cost. The DAPP schedule that outperforms in terms of cost and duration is selected as the final candidate and further conceptualised as an example of adaptive construction scheduling under deep uncertainty.

4 Results

Section 4.1 addresses sub-question 1 and provides the theoretical foundation for the modelling work. The modelling results start in Section 4.2 and answer sub-questions 2 through 4.

4.1 Deep Uncertainty in Infrastructure Construction

The infrastructure construction sector has long acknowledged the importance of risk management, but *uncertainty* often remains misunderstood or misapplied (Fang et al., 2013). This confusion can lead to overconfidence in cost and schedule estimates, especially when rare or unexpected events are incorrectly formalised into conventional risk categories (Taleb, 2020; Sadeghi et al., 2015). In practice, the distinction between risk and uncertainty is frequently overlooked, with consequences for how project planning and mitigation strategies are designed. As discussed by Wied et al. (2021), unexpected events are especially disruptive to project outcomes and while practitioners often understand the variability in task durations (aleatoric uncertainty), they may not fully account for deeper structural uncertainties that arise during early planning stages. Kim and Reinschmidt (2009) note that epistemic uncertainty plays a dominant role at this stage, when little observational data is available. Although Bayesian methods can reduce some of this uncertainty, their effectiveness is constrained when sufficient knowledge cannot be gathered before decisions must be made.

Infrastructure construction planning is therefore especially vulnerable to deep uncertainty. Project risk registers, for example, often rely on expert judgments and fixed assumptions to define risk events. While such methods can identify plausible disruptions—like wind delaying a bridge lift—they often lack clarity about when a disruption becomes critical, how long it will last, and what trade-offs are associated with mitigation. These challenges reflect deeper issues: how to model event impacts, how to estimate likelihoods, and how to evaluate consequences. Tegeltija et al. (2016) argue that traditional probabilistic methods work well for uncertainties within levels 1 to 3 (see Table 2.1), but break down under deep uncertainty.

This has led to calls for new approaches that go beyond conventional risk management frameworks (Ruckert et al., 2019; Stanton and Roelich, 2021; Feng et al., 2022).

Deep uncertainty is particularly important to address in the early phases of infrastructure projects, when decisions shape long-term performance but knowledge is most limited (Lau et al., 2018; Williams and Samset, 2010; Mohd Nasir et al., 2016). Although a growing body of empirical work confirms that deep uncertainty significantly affects construction schedules (Luo et al., 2017; Qiao et al., 2019; Wang et al., 2023), methodological responses remain underdeveloped (Feng et al., 2022). This study contributes to that gap by explicitly operationalising the distinction between risk and uncertainty in construction scheduling.

In this research, risks are represented as discrete events that either occur or do not, often based on probability thresholds. Even these can be classified as deeply uncertain if their underlying assumptions are disputed. By contrast, uncertainties are modelled as continuous variables. For example, heavy wind during a critical lift is modelled as a boolean risk, while fuel prices are modelled as a continuous stressor. Following ECM, this study acknowledges that continuous uncertainties only become critical when enforced by a connected risk event. This event-driven structure closely aligns with the definition of deep uncertainty, particularly in cases where the likelihood, timing, or system impact of the triggering risk events is poorly understood or fundamentally disputed. In such cases, it is not only the continuous variable that becomes uncertain, but the very conditions under which it becomes disruptive. This blurs the line between risk and uncertainty and reinforces the need for planning approaches that explicitly acknowledge deep uncertainty.

To put this into perspective, an example specifically tied to the delivery of the Schiphol Bridge is given. Hoisting the flaps of the bridge involves activities that lie on the critical path of the schedule. The highway must be closed during the weekend to allow these operations to proceed, making such lift operations tightly scheduled and highly sensitive to external conditions, such as the risk for heavy wind. From a traditional risk perspective, planners may assign a wind speed threshold and a fixed probability of delay based on historical weather data. However, under deep uncertainty, it is not just the severity of the wind that is in question — but the ability to forecast such conditions with sufficient lead time, the reliability of long-range predictions, and the interaction with other project constraints such as resource availability or permitting. The structural uncertainty lies not in whether heavy wind can occur, but in whether it can be anticipated early enough to reschedule the lift safely and cost-effectively, while also balancing potential penalties for changes to the schedule. In this sense, the project is not only exposed to variability in wind conditions (aleatoric uncertainty), but also to fundamental ambiguity regarding the timing, detectability, and operational consequences of weather-related disruptions. This illustrates why conventional risk methods may fall short and reinforces the need for planning approaches that explicitly account for deep uncertainty.

4.2 Modelling and Simulation Results

4.2.1 Slack Time and Critical Path

The slack sensitivity analysis reveals that in the first 5% of slack, the number of activities increases by a factor of 3 relative to the critical path. In contrast, expanding the threshold from 5% to 20% adds only an additional factor 1.5. This indicates that the lower slack range (0–5%) is significantly more sensitive, capturing a concentrated set of near-critical tasks. Beyond 5%, the curve flattens, suggesting diminishing marginal insight from including additional tasks. The critical path plus 5% slack totals 76 tasks that are passed along into the next stage of analysis. The graph with the sensitivity results is presented in Figure A.2 in Appendix Section A.2.1.

The final NetworkX graph of the critical path plus slack is portrait in Figure 4.1. Each circle represents a task and each line represents the relationship between the tasks. The Figure shows similarities to the classic Gantt waterfall structure as modelled in Primavera P6. Uncertainty and risk events, measures, and outcomes are added to the NetworkX model to make it a compatible EMA-workbench model that can be used for the further stages of analysis.

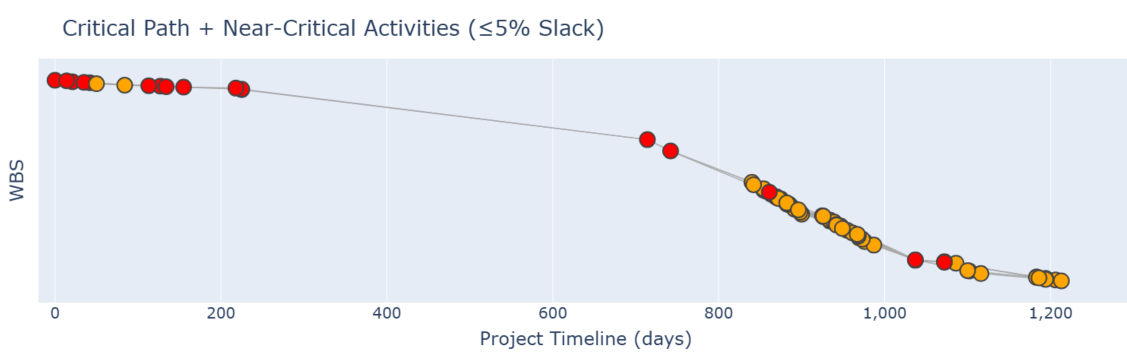

Figure 4.1: NetworkX model that shows the critical path in red and the extra slack in orange. The circles represent activities and the lines the relationships between them. On the X-axis the project duration is plotted while on the Y-axis the Work Breakdown Structure (WBS) vertically groups tasks based on their hierarchical project phase or functional component.

4.2.2 Scenario Discovery

The EMA model is sampled 10,000 times. Following the iterative PRIM procedure introduced by Guivarch et al. (2016), three scenario families are identified. Figure A.3 in Appendix A.2.2.1 presents the results of these three iterations.

Figure 4.2 shows how the scenarios captured in the three PRIM boxes—each representing a scenario family defined by distinct uncertainty drivers—are positioned within the outcome space of all 10,000 samples, plotted in terms of total cost and project duration. Box 1 results from the first iteration and captures the most severe scenarios in terms of duration. Box 2 spans a broader region, capturing both high-cost and high-duration scenarios as well as some of the shortest severe scenarios in the outcome space. Box 3 covers a structurally narrower region and appears to define the lower boundary of severe scenarios in terms of cost.

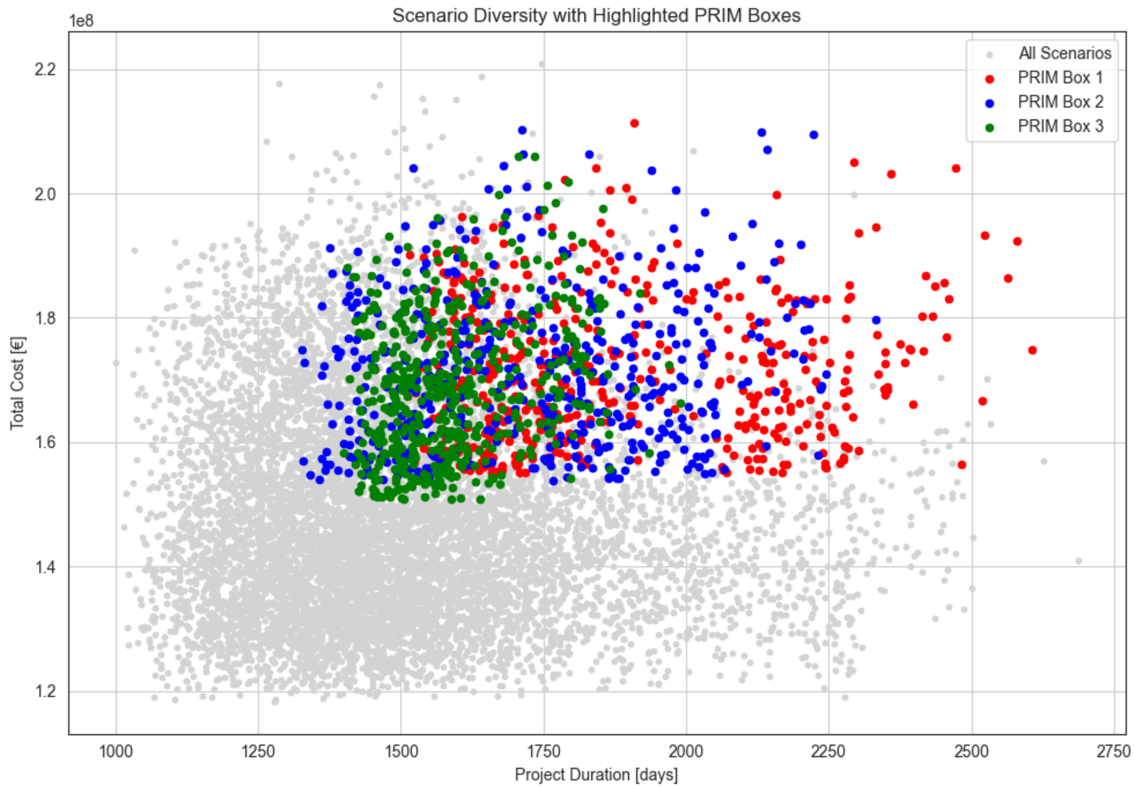

Figure 4.2: The three PRIM boxes plotted against all 10,000 sampled scenarios in outcome space. All boxes capture scenarios with severe values in both cost and duration. The severity of outcomes decreases with each iteration, as the most extreme cases are filtered out in earlier steps.

While the PRIM boxes are defined by poor performance in outcome space, it is also relevant to assess whether they are distinguishable in terms of their underlying uncertainty configurations. To this end, a PCA is applied to the full set of uncertainty variables. Figure 4.3 presents the resulting 2D projection of all PRIM box members. A clear visual separation is observed between the three boxes, particularly along Principal Component 1 (PC1). The PCA loadings in Table 4.1 indicate that PC1 is primarily driven by RAI and delay factor. Box 3 scenarios cluster toward the higher end of PC1, suggesting that these uncertainties are dominant in that group. Box 1 and Box 2 are also clearly separated, with Box 1 tending toward the positive side of PC1 and Box 2 positioned more toward the negative side. In contrast, Principal Component 2 (PC2) is almost entirely dominated by the fuel cost multiplier, which has a loading of 0.998 (highlighted in bold in Table 4.1). This indicates that PC2 effectively only captures variation in fuel cost assumptions across scenario families.

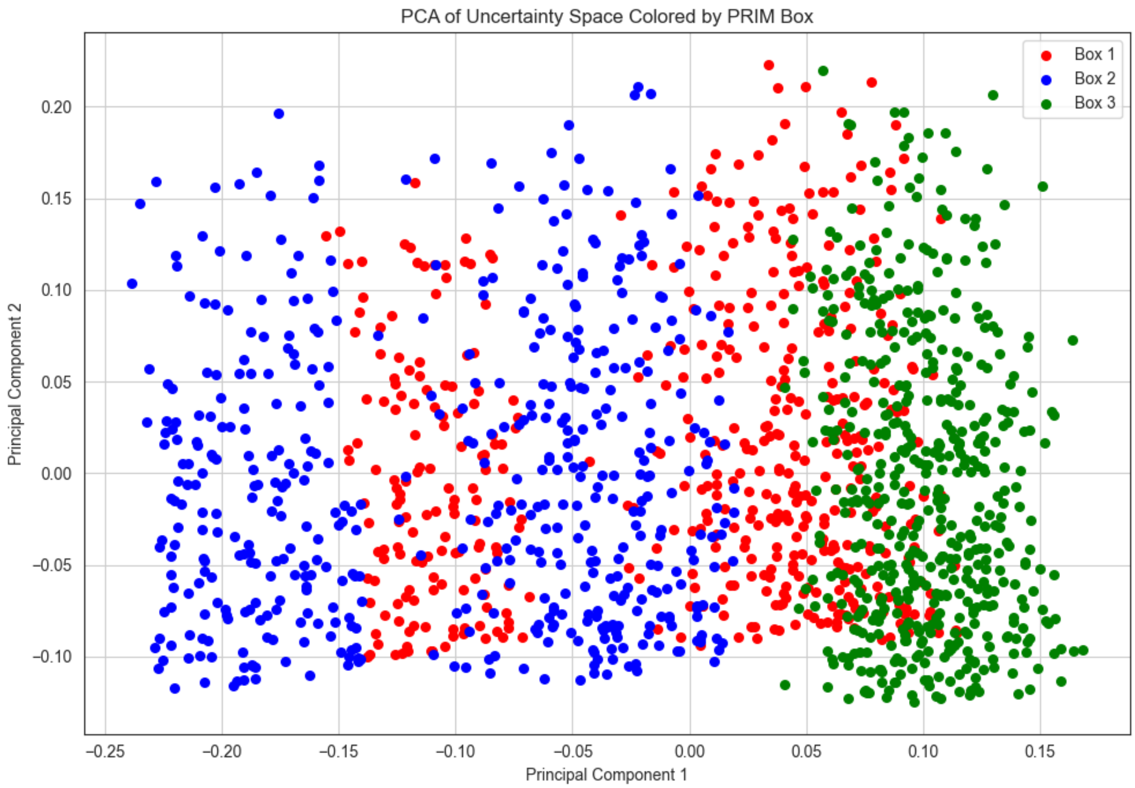

Figure 4.3: Principal Component Analysis (PCA) projection of the three scenario families discovered through iterative PRIM. Each dot represents a scenario, colored by its corresponding PRIM box.

Table 4.1: PCA loadings for Principal Components 1 and 2. The dominant variables in each component reflect the primary directions of variation in the uncertainty space.

Uncertainty Variable	**PC1**	**PC2**
RAI	0.844	0.034
Delay factor	0.526	0.051
Heavy wind (probability)	0.017	-0.003
Oversized foundation pillars (prob.)	0.003	0.009
UXO found (probability)	-0.001	0.004
Material failure (probability)	-0.006	0.009
Hard layer found (probability)	-0.007	0.029
Fuel cost multiplier	-0.055	**0.998**
Influenza wave (probability)	-0.081	0.003

The structural separation between the scenario families supports the conclusion that the iterative PRIM procedure identifies distinct regions of the uncertainty space associated with poor outcomes. However, visual distinction alone does not imply that all included scenarios are equally impactful. To focus the subsequent analysis on the most stressing cases, the final filtering step is applied based on the scalar score discussed in 3.3.2.4.

4.2.2.1 High-Impact Scenarios

The three distinct scenario families filtered out 1,504 scenarios of the initial sampled 10,000. The scaling function is applied to score the scenarios in terms of severity on project objectives cost and duration. The highest scoring scenario is selected for each family, as well as the scenario with the greatest Euclidean distance from the severest scenario in the top 15 highest scalar scores. The six final high-impact scenarios are presented in Table 4.2.

Table 4.2: Values of the risks, uncertainties, and outcomes for the high-impact scenarios. The outcomes and scores are separated from the risks and uncertainties in the lower part of the Table.

Variable	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**
Fuel Cost Multiplier	1.405	1.444	1.186	1.398	1.253	1.189
Resource Availability Index (RAI)	0.600	0.600	0.600	0.857	0.828	0.600
Task Duration Variability	1.226	1.109	1.127	1.165	1.142	1.113
Critical Material Failure	0	0	1	0	1	0
Hard Soil Layer	0	0	0	0	0	0
Heavy Wind	1	1	1	1	1	1
UXO Found	0	0	0	0	0	0
Foundation Size Issue	0	0	1	0	0	1
Influenza Wave	1	1	1	0	0	1
Project Duration (days)	2472.9	2222.7	2564.0	1734.6	1918.4	2331.5
Total Project Cost (€)	204,111,200	209,587,300	186,380,100	206,022,100	183,979,000	179,712,900
Scenario Family	1	2	1	3	3	2
Normalised Cost	0.880	0.971	0.587	0.912	0.548	0.477
Normalised Duration	0.897	0.701	0.968	0.319	0.463	0.786
Scalar Score	**0.888**	**0.836**	**0.778**	**0.615**	**0.505**	**0.632**

The continuous uncertainty values listed in this table are sampled at the start of each simulation and remain fixed throughout the project duration. Consequently, the values shown for these variables represent their initial state rather than a time-dependent trigger. In this implementation, the recorded ATPs for continuous uncertainties reflect these static inputs, rather than the moment at which the system transitions into a stressed condition during execution. Figure 4.4 shows the 6 high-impact scenarios plotted as a Parallel coordinates plot.

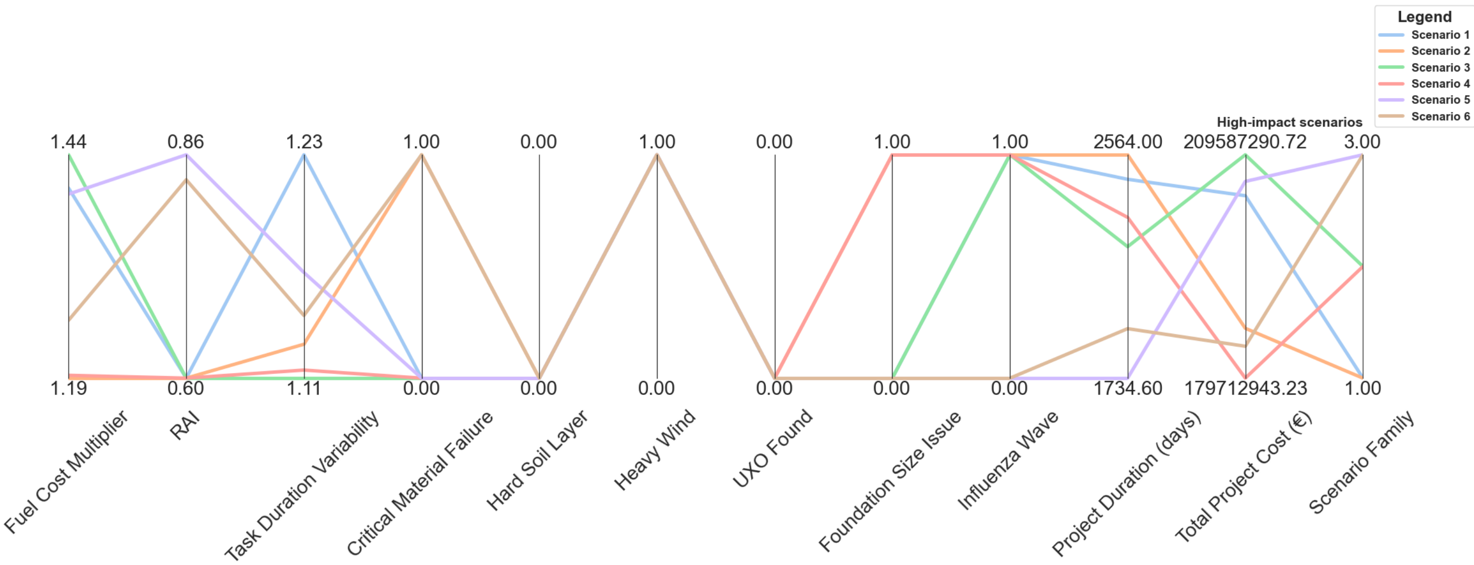

Figure 4.4: Parallel coordinates plot of the six high-impact scenarios. Each coloured line represents one scenario. On the X-axis the uncertainties and outcomes are displayed. Each scenario has its unique set of uncertainty values that influence the outcomes. Under the high-impact scenarios the project outcomes duration and cost vary between 1735–2564 days and €179,712,943–€209,587,290. The final axis *Scenario Family* refers to the PRIM cycle in which the scenario was discovered.

For reference, compared to the project outcomes under the high-impact scenarios, the length of the critical path is 1290 days against a cost of €56,460,000. The tasks included under the extra 5% slack time do not extend the overall project duration, as all additional activities fall within the time window already defined by the critical path. However, the associated cost increases to €74,820,000.

- **Scenario 1 – Most severe overall (Blue)**: This scenario has the second longest project duration at 2,473 days and the third highest total costs at €204,111,200. All three continuous uncertainties are high, meaning they form stressors on the schedule: with high *task duration variability* (1.226), the lowest possible *RAI* value (0.600) triggered by the *Influenza Wave*, and elevated *fuel costs* (1.405). Additionally, a *Heavy* Wind event is triggered, compounding the pressure on project execution. The combination of high project outcomes establishes the highest scalar score for Scenario 1.
- **Scenario 2 – Most costly (Orange)**: Scenario 2 results in the highest total project cost at €209,587,300, with a moderately long project duration of

2,223 days. This scenario has the highest value for *fuel cost multiplier* (1.444). Its *task duration variability* (1.109) is lowest while *RAI* is (0.600) due to an *Influenza Wave*. This scenario also includes a *Heavy Wind event*, further burdening the project.

- **Scenario 3 – Longest project duration (Green)**: With a duration of 2,564.0 days and a moderate cost of €186,380,100, this scenario presents the longest timeline. It features the lowest *fuel cost multiplier* (1.186) among high-impact cases, but this is offset by the activation of three discrete events: *Critical Material Failure*, *Foundation Size Issue*, and *Heavy Wind*. The *RAI* is again reduced to 0.600 due to an *Influenza Wave*, increasing schedule strain.

- **Scenario 4 – Most efficient under stress (Red)**: This scenario yields the shortest project duration among the group at 1,734.6 days. This is traded against the second highest cost of €206,022,100. The *RAI* is 0.857, which is the highest value amongst the high-impact scenarios.

- **Scenario 5 – Cost-efficient but risk-exposed (Purple)**: Scenario 5 yields a relatively low total cost of €183,979,000 and a moderate duration of 1,918.4 days. It includes a *Critical Material Failure* and a *Heavy Wind* risk. The *RAI* remains stable at 0.828 due to the absence of an *Influenza Wave*. This scenario scores the lowest overall scalar score.

- **Scenario 6 – Cost effective (Brown)**: With a duration of 2,331.5 days and the lowest total cost at €179,712,900, this scenario is structurally less extreme. However, it still involves multiple interacting stressors, including low *RAI* (0.600) from an *Influenza Wave*, a *Foundation Size Issue*, and a *Heavy Wind* event.

Across the six high-impact scenarios, a number of patterns emerge. Scenarios with high values for the continuous uncertainties consistently score higher on the scalar function. This is not unexpected, as these variables are modelled globally and influence the project from the outset of each simulation. Their role as persistent stressors highlight the effects of discrete risk events, increasing overall scenario severity. While each scenario includes at least one triggering risk, the underlying pressure from continuous conditions often drives the most substantial outcome differences. This reinforces the theoretical distinction between risk and uncertainty explored earlier and confirms that robust planning under deep uncertainty must consider their interaction explicitly. In this way, the scenario outcomes align with insights from the literature review in Section 4.1, which emphasise that severe disruptions often arise from the interaction between persistent uncertainties and discrete events, rather than from either source alone.

As discussed in Section 3.3.4, the values of the risk events and uncertainties serve as ATPs in the creation of the DAPP schedule. The combination of the values in Table 4.2 and the visual display of the scenarios in Figure 4.4 concludes how different pathways can be created as the foundation of a DAPP-based schedule.

4.2.3 Directed Search

For the directed search phase, the results of the PRIM experiment are presented. At each step leading up to the final DAPP-based schedule, two variants are examined. For the outcomes of cost and duration, both a joint PRIM threshold and separated PRIM thresholds are applied, with the aim of determining which approach yields the most effective robust policies. In Appendix A.2.4.3, the results of the PRIM experiment are discussed. In this section, the Pareto front is presented followed up by the parallel coordinates plot of the robust policies. At last, the solution preference is presented, which provides the final step for the design of the DAPP.

4.2.3.1 Pareto Front

Figure 4.5 presents all the policies found in the selected PRIM box of the 80^{th} percentile. The box consists of 492 policies of which two are non-dominated. The two final policies are shown as red dots.

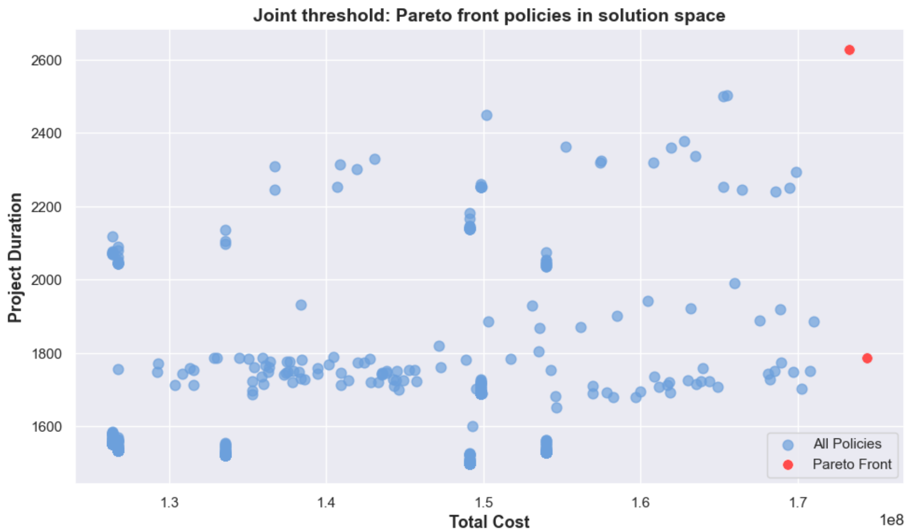

Figure 4.5: Scatter plot of all policies in the selected seventh PRIM box under the 80$^{\text{th}}$ percentile threshold, evaluated in the solution space. The box contains 492 scenario–policy combinations, shown as blue dots. Among these, the two non-dominated Pareto-optimal policies are highlighted in red. The distribution suggests that the Pareto front is more tightly bounded in terms of cost (approximately €175 million), while allowing greater variation in project duration. This indicates that cost plays a more constraining role in defining Pareto-optimality within this region.

The PRIM boxes resulting from the separated thresholds are combined and presented in Figure 4.6. The combined boxes carry 521 policies. The sum of the two boxes is 678 policies, indicating some overlap in the found solutions despite the different thresholds. In the combined solution space there are in total five Pareto optimal policies. Four come from the duration threshold and one is added from the cost threshold. The five remaining policies are robust following the definition used in this study: Pareto optimal performance when evaluated across all identified high-impact scenarios.

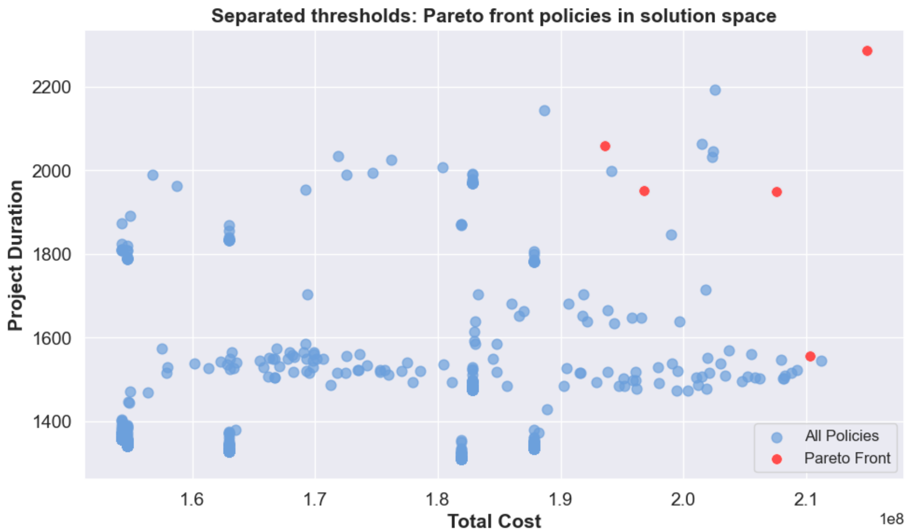

Figure 4.6: Scatter plot of all policies selected through the separated threshold approach, evaluated in the solution space. The blue dots represent 521 scenario–policy combinations resulting from the union of the cost-based and duration-based PRIM boxes. The red dots represent the five non-dominated Pareto-optimal policies within this combined set. Compared to the joint threshold plot, the Pareto front here spans a broader range in both cost (approximately €193–€215 million) and project duration (1,555–2,286 days). This wider spread reflects the more relaxed robustness conditions applied in the separate threshold approach, allowing policies that are robust with respect to only one objective to enter the final evaluation.

4.2.3.2 Final Robust Policies

In this section, the robust policies identified from both the joint and separated PRIM threshold approaches are presented. These policies are optimised to remain viable under the high-impact scenarios described in the previous section. From the joint threshold analysis, two robust policies were identified, while the separated threshold approach yielded five. The selected policies are visualised in Figures 4.7 and 4.8.

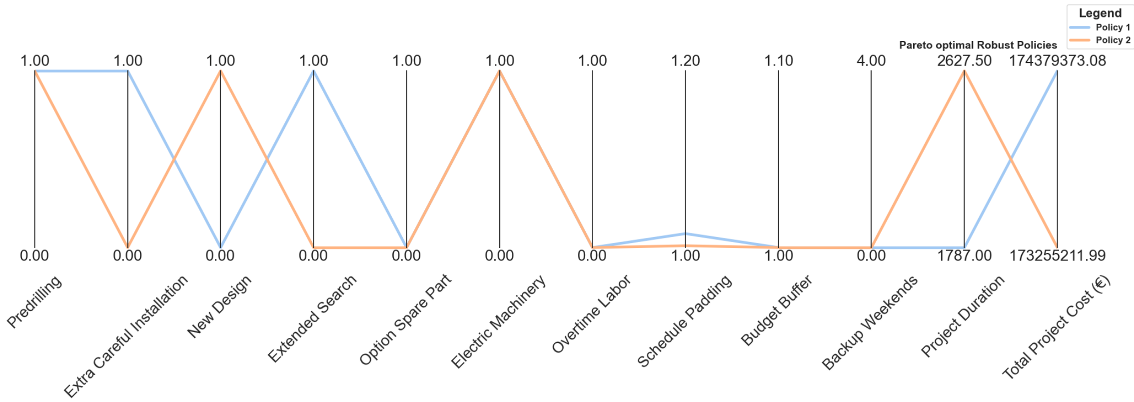

Figure 4.7: Parallel coordinates plot of the joint threshold robust policies. Both policies rely on *electric machinery* as well as *predrilling*. Policy 2 (orange) applies a *new design* whereas Policy 1 builds on *Extra careful installation* and an *extended search*

- **Policy 1 - Safety First (Blue)**: This policy incorporates *predrilling* as a measure against a hard soil layer while also opting to take extra time for a *careful installation* of the foundation pillars. An *extended search* is performed to not find an UXO by surprise while already building. This Policy build on the use of *electric machinery* to be less vulnerable to sudden spikes in fuel prices.

- **Policy 5 - Think first (Orange)**: This Policy incorporates a *new design* into its set of measures. It takes a hit in extra time upfront but prevents problems later on in execution. *Predrilling* and *electric machinery* are also included.

Although both Policies are based on similar drivers, including the use of *electric machinery* and *predrilling*, their project durations differ by more than 800 days. In contrast, their total costs are relatively close, with a difference of approximately €1 million.

It is further observed that the duration of Policy 1 - Safety First, which yields the shortest completion time among the two, exceeds that of the *cost-efficient but risk-exposed* high-impact scenario. No policy intervention was applied in this stage of the study. Both policies, when applied are cheaper than the outcomes of any of the high-impact scenarios.

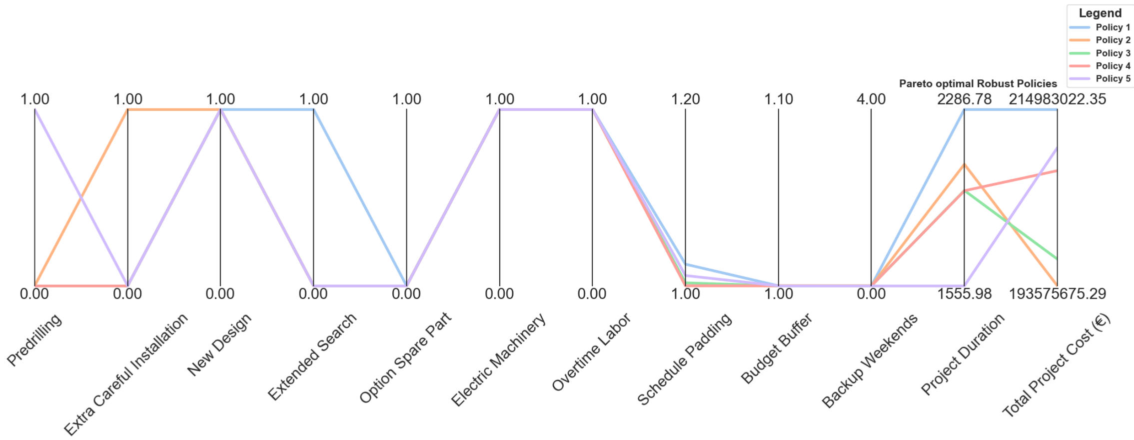

Figure 4.8: Parallel coordinates plot of the separated threshold robust policies. All five rely on the use of *electric machinery* , *overtime labour*, and *new design* as measures. *predrilling* is used in Policies 1,3, and 5. *Extra careful installation* is used in Policy 2. Only Policy 1 makes use of and *extended search*

.

- **Policy 1 - UXO emphasis (Blue)**: Besides the measures that all five policies share, Policy 1 puts extra emphasis on an *extended search* plan for unexploded ordnances.

- **Policy 2 - Care first (Orange)**: Policy 2 distinguishes by the use of an *extra careful installation* of the foundation pillars of Schiphol bridge. It shares the measures for *electric machinery*, *new design*, and *overtime labour* with the other Policies.

- **Policy 3 & 4 - the Median Policies (Green, Red)**: Policies 3 and 4 are characterised by the exact same set of measures: *New design* to prevent unpleasant surprises later on, *electric machinery* to decrease vulnerability against fuel price fluctuations, and *overtime labour* to catch up on delayed work.

- **Policy 5 - No Excavation Trouble (Purple)** Policy 5 is the only policy that takes on *predrilling* as a measure besides the measures shared by all other policies.

Compared to the joint threshold, the policies identified under the separated PRIM threshold are consistently more expensive by at least €20 million. However, they result in notably shorter project durations, with improvements ranging from 200 days for the shortest policy to nearly 400 days for the longest. Regardless of the underlying threshold for PRIM, the three measures that are not modelled as boolean variables are all all show values at or near their lower bounds. The highest percentage for *schedule padding* does not exceed 3% while *budget buffer* and *back-up weekends* are not used at all.

The consistent exclusion of certain levers, such as *backup weekends* and *budget buffers*, highlights a structural feature of the model: anticipatory measures that impose up-front cost or delay without directly improving measurable outcomes are systematically penalised. This reflects how the input parametrisation and scalar weighting shape optimisation behaviour, favouring reactive over preventive strategies. These findings underscore the sensitivity of policy selection to cost assumptions and reinforce the importance of aligning model incentives with practical project priorities.

4.2.3.3 Solution Preference per High-Impact Scenario

Formula 3.3.3.2 is used on the different *cost:duration* ratios to analyse which policies score best for each scenario. For both policies from the joint and separated PRIM thresholds, a heatmap is presented to show potential sensitivity between the different ratios

Figure 4.9 of the Joint threshold policies shows that *Safety first* is the preferred Policy (1) for the *longest project duration* Scenario (3). It furthermore highlights that Scenarios *Most efficient under stress* (4) and *Cost effective* (6) are sensitive to a change in cost and duration weighting. From an even split upwards to a duration heavy preference, *Safety first* (1) is the preferred policy. When cost weighs more, Policy *Think first* (2) is preferred.

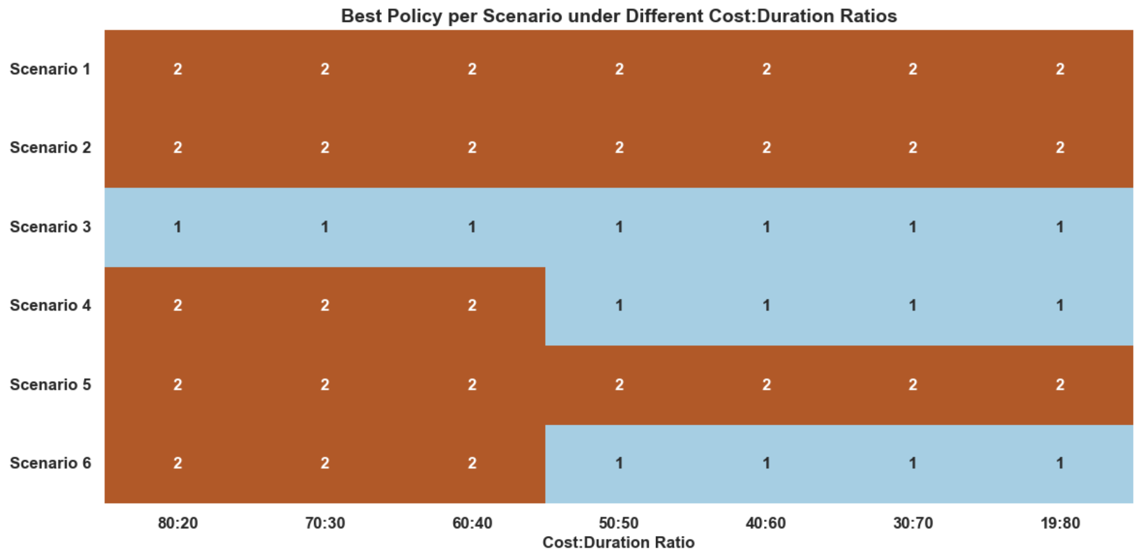

Figure 4.9: Heatmap of the joint threshold policies. Scenario 4 and 6 show sensitivity to cost and duration weighting. Policy 2 is the preferred solution set for Scenarios 1, 2, and 5. Policy 1 is preferred for Scenario 3.

The heatmap for the five policies identified through the separated PRIM thresholds is presented in Figure 4.10. Since the *Median Policies* were based on an identical set of measures, the scalar function filtered out Policy 3 due to redundancy. The *UXO Emphasis* Policy (1) emerges as the preferred option for Scenarios *Most Severe Overall* (1) and *Cost-Efficient but Risk-Exposed* (5). The *Care First* Policy (2) is selected for the *Cost Effective* Scenario (6), while the *Longest Project Duration* Scenario (3) is best mitigated by the *Median Policy* (4). Scenario *Most Efficient Under Stress* (4) aligns most effectively with the *No Excavation Trouble* Policy (5). Scenario *Most Costly* (2) is sensitive to the selected weighting scheme: if cost is weighted more heavily, the *Care First* Policy (2) is preferred; if duration dominates, the *Median Policy* (4) performs best. Under more balanced trade-offs, the *No Excavation Trouble* Policy (5) proves most effective.

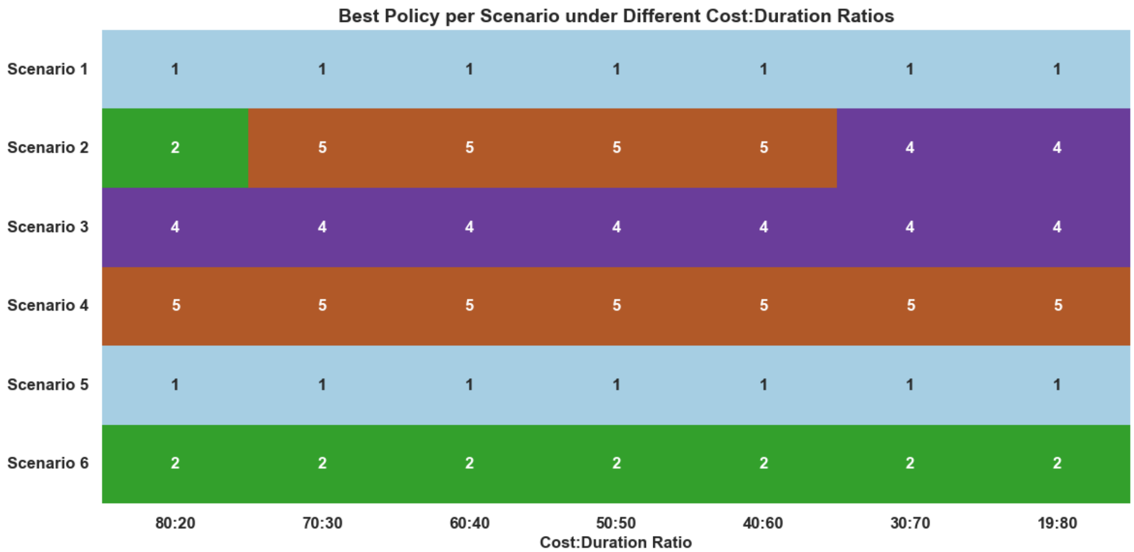

Figure 4.10: Best-performing policy per scenario across different cost:duration weighting ratios for the policies found under the separated PRIM thresholds. The heatmap highlights how policy preference shifts depending on the trade-off between cost and duration. Stable horizontal bands indicate insensitive policies, while transitions reflect sensitivity to weighting schemes.

To proceed with the DAPP-based schedule, a policy configuration must be selected for Scenario *Most Costly* (2)—the only scenario-policy combination that shows sensitivity. A 50/50 weighting between cost and duration is applied, leading to the selection of *No Excavation Trouble* Policy (5) as the designated policy.

4.2.4 Dynamic Adaptive Policy Pathways

4.2.4.1 Quantitative Assessment of Project Outcomes

Figure 4.11 shows that all six high-impact scenarios were sampled at least once across the five experimental seeds. While Scenario 5 and Scenario 6 are sampled most frequently, the full set of high-impact scenarios is represented. This distribution indicates that the experimental setup was capable of generating a diverse range of severe futures. Although ATPs were only triggered in approximately 5% of the runs, the fact that each high-impact scenario appeared as a trigger suggests that any of them could develop under the given uncertainty space. This supports the relevance of exploring adaptive measures throughout the schedule.

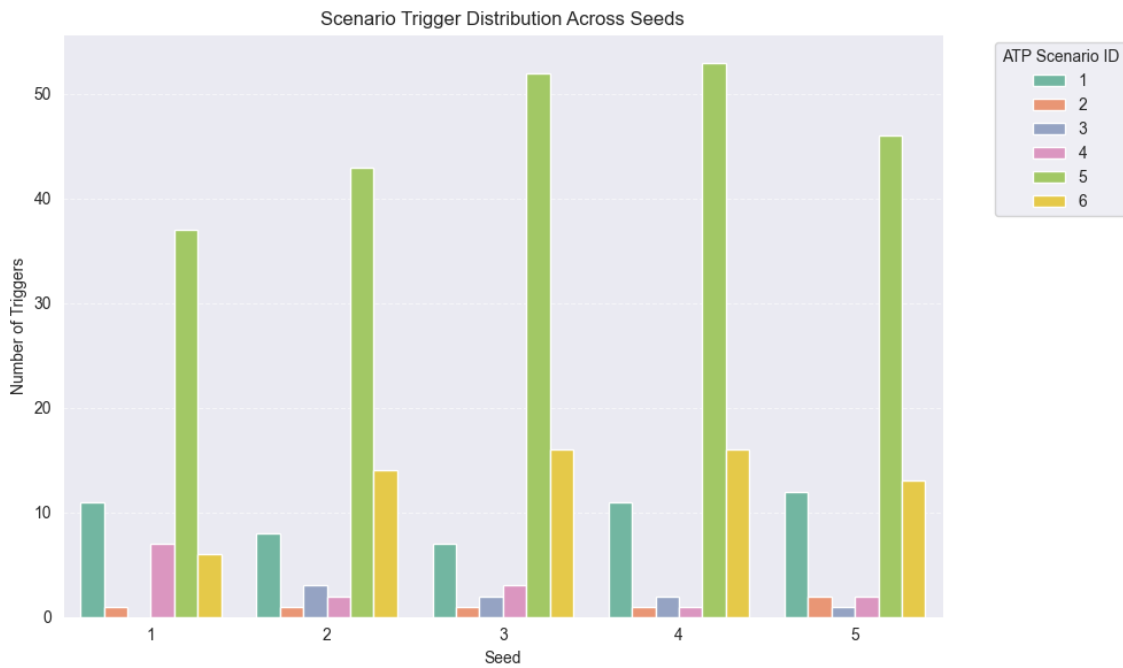

Figure 4.11: Distribution of high-impact scenarios sampled in the DES environment. Each seed is sampled 5,000 times. The bars represent the runs for which the ATPs are triggered.

To continue the analysis, the mean amount of samples per scenario across the seeds is taken, after which a sample of twenty triggered runs is selected for the comparison of the two DAPP-based schedules against the baseline. Twenty is selected because it is a large enough sample to have each scenario represented while being small enough for practical visualisation.

For the joint PRIM threshold of cost and duration, there were two final policies used in the DAPP-based on the scalar score. The baseline resulted in a shorter duration in 55% of the 20 runs. This can be seen in the upper graph of Figure 4.12, as the red dots appear on the left of the blue dots several times. In terms of cost, the DAPP-based schedule outperforms the baseline by approximately €100 million across the 20 samples.

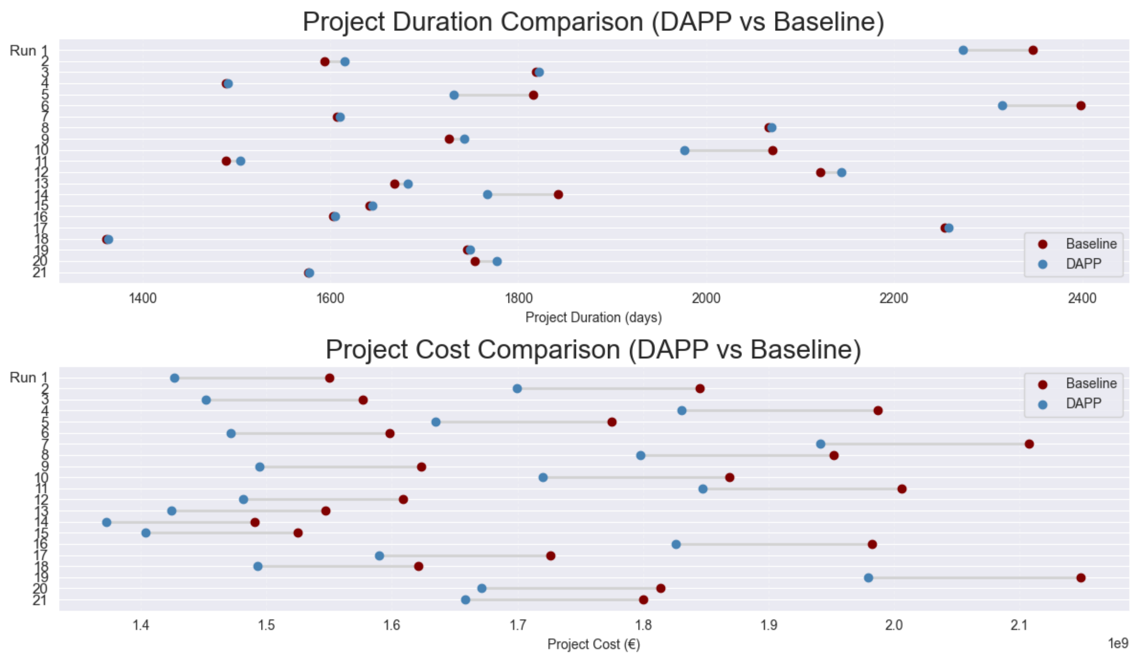

Figure 4.12: Project outcomes comparison between the DAPP schedule based on the joint PRIM threshold and the baseline schedule. On the axis duration and cost are set out against 20 DES runs for both schedules, The red dots represent the baseline schedule and the blue dots represent the DAPP schedule. DAPP is consistently cheaper while the baseline is shorter 55% of the runs.

The comparison for the DAPP schedule created with the policies found through the separated PRIM thresholds shows that the DAPP-based schedule outperforms the baseline schedule on both project objectives. Figure 4.13 shows consequently for all 20 samples that the DAPP schedule has a shorter duration and a cheaper cost.

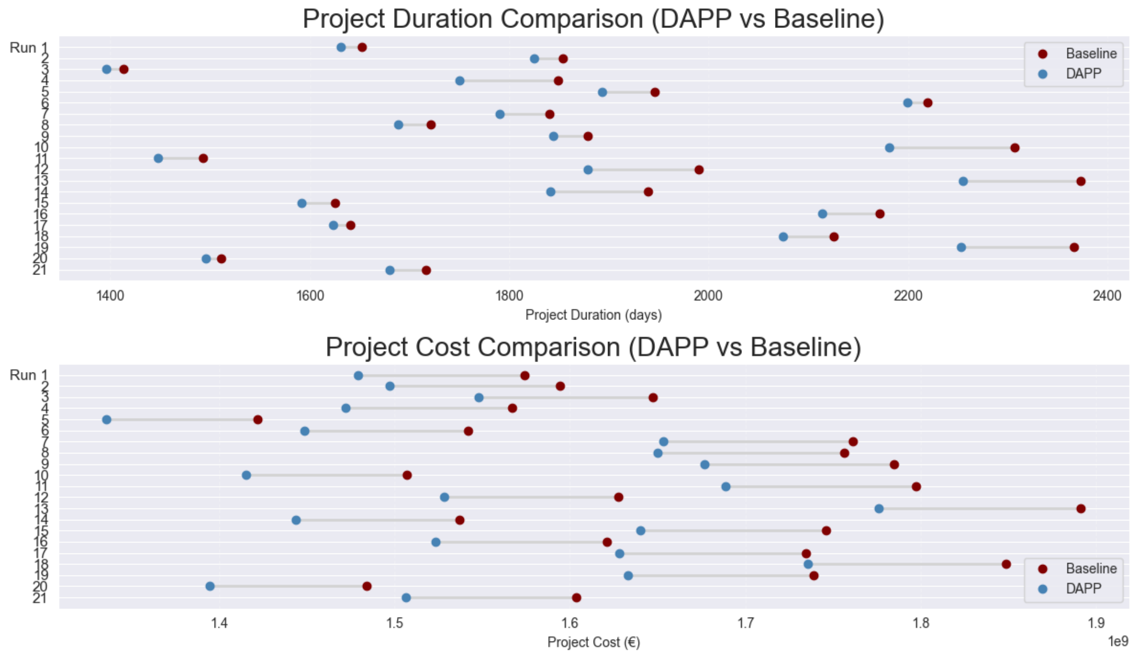

Figure 4.13: Comparison between the baseline schedule and the separate PRIM thresholds DAPP schedule in project outcomes. The DAPP-based schedule is both cheaper and quicker than the baseline schedule for the 20 simulated runs.

In the isolated experimental setup, the policies derived from the separated PRIM thresholds reduced project duration by an average of 67 days and cost by approximately €97,5 million across the 20 simulation runs. While the joint threshold runs reduced the cost on average by approximately €139 million, the reduction in duration was only 6 days.

while cost was added to the analysis to force a trade-off, the duration is based on a professionally created and used project planning for Schiphol bridge. Therefore, the final policy selection puts more emphasis on duration than on cost. This results in the policies from the separated PRIM thresholds to be selected for the final DAPP-based schedule.

4.2.4.2 Conceptual DAPP-Schedule

Figure 4.14 presents a conceptualisation of the final DAPP schedule, constructed around the six high-impact scenarios and their associated robust policies. The project starts at day 0, corresponding to August 30^{th}, 2021. Each pathway in the figure represents a distinct high-impact scenario identified in the open exploration. The values of the risk events and uncertainties — as defined in Table 4.2 — serve as ATPs. While the switching logic is condition-based, the occurrence of certain events is tied to specific activities in the schedule, making the triggering mechanism effectively time-dependent in some cases.

For each high-impact scenario, the associated risk events and uncertainties collectively function as a single ATP. In practice, these events are unlikely to occur simultaneously. Events such as *Critical Material Failure*, *Hard Soil Layer*, *Heavy Wind*, *UXO Found*, and *Foundation Size Issue* are modelled as task-specific risks, meaning they can only occur at particular points in the schedule.

In the conceptual DAPP schedule shown in Figure 4.14, an ATP is considered triggered when the final relevant task-specific risk within a scenario occurs. Each pathway formally begins with the deterministic Primavera baseline schedule and only diverges once a policy is activated to prevent the ATP from being triggered. Because timing is only defined by task-specific risks, the global risks and uncertainties—*influenza wave*, *RAI*, *task variability*, and *fuel cost multiplier*—do not influence the timing or structure of the pathways in this conceptual representation, even though they have real-world consequences. To enhance visual clarity, Figure 4.14 assigns each pathway a distinct colour and name from the initial white dot at Time 0, even though, in practice, the baseline Primavera schedule is shared across all pathways until the moment of divergence. Each dot along the timeline represents a risk realisation, but only the final task-specific risk triggers the ATP and thus determines the moment of adaptation. From that point onward, the baseline pathway and the policy pathway diverge, allowing for a direct comparison of the duration saved through the application of DAPP. In alignment with the simulation logic, the switch illustrated in the conceptual figure also occurs slightly before the ATP is reached—reflecting the 20% proximity threshold used to anticipate tipping points and initiate policies pre-emptively.

The final project durations shown in the Metro map are based on the results of the 20 comparative DES runs. Due to the mix of stochastic risks and continuous uncertainties, project outcomes vary slightly — even under identical Scenario-Policy combinations. To capture this variation while maintaining clarity, the final end dates for each DAPP pathway are based on the mean outcomes observed across these runs.

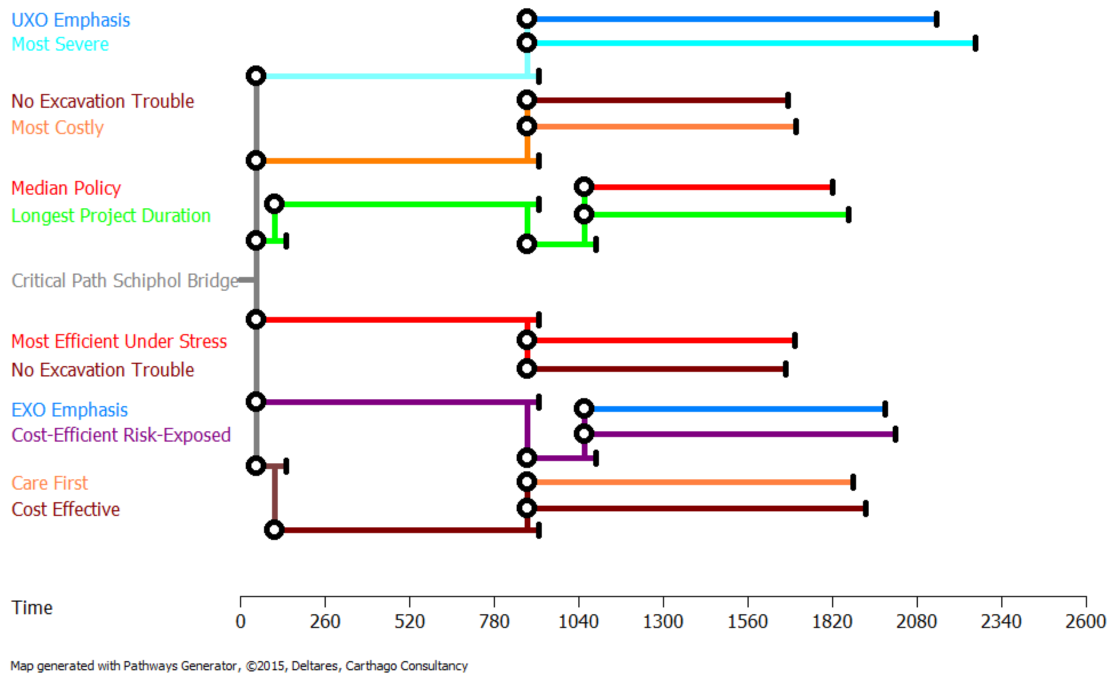

Figure 4.14: The DAPP starts with the baseline deterministic schedule as this is the desired project course. The ATPs equal the risk and uncertainty values specifically tied to each high-impact scenario. Over time an ATP can be triggered, recognising the developing high-impact scenario and its corresponding mitigation policy, initiating a pathway switch.

Table 4.3 summarises the benefits of adaptive scheduling by comparing project duration and cost under baseline and DAPP conditions for each high-impact scenario. Across all six scenarios, the application of robust policies consistently reduced both total project time and cost. These reductions demonstrate the practical value of incorporating scenario-specific adaptation strategies into the planning process. While variability remains due to stochastic risk dynamics, the average savings observed across the 20 DES runs highlight the potential of DAPP to enhance schedule robustness under deep uncertainty.

Table 4.3: Average Duration and Cost per High-Impact Scenario for the baseline and DAPP schedule on the entire project. The duration is in days. The cost is in euros.

Scenario	**Baseline Duration**	**Baseline Cost € (Billion)**	**DAPP Duration**	**DAPP Cost € (Billion)**	**Δ Duration (days)**	**Δ Cost (€ Million)**
1	2224	1.68	2105	1.57	119	101.89
2	1672	1.69	1648	1.59	25	102.75
3	1835	1.73	1784	1.62	51	105.04
4	1670	1.68	1641	1.58	29	101.69
5	1977	1.57	1948	1.48	30	93.88
6	1888	1.65	1847	1.55	42	100.50

4.3 Validation

For validation, the results of this study were presented to two Senior Project Leads at Count & Cooper who are actively involved in the Schiphol Bridge project. The objective was to evaluate whether the proposed methods could meaningfully improve schedule robustness under deep uncertainty, based on their professional judgement and experience.

During the discussion, two different strategies for applying the study's findings in practice were explored. The first focused on the use of the identified high-impact scenarios and their corresponding robust policies as a supporting tool within the project's risk management process. Currently, schedulers and project managers use stochastic risk analysis — typically in the form of Monte Carlo simulations — to calculate P-85 values, which estimate the level of buffer needed to ensure, with 85% probability, that the project will meet a given delivery date under modelled risks. While the quantitative DES assessment in this study could potentially offer similar insights, its objective was different.

Whereas Monte Carlo methods primarily confirm risks already identified during expert sessions, the exploratory modelling approach in this study was designed to discover high-impact scenarios that may not have been initially foreseen. Furthermore, the directed search phase enabled the identification of robust policies tailored to these scenarios. The Senior Project Leads recognised value in this exploratory capability, especially as a complement to existing workflows. In their view, integrating EMA alongside current risk management practices could provide deeper insight into the effects of deep uncertainty on project schedules and serve as a valuable extension of the tools already in use.

The second implementation strategy that was discussed involved using DAPP as a truly dynamic scheduling tool during project execution. This idea raised important questions from the Senior Project Leads, particularly regarding the practical feasibility of detecting and acting on ATPs in real time. Even if ATPs could be defined hypothetically, there remains considerable ambiguity around how such triggers would be operationalised to enable dynamic re-scheduling.

To illustrate this point, the *Heavy Wind* risk was used as a hypothetical example. In the current modelling setup, each ATP is defined by a combination of risks and uncertainties, with the latest occurring risk in the sequence serving as the actual trigger. In this example, *Heavy Wind* is assumed to be the final condition that must materialise for the ATP to be activated. Project leads raised a crucial concern: the challenge is not in measuring wind conditions themselves, which is technically feasible, but rather in forecasting such conditions early enough to allow for operational changes to the schedule.

In dynamic project environments, rescheduling takes time — work must be rescheduled, resources reassigned, and coordination adjusted. Therefore, for DAPP to be effective in practice, ATP detection must offer sufficient lead time, not just real-time confirmation. If a risk like *Heavy Wind* is only identifiable shortly before it occurs, it may be too late to adapt the schedule meaningfully, undermining the intended benefits of a DAPP-based approach.
Even under the assumption that ATPs can be monitored, applying a DAPP schedule at the entire project level may be too abstract to support meaningful, real-time decision-making. Greater potential may lie in applying DAPP to more localised, tactical detailed schedules — parts of the project that are still exposed to uncertainty but allow for more direct monitoring and intervention. In their professional opinion, however, DAPP in its current form is not yet ready to function as a dynamic control tool in operational construction scheduling.

Given the scope and level of abstraction of the simulation model, direct validation against the actual course of the Schiphol Bridge project was not pursued. The modelling framework used in this study is intentionally simplified to facilitate broad exploratory analysis under deep uncertainty. As such, it does not capture the full complexity of real-world project dynamics. Any resemblance between simulated scenarios and the realised project trajectory would therefore be coincidental rather than indicative of model accuracy. In light of this, validation through expert judgement was considered more appropriate and informative, aligning with the study's objective to assess methodological potential rather than reproduce historical outcomes.

5 Discussion

5.1 Model Limitations

This section examines how structural assumptions and simplifications in the model influence both the simulation outcomes and their interpretation. These limitations result from deliberate simplifications to manage model complexity, assumptions made in cost modelling, and the decision to represent certain risk events globally. Together, these factors affect the robustness and realism of the results produced by the simulation framework.

Although this study explicitly incorporates deep uncertainty into the modelling framework, the underlying logic that defines how uncertainties interact with measures and outcomes remains relatively underdeveloped. The design of the uncertainty space was based primarily on risk register inputs from the Schiphol bridge project, which included estimated cost impacts and odds of occurrence for various risks. While these ranges were informed by expert judgment, they reflect subjective assessments rather than empirical frequencies. As such, the precise likelihood of each risk occurring remains fundamentally uncertain.

To create a workable cost outcome within the simulation, daily rates had to be manually adjusted to fall within a comparable order of magnitude. This calibration step was necessary to ensure that cost and duration outcomes could be meaningfully compared within the model. As a result, the cost outputs used in this study should not be interpreted as realistic estimates of actual project expense. Instead, they function as a relative trade-off dimension within the MORDM framework. Their primary purpose is to demonstrate how robust policy search can be used to address deep uncertainty, rather than to provide precise financial forecasting.

Including slack time in this study follows industry standards and allows for critical path switches under deep uncertainty. However, this decision introduces an asymmetry in the modelling process: while project duration is only penalised after the 5% slack buffer is exceeded, cost begins to accumulate from the start of any disruption. This makes cost a more sensitive metric in marginal cases. While duration remains masked until critical delays occur, cost reflects incremental impacts immediately.

As a consequence, cost tends to weigh more heavily in both the scenario discovery and the directed search phases, influencing the peeling behaviour of PRIM and the dominance of cost in scalar scoring. This effect is a structural consequence of incorporating slack and should be taken into account when interpreting results.

In the scenario discovery phase, four out of the six high-impact scenarios included an influenza wave. The scope of the model is deliberately simplified, meaning that fewer variables are analysed than would be relevant in actual construction projects. Within this limited scope, the dynamics of the real system are further simplified to enhance the transparency and interpretability of the model logic. One example of this is the treatment of the influenza wave variable. Following the principles of ECM, the influenza wave is modelled as an event that reduces the RAI to 0.6 — a value that lies outside the regular bounds of its distribution (see Table 3.2). In the current implementation, this value is applied at the beginning of each simulation run, resulting in a constant 40% reduction in available resources throughout the entire project. This static treatment does not reflect the temporary and often uncertain nature of such disruptions in practice.

The use of the iterative PRIM cycle proposed by Guivarch et al. (2016) helped to counterbalance this modelling bias by identifying a third scenario family in which the influenza wave did not occur. Nevertheless, the first four high impact scenarios were discovered under a simplification that does not realistically reflect how such events unfold in practice. As a result, the influence of the influenza variable may be overstated, introducing a skew in how scenarios are classified as high-impact.

Three measures are modelled as non-boolean variables in the EMA workbench model. *budget buffer* and *schedule padding* are continuous whereas *back-up weekends* is categorical. As intended in the directed search, the MOEA searches for policies that are minimised on cost and duration. In this study, this resulted in the consistent exclusion of budget buffer and schedule padding from the resulting robust policies. This outcome can be explained by the inherent bias introduced through the minimisation objective. Both measures are anticipatory in nature, designed to provide resilience against unforeseen disruptions by allocating additional resources beforehand. However, when these measures are explicitly modelled as increasing baseline cost or project duration, without a mechanism to capture their conditional benefits, they are systematically penalised by the optimisation algorithm. In effect, they introduce guaranteed costs or delays in exchange for uncertain future gains. As a result, the evolutionary algorithm excludes them from the solution space under minimisation objectives.

As discussed in the validation chapter, the current workflow still relies on Monte Carlo simulations for risk management. Such techniques specifically calculate success rates depending on the aforementioned buffers, separating the two techniques in use. This contrast helps explain why *budget buffer* and *schedule padding* are consistently absent from the robust policies found in this study. Their exclusion does not imply irrelevance, but rather reflects how they were modelled: as cost and duration inflators without corresponding benefits. Because the optimisation seeks to minimise these objectives, and no compensating value is assigned to buffers within the model, they are systematically filtered out. This creates a gap between the theoretical results of this study and the practical preferences observed in real-world project management, where such buffers are considered crucial for absorbing unforeseen disruptions.

The exclusion of *back-up weekends* reflects different dynamics within the model. All six high-impact scenarios include the occurrence of a heavy wind risk event. Although back-up weekends were specifically introduced to mitigate this type of disruption, none of the robust policies included the measure. A plausible explanation is that the input cost assigned to back-up weekends was disproportionately high relative to the expected cost of wind-related delays. As a result, the optimiser consistently avoids selecting it. This outcome underscores the sensitivity of policy selection to cost assumptions and reveals a potential disconnect between model logic and practical judgement. In reality, a back-up weekend was used during the execution of the project, suggesting that practitioners viewed the measure as both viable and necessary. This discrepancy reinforces the idea that the model's cost assumptions may have biased the optimisation procedure against a mitigation strategy that holds practical value.

A further consequence of the model simplifications is that knock-on effects within the schedule are not explicitly represented. The SimPy-based simulation environment captures primary effects by applying both global and task-specific uncertainties directly onto the project schedule. These uncertainties influence task durations, resource availability, and weather-related disruptions, but their effects are confined to the immediate tasks they impact. In reality, however, delays often propagate through indirect channels—for example, through rescheduled subcontractors, delayed permits, or misaligned equipment logistics—causing broader disruptions across the project. The current model structure does not simulate such systemic or cascading effects, which means the true scale of disruption in high-impact scenarios may be under-represented, particularly in terms of cumulative cost and duration impacts.

5.2 Methodological Challenges

Beyond the model design, this section reflects on the practical and conceptual challenges in operationalising adaptive strategies. These challenges include ATP monitoring, the mismatch between conceptual and executable pathways, and the feasibility of implementing DAPP logic in real-world infrastructure settings.

The representation of ATPs differs in subtle but important ways between the conceptual DAPP schedule and the operational DES engine. In the metro map in Figure 4.14, policy switches are shown to occur only after the final task-specific risk materialises. This visualisation assumes that adaptation is triggered only once all the discrete risk conditions of a high-impact scenario are realised, effectively ignoring the influence of continuous uncertainties. In contrast, the DES engine evaluates both continuous uncertainties, which are modelled as fixed values at the start of each run, and discrete risks, which materialise dynamically during execution. An ATP is triggered when 80% of the defining conditions of a high-impact scenario are met. As a result within the simulation, continuous uncertainties can push the system toward an ATP earlier than the conceptual map suggests.

However, even when this threshold is met, switching can only occur at predefined task-linked nodes. Given that this study models a limited number of task-specific risk events, the set of possible ATP trigger points is relatively sparse. As a result, the discrepancy between the conceptual and operational ATP timing in this current setting is small in terms of the number of tasks. Still, the difference in actual timing in days may be more pronounced, since some risk-related tasks span longer durations than others. This relationship is further influenced by how the critical path is structured. A richer set of task-dependent risk events would allow for more granular ATP triggering and a closer alignment between conceptual design and implementation.

The limited number of task-specific risks in this study simplifies ATP switching logic but also highlights a deeper implementation issue: the practical difficulty of detecting ATP conditions in real time. While the simulation framework defines adaptation points based on scenario proximity and task triggers, applying this approach in actual projects would require the ability to monitor evolving conditions across a granular, network-wide schedule. This level of detail, while valuable for analysis, exposes the challenge of project-wide ATP recognition—especially when uncertainties span both continuous variables and discrete events.

Another challenge concerns the feasibility of monitoring ATP conditions during project execution. While the analysis identifies which measures are effective under severe scenarios, it remains unclear how these conditions could be reliably recognised in real time. In the absence of a mechanism for scenario detection or ATP evaluation during execution, dynamic adaptation may fail to respond proactively and instead default to reactive behaviour. As a result, although this study advances the conceptual understanding of adaptive planning, it offers only a partial foundation for operationalising DAPP in real-world infrastructure delivery.
The results of the quantitative assessment show that the DAPP-based schedule consistently outperforms the baseline in terms of both cost and duration. However, this finding must be interpreted in light of how the baseline was constructed. It includes no mitigation measures and represents a purely reactive planning approach under deep uncertainty. Because the baseline simulation lacks a mechanism for monitoring the evolution of scenarios, generic contingency planning is infeasible; responses must be tailored to specific unfolding events. While it would have been possible to model multiple static baseline strategies, doing so would merely replicate conventional practice. The aim of this study was not to compare static alternatives, but to demonstrate how EMA can identify critical scenarios and support the design of robust adaptive strategies capable of responding to them.

This study offers a novel methodological contribution by applying EMA and DAPP to a highly granular, task-based infrastructure schedule. Prior applications of these frameworks—such as those by Haasnoot et al. (2013) and Michas et al. (2020)—have focused on long-term or conceptual planning contexts. In contrast, the current study embeds adaptive pathways within a detailed discrete event simulation model based on an operational infrastructure construction schedule. As noted by Stanton and Roelich (2021), most documented DMDU applications are found in high-level planning domains such as water and energy systems, with little to no prior work at the level of detailed project scheduling. A recent exception is the study by Feng et al. (2023), which demonstrates the feasibility of combining discrete event simulation and scenario discovery in a construction context. Their work highlights the conceptual potential of DMDU methods for project scheduling, using a stylised model and binary disruption outcomes to explore scenario sensitivity. Building on this foundation, the current study extends the approach to an operational project schedule, introducing continuous stressors and embedding adaptive logic within a granular EMA pipeline. By structuring adaptation around task-specific risk events and simulating their effects across a full project network, this study reveals new challenges in both modelling and implementation. These findings suggest that while EMA and DAPP hold significant promise for infrastructure planning, their operationalisation at the task level demands new methods for monitoring, triggering, and real-time policy adjustment.

6 Conclusion

Deep uncertainty in infrastructure construction scheduling arises when planners cannot confidently describe system behaviour, assign probabilities to outcomes, or agree on how those outcomes should be evaluated. While traditional risk management addresses discrete, probabilistic events, deep uncertainty emerges in early planning stages where knowledge is limited or contested. This study distinguishes between aleatoric uncertainty, reflecting natural variability, and epistemic uncertainty, arising from incomplete knowledge. When epistemic uncertainty exceeds the scope of existing planning tools, deep uncertainty emerges. This is particularly relevant for early-stage decisions that influence long-term project outcomes, which is why this study focuses on infrastructure construction scheduling.

To address deep uncertainty, this study applies EMA, a framework designed to explore a wide range of plausible futures without relying on fixed models or known probability distributions. Instead of optimising for a single expected outcome, EMA supports the development of strategies that remain effective across many possible conditions. Scenario discovery is used to identify combinations of uncertainties consistently associated with poor project outcomes in terms of both cost and duration. By applying Latin hypercube sampling to a highly granular, task-level construction schedule—focusing specifically on tasks along the critical path and within a 5% slack threshold—10,000 scenario runs were generated with embedded risks and uncertainties derived from the project's risk register. This process led to the identification of six high-impact scenarios. Each high-impact scenario comprises a distinct combination of risk events and uncertainty values, which together define an ATP. These ATPs form the basis for constructing a DAPP schedule.

In the directed search phase, predefined measures are optimised separately for each high-impact scenario, resulting in six tailored policies, each of which performs well under its corresponding scenario but tends to fail when applied to others. Given that deep uncertainty makes it impossible to reliably predict how the future will unfold, this reinforces the impracticality of relying on scenario-specific strategies. Instead, robust policies are designed to perform reasonably well across a wide range of conditions, offering greater resilience against the unknown. Using MORDM, this study evaluates cost–duration trade-offs to identify policies that perform acceptably across all six high-impact scenarios. A scalar function was applied to normalised cost and duration outcomes to help decision-makers prioritise policies based on their implicit preferences between the two objectives. Based on this process, four robust policies were retained. Notably, three measures—*new Design*, *overtime labour*, and *electric machinery*—appear consistently across the retained policies, suggesting that a focused subset of actions can meaningfully improve schedule robustness.

These robust policies were then embedded into a DES environment, enabling the construction of a DAPP schedule capable of adapting to unfolding uncertainty. To enable adaptation, a simplified switching logic was implemented, activating a policy when a scenario's conditions were matched within 20% of the defined ATP threshold. While this does not fully solve the challenge of ATP detection, it allows for anticipatory switching before a tipping point is crossed. This operationalisation demonstrates how DAPP can function within a simulated environment, providing not only performance metrics but also insights into how often specific scenarios are triggered. Although these frequencies are not real-world probabilities, they serve as proxies for prioritising mitigation strategies during early planning—adding a supplementary decision variable alongside robustness and cost-duration preferences.

This study demonstrates that the EMA and DAPP frameworks can be used to improve schedule robustness in infrastructure construction projects. When used alongside existing risk management practices, EMA and DAPP can enhance schedule robustness by uncovering high-impact scenarios that are structurally distinct from traditional risk events. This insight is supported by principal component analysis on the LHS-generated uncertainty space. Through the PRIM experiments and robust policy evaluation in the directed search, a consistent set of adaptive measures concerning both cost and duration was identified and structured into a conceptual DAPP schedule. This offers decision-makers a method for treating deep uncertainty not as background noise, but as a central component of early-stage risk analysis. At the same time, the study highlights the current limitations of dynamic scheduling. While ATPs offer a promising way to structure pathway switching, the actual use of scenario-based triggers remains far from operational. To implement DAPP dynamically, planners would require mechanisms for initiating policies from the first relevant signal and continuously monitoring the unfolding uncertainty space—capabilities that are not yet embedded in current project delivery systems. As such, the results provide a functional modelling framework and decision support tool, while also clarifying what gaps must be addressed to move from conceptual insight to practical implementation.

6.1 Recommended Work

6.1.1 Selecting Robust Policies in Discrete Solution Spaces

This study explored the use of the MSMOP approach for identifying robust policies across a set of high-impact scenarios. While the method is conceptually well-aligned with the goal of cross-scenario robustness, its practical implementation was limited by the structure of the policy space. Many of the available measures in construction planning are discrete or binary, which MSMOP is not inherently equipped to handle. As a result, the technique proved difficult to apply directly in a solution space composed of boolean decision variables.

By contrast, the scenario discovery phase performed well. Out of 10,000 Latin hypercube samples, the iterative approach based on Guivarch et al. (2016) successfully revealed distinct families of high-impact scenarios, as confirmed through principal component analysis. These scenarios were then used to generate six optimised policies, each tailored to one scenario. To assess their robustness, these policies were evaluated in a full factorial design across all six scenarios. This generated a much smaller solution space than the original 10,000-sample uncertainty space and allowed for systematic performance comparison.

However, the policy evaluation phase lacked a robust selection method for boolean policy spaces. The iterative PRIM method could not be repeated effectively in the solution space, and as such, only conventional PRIM analysis was applied using cost and duration as outcome variables. This process involved manual experimentation with different thresholds and aggregation techniques, which limited its scalability and interpretability.

Future research could explore how robust policy selection can be better structured in small, discrete solution spaces. One direction would be to adapt the MSMOP method to support boolean or categorical decision variables, making it more suitable for real-world construction planning. Alternatively, new methods could be developed or repurposed that explicitly target robustness in binary measure configurations. Either pathway would advance the applicability of scenario-informed robust planning in domains where decision levers are limited and non-continuous.

6.1.2 Iron Triangle of Infrastructure Project Management

Incorporating scope as a third objective would complete the *iron triangle* of project management (Egboga and Daniel, 2022). This classic framework describes the three fundamental and interrelated constraints in project delivery: time, cost, and scope. In construction, it reflects the reality that gains in one area often require compromises in the others—for example, expanding the project scope may increase both duration and cost. Including scope would significantly improve the model's resemblance to real-world decision-making, where such trade-offs are central during early planning and tendering. In this study, duration was the primary metric, directly derived from the project schedule. Cost, by contrast, was introduced mainly to enable trade-off analysis, but was not firmly based on real project budgets. If cost can be more closely tied to actual financial structures, just as duration was based on task logic, and then meaningfully connected to scope or quality, the model would offer even greater value for strategic planning and schedule development under deep uncertainty.

6.1.3 Knock-on Effects

While this study focused on primary scheduling effects using a DES structure, it did not fully capture the broader impact of resource constraints and knock-on effects. In real-world projects, such constraints often trigger secondary and tertiary effects, such as delays in parallel activities or disruptions in subcontractor coordination. These effects go beyond direct task logic and require more dynamic modelling. Future research could integrate Agent-Based Modelling (ABM) into the EMA pipeline of this study to simulate interactions between agents like workers, equipment, and planners. ABM would allow for a more realistic representation of cascading delays and adaptive behaviours, extending the current framework to reflect the operational complexity of construction projects.

Declaration of Interest

This research was conducted under a tripartite agreement between TU Delft, Count & Cooper, and the author, Daan Nienhuis. The author has been employed part-time at Count & Cooper during his studies at TU Delft. Aside from an internship allowance, this research received no external funding. The results are publicly available, and the author declares no competing interests.

Declaration on the use of Artificial Intelligence

Throughout this study artificial intelligence has been used to assist with finding sources, coding as well as writing. At no point was AI used to generate original research findings or to fabricate results. All conceptual insights, modelling decisions, and data analysis were developed and verified by the author. The final responsibility for the content, interpretation, and conclusions of this work lies entirely with the author.

Bibliography

Ahmad, H. S., Ayoush, M. D., and Al-Alwan, M. S. (2020). Causes of delay to public infrastructure projects according to engineers representing different contract parties. *Built Environment Project and Asset Management*, 10:153–179.

Ahmadu, H. A., Ibrahim, A. D., Ibrahim, Y. M., and Adogbo, K. J. (2020). Incorporating aleatory and epistemic uncertainties in the modelling of construction duration. *Engineering, Construction and Architectural Management*, 27:2199–2219.

Aven, T. (2012). *How to Assess Uncertainties and Specify Probabilities*, pages 71–109. John Wiley and Sons, Ltd. Wiley Online Books.

Aven, T. (2013). On the meaning of a black swan in a risk context. *Safety Science*, 57:44–51.

Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, 253:1–13.

Bankes, S. (1993). Exploratory modeling for policy analysis. *Operations Research*, 41:435–449.

Bankes, S., Walker, W. E., and Kwakkel, J. H. (2013). *Encyclopedia of Operations Research and Management Science*. Springer US.

Bankes, S. C., Lempert, R. J., and Popper, S. W. (2001). Computer-assisted reasoning. *Computing in Science and Engineering*, 3:71–72.

Bryant, B. P. and Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting and Social Change*, 77:34–49.

Chen, D., Xiang, P., Jia, F., and Guo, J. (2022). A systematic review of current status and trends of mega-infrastructure projects. *Ain Shams Engineering Journal*, 13(6):101773.

Chung, T. H., Mohamed, Y., and AbouRizk, S. (2004). Simulation input updating using bayesian techniques. In *Proceedings of the 2004 Winter Simulation Conference, 2004.*, volume 2, pages 1238–1243 vol.2.

Count & Cooper (2023). Areal overview schiphol bridge. Unpublished internal document.

Dutta, S. and Gandomi, A. H. (2020). *Chapter 15 - Design of experiments for uncertainty quantification based on polynomial chaos expansion metamodels*, pages 369–381. Butterworth-Heinemann.

Egboga, I. and Daniel, C. (2022). How relevant is the iron triangle as a measure of construction project performance. *World Journal of Management and Business Studies*, 2:1–9.

Fang, C., Marle, F., Xie, M., and Zio, E. (2013). An integrated framework for risk response planning under resource constraints in large engineering projects. *IEEE Transactions on Engineering Management*, 60:627–639.

Feng, K., Chen, S., Lu, W., Wang, S., Yang, B., Sun, C., and Wang, Y. (2023). Embedding ensemble learning into simulation-based optimisation: a learning-based optimisation approach for construction planning. *Engineering, Construction and Architectural Management*, 30:259–295.

Feng, K., Wang, S., Lu, W., Liu, C., and Wang, Y. (2022). Planning construction projects in deep uncertainty: A data-driven uncertainty analysis approach. *Journal of Construction Engineering and Management*, 148.

Flyvbjerg, B. (2014). What you should know about megaprojects and why: An overview.

Flyvbjerg, B., Buhl, S., and Skamris, M. K. (2002). Editorial board page. *Journal of the American Planning Association*, 68:C–C.

Flyvbjerg, B., Buhl, S. L., and Skamris, M. K. (2005). How (in)accurate are demand forecasts in public works projects. *Journal of the American Planning Association*, 71:131–146.

Fox, C. and Ülkümen, G. (2011). *Distinguishing Two Dimensions of Uncertainty*, chapter 1. SSRN.

Friedman, J. H. and Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, 9:123–143.

Giudici, F., Castelletti, A., Giuliani, M., and Maier, H. R. (2020). An active learning approach for identifying the smallest subset of informative scenarios for robust planning under deep uncertainty. *Environmental Modelling and Software*, 127.

Guivarch, C., Rozenberg, J., and Schweizer, V. (2016). The diversity of socio-economic pathways and co2 emissions scenarios: Insights from the investigation of a scenarios database. *Environmental Modelling and Software*, 80:336–353.

Gómez Sánchez, M., Lalla-Ruiz, E., Fernández Gil, A., Castro, C., and Voß, S. (2023). Resource-constrained multi-project scheduling problem: A survey. *European Journal of Operational Research*, 309(3):958–976.

Haasnoot, M., Kwakkel, J. H., Walker, W. E., and ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change*, 23:485–498.

Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using networkx. In Varoquaux, G., Vaught, T., and Millman, J., editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA.

Hamarat, C., Kwakkel, J., and Pruyt, E. (2013). Adaptive robust design under deep uncertainty. *Technological Forecasting and Social Change*, 80(3):408–418.

Kasprzyk, J. R., Nataraj, S., Reed, P. M., and Lempert, R. J. (2013). Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling & Software*, 42:55–71.

Kim, B.-C. and Reinschmidt, K. F. (2009). Probabilistic forecasting of project duration using bayesian inference and the beta distribution. *Journal of Construction Engineering and Management*, 135(3):178–186.

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Houghton Mifflin Co.

Kwadijk, J. C., Haasnoot, M., Mulder, J. P., Hoogvliet, M. M., Jeuken, A. B., van der Krogt, R. A., van Oostrom, N. G., Schelfhout, H. A., van Velzen, E. H., van Waveren, H., and de Wit, M. J. (2010). Using adaptation tipping points to prepare for climate change and sea level rise: A case study in the netherlands. *Wiley Interdisciplinary Reviews: Climate Change*, 1:729–740.

Kwakkel, J. H. (2017). The exploratory modeling workbench: An open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. *Environmental Modelling and Software*, 96:239–250.

Kwakkel, J. H. and Pruyt, E. (2013). Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty. *Technological Forecasting and Social Change*, 80:419–431.

Larsson, J., Lu, W., Krantz, J., and Olofsson, T. (2016). Discrete event simulation analysis of product and process platforms: A bridge construction case study. *Journal of Construction Engineering and Management*, 142.

Lau, S. E. N., Zakaria, R., Aminudin, E., Saar, C. C., Yusof, A., and Wahid, C. M. F. H. C. (2018). A review of application building information modeling (bim) during pre-construction stage: Retrospective and future directions. In *IOP Conference Series: Earth and Environmental Science*, volume 143. Institute of Physics Publishing.

Lempert, R. J., Popper, S. W., and Bankes, S. C. (2003). *Shaping the next one hundred years : new methods for quantitative, long-term policy analysis and bibliography.* RAND.

Luo, L., He, Q., Jaselskis, E. J., and Xie, J. (2017). Construction project complexity: Research trends and implications. *Journal of Construction Engineering and Management*, 143.

Maier, H. R., Guillaume, J. H., van Delden, H., Riddell, G. A., Haasnoot, M., and Kwakkel, J. H. (2016). An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? *Environmental Modelling and Software*, 81:154–164.

Marchau, V. A. W. J., Walker, W. E., Bloemen, P. J. T. M., and Popper, S. W., editors (2019a). *Dynamic Adaptive Policy Pathways (DAPP)*, pages 71–92. Springer International Publishing, Cham.

Marchau, V. A. W. J., Walker, W. E., Bloemen, P. J. T. M., and Popper, S. W., editors (2019b). *Introduction*, pages 1–22. Springer International Publishing.

McCann, B. T. (2020). Using bayesian updating to improve decisions under uncertainty. *California Management Review*, 63:26–40. doi: 10.1177/0008125620948264.

Meurer, A., Smith, C. P., Paprocki, M., Čertík, O., Kirpichev, S. B., Rocklin, M., Kumar, A., Ivanov, S., Moore, J. K., Singh, S., Rathnayake, T., Vig, S., Granger, B. E., Muller, R. P., Bonazzi, F., Gupta, H., Vats, S., Johansson, F., Pedregosa, F., Curry, M. J., Terrel, A. R., Roučka, v., Saboo, A., Fernando, I., Kulal, S., Cimrman, R., and Scopatz, A. (2017). Sympy: symbolic computing in python. *PeerJ Computer Science*, 3:e103.

Michas, S., Stavrakas, V., Papadelis, S., and Flamos, A. (2020). A transdisciplinary modeling framework for the participatory design of dynamic adaptive policy pathways. *Energy Policy*, 139.

Mohd Nasir, N., Mohd Nawi, M. N., and Tapa, A. (2016). Dilemma in decision making at pre-construction phase -a state of art review dilemma in decision making at pre-construction phase -a state of art review. In *Dilemma in decision making at pre-construction phase -A state of art review Dilemma in Decision Making at Pre-construction Phase -A state of art review*, volume 1761.

Moret, Y. and Einstein, H. H. (2016). Construction cost and duration uncertainty model: Application to high-speed rail line project. *Journal of Construction Engineering and Management*, 142(10):2–3.

Okmen and Oztas, A. (2014). Uncertainty evaluation with fuzzy schedule risk analysis model in activity networks of construction projects. *Journal of the South African Institution of Civil Engineering*, 56:9–20.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.

Project Management Institute (2021). *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)*. Project Management Institute, 7th edition.

Qiao, Y., Labi, S., and Fricker, J. D. (2019). Hazard-based duration models for predicting actual duration of highway projects using nonparametric and parametric survival analysis. *Journal of Management in Engineering*, 35.

Ruckert, K. L., Srikrishnan, V., and Keller, K. (2019). Characterizing the deep uncertainties surrounding coastal flood hazard projections: A case study for norfolk, va. *Scientific Reports*, 9.

Sadeghi, N., Fayek, A. R., and Pedrycz, W. (2010). Fuzzy monte carlo simulation and risk assessment in construction. *Computer-Aided Civil and Infrastructure Engineering*, 25:238–252.

Sadeghi, N., Fayek, A. R., and Seresht, N. G. (2015). Queue performance measures in construction simulation models containing subjective uncertainty. *Automation in Construction*, 60:1–11.

Saltelli, A., Chan, K., and Scott, E. (2000). *Sensitivity Analysis*, volume 134. Wiley.

Shavazipour, B., Kwakkel, J. H., and Miettinen, K. (2021). Multi-scenario multi-objective robust optimization under deep uncertainty: A posteriori approach. *Environmental Modelling and Software*, 144.

Stanton, M. C. and Roelich, K. (2021). Decision making under deep uncertainties: A review of the applicability of methods in practice. *Technological Forecasting and Social Change*, 171.

Taleb, N. N. (2020). *Statistical consequences of fat tails : real world preasymptotics, epistemology, and applications : papers and commentary.* STEM Academic Press.

Tegeltija, M., Oehmen, J., Kozine, I., and Geraldi, J. (2016). Post-probabilistic uncertainty quantification: Discussion of potential use in product development risk management. Technical report, institution.

Virine, L. and Trumper, M. (2013). Event chain methodology: Managing event chains. In Virine, L. and Trumper, M., editors, *Project Risk Analysis Made Ridiculously Simple*, chapter 8, pages 121–136. World Scientific Publishing Co., Singapore.

Walker, W. E., Lempert, R. J., and Kwakkel, J. H. (2013). *Encyclopedia of Operations Research and Management Science*. Springer US.

Wang, J. and Yuan, H. (2017). System dynamics approach for investigating the risk effects on schedule delay in infrastructure projects. *Journal of Management in Engineering*, 33.

Wang, Z., Zhang, Y., Zheng, K., Zeng, R., Yuan, H., and Liu, J. (2023). A review of mega-project management research from an organization science perspective: Current status and future directions.

Wied, M., Oehmen, J., Welo, T., and Pikas, E. (2021). Wrong, but not failed? a study of unexpected events and project performance in 21 engineering projects. *International Journal of Managing Projects in Business*, 14.

Williams, T. and Samset, K. (2010). Issues in front-end decision making on projects. *Project Management Journal*, 41:38–49.

Zamani, V., Yavari, E., and Taghaddos, H. (2024). A science mapping lens on discrete event simulation applications in construction engineering and management.

A Appendix

A.1 Research design

A.1.1 Extended flow chart of research design

This flow chart represents the entire research design. The rectangles represent intermediary products in the sequential process and the diamonds represent techniques applied to go from one product to the next. The green boxes are the products of the open exploration and the directed search and are the two main inputs for creating the DAPP. At step 3 and step 4 information is added to the model

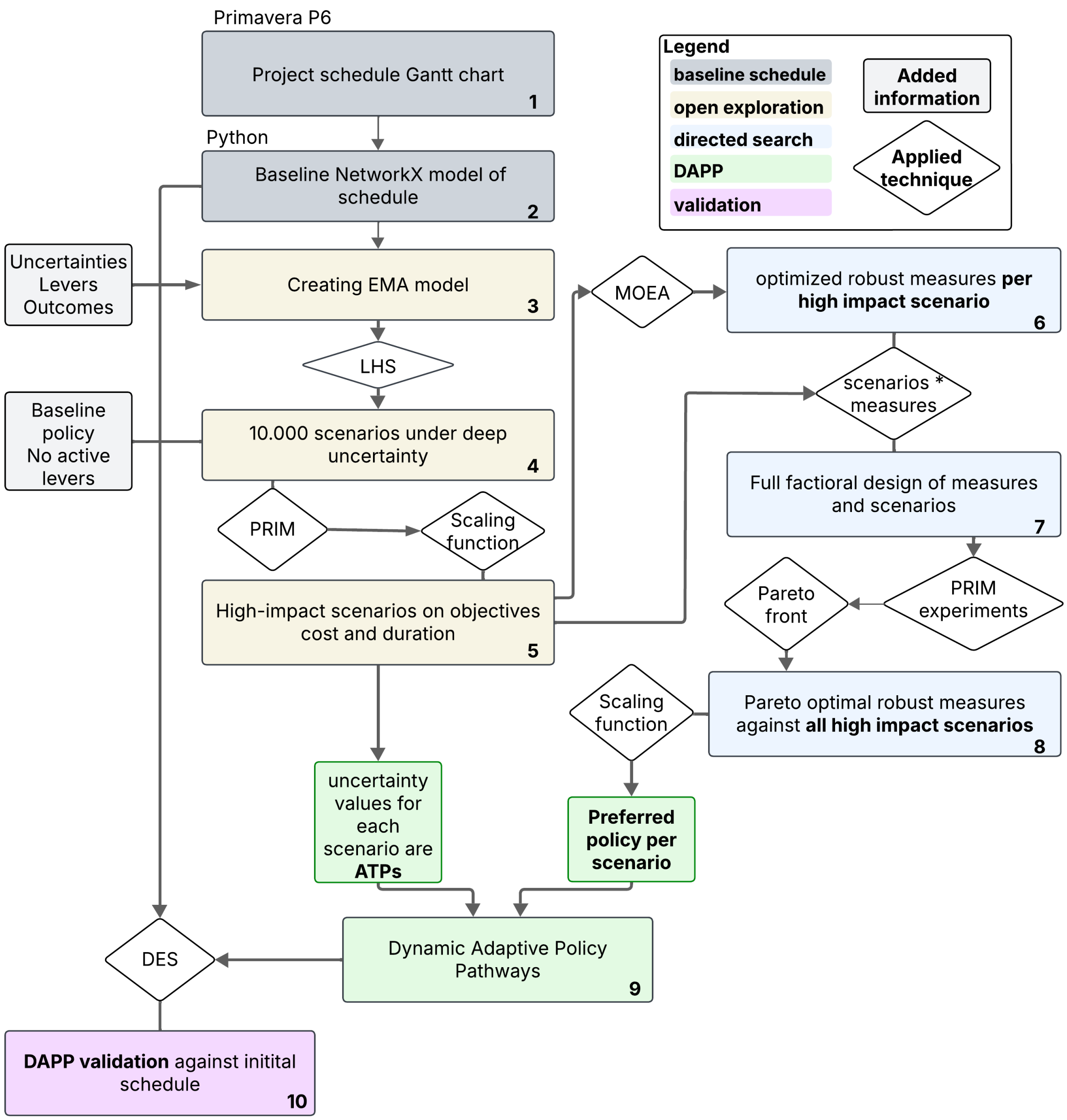

Figure A.1: Flow chart of the full research design. Each rectangle box represents one of the sequential steps in this project pipeline. Each rectangle box produces an intermediary product that is the input for the following step. The diamonds represent the used technique between the steps. In step 3 and 4, information is added to the model.

A.2 Results

A.2.1 Critical path and slack time

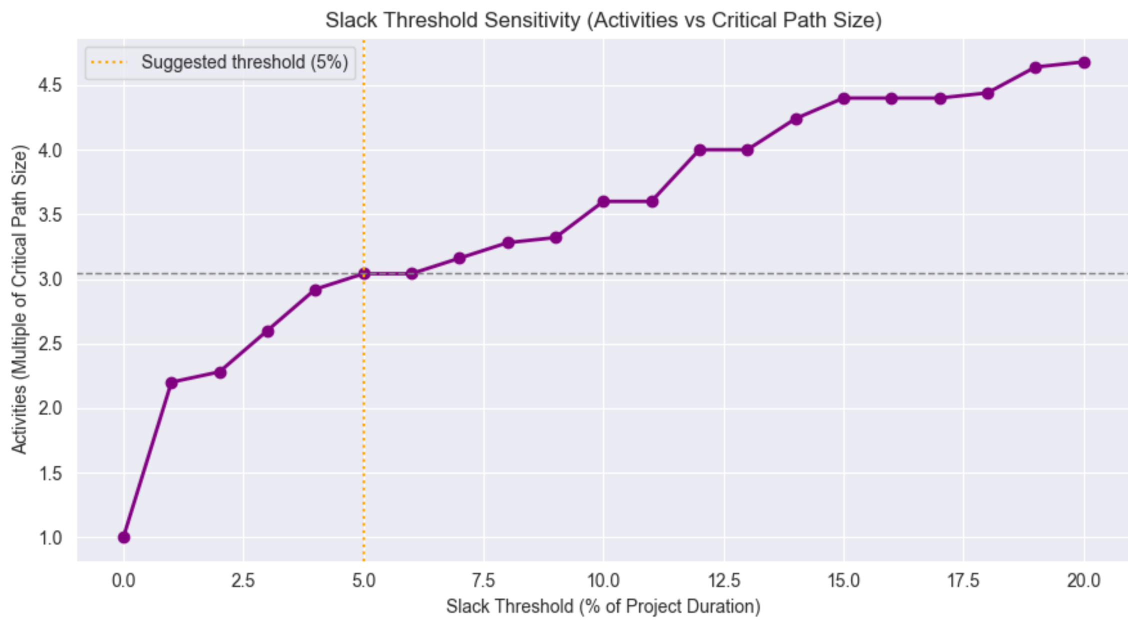

Figure A.2: Sensitivity analysis of added slack time. The X-axis shows the slack percentage, and the Y-axis presents the multiplier of activities relative to the critical path. After 5% slack, the curve flattens, indicating that the number of near-critical tasks increases only marginally. A threshold of 5% is selected as the best trade-off between scope and marginal insight.

A.2.2 Scenario discovery

A.2.2.1 PRIM iterations

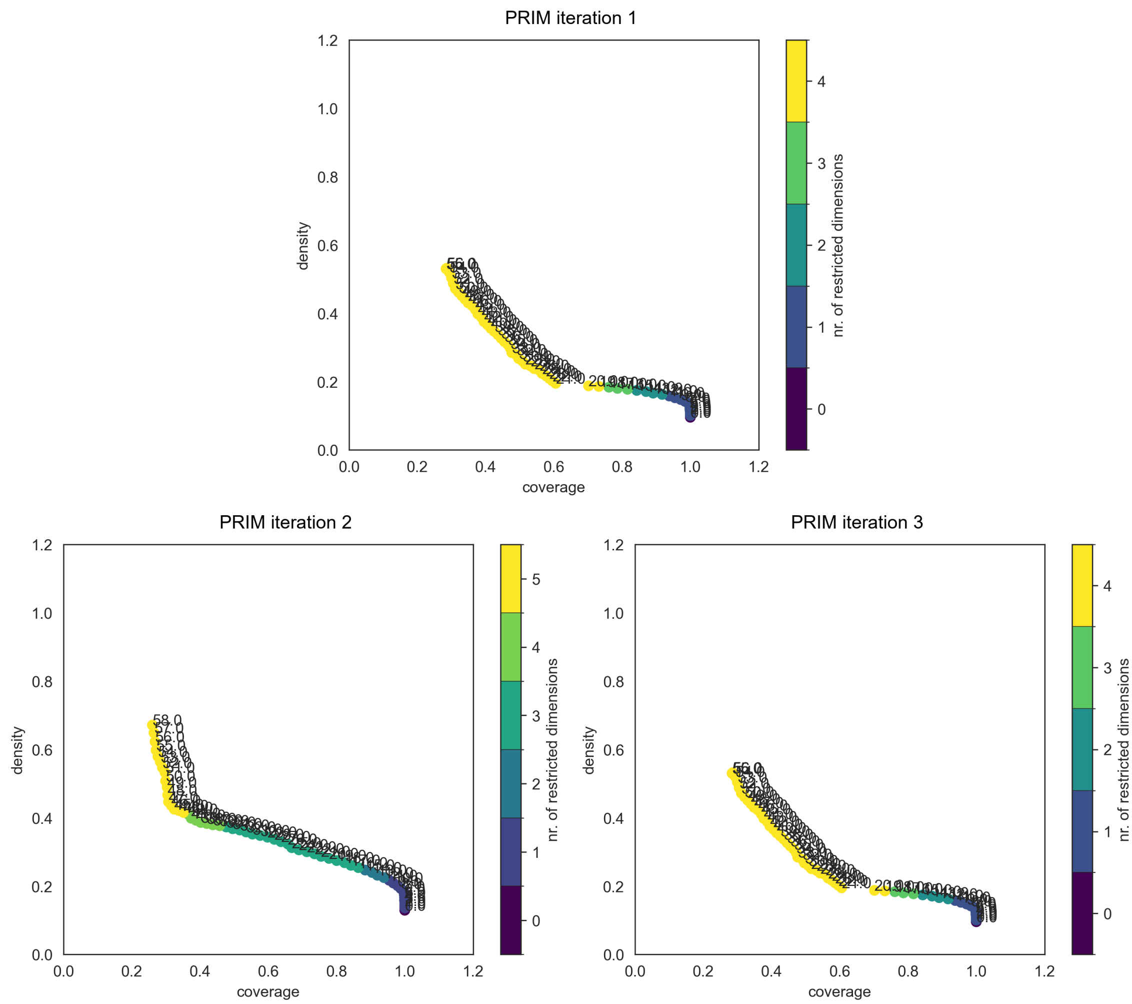

Figure A.3: The three iterations of PRIM to find scenario families that stress the schedule on cost and duration. Each iteration follows up the other. Each time the scenario family that makes the most impact is filtered out to force PRIM to find a new set of rules for the following iteration.

A.2.3 Directed search

A.2.4 Convergence metrics

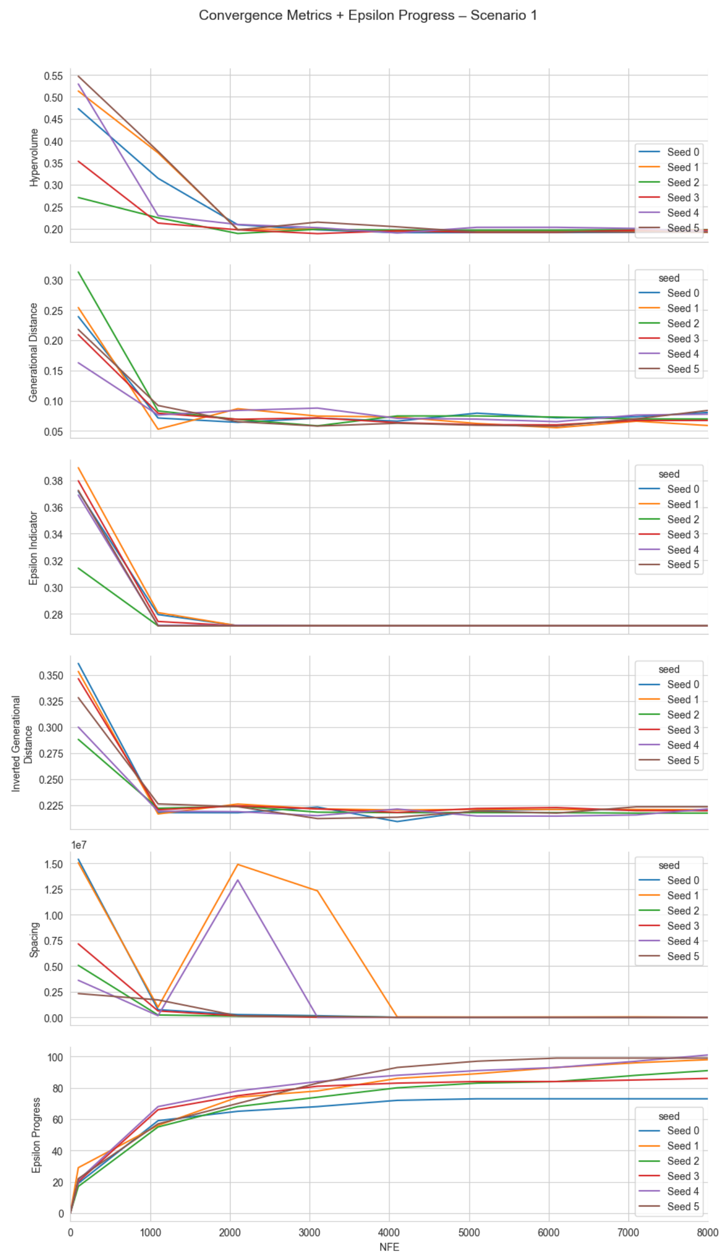

Figure A.4: Hypervolume, Generational distance, Epsilon indicator, Inverted generational distance, spacing, and Epsilon progress are measured during optimisation. Once all metrics stabilise, the algorithm can stop running evaluations. 8,000 nfe is chosen for the optimisation. These metrics are from scenario 1 across all five seeds.

A.2.4.1 Directed Search and the role of PRIM in policy space

MSMOP integrates robustness directly into the optimisation process by simultaneously evaluating multiple objectives across multiple plausible futures (Shavazipour et al., 2021). Unlike post-sampling approaches, which evaluate solutions after scenario-specific optimisation, MSMOP searches directly for policies that are consistently effective—if not dominant—in all scenarios.

MSMOP is designed for continuous decision spaces and does not support discrete or mixed-variable optimisation without significant adaptation. In this study, several strategic decisions—such as predrilling, redesign, and spare part options—are modelled as boolean measures. Applying MSMOP would require fixing these discrete variables in advance, thereby restricting the search space and limiting the ability to explore trade-offs among key decisions.

As an alternative to MSMOP, this study applies PRIM in the directed search phase to identify input conditions consistently associated with robust policy performance. Although PRIM is more commonly used for scenario discovery, this study applies it twice: first, to identify high-impact regions of the uncertainty space; and later, within the directed search phase, to explain why certain robust policies perform well within those same scenario-defined conditions. As discussed earlier, the sequence of scenario discovery and policy search in EMA is flexible; this study adopts a scenario-first strategy throughout, with scenario discovery guiding both policy design and its post-hoc explanation. A related but inverted approach is found in Hamarat et al. (2013), where a policy-first strategy is followed: candidate policies are evaluated first, and PRIM is then used to uncover the scenario conditions under which they fail. While their use of PRIM helps map vulnerabilities after policy design, this study uses PRIM to reveal which policies of measures consistently yield robust outcomes across all high-impact scenarios. By treating the policy measures explanatory variables, PRIM helps expose generalisable design patterns—linking policy structure to multi-scenario performance. This demonstrates that, within this context, PRIM offers a feasible alternative to MSMOP for exploring robust policy designs when mixed-variable constraints exclude more complex optimisation techniques.

A.2.4.2 Selection of robust policies

The selection of the percentiles in this PRIM sensitivity experiment is discussed in Section 3.3.3.1. In this Section in the Appendix the reasoning for the selection of the final PRIM box is discussed. The analysis of policies from the Pareto front onwards is presented in Section 4.2.3 in the main body.

For the joint PRIM threshold the three different runs at the different percentile all show a peeling trajectory. From the 80^{th}percentile Box 7 is selected due to its combination of coverage, density, and number of restricted dimensions. This PRIM box counted 492 solutions with a density of 0.76 and a coverage of 0.42. The 4 measures in this box are: *schedule padding, budget buffer, a spare part option, and overtime labour.*

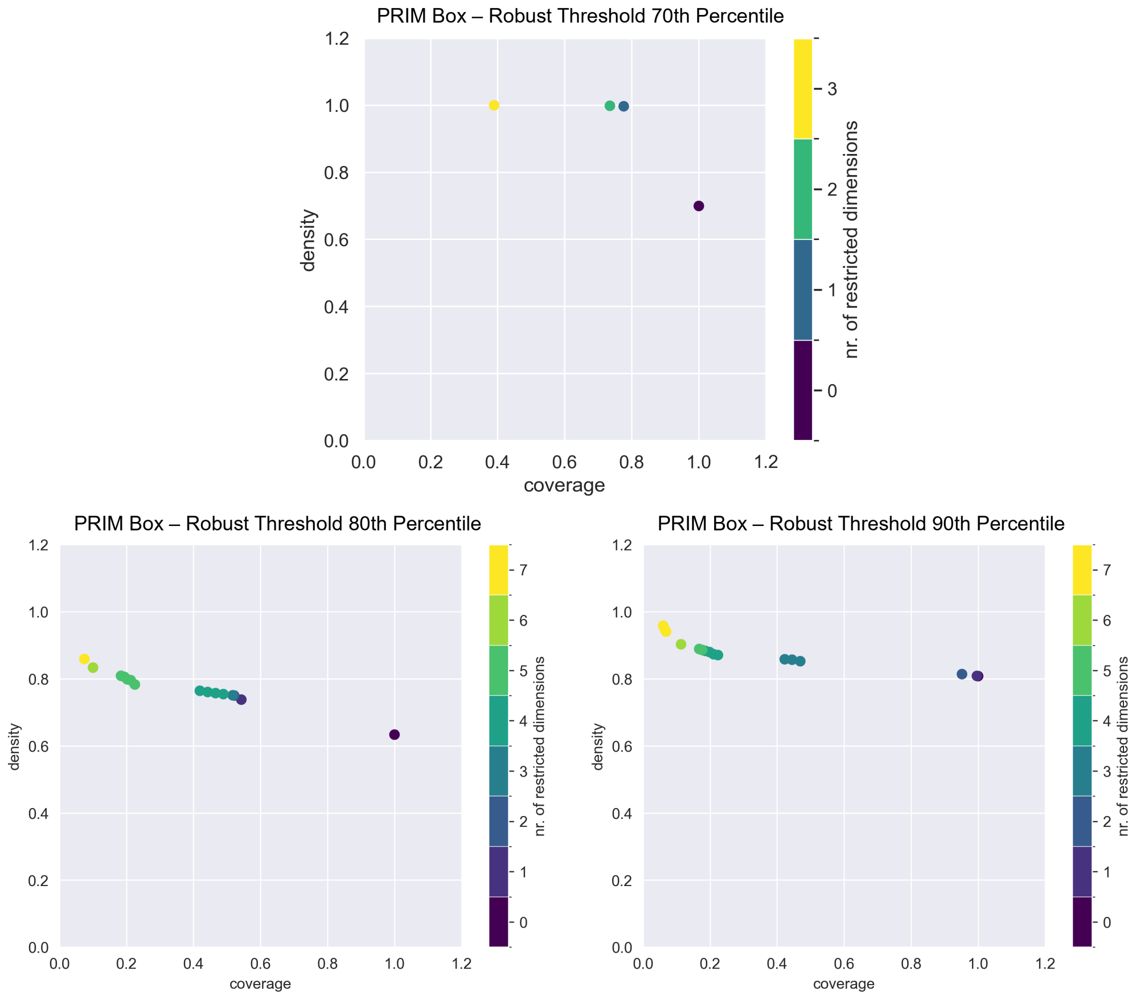

Figure A.5: The figure shows the PRIM of the 70$^{\text{th}}$, 80$^{\text{th}}$, and 90$^{\text{th}}$ percentiles for the joint threshold.

For the separated conditions the 90^{th} percentile is selected for duration. Box 6 is chosen for its density of 0.97 and coverage of 0.23. The 5 measures in this box are: *schedule padding, budget buffer, a spare part option, overtime labour, and new design.*

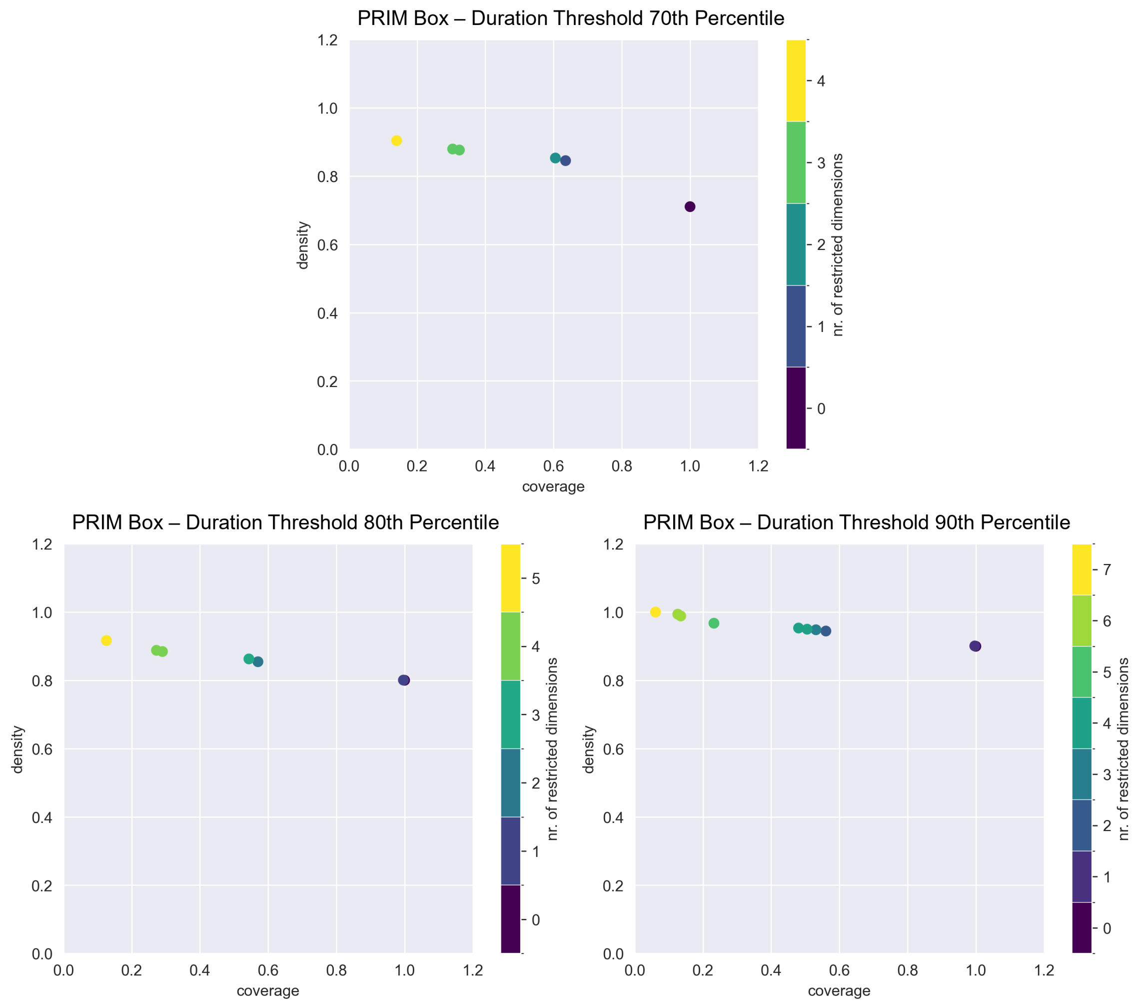

Figure A.6: PRIM outcomes of the 70^{th}, 80^{th}, and 90^{th} percentiles for the duration condition.

For the cost threshold, the 3rd box was selected from the 70th percentile. This box has a coverage of 0.32 and a density of 0.87 with 372 policies. The driving measures are: *budget buffer, overtime labour, and extended search.*

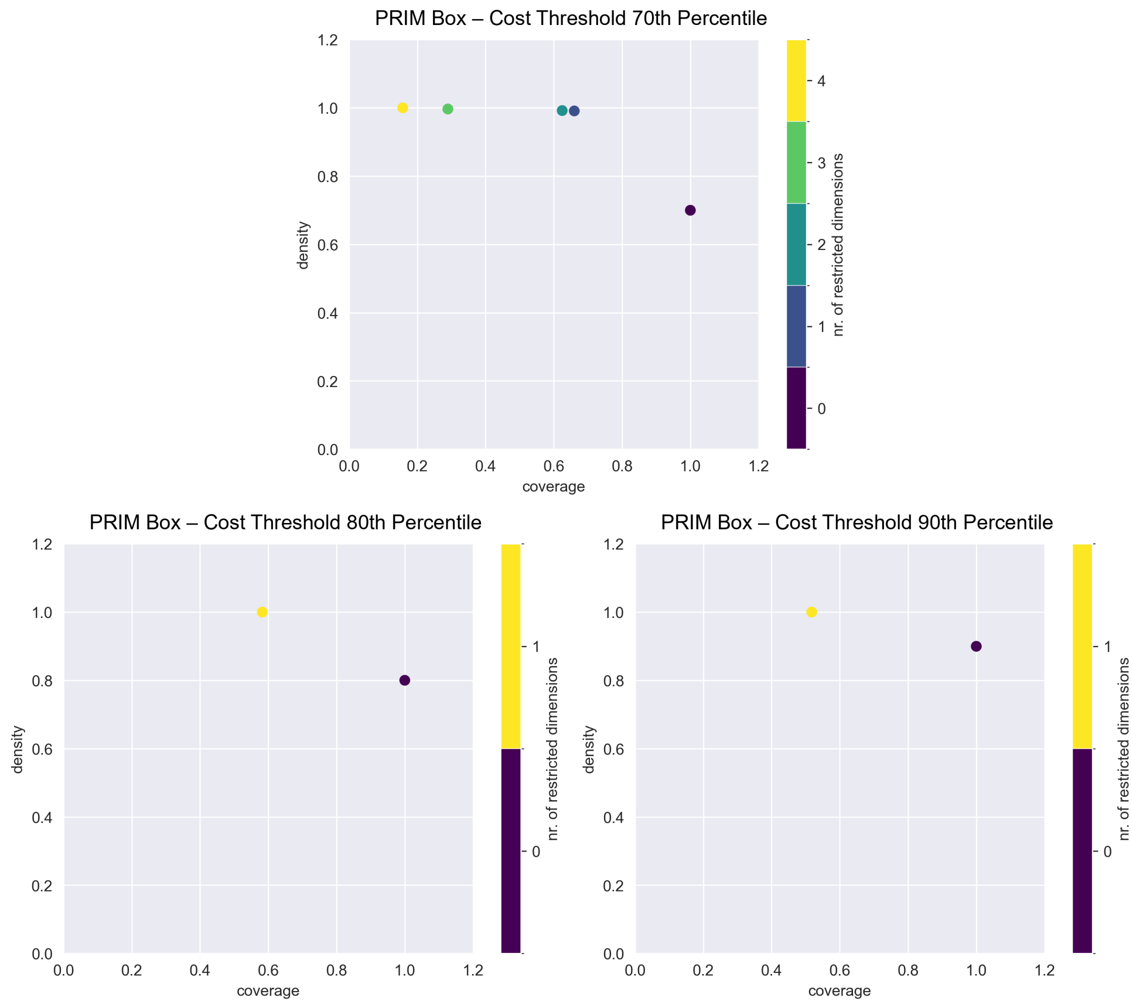

Figure A.7: PRIM outcomes of the 70th, 80th, and 90th percentiles for the cost condition.

A.2.4.3 Robust policies

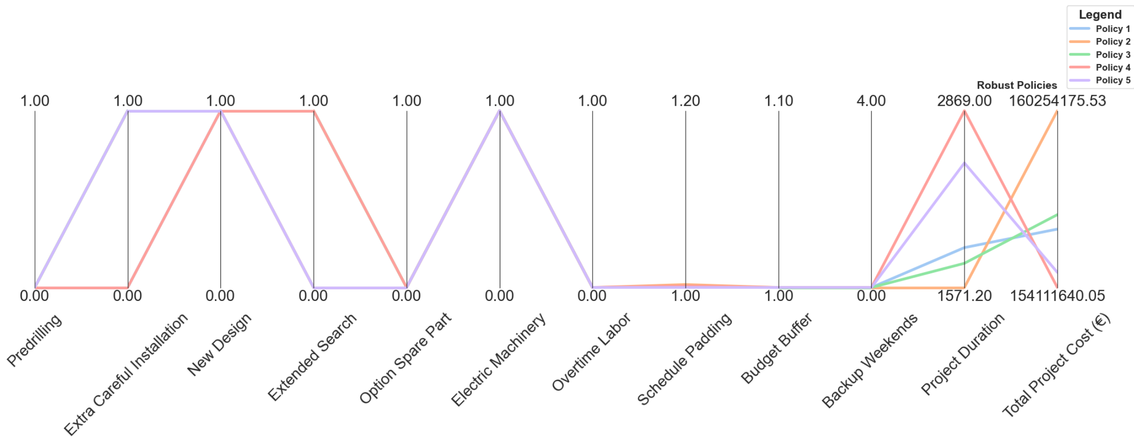

Figure A.8: Parallel coordinates plot of the joint threshold robust policies.

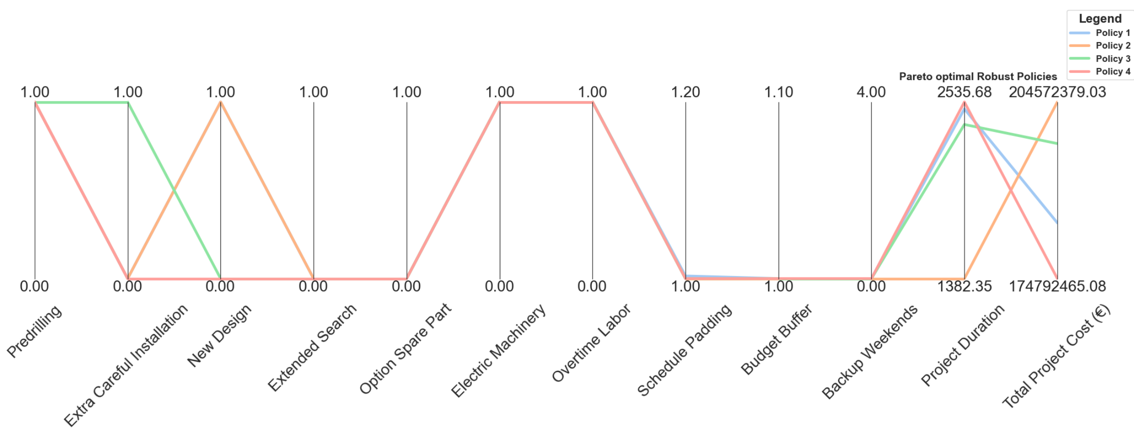

Figure A.9: Parallel coordinates plot of the separated threshold robust policies.

A.2.5 DAPP

A.2.5.1 Distribution of Sampled Scenarios

Figure A.10 shows the amount of triggered DES runs distributed over the high-impact scenarios. The bar plots show that for each seed, approximately 5% of the 5,000 runs the ATP trigger measured a high-impact scenario. Each of the six high-impact scenarios is triggered, meaning that subsequently each policy is allowed to be applied.

While the parallel coordinates plot in Figure 4.4 provides an overview of the high-impact scenarios, the accompanying bar plot offers additional insights to how they are sampled.
Scenario 5 *Cost-efficient but risk-exposed* accounts for more than 50% of the triggers for each seed. This Scenario is one of only two where no influenza wave has occurred. The lowest overall scalar score from the Scalar function in Section 3.3.2.4 is connected to this scenario.

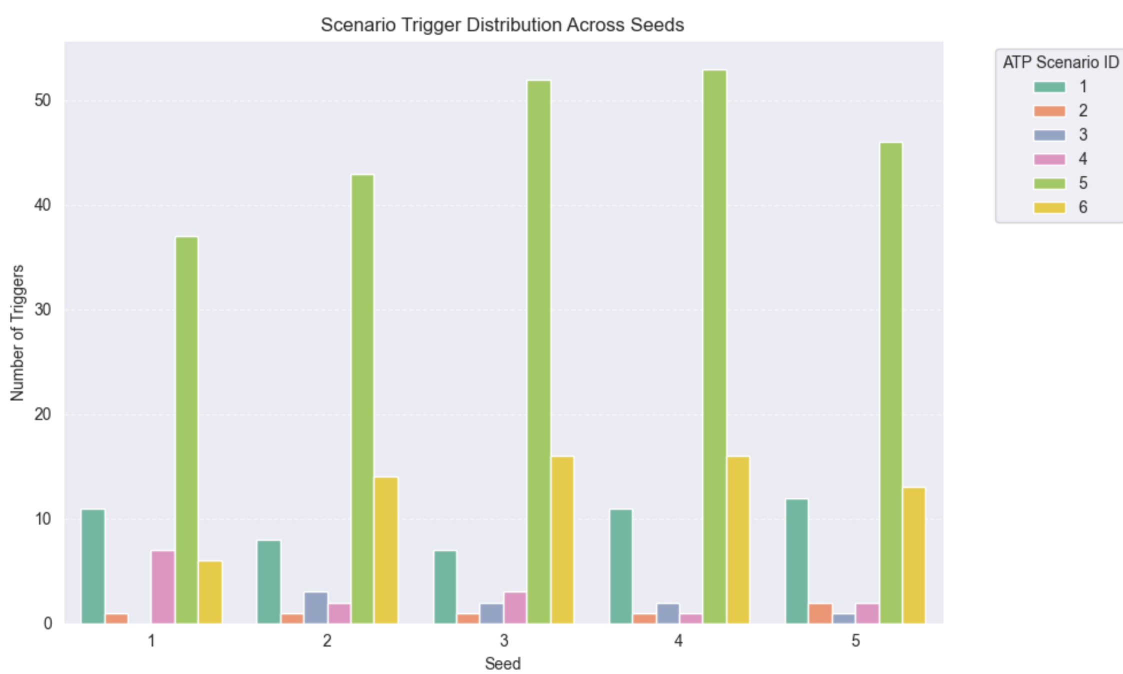

Figure A.10: Distribution of high-impact scenarios sampled in the DES environment. Each seed is sampled 5,000 times. The bars represent the runs for which the ATPs were triggered.

This distribution gives additional insight into the likeliness of pathway development. High-impact Scenarios 5 and 6 account for an average of 79% of all sampled scenarios over the seeds. As Scenario 5 has the lowest scalar score, and Scenario 6 the third lowest, it could be argued that the scalar score in Section 3.3.3.2 is correlated with the occurrence in the quantitative analysis. When adding the third most sampled scenario, high-impact Scenario 1, it shows that this suggested correlation is untrue. Scenario 1 is sampled 12% of the time but has the highest scalar score. This information confirms that it is good practice to prepare for even the most severe scenarios even they appear very infrequently.