Multi-Metrics Robustness Evaluation of Water Allocation Policies in Nile River Basin

Navigating Deep Uncertainties

Master Thesis

Fahira Gearahmani Gatot

Delft University of Technology

Multi-Metrics Robustness Evaluation of Water Allocation Policies in Nile River Basin

Navigating Deep Uncertainties

by

Fahira Gearahmani Gatot

Student Name	Student Number
Fahira Gearahmani Gatot	5427444

Chairperson:	Dr. J.A. (Jan Anne) Annema
First Supervisor:	Dr. J. (Jazmin) Zatarain Salazar
Faculty:	Faculty of Technology, Policy and Management

Code and data files for this thesis are available at
https://github.com/fahiragea/masterthesis/

TU Delft

Acknowledgment

I never imagined that I would find myself diving into the world of water, let alone writing a thesis on it. This field is completely new to me, and before starting my master's at TU Delft, I had never even been exposed to Water and Delta Systems Management. I chose this path solely based on my interest and moral compass, wanting to contribute to ensuring that everyone has access to their most basic need—water—for a better future. Such a journey has been life-changing, and this thesis marks the end of my time as a Master's student in Complex Systems Engineering and Management at TU Delft. I can say with absolute certainty that my experience here has truly changed my life. So, I'm incredibly hopeful and excited about the new adventures that await me from this point on.

First and foremost, I dedicate this thesis to my mom, dad, and sister in Bandung, who have patiently waited for me to pursue my studies in the Netherlands, equipping me with their blessings and love. For all the endless encouragement and prayers they've sent me every single day for their eldest daughter, I'm beyond grateful. In everything I do, I carry their support with me.

Secondly, I want to thank my supervisor, Jazmin Zatarain Salazar, who provided me with so much valuable insight and inspiration. Only through her guidance and support was I able to learn so much from scratch. Really, I'm deeply grateful that she trusted me to be her student. Thank you so much for your patience with me, and for guiding me on where to go and how to navigate my work. You gave me the confidence to tackle challenges on my own. Without your guidance, I'm certain this thesis wouldn't have been possible.

I also want to express my gratitude to Jan Anne Annema, both as a supervisor and a lecturer I've frequently met since my first year. Our lighthearted discussions greatly influenced how I approached writing my thesis, and more importantly, how I ensured that my research and work could be understood not just by myself or a small group of experts, but by everyone, because the impact of this work will be felt by all. Understanding how to communicate effectively was one of the most valuable lessons I learned. And, I'm also very thankful for his encouragement in my writing, something he's done since my first year, which has consistently driven me to improve and to become a better writer and researcher.

In line with the theme of this thesis—water—I'd also like to extend my thanks to the lecturers in the Water and Delta Systems domain: Tineke Ruijgh-van der Ploeg, Lisa Scholten, and Neelke Doorn. Learning about Water and Delta Systems at TU Delft has truly opened my eyes and changed the trajectory of my life for the better. Dedication to the craft and genuine support from my mentors have made this journey possible. Your guidance has given me endless opportunities to grow.

Of course, I owe a huge thank you to my Indonesian sisters and international friends at CoSEM. Thank you for always offering advice, giving me a space to confide in when I was struggling, and for being such great friends, always pushing each other to the top! And to the Indonesian community at TU Delft, thank you so much for making my life in the Netherlands brighter!

I am very thankful for every single friend and family member that I am not able to mention in this acknowledgment. Your support, even from afar, has been invaluable throughout this journey. I would like to also thank LPDP (Indonesia Endowment Fund for Education) for sponsoring my studies and making this all possible. Your financial support has been a cornerstone in enabling me to pursue and complete my education. Finally, I'd like to dedicate this thesis to my beloved grandfather, who sadly passed away after I arrived in the Netherlands to study.

I present my work, Multi-Metrics Robustness Evaluation of Water Allocation Policies in the Nile River Basin. I hope that my work will be useful not only in the Nile Basin but also in other water systems around the world, helping us plan well for the future and ensuring that we all have equitable access to water.

Fahira Gearahmani Gatot
Delft, August 2024

Summary

The construction of the Grand Ethiopian Renaissance Dam (GERD), a massive hydroelectric project on the Blue Nile, is part of the broader Nile River basin, which spans across 11 countries in northeastern Africa, including Ethiopia, Sudan, and Egypt. The GERD is designed to be the largest dam in Africa, with the primary purpose of generating electricity to power Ethiopia's development and provide energy to neighbouring countries. However, its construction has ignited significant debate, particularly due to concerns from downstream countries, Sudan and Egypt, which rely heavily on the Nile's waters for agriculture, drinking water, and energy production. The dam's operation and management have far-reaching implications for the entire Nile basin region. Adding to this complexity is the uncertainty of future climate conditions, which could drastically change water availability in the region. To navigate these challenges, it is essential to develop robust reservoir management strategies—policies that are resilient and adaptable to various future scenarios. This thesis explores a broader and more detailed approach to evaluating the robustness of water management policies, focusing on the case of the Nile River. Instead of relying on a single metric, we use multiple robustness metrics; Percentile-based Skewness, Mean-Variance, Undesirable Deviations, and Minimax Regret to see how they might lead to different conclusions about the best strategies.

To evaluate the resilience of these policies against future uncertainties, a series of steps are involved. First, policies are generated using optimization techniques designed to identify the best strategies that could foster cooperation among the countries, while also addressing their individual objectives. These strategies are then tested for their effectiveness under various future scenarios, which is done by applying robustness metrics. A robustness metric is a quantitative measure used to assess the resilience and stability of a system or process in the face of uncertainties, disturbances, or perturbations. It provides a way to quantify how well a system can maintain its performance or functionality under varying conditions. These metrics can range from assessing absolute performance or regret prioritizing risk-aversion, maximizing performance, or minimizing variance, depending on the specific uncertainties and the decision-maker's risk tolerance. However, many previous studies done on robustness of reservoir control rely on the 90th Percentile Regret metric which looks at how much worse a given strategy performs compared to the best possible outcome. While useful and practical, this approach doesn't fully capture the range of different ways to test against future scenarios.

This research reveals that the choice of robustness metric can greatly affect the evaluation of different policies. Using four different metrics, we were able to conclude different "most robust" policies. Using minimax regret metric, Compromise Policie(s) are the most robust. However, using Undesirable Deviations metric, Best Egypt Irrigation Policy is the most robust. Using Percentile-Based Skewness yields Compromise: Percentile and Best Sudan Irrigation Policy as the most robust, and using Mean-Variance metric, only Best Sudan Irrigation policy emerged as the most robust. This variation occurs because different robustness metric prioritizes different aspects of policy performance under uncertainty. While the regret metric focuses on minimizing the worst-case scenario outcomes, other metrics like Percentile-Based Skewness emphasize the consistency of outcomes across a range of scenarios.

These findings emphasize the importance of using a diverse set of robustness metrics in policy evaluation. A diverse set of metrics allows for a more comprehensive assessment of policies by capturing different aspects of performance and risk under uncertainty. Depending on which metric is used, different strategies may be recommended, leading to potentially different outcomes for the countries involved. While strategies may seem mutually exclusive, a more nuanced approach can involve balancing the trade-offs highlighted by the various metrics, leading to a more informed and robust decision-making process.

Contents

1 Introduction

The management of transboundary river systems is a complex and challenging tasks in global water governance (Lorenz et al., 2001). Transboundary river systems, which span multiple countries, often serve as critical lifelines for millions of people, providing essential resources such as water for agriculture, industry, and domestic use, as well as supporting energy generation through hydropower (Varis et al., 2008; Zeitoun et al., 2017; Wheeler et al., 2018).

The Nile River, the world's longest river, is a quintessential example of such a system (Swain, 1997). Flowing through eleven countries in northeastern Africa, the Nile has been a cornerstone of civilization in the region for millennia (Paisley and Henshaw, 2013). Today, it remains vital to the livelihoods and economies of the nations it traverses (Yitayew and Melesse, 2011).

The construction of the Grand Ethiopian Renaissance Dam (GERD) on the Blue Nile, one of the Nile's major tributaries, has brought the complexities of transboundary water management to the forefront (Wheeler et al., 2018). As Ethiopia seeks to harness the river's potential for hydropower, downstream countries, particularly Sudan and Egypt, have expressed concerns over the implications for their own water security and agricultural productivity (Elsayed et al., 2020). These concerns have led to heightened tensions and stalled negotiations, underlining the intricate balance of interests that must be maintained in managing the Nile's waters. This dispute, however, is not new; it is a reopening of longstanding conflicts over Nile water rights that date back to the Anglo-Egyptian Treaty of 1929 (Mohyeldeen, 2021).

In addition to the political and social challenges, the management of the Nile Basin is further complicated by environmental factors, including climate change, and socio-economic changes such as population growth and urbanization (Tariku and Gan, 2018). The potential for altered precipitation patterns and increased frequency of extreme weather events introduces significant uncertainty into the future availability of water resources (Tariku and Gan, 2018). This makes it imperative to develop reservoir management strategies that are not only effective under current conditions but are also resilient to a wide range of possible future scenarios (Zaitchik et al., 2012).

Given these challenges, there is a critical need for robust water management policies that can address the diverse and often conflicting objectives of the countries within the Nile Basin. Such policies must be capable of withstanding the uncertainties associated with both political dynamics and climate variability. This thesis aims to explore the development and evaluation of these policies, focusing on the use of multiple robustness metrics to assess their effectiveness. The rationale for using multiple metrics is to capture the diverse challenges and uncertainties inherent in managing the Nile Basin, as will be further explained in Section 1.4. By doing so, this study seeks to contribute to the ongoing discourse on how best to manage one of the world's most important and contested river systems.

1.1. On the Nile Basin

The Nile River is a crucial water resource in northeast Africa and plays a significant role in nurturing one of the earliest human civilizations (Bunbury et al., 2023). Its basin spans over 11 nations, and the river serves essential purposes such as hydropower, municipal, industrial, and agricultural needs (Wheeler et al., 2020). Egypt, as the most downstream country, has historically been the primary user of the Nile's water, along with Sudan as also one of its main users (Kendie, 1999). However, the initiation of the Grand Ethiopian Renaissance DAM (GERD) by Ethiopia in 2011 has introduced a new dynamic to the region (Tsega, 2017).

The GERD, designed to harness Ethiopia's hydropower potential, has raised concerns among downstream countries like Sudan and Egypt, who perceive it as a threat to their water security and sovereignty (Salman, 2018; Mohyeldeen, 2021). The filling of the GERD's reservoir has become a point of dispute, with negotiations for a filling agreement deadlocked, leading to escalated tensions (Attia and Saleh, 2021). This situation is further complicated by environmental and demographic challenges in the Nile Basin, including water variability, population growth, and increased economic activities (Swain, 1997).

Developing operational strategies for the basin's reservoirs is a complex task that necessitates considerations of physical infrastructure, geopolitics, socioeconomic factors, and hydro-climatic uncertainties. The management of the GERD and its impacts under current and future climates is a critical issue that requires careful assessment, as the potential implications of the GERD on downstream countries is an urgent need for agreements on filling and operational rules, especially during periods of drought (Aty, 2022).

The construction and operation of the GERD have introduced significant political tensions in the eastern Nile Basin, making it important to develop of cooperative transboundary river management strategies to address the challenges posed by the dam. Resolving the complexities surrounding the GERD requires a comprehensive consideration of its impacts on water security, hydrology, and regional stability, emphasizing the need for collaborative efforts among all involved nations for the long-term.

1.2. On Robustness

The concept of robustness in water management is grounded in the need to ensure reliable and resilient access to water resources despite uncertainties such as climate variability, hydrologic response, and socio-economic developments (Haasnoot et al., 2011; Salazar et al., 2016). Unlike optimal strategies which may perform best under a single predicted scenario, robust strategies are designed to perform satisfactorily across a broad spectrum of potential futures (Beh et al., 2017; Matrosov et al., 2013), given the characteristics of future conditions, especially climate change, to be nonstationary (Milly et al., 2008), and deeply uncertain (Lempert, 2014; Dittrich et al., 2016). Nonstationary refers to changes in the statistical properties of random variables with shifts in time (Razavi and Gupta, 2016), while deep uncertainty refers to situations in which analysts don't know and cannot agree on probabilistic distributions of key random variables and parameters (Lempert et al., 2003; Ben-Haim, 2006; Hallegatte et al., 2012; Brown et al., 2020).

Failure to incorporate robustness in transboundary rivers can lead to detrimental consequences. Without robust strategies, there is an increased risk of water scarcity, conflicts over water allocation, and instability in the region. Lack of robustness can lead to ineffective water management policies that cannot successfully meet the growing difficulties given by climate change and rising water demands (Quinn et al., 2020; Hadjimichael et al., 2020).

Uncertainties in environmental decision-making has traditionally been managed by focusing on localized uncertainties around expected future conditions, using metrics such as reliability, vulnerability, and resiliency (Howard, 1966; Howard and Matheson, 2005; Hashimoto et al., 1982). However, as climatic, sociopolitical, and technological changes increasingly challenge the localized assumptions, there has been a shift towards addressing resiliency from a deep uncertainty perspective – situations where future conditions cannot be predicted or agreed upon as its no longer feasible to determine a single best guess of how future conditions might evolve, especially when considering longer planning horizon (Döll and Romero-Lankao, 2017; Kwakkel et al., 2010; Walker et al., 2013).

1.2.1. Knowledge Gap

Based on the literature review that will be explained thoroughly in Chapter 2, significant gaps exist in the study of robust reservoir operations in the Nile Basin. Although multi-objective optimization (MOO) has been widely applied, few studies have focused on the robustness of these operations postoptimization, which is crucial for ensuring effectiveness under various future scenarios. The existing robustness analyses within MOO typically use single metrics like the 90th percentile or minimax regret, offering a limited perspective. Employing multiple robustness metrics, as suggested by approaches like Evolutionary Multi-Objective Direct Policy Search (EMODPS), would provide a more comprehensive assessment. However, no studies in the Nile Basin have yet adopted this multi-metric approach, leaving a critical gap in current research. This knowledge gap is further explained in Chapter 2.

1.3. Research Objective

The objective of this study is to evaluate and compare the robustness of reservoir control strategies in the Nile Basin by applying multiple robustness metric approach. The study aims to explore the impact associated with using different robustness metrics in assessing optimized reservoir control policy alternatives under conditions of uncertainty. By analyzing the robustness of selected policies through a multi-metrics approach, this research seeks to enhance the understanding of how varying metrics influence the perceived resiliency of policy choices. Ultimately, the study will provide decision-makers with a more comprehensive and nuanced framework for robust water resource management in the face of diverse future scenarios.

1.4. Research Question

Incorporating a multi-metrics approach of robustness metric into many-objective optimization within the context of the Nile River Basin remains largely unexplored, especially when considering the combination of deep uncertainty variables and their impacts. Conducting a case study in this context can provide valuable insights and inform methodological advancements. Therefore, this research aims to address the following question:

What is the consequence of applying multiple robustness metrics within many-objective optimization models to address water allocation issues under deep uncertainty in the Nile River Basin?

1.4.1. Research Sub-questions

SQ1: What are the trade-offs of the Pareto-optimal policy alternatives of the optimal reservoir control in the Nile River Basin?

This question aims to explore the balance between different goals (such as maximizing water supply, generating hydroelectric power, and preserving the environment) in the best possible reservoir management policies. It seeks to understand the compromises that need to be made when trying to achieve these competing objectives simultaneously.

SQ2: How do different robustness metrics influence the selection of optimal policy alternatives in the Nile River Basin?

This sub-question will investigate how the choice of robustness metrics affects the identification and prioritization of policy alternatives.

SQ3: What are the implications of using different robustness metrics for stakeholder decision-making in the Nile River Basin?

This sub-question will explore how the application of various robustness metrics can impact the decision-making process for different stakeholders involved in Nile River Basin management.

1.5. Organization of the Report

The methods of the study will be presented in Chapter 3. Section 3.1 will cover the Modelling Approach, explaining the theoretical framework and choose methods. Section 3.3 will detail how the system is modelled, including model components, assumptions, and data sources. Section 3.5 will outline the experimental strategy, describing the design, scenarios, and evaluation criteria. Chapter 4 will present the experimental results, highlighting the performance of different policy alternatives. Chapter 5 will discuss these results, interpret the findings, and explore their implications for water management in the Nile River Basin. Finally, Chapter 6 will revisit the research questions, summarize the study's contributions, and provide policy recommendations and suggestions for future research.

2 Literature Review

This chapter aims to review the evolution of reservoir operations and robustness analysis in water resource management. It explores key methodologies, from traditional to advanced approaches, and examines how these concepts have been applied, particularly in the Nile Basin. The chapter also discusses various robustness metrics and their impact on decision-making under uncertainty, identifying gaps and opportunities to enhance reservoir management practices.

2.1. Optimal Reservoir Operations

Reservoir operations refer to the process of managing the storage and release of water in a reservoir to achieve specific objectives, and are critical to managing water resources globally, playing a vital role in enhancing water availability, generating renewable electricity, and mitigating flood risks. Since the 1960s, reservoir operations have been one of the most active areas of research in the water systems literature (Yeh, 1985; Kim et al., 2021), focused on two main topic branches: forecast-informed reservoir operations (Tejada-Guibert et al., 1995; Kim and Palmer, 1997; Zhao et al., 2012; Turner et al., 2017).

Historically, reservoir control progressed from deterministic methods based on rule curves and linear programming to more adaptive approaches. Traditional methods frequently fail to account for real-world uncertainties (Giuliani et al., 2014; Zarfl et al., 2015). Stochastic Dynamic Programming (SDP) increased realism but had limitations such as dimensionality and handling multiple objectives (Esogbue and Kacprzyk, 1998; Giuliani et al., 2014; Tsitsiklis and Van Roy, 1996). The Parametrization-Simulation-Optimization (PSO) approach, which uses systematic simulation runs, evolved into the Direct Policy Search (DPS) method, which combines machine learning and evolutionary algorithms for adaptive policy optimization (Koutsoyiannis and Economou, 2003; Maier et al., 2014).

The Evolutionary Multi-Objective Direct Policy Search (EMODPS) framework improves computational efficiency, adaptability, and effectiveness when managing large-scale, multi-objective systems, allowing for real-time adjustments and balancing competing demands (Giuliani et al., 2015). Recent advancements, including artificial intelligence and advanced simulation techniques, enhance the robustness of reservoir operations against climate change and socioeconomic variability. Despite its benefits, EMODPS demands significant data and computational resources, studies may benefit when concentrating on improving data integration and creating comprehensive frameworks for managing the complexities of water systems under extreme uncertainty (Giuliani et al., 2015).

2.2. Robustness Analysis in Reservoir Control Studies

Robustness in reservoir operations refers to the ability of a system to maintain acceptable performance even when faced with uncertainties and unexpected conditions (Huang et al., 2022). To address uncertainties in hydrologic processes, such as changes in precipitation, streamflow, or water demands, stochastic optimization techniques have been developed. These techniques incorporate random or unknown variables to create effective operational policies for both single reservoirs (Butcher, 1971; Saadat and Asghari, 2019) and multi-reservoir systems (Stedinger et al., 2013; Salazar et al., 2017).

The development of these techniques began with the Linear Decision Rule (LDR), introduced by (Revelle et al., 1969), which set guidelines for reservoir releases based on predefined parameters like hydropower targets or minimum flow requirements. While LDR laid the groundwork for managing uncertainty, it was limited by its linear approach and focus on a single objective, often treating environmental needs as secondary concerns (Butcher, 1971; Saadat and Asghari, 2019).

Over time, more advanced non-linear stochastic methods were developed. For example, Stochastic Dual Dynamic Programming (SDDP) has been used in multi-reservoir systems to create robust policies that account for uncertainties in water inflows (Pereira and Pinto, 1991; Tilmant and Kelman, 2007). Further improvements, such as incorporating stream-aquifer interactions (Macian-Sorribes and Pulido-Velazquez, 2020), have made these models more complex and robust. Similarly, Sampling Stochastic Dynamic Programming (SSDP) has shown better performance by explicitly considering inflow uncertainties, proving to be more effective than deterministic methods (Kim et al., 2007; Eum and Kim, 2010).

More recently, the Iterative Linear Decision Rule (ILDR) has emerged as a more manageable robust optimization model. ILDR, used in systems like the Three Gorges Dam and Shasta-Trinity reservoirs, integrates non-linear objectives and offers robustness comparable to advanced stochastic methods like SSDP (Pan et al., 2015). However, these methods often rely on a single or limited robustness metric, like minimizing the risk of poor decisions (Gaivoronski et al., 2012), which doesn't fully capture all dimensions of robustness. This leaves a gap in providing a more complete view of system performance under uncertainty.

Comparative studies have looked at different approaches to evaluating robustness in reservoir operations. The "a posteriori" decision support approach generates decision alternatives through computational search before considering stakeholder preferences, allowing for more flexibility in evaluating robustness (Cohon and Marks, 1973, 1975). On the other hand, "a priori" methods require preferences to be set before generating alternatives, which can limit the robustness of the chosen solutions (Reuss, 2003; Banzhaf et al., 2009). "A posteriori" methods, such as Multi-Objective Robust Decision Making (MORDM), help identify near-optimal solutions under expected conditions and then evaluate them under uncertainty, supporting adaptive and resilient decision-making (Kasprzyk et al., 2013; Paton et al., 2014). In contrast, "a priori" methods may provide optimal solutions for assumed conditions but often fail when conditions change, highlighting the need for flexibility and adaptability in robustness analysis.

2.3. Robust Reservoir Operations in Nile Basin

Robust reservoir operations have been a critical focus of research in the Nile Basin, given the complex hydrological, geopolitical, and climatic challenges that the region faces. Early work by Stedinger et al. (1984) pioneered the application of Stochastic Dynamic Programming (SDP) using data from the High Aswan Dam (HAD). This foundational study laid the groundwork for later research by introducing probabilistic methods to address uncertainties in water management.

Jeuland and Whittington (2014) introduced a combined approach using real options and robust decision-making for the Blue Nile. Their method identified key uncertainties and employed Monte Carlo simulations to evaluate different investment alternatives, using a 90^{th} percentile Net Present Value (NPV) distribution-based metric to assess the robustness of these investments under varying future scenarios. Alexander et al. (2021) enhanced the robustness of reservoir operations by integrating forecast-informed frameworks with seasonal streamflow predictions, optimizing water releases to improve hydropower and agricultural outcomes in the Blue Nile Basin.

Recent studies have further advanced the field by exploring multi-objective optimization approaches. Geressu and Harou (2015) and their later work in 2019 applied Pareto-optimal design principles to address trade-offs between conflicting objectives in the Blue Nile's reservoir operations, with a particular focus on the impacts of dam filling periods. Wheeler et al. (2018) and Sari (2022) also investigated the robustness of cooperative reservoir operations in the Eastern Nile using multi-objective evolutionary algorithms, applying minimax regret and 90^{th} percentile regret scores, respectively, to evaluate and minimize potential losses under worst-case scenarios.

2.4. Robustness Metrics

Measuring robustness in water management involves evaluating how well different strategies perform under a wide range of plausible future conditions (Herman et al., 2015). This evaluation is typically done using various robustness metrics, each providing different insights into the resilience and adaptability of management approaches. The choice of metric is crucial, as it can significantly influence the outcomes of the decision-making process (McPhail et al., 2018).

A variety of robustness metrics type can be used to measure system performance under deep uncertainty (McPhail et al., 2018), including:

1. Expected Value Metrics Type: Metrics that falls under this type assesses the expected level of performance across a range of scenarios, providing a baseline for evaluating robustness (Wald and Wolfowitz, 1950).
2. Higher-Order Moment Metrics Type: Metrics that falls under this type, such as variance, skewness, offers insights into how the expected performance varies across multiple scenarios, helping to capture the range and distribution of potential outcomes (Kwakkel et al., 2016).
3. Regret-based Metrics Type: Metrics that falls under this type evaluates robustness by comparing the performance of a chosen strategy against the best possible performance across various scenarios, with lower regret indicating a more robust strategy (Savage, 1951).
4. Satisficing Metrics Type: Metrics that falls under these types focus on the range of scenarios where a strategy meets a predefined performance threshold, with a strategy considered robust if it performs satisfactorily across a wide range of plausible futures (Simon, 1956).

Each of these metrics reflects different aspects of robustness, thus the choice of metric is therefore not only a technical decision but also a reflection of the decision-makers' risk preference and priorities (Lempert and Collins, 2007; Herman et al., 2015).

(McPhail et al., 2018) identifies 11 robustness metrics for environmental systems, including:

Table 2.1: Classification of Robustness Metrics from McPhail et al. (2018)

Metrics	Reference	Type
Maximin	Wald and Wolfowitz (1950)	Expected Value Metrics
Maximax	Wald and Wolfowitz (1950)	Expected Value Metrics
Hurwicz Optimism-Pessimism Rule	Hurwicz (1953)	Expected Value Metrics
Laplace's Principle of Inufficient Reason	Laplace and Simon (1951)	Expected Value Metrics
Minimax Regret	Savage (1951) and Giuliani and Castelletti (2016)	Regret-Based Metrics
90th Percentile Minimax Regret	Savage (1951)	Regret-Based Metrics
Mean-Variance	Hamarat et al. (2014)	Higher-Order Moment Metrics
Undesirable Deviations	Kwakkel et al. (2016)	Regret-Based Metrics
Percentile-based skewness	Voudouris et al. (2014) and Kwakkel et al. (2016)	Higher-Order Moment Metrics
Percentile-based peakedness	Voudouris et al. (2016) and Kwakkel et al. (2016)	Higher-Order Moment Metrics
Starr's domain criterion	Starr (1963) and Schneller and Sphicas (1983)	Satisficing Metrics

Which metrics are chosen has different relative level of inherent risk aversion. Maximax has the lowest risk aversion, as it considers only the best-case scenario, while Maximin has the highest, focusing solely on the worst-case scenario. Metrics like Minimax Regret are also highly risk-averse, emphasizing minimizing the worst relative performance. In contrast, metrics such as Mean-Variance, Percentile-based Skewness, and Laplace's Principle of Insufficient Reason fall in the middle, providing a balanced perspective by using a range of scenarios. The Hurwicz Optimism-Pessimism Rule can vary widely on the risk aversion scale depending on the chosen weighting between the best and worst scenarios. Similarly, Starr's Domain Criterion is adaptable, as its placement depends on the user-defined performance threshold, which can lead to high or low risk aversion. Metrics like Undesirable Deviations and 90^{th} Percentile Minimax Regret lean towards higher risk aversion by focusing on specific subsets of scenarios that are less balanced, often nearer to worst-case conditions.

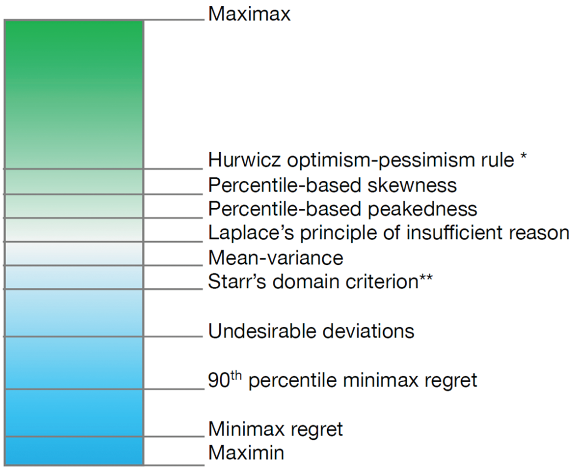

Figure 2.1: Classification of Robustness Metrics based on Risk Aversion Level (Taken from McPhail et al. (2018)).

However, it should be noted that the metrics all measure system performance over a set of future states of the world, they do so in different ways, which makes the assessment of how robust a system's performance truly complicated. Different robustness metrics highlight different aspects of what is robustness, making it difficult to determine the absolute robustness of an alternative or to compare alternatives objectively. Choice of robustness metrics can have a large effect, with some robustness metrics may disagree, and some may show similar performance depending on the transformation, or the calculation used in the robustness metric itself (Giuliani and Castelletti, 2016; Herman et al., 2015). There exists "interutility robustness tradeoffs", and ultimately settling on a single most robust policy would require negotiation and compromise between decision makers (Herman et al., 2014).

When multiple robustness metrics yield different most robust solutions for a set of policies, choosing the best solution needs to consider a few factors, including stability of performance across multiple metrics, and alignment with decision context and priorities. It may be beneficial to consider a composite or weighted approach, to gain insight from multiple metrics rather than relying on a single metric (McPhail et al., 2018).

2.5. Literature Review: A Conclusion

Several significant gaps in the study of robust reservoir operations in the Nile Basin have been identified. Firstly, while multi-objective optimization (MOO) has been extensively applied to reservoir operations in the Nile Basin, only a few studies have focused specifically on the robustness aspect of these operations after the optimization process. Robustness is key to ensuring that reservoir operations remain effective under a wide range of future scenarios, making this an oversight. Secondly, the robustness analyses that have been conducted within MOO predominantly rely on single metrics, such as the 90th percentile regret or minimax regret. While these metrics are valuable, they provide only a limited view of robustness.

Given that the Evolutionary Multi-Objective Direct Policy Search (EMODPS) and other multi-objective optimization approaches inherently consider robustness, it would be highly beneficial to employ multiple robustness metrics for comparison. Different robustness metrics can highlight different facets of what makes a solution robust. The use of multiple metrics would provide a more comprehensive assessment of robustness and help ensure that policy alternatives are truly resilient. Despite the importance of the approach, no studies on the Nile Basin have yet utilized a multi-metrics approach to analyse robustness post-MOO. This represents a significant gap in the current research, as relying on a single metric may not capture the full complexity of robustness in reservoir operations.

3 Methods

3.1. Methodological Framework

The methodology of this thesis is inspired and adapted from the study conducted by Sari (2022) that explores trade-offs in reservoir operations through multi-objective optimization, using the same case as our focus, the Nile River Basin. The study builds a simulation model using the EMODPS (Evolutionary Multi-Objective Direct Policy Search) framework. A set of policies are then selected from the Pareto-front generated by the optimisation, and simulations under varying uncertain scenarios are carried out.

Our methodology in this study diverges from Sari's study, as we focus on the robustness analysis of the policies, hence we take out the scenario discovery approach, and instead focus on the robustness analysis which is loosely inspired by post-MORDM robustness policy mapping (Bonham et al., 2022), which presents decision makers with a visual representation of decision variable values, objective values, and robustness values to ease with decision making, from a multi-metrics perspective.

Following the methodological framework laid out in Sari (2022)'s work, the methodological framework of this study is adapted from the Multi-Objective Robust Decision-Making framework by Kasprzyk et al. (2013), adapted to our research objective. The framework consists of Data Collection, Problem Formulation, Policy Generation, and Robustness Analysis.

Our first step, data collection, entails collection of data that is going to be used for our research. Three types of data are required, including physical system quantities, hydroclimatic data, and water demand. Physical System Quantities entails quantitative data about the reservoir operations. Hydroclimatic data consists of flow rates of the streamflows included in the scope of our water system from catchments, as well as the rate of evaporation and precipitation of the reservoirs. The final data required is the water demand of irrigation districts. The data in this study utilizes the data from Sari (2022), which mainly uses data from Wheeler et al. (2016), Wheeler et al. (2018), UN Global Runoff Data Center, and other sources.

The problem formulation defines the scope of the decision problem in terms of objective formulation and exogenous uncertainties. Exogenous uncertainties refer to factors outside the explicit control of the decision maker, such as hydroclimatic factors and socioeconomic factors that will be included in the uncertainty parameter. In this stage we will also define the robustness metrics that will be utilized for our robustness analysis.

After the problem formulation and system modelling, policy generation is performed using multi-objective simulation-based optimization. The algorithm used will generate policies, and the model will evaluate the performance objectives. The outcome of the solutions generated by the algorithm are referred as 'policies', consisting of vectors of decision variable values. Building on the result of the policy generation, to incorporate uncertainty values for variables that are not included in the optimisation space, we conduct additional analysis by sampling over the uncertainty space and re-evaluating the optimised solutions.

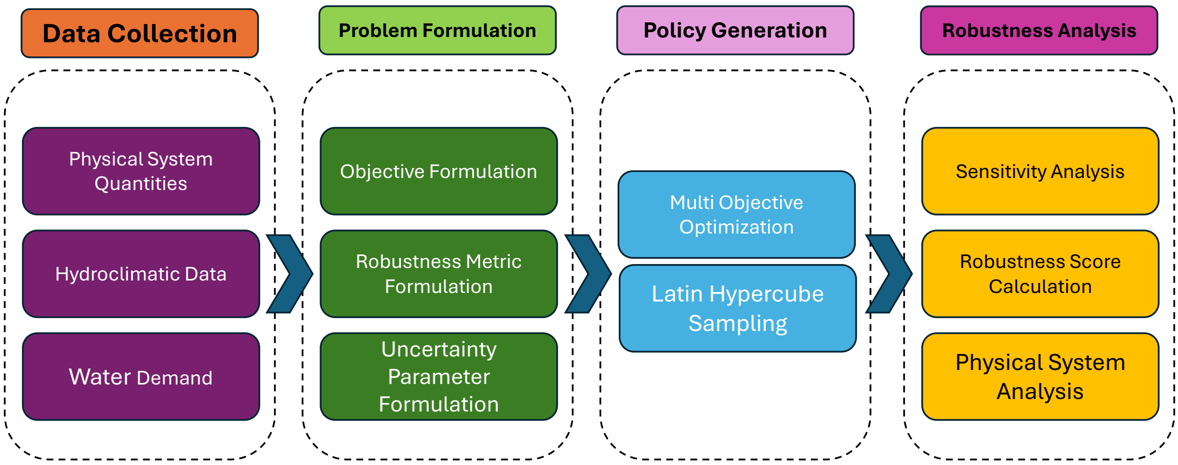

Figure 3.1: Methodological Framework Schematic Workflow.

Lastly, we analyzed how the performance of the policies changes under different SOWs (State of Worlds), a multivariate sample of the uncertainty factors that is created to sample each factor's bounds. The robustness of the policies is selected using a selection of robustness metrics, to follow the multi-metrics research objective, and based on the robustness scores we will examine the possible trade-offs among objectives, and based on the most robust policies we will also analyze how the policies impact the physical water system.

3.2. EMODPS

EMODPS method used in this study is chosen for its flexibility in addressing complex many-objective reservoir control problems. It can simultaneously handle various objective functions, such as minimizing worst-case scenarios or optimizing for expected values (Giuliani and Castelletti, 2016). EMODPS also effectively manages uncertainties through simulation-based methods (Giuliani et al., 2014) and offers lower computational costs compared to methods like SDP. It provides an approximation of the Pareto front from the first generation of solutions, making it well-suited for complex systems like the Nile (Giuliani et al., 2015). Its adaptive policy decisions enhance robustness against changing conditions.

The success of EMODPS, however, relies on the accurate representation operating policies and the capability of the many-objective evolutionary algorithm (MOEA) to optimize them. To address potential challenges, we selected ϵ-NSGA-II for its search diversity and ability to escape local optima (Kasprzyk et al., 2013; Salazar et al., 2016; Kollat and Reed, 2005). EMODPS involves integrating a nonlinear network that approximates decision-making rules directly into a simulation model. This network, acting as a policy function, determines the appropriate release decisions at each time step during the simulation based on the current state of the system at that given moment.

In the optimization process, the parameters of the policy function are gradually refined with each generation of a many-objective evolutionary algorithm run (Giuliani et al., 2015). Ideally, this process should lead to a set of solutions that represent the best possible trade-offs among the different objectives, known as a Pareto-optimal set.

We present a visual summary of the EMODPS methodology, taken from Kwakkel (2017), in a XLRM framework that identifies four possible locations where uncertainties exist in any system (Lempert et al., 2003). In this figure:

1. External Factors are labeled as "X" (these are elements outside the system that can affect its performance).
2. The Policy Function is shown as the "Relationship" within the System (R).
3. The Multi-objective Evolutionary Algorithm (MOEA) is used to measure performance ("M"), which then feeds back into the Policy Function and influences the Policy Levers ("L").

Ultimately, this process is designed to converge on a set of near-optimal policy alternatives, known as the Pareto-Approximate Policy Alternative(s).

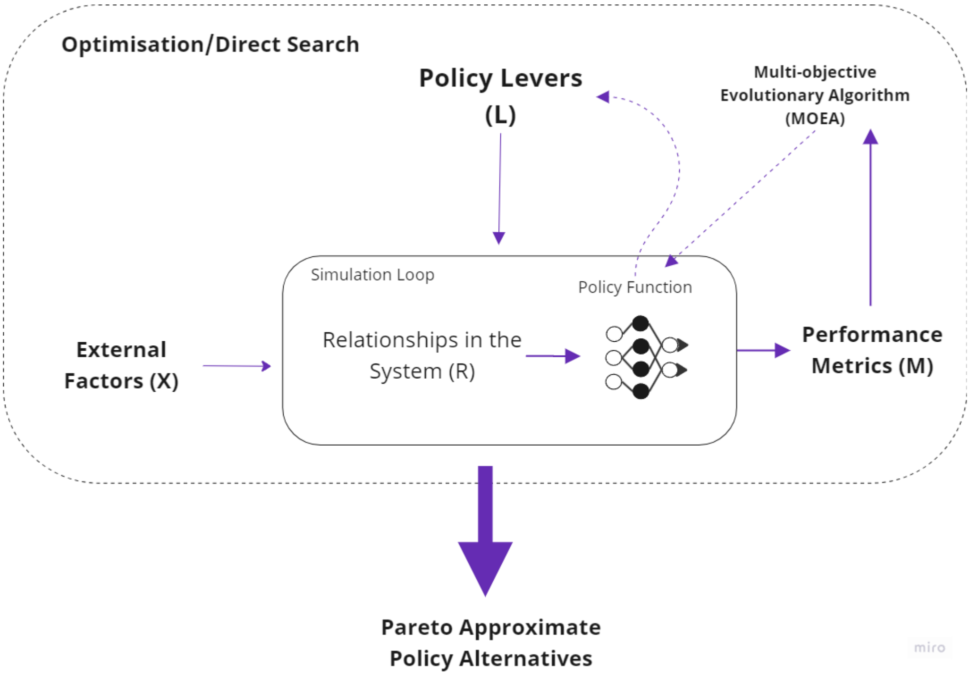

Figure 3.2: Illustration of EMODPS in an XLRM Framework, adapted from Kwakkel (2017).

The EMODPS method used in this study uses a mechanistis model for the water system adapted from Soncini-Sessa et al. (2007) to quantitatively explain the physical water system's operation, with the following formula:

3.2.1. System Modelling

The EMODPS method used in this study uses a mechanistis model for the water system adapted from Soncini-Sessa et al. (2007) to quantitatively explain the physical water system's operation, with the following mass-balance equation:

$$S_{T+1} = S_T + Q_{T+1} - E_{T+1} - R_{T+1} \tag{3.1}$$

where:

- S_T is the volume of water at the beginning of the decision step τ
- Q_{T+1} is the total inflow to the reservoir,
- E_{T+1} is the total net evaporation,
- R_{T+1} is the total volume of water released from the reservoir between decision steps τ and $\tau + 1$

We use a monthly decision interval, which balances capturing seasonal variations with the feasibility of long-term simulations. Each component of the mass-balance equation is calculated using known data at the time.

3.2.2. Policy Function

EMODPS in reservoir operations enhances the resilience and adaptability of reservoir systems by optimizing the parameters of control policies within a specified family of functions, including artificial neural networks (ANN) and radial basis functions (RBFs) (Quinn et al., 2017). A study by Giuliani et al. (2015) comparatively analyzed the potential of ANN and RBF policy parameterizations in solving EMODPS, and the result shows that RBF solutions are more effective in designing Pareto-approximate policies with better performance in terms of convergence, consistency, reliability, interpretability and diversity. Thus, RBFs are chosen over ANN for this study. We specifically use a structure adopted in Giuliani et al. (2020), a weighted sum of Gaussian RBFs, as the function that returns the release decisions. Gaussian distribution of the RBF parameters is used for its unbounded support, and effectivity in solving multi-objective control problems (Salazar et al., 2024; Giuliani and Castelletti, 2016).

The mathematical representation of this structure is the following:

$$u_\tau^k \; = \; u_\theta^k\left(\bar{z}_\tau\right) = \; \sum_{i=1}^{n}\left(w_i^k \varphi_i\left(\bar{z}_\tau\right) + \; \alpha_k\right) \tag{3.2}$$

Where:

- u_τ^k is the k-th release decision at month τ,
- θ is the parameter vector for RBF$_s$, $\theta = \; \left|c_{j,i}, b_{j,i}, w_i^k\right|$,
- Z_τ is the input vector,
- α^k is the constant adjustment parameter that fine-tunes the release decision

The total number of free parameters in a Gaussian RBF network is calculated as $(2mn) + (nK) + K$, where m represents the number of inputs, n represents the number of RBF$_s$, and K represents the number of release decisions.

3.3. Simulation Model

After the study area and the modelling approach has been laid out, we now define the simulation model for running experiments. This model explores the system behaviour of the Nile Basin under various conditions and policy alternatives. We implement the model in Python for its user-friendliness and extensive supplementary packages, for a transparent and reproducible process.

3.3.1. Model Conceptualization

The scope of the model focuses on the Blue Nile, incorporating the GERD Dam while treating the White Nile and Atbara rivers as exogenous inputs. The simulation covers a period from 2022 to 2042 (20 years) and captures both the filling period of the GERD and subsequent operational impacts. We employ an assumption that there is a cooperative setup in place where a hypothetical central authority coordinates release decisions for all reservoirs to enhance the flexibility and simplification in our policy exploration. Our model includes reservoirs, hydropower plants, irrigation districts, and catchments for determining system state and calculating our problem objectives.

The model components include:

1. Reservoirs
   The model includes four reservoirs: GERD, Roseires, Sennar, and the High Aswan Dam (HAD). These reservoirs influence the system state by storing and releasing water, with significant evaporation losses from their surfaces. Roseires (5.9 BCM) and Sennar (0.4 BCM) are vital for Sudan's agricultural irrigation schemes, while the HAD (148 BCM) is crucial for Egypt's water supply and serves as an indicator of downstream impacts from GERD.
2. Hydropower Plants
   Hydropower plants are modelled to calculate the amont the hydroelectric energy generated by the dams. All four dams included in this model have hydroenergy generation capacities, with GERD being a major focus for Ethiopia due to its significant potential for electricity production.

3. Irrigation Districts
   Irrigation districts represent major sources of water consumption such as agricultural irrigation schemes and urban districts. The model includes five irrigation districts in Sudan and one aggregate irrigation district for Egypt. The irrigation districts in Sudan includes Upstream Sennar, Gezira-Managil, Downstream Sennar, Tamaniat to Hassanab, and Hassanab to Dongola.
4. Catchments
   Catchments are areas where precipitation accumulates and contributes inflows to the system. In addition to the main tributaries (Blue Nile, White Nile, and Atbara), the model includes five more catchments based on the detailed schematic provided by Wheeler et al. (2018), including Dinder River, Rahad River, Lower Atbara, Upper Blue Nile, and Central Blue Nile.

The following topological chart illustrates the components of the system, showing the layout of reservoirs, irrigation districts and catchments. The main paths, smaller branches, and minor catchments are shown in purple. Water usage by irrigation districts is indicated with dashed arrows.

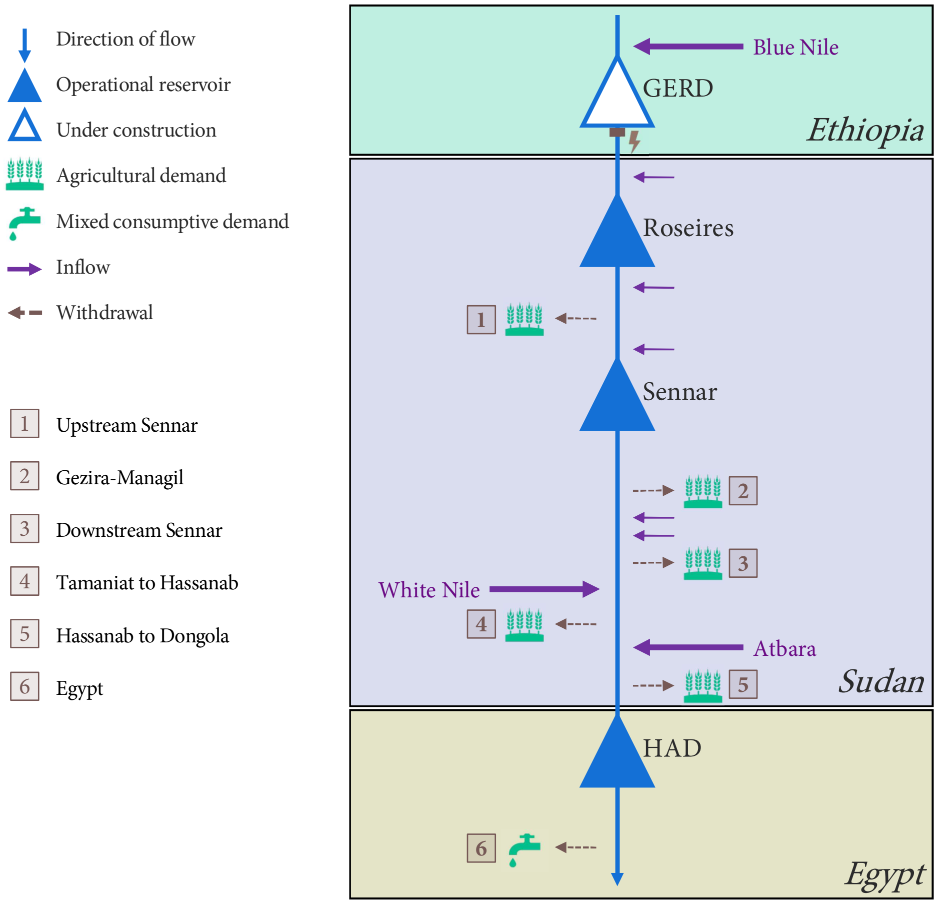

Figure 3.3: Topology of the Conceptual Model, taken from Sari (2022).

3.3.2. Model Formulation

The model's mathematical framework uses mass-balance equations to represent water storage dynamics and release decisions. The policy function for release decisions is based on radial basis functions (RBFs), with the release decision u_τ^k for reservoir k at month τ calculated as:

$$u_\tau^k = \sum_i \left[w_i^k \varphi_i(S_\tau, \sum_p Q_{\tau-1}^p, \tau \ mod\ 12) + \alpha_k \right] \tag{3.3}$$

where φ_i are the RBF$_s$, S_τ is the storage vector, $Q^p_{\tau-1}$ are the catchment inflows, and α_k are constant adjustment parameters. The mass-balance equations track changes in the reservoir storage, with the following formula:

$$s^k_{\tau+1} = s^k_\tau + \int_0^1 f(s^k_{\tau+t},\ q^k_\tau, e^k_\tau,\ u^k_\tau)\, dt \tag{3.4}$$

$$f = \frac{\Delta s^k_{\tau+t}}{\Delta t} = q^k_\tau - A^k(s^k_{\tau+t})e^k_\tau - r^k(s^k_{\tau+t}, u^k_\tau) \tag{3.5}$$

where f includes net inflow, evaporation, and actual release calculations.

3.3.3. Objectives

Our model is streamlining the objectives adapted from Sari (2022) from six objectives to four objectives to reduce computational complexity. The decision to reduce the objectives from six to four was driven by several considerations. Firstly, the 90^{th} percentile irrigation deficits for Egypt and Sudan did not significantly differ from the average irrigation deficits, making the additional complexity unnecessary. Secondly, the 90^{th} percentile Egypt irrigation deficit resulted in multiple best-performing policies under all scenarios, indicating redundancy in evaluating both average and 90*th* percentile metrics separately. Finally, reducing the number of objectives helps save computational resources to make the modelling process more efficient.

The selected objectives are:

1. Egypt Average yearly demand deficit (Egypt Irrigation Deficit): Minimizes the yearly average value of unmet demand in Egypt.

$$\frac{1}{20}\sum_{\tau}^{240} \max(0, D^{Egypt}_\tau - V^{Egypt}_\tau) \tag{3.6}$$

   where D^{Egypt}_τ is the water demand of Egypt and V^{Egypt}_τ is the received water flow by the irrigation district of Egypt in month τ. Egypt average yearly demand deficit is expressed in BCM/year for 20 years of the model run.

2. HAD frequence of months below minimum power generation level (Egypt Low HAD) minimizes the frequency of months when the HAD water level falls below the minimum power generation level (159 masl).

$$HAD_\tau = \begin{cases} 1, & \text{if } h^{HAD}(S_\tau) < 159. \\ 0, & \text{otherwise.} \end{cases} \tag{3.7}$$

3. Sudan Average yearly demand deficit (Sudan Irrigation Deficit) minimizes the yearly average value of unmet demand in Sudan, aggregated across multiple irrigation districts.

$$\frac{1}{20}\sum_{\tau}^{240}\sum_{j\in SD} \max(0, D^{Egypt}_\tau - V^{Egypt}_\tau) \tag{3.8}$$

   where SD is the set of irrigation districts in Sudan.

4. Ethiopia yearly hydroenergy generation from GERD (Ethiopia Hydropower Generation) maximize the yearly hydroelectric energy generation from GERD.

$$\frac{1}{20}\sum_{\tau}^{240} P^{GERD}_\tau \cdot d(\tau mod 12) \cdot 24 \tag{3.9}$$

$$P_{\tau}^{GERD} = min\left(r_{\tau}^{GERD},\ r_{max}^{GERD}\right) \cdot g \cdot \max(0, h_{S_{\tau}}^{GERD} - h_{turbine}^{GERD}) \cdot \eta^{GERD} \tag{3.10}$$

where r_{τ}^{GERD} is the flow corresponding to the maximum rotational speed of the turbine, $h_{turbine}^{GERD}$ is the height of the GERD turbines, and η^{GERD} is the efficiency of the hydropower generation.

To summarize, the following are the objective formulations we'll use in the study:

Table 3.1: Summary of Optimisation Objectives

Country	Objective	Aggregation Level	Unit	Direction
Egypt	Demand Deficit	Yearly Average	BCM/year	Minimise
Egypt	HAD level Reliability	Frequency over 20 years	%	Minimise
Sudan	Demand Deficit	Yearly Average	BCM/year	Minimise
Ethiopia	Hydroenergy Generation	Yearly Average	TWh/year	Maximise

3.3.4. Data Requirements

Three data types are required, including physical system quantities (such as storage-level-surface conversions, release constraints, and hydropower plant parameters), hydroclimatic data (including streamflows of major tributaries, minor catchments, and monthly evaporation rates), and demand data (considering water demand growth based on population growth rates in Egypt, Sudan, and Ethiopia). The detailed specifications and sources of this data are provided in Sari (2022).

3.3.5. Key Assumptions

Our simulation model incorporates several key assumptions for simplicity and focus. Firstly, the lag times between components are approximated, generally assuming water transitions occurs within the same month, except for a one-month delay between the Tamaniat and Hassanab irrigation districts. Sceond, no specific filling rule for the GERD is imposed, allowing exploration of various filling patterns to determine the optimal strategy for maximizing hydroelectric power generation. Third, the number of Radial Basis Functions (RBFs) is determined using heuristics from previous studies, resulting in ten RBFs. This is based on the sum of six input variables (storage values of four dams, total inflow of the previous month, and the current month index) and four output variables (release decisions for each dam). These assumptions are placed to streamline the model while maintaining its flexibility and robustness.

3.4. Uncertainty and Robustness Analysis

3.4.1. Quantifying Robustness of Policy Alternatives

The mechanisms of robustness analysis in this study are adapted from the many-objective robust decision making (MORDM) framework. After problem formulation and the policy generation, it is followed by robustness analysis to determine how well policies perform in uncertainties – which can be the basis for decision makers to settle on a policy to apply in real life.

After running the simulation model and generating our Policy Alternative(s), we conduct a robustness analysis to evaluate its effectiveness, ensuring that the policy alternatives we develop can perform well under a variety of uncertain future conditions. It should be noted that the policy alternative(s) obtained in the simulation model exhibits performance trade-offs, where improving in one objective necessitates inferior performance in one or more objectives. However, it is non-dominated, or no policies are ultimately better than the others. In this stage, we account for uncertainties in the system by testing each selected policy against a wide range of possible future scenarios, known as States of the World (SOWs). These scenarios represent different plausible futures. We then use robustness metrics, to measure how well each policy performs across all these scenarios, giving us a clear understanding of its reliability under varying conditions (Bonham et al., 2022).

There exists different robustness metrics which uses varying transformations and statistical calculations across the sampled SOWs. Each performance metric reflects different prioritization of objectives, thresholds, and risk aversion level of decision makers (McPhail et al., 2018). It should also be noted that different robustness metrics define robustness differently based on their respective performance criteria, and there exists inter-utility robustness tradeoffs (Herman et al., 2014).

Our goal in this study is to test multiple robustness metrics on the selected policies. Consequently, each policy will yield multiple robustness scores. To facilitate comparison, we will quantify these scores into a single, comparable value by calculating a weighted sum robustness score for each policy, with equal weights assigned to each robustness metric (Sahabuddin and Khan, 2021). This will provide a single robustness score per policy, simplifying the comparison of their overall performance.

3.4.2. Robustness Metrics

McPhail et al. (2018) defines 11 classifications of robustness metrics, each with varying levels of risk aversion and each with their own transformations, laid out in Table 3.4. Due to computational and scope limitation, we further narrow down the robustness metrics into four metrics for selection. Each robustness metrics have different prioritization, especially risk aversion level of decision makers. We aim to select robustness metrics with varying risk aversion level for a comprehensive analysis and to capture a broad spectrum of decision-making preferences.

We ultimately selected four metrics: Percentile-based Skewness, which measures the asymmetry in outcome distribution (Voudouris et al., 2014; Kwakkel et al., 2016); Mean-Variance, which assesses average performance and its variability (Hamarat et al., 2014); Undesirable Deviations, which quantifies the extent of negative deviations from a target (Kwakkel et al., 2016); and Minimax Regret, which focuses on minimizing the worst-case regret (Savage, 1951; Giuliani et al., 2016).

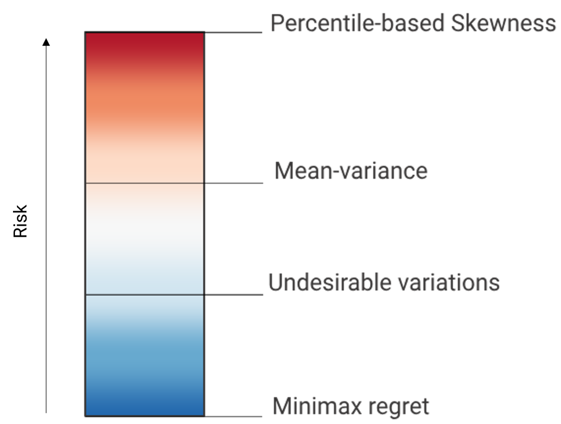

Figure 3.4: Selected Robustness Metrics based on Risk Level.

The spectrum of robustness metrics ranges from lower to higher risk level. Metrics at the top, like Percentile-based Skewness, have lower risk aversion, thus allowing for higher risk by considering extreme outcomes. As we move down, the metrics show increasing risk aversion and lower level of risk, favoring more cautious decision-making.

Percentile-based Skewness is positioned at the top of the spectrum, emphasizing the asymmetry in outcome distribution. It has the lowest risk aversion and highest risk level, allowing for greater variability and the possibility of extreme results. Mean-Variance sits in the middle of the spectrum, balancing the average performance with the variability of outcomes. It represents a moderate approach to risk, considering both the potential rewards and the associated risks. Undesirable Deviations is further down the spectrum, prioritizing the minimization of negative deviations from a target, which aligns with a more risk-averse, cautious strategy. Finally, Minimax Regret lies at the bottom of the spectrum, embodying the highest level of risk aversion and lower risk. It focuses on minimizing the worst possible regret, making it the most conservative choice.

We can calculate the respective robustness metrics with the following equations:

1. **Minimax Regret** bases decisions based on the concept of regret, defined as the difference between the performance of the best possible alternative when w_j is the true state of the world, and the performance of the chosen alternative a under the same state w_j.

$$r_j(a) = \max_a \left(f\left(a, w_j\right)\right) - f(a, w_j) \tag{3.11}$$

   The optimal alternative is then selected by adopting a pessimistic approach, which involves minimizing the maximum regret across all possible states of the world:

$$a^* = \arg\min_a(\max_{\Xi} r\,(a)) \tag{3.12}$$

2. **Undesirable Deviations** is variant of the approach where the mean and undesirable deviations from a target value are measured as separate objectives. This metric is formulated by the following:

   - For minimization:

$$f_i\,(x) = -\mu_i, \sum_{k=1}^{k} (x_k - q_{50})^2\,[x_k > q_{50}] \tag{3.13}$$

   - For maximization

$$f_i\,(x) = -\mu_i, \sum_{k=1}^{k} (x_k - q_{50})^2\,[x_k < q_{50}] \tag{3.14}$$

   q_{50} represents the median performance, k is a scenario, and x_k is the score for the i-th outcome indicator in scenario k. The summation is only performed for cases that meet the specified condition. This approach sums the squared differences from the median in the undesirable direction, effectively using these differences as a proxy for the skewness of the distribution.

3. **Percentile-based Skewness** is a method to measure the skewness of outcome distributions, utilizing a quantile-based approach as described by Voudouris et al. This approach addresses the potential unreliability of moment-based skewness definitions, particularly when the distribution is fat-tailed, meaning there are many data points in the tail regions.

   - For minimization:

$$f_i\,(x) = -\mu_i\left(\frac{(q_{90} + q_{10})/2 - q_{50}}{(q_{90} - q_{10})/2}\right) \tag{3.15}$$

   - For maximization:

$$f_i\,(x) = \mu_i\left(\frac{(q_{90} + q_{10})/2 - q_{50}}{(q_{90} - q_{10})/2}\right) \tag{3.16}$$

   For these equations, $q10$, $q50$, and $q90$ represents the 10^{th}, 50^{th}, and 90^{th} percentiles of the outcome distribution for the i-th indicator. A more positive value of this metric indicates that the density estimate is skewed to the right (higher values), while a more negative value indicates left skewness (lower values).

4. **Mean-Variance** aims to enhance the expected outcomes of a policy while reducing sensitivity to uncertainties. This approach increases the certainty of expected outcomes across various scenarios. For maximization objectives, robustness is defined as the mean divided by the standard deviation, where a higher mean and lower standard deviation indicate higher robustness. For minimization objectives, robustness is the mean multiplied by the standard deviation, with lower values indicating higher robustness.

- For minimization:

$$\frac{mean}{standard\ deviation} \tag{3.17}$$

- For maximization:

$$mean \times standard\ deviation \tag{3.18}$$

3.5. Experimental Setup

3.5.1. Baseline Optimisation

After developing the simulation mode, we execute the evolutionary optimization within the policy lever space, which includes 164 decision variables. This is calculated based on the formula of free parameters in a Gaussian RBF in chapter 2.2.1., as $(2mn)+(nK)+K$, thus $(2\ \times 10\ \times 6)+(10\times 4)+4 = 164$. We utilize the ϵ-NSGA-II algorithm, running 50,000 function evaluations while tracking epsilon progress and hypervolume metrics to ensure convergence. Epsilon values for all objectives, as well as the convergence result are detailed in the Appendix.

Our model is connected to the Python library Exploratory Modelling and Analysis (EMA) Workbench to run the ϵ-NSGA-II optimization algorithm (Kwakkel, 2017). The experiments are conducted using a 48-CPU computing node from the Delft High Performance Computing Centre (DHPC) / DelftBlue.

3.5.2. Uncertainty and Robustness Analysis

Although the model does incorporate uncertain hydro-climatic and socioeconomic factors, we must assign a specific value to every variable not included in the optimization search space to run the optimization. These requirement(s) mean that the optimization is conditioned on a reference scenario (Watson and Kaspryzk, 2017). Consequently, trade-offs identified in one scenario may not be significant if the process is repeated under varying SOWs (State of Worlds).

To address this limitation, we performed additional analyses based on the results of the baseline optimization. The main idea behind these analyses is to sample across the uncertainty space, generate alternative SOWs, and re-evaluate the optimized solutions under each scenario. The uncertain variables considered in this study, along with their sampling ranges, are presented in the following table.

Table 3.2: Uncertainty Variables and Ranges

Uncertainty Variable	Baseline Value	Range
Yearly Demand Growth Rate	0.0216	0.01-0.03
Blue Nile Mean Coefficient	1	0.75-1.25
White Nile Mean Coefficient	1	0.75-1.25
Atbara Nile Mean Coefficient	1	0.75-1.25
Blue Nile Deviation Coefficient	1	0.5-1.5
White Nile Deviation Coefficient	1	0.5-1.5
Atbara Nile Deviation Coefficient	1	0.5-1.5

The baseline annual demand growth rate of 2.16% for the Nile Basin countries is supported by population growth trends (World Bank, 2022). Agricultural expansion, driven by irrigation, is projected to grow annually by 1.72% and 1.43% until 2050 (IWMI, 2021), though these projections do not fully account for urban freshwater needs. Egypt and Ethiopia are expected to be among the top eight countries globally in population growth, with the Nile Basin's population nearing one billion by 2050 (Allan et al., 2019; McCartney and Menker Girma, 2012; Siam and Eltahir, 2017). To account for various possibilities, annual demand growth rates range between 0.01 and 0.03.

Among seven uncertain variables, six pertain to the hydrology of the Nile's major tributaries, except for the yearly demand growth rate. Probability distributions for monthly flow values, derived from historical data, introduce uncertainty influenced by inter-annual variability and climate change. Water consumption patterns beyond the focus area, such as increased usage in Ethiopia and upper White Nile countries, could significantly impact inflows. To cover a range of potential scenarios, adjustments were made to the mean and standard deviation of the probability distributions generating monthly streamflows, with a range of ±25% for the mean and ±50% for the standard deviation of baseline values. Using the Latin hypercube sampling technique, 5,000 alternative States of the World (SOWs) were created to efficiently cover the uncertainty space (Helton and Davis, 2003), forming the basis for further uncertainty analysis.

For computational and time limitation, after the Latin hypercube sampling is run, we will narrow down the policies to further analyze down to six policies, consisting of four extreme policies (best performing in the given objective), and two compromise policies.

Global Sensitivity Analysis

To investigate the influence of uncertainty variables and policy selection in explaining the variability in the outcomes of interest, we conduct a global sensitivity analysis. The objective values are evaluated through resimulation of a selected set of policies under the State of Worlds (SOWs) obtained from Latin Hypercube Sampling. We use the extra trees algorithm utilizing multiple regression trees for analysis, for assessing the significance of each uncertain variable and policy selection. Through this approach, the influence of each input on the various outcomes of interest can be quantified.

Robustness of Policy Analysis

The primary objective of this research is to evaluate how well policy alternatives performs under different definitions of robustness. We use four selected robustness metrics: Percentile-based Skewness, Mean-variance, Undesirable Deviations, and Minimax Regret. Overall Robustness Scores are the weighted sum of each policy performance under different robustness metrics.

4 Results

4.1. Baseline Optimization

We run the EMODPS optimization under the baseline scenario without the uncertainties accounted in. This results in 748 Pareto-optimal solution alternatives. The convergence analysis for this optimisation run can be found in the Appendix.

4.1.1. Objective Trade-Offs

The following presents the Parallel Coordinate plot, representation of the optimization results. Each line in the parallel plot represents a different policy alternative generated by the optimization. The plot itself shows how each policy alternative performs across multiple objectives, displayed on the vertical axis. Each axis corresponds to a specific performance metric, such as Egypt Irrigation Deficit and Ethiopia Hydropower. The position of the lines on each axis indicates the value of that objective for a particular policy, with the direction of preference indicated by the arrow on the right side of the plot – the higher the position of the line, the preferable it will be, in our case.

The highlighted, colour-coded lines highlight specific policy alternatives that performs best for objectives, as indicated in the legend. For example, the light pink line represents the policy with the best performance for Egypt's Irrigation (Best Egypt Irr).

The main goal of the parallel plot is to show how performance of each policy alternative changes relative to other objectives, as indicated by the lines moving across the axes. A steep upward or downward movement between axes shows a significant trade-off between objectives.

Trade-Offs Extreme and Compromise Policies

From the plot, we highlight six policies, which consist of four extreme policies and two compromise policies. The extreme policies are:

1. Best Egypt Irrigation Deficit (policy with the best performance for Egypt Irrigation Deficit, represented in Pink).
2. Best Egypt Low HAD (policy with the best performance for Egypt HAD Level Reliability, represented in Purple).
3. Best Sudan Irrigation Deficit (policy with the best performance for Sudan Irrigation Deficit, represented in Green).
4. Best Ethiopia Hydropower (policy with the best performance for Ethiopia Hydropower Level, represented in Blue).

We then add two compromise solutions to the extreme policy set, in which both were obtained through imposing a constraint on the objective value sets. The policy Compromise: Percentile Threshold represented in Orange makes all objectives attain a value that is more desirable than the 45^{th} percentile (100^{th} percentile is the most desirable or best objective value). Meanwhile, the policy Compromise:

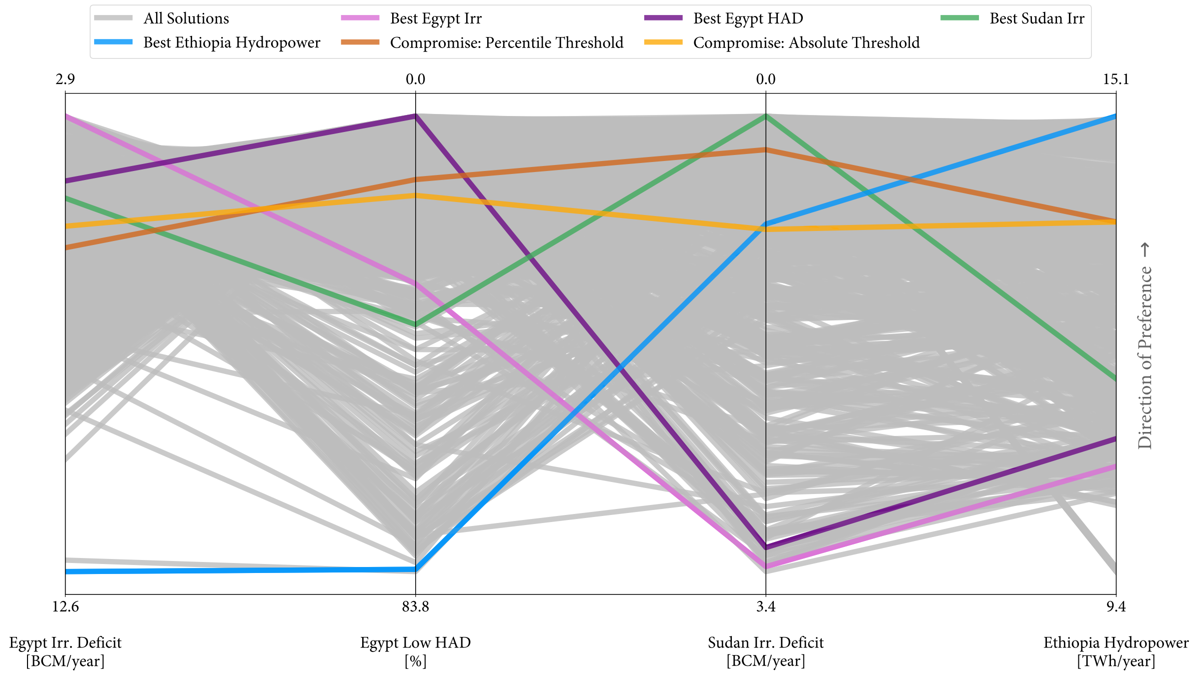

Figure 4.1: Parallel coordinates plot of Pareto-optimal solutions of baseline optimisation with Compromise Policies.

Absolute Threshold represented in Yellow corresponds to all objectives that attain a value that is higher than 0.75 on the normalised scale where the best objective value corresponds to 1.

For the extreme solutions, a similar theme is observed, where the solutions perform best at their respected extreme objectives, however always at the expense of other objectives. Best Egypt Irrigation Deficit Policy prioritizes minimizing Egypt's irrigation deficit, but this comes at the expense of other objectives, such as Ethiopia's Hydropower generation and Sudan Irrigation Needs. Best Egypt Low HAD Policy focuses on maximizing the reliability of Egypt's High Aswan Dam (HAD) levels but leads to deficits in irrigation and reduced Ethiopia hydropower generation and performs very poorly in Sudan Irrigation deficit goals.

Best Sudan Irrigation Deficit optimizes Sudan's irrigation deficit, but poorly compromises Ethiopia's Hydropower and other objectives, although its Egypt Irrigation Deficit has a decent performance. Best Ethiopia Hydropower maximizes Ethiopia's hydropower generation well, however, performs very poorly in Egypt Irrigation Deficit and Egypt Low HAD.

For compromise solutions, we see a more balanced approach. Both Compromise: Percentile Threshold and Compromise: Absolute Threshold policies all achieve objective performance that are more balanced and has lower difference. Although the performance in all objectives is never the best, or not optimizing any single objectives compared to the extreme policies, its performance is still scoring quite well and competitively across all objectives.

Pairplot Comparison of Objectives

To examine the overall trade-off patterns within the full Pareto front, in addition to the differences between individual performance maximising policies, we visualize two-way comparisons between objectives with a pair plot as shown in figure 4.2. Each scatter plot is fitted a regression line which informs about the trend between two objectives. When the angle between the regression line and the x-axis is approximately 45° in the positive direction, it is an indication that there is an alignment between two objectives. On the other hand, when the angle is -45°, it indicates a trade-off in a similar manner. The slope of the regression fitted for each objective pair is shown in table 1. In these scatter plots, each colored point represents a specific policy alternative, as indicated in the legend. For the Ethiopia Hydropower objective (maximization), the axes are reversed so that objective values get more desirable towards the bottom on the y-axis, and the left on the x-axis.

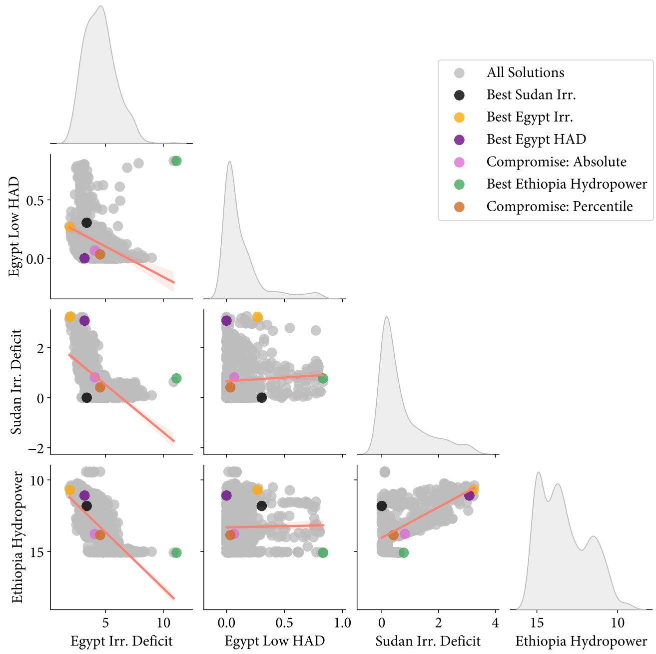

Figure 4.2: Pairplot of Objective Values.

It should be noted that the slope itself is not enough to draw conclusions on about the significance of the relation. If the points are scattered dispersedly and there is no substantial accumulation around the regression line, the slope may be misleading. In this regard, we also use Pearson correlation coefficient to quantify the significance of the relation between the two objectives. The closer the coefficient is to +1 or to -1, the more stable the relation is. The Pearson correlation coefficients calculated for each objective pairs can be found in table 2.

The slopes and Pearson correlation coefficients reveal important insights about the relationships between the objectives. The most prominent relation among all objective pairs is the trade-off between Egypt's irrigation deficit and Sudan's irrigation deficit. Tables 1 show that the slope is -0.943137 and the Pearson correlation coefficient is -0.625568. This strong negative alignment suggests that policies aiming to reduce Egypt's irrigation deficit are likely to increase Sudan's irrigation deficit. This result highlights a significant trade-off that policymakers need to consider when formulating strategies for water management in these regions.

Another notable pattern is the relationship between Egypt's irrigation deficit and Ethiopia's hydropower. The slope of -0.794570 and the Pearson correlation coefficient of -0.711667 indicate a strong negative relation, suggesting that improvements in Egypt's irrigation deficit are associated with reductions in Ethiopia's hydropower generation. This trade-off is critical as it shows the potential conflict between agricultural water use in Egypt and energy production in Ethiopia.

In contrast, the relationship between Sudan's irrigation deficit and Ethiopia's hydropower exhibits a moderate positive slope of 0.611221 and a Pearson correlation coefficient of 0.585811. This indicates a potential synergy where policies that improve Sudan's irrigation deficit could also benefit Ethiopia's hydropower generation. This positive relationship suggests that cooperative water management strategies could be developed to enhance both irrigation and hydropower objectives.

Interestingly, Egypt's Low HAD objective shows minimal interaction with other objectives. For instance, the slope of 0.074666 and the Pearson correlation coefficient of 0.063587 between Egypt's Low HAD and Sudan's irrigation deficit suggest that policies affecting Egypt's Low HAD have little to no impact on Sudan's irrigation deficit. This weak relationship indicates that Egypt's reservoir management can be somewhat decoupled from Sudan's irrigation outcomes.

Moreover, Ethiopia's hydropower objective generally has negative relations with the irrigation deficit objectives of both Egypt and Sudan. However, this relationship is more pronounced with Egypt, as shown by the slope and correlation values. Despite these negative correlations, the spread of deficit outcomes over a wide range of values for Egypt and Sudan suggests that there might be room for negotiation and finding mutually acceptable policies.

Although individual color-coded policies might suggest an alignment between Ethiopia Hydropower and Egypt Low HAD, the overall Pareto-optimal set indicates that these two objectives are uncorrelated. This lack of correlation is because both objectives are influenced by major decision points like releases from the GERD and the HAD. The Pareto-approximate set flexibly represents various possibilities for these decision points, covering scenarios where the GERD releases are not optimal from Egypt's perspective. This flexibility points to an internal trade off for Egypt between maintaining reservoir levels and managing irrigation deficits downstream.

Table 4.1: Regression Slopes between Objective.

	Egypt Irr. Deficit	Egypt Low HAD	Sudan Irr. Deficit	Ethiopia Hydropower
Egypt Irr. Deficit	1	-0.522202	-0.943137	-0.79457
Egypt Low HAD	-0.522202	1	0.074666	0.01991
Sudan Irr. Deficit	-0.943137	0.074666	1	0.611221
Ethiopia Hydropower	-0.79457	0.01991	0.611221	1

Table 4.2: Pearson Correlation Coefficient between Objectives.

	Egypt Irr. Deficit	Egypt Low HAD	Sudan Irr. Deficit	Ethiopia Hydropower
Egypt Irr. Deficit	1	-0.361773	-0.625568	-0.711667
Egypt Low HAD	-0.361773	1	0.063587	0.016251
Sudan Irr. Deficit	-0.625568	0.063587	1	0.585811
Ethiopia Hydropower	-0.711667	0.016251	0.585811	1

4.2. Uncertainty Analysis

4.2.1. Global Sensitivity Analysis

In table 3, we have the importance of uncertain factors in explaining the variability in outcomes of interest. These values were calculated using the extra trees algorithm and fitting 1,000 regression trees for each objective.

The table indicates that the selected policy is influential on the outcome for all objectives, particularly for Sudan, where it significantly impacts the irrigation deficit. The mean Blue Nile inflow is of utmost importance for hydroenergy generation in Ethiopia, demonstrating the highest impact among all factors. Similarly, average inflows of tributaries, especially the Blue Nile, have a substantial effect on the elevation of the HAD in Egypt.

The irrigation deficit in Egypt is greatly influenced by the demand growth rate, showing the highest importance among the factors for this objective. Inflow deviations of major tributaries did not turn out to be impactful for any of the objectives. This lack of impact can be attributed to the high level of storage and regulation provided by the four reservoirs in the system. The Atbara and White Nile mean inflows showed minimal importance across the objectives, with the White Nile mean inflow only moderately affecting the low HAD scenario in Egypt.

Table 4.3: Feature score of uncertain parameters on each outcome calculated using extra trees algorithm.

	Egypt Irr. Deficit	Egypt Low HAD	Sudan Irr. Deficit	Ethiopia Hydropower
Policy	0.18	0.09	0.6	0.17
Atbara Deviation	0	0	0	0
Atbara Mean	0.01	0.04	0	0
Blue Nile Deviation	0	0	0	0
Blue Nile Mean	0.12	0.6	0.3	0.8
White Nile Deviation	0	0	0	0
White Nile Mean	0.06	0.21	0.01	0
Demand Growth	0.59	0.01	0.06	0

4.2.2. Robustness Analysis

We performed the robustness analysis using four robustness metrics with varying risk-averse levels, Minimax Regret, Percentile-based Skewness, Mean-Variance, and Undesirable Deviations, which was explained in detail in section 3.4.

The general approach for calculating the robustness scores involves analyzing the performance of policies across multiple scenarios, considering both the direction of the outcome (whether higher or lower values are preferable). Echoing the objective trade-offs of the baseline optimization performed in subsection 4.1, in this robustness analysis we are only analyzing six policies, including 4 extreme policies for each of the objectives and two compromise policies, absolute and percentile.

This decision is made to streamline the analysis due to time constraints and the scope of the thesis. By focusing on just six policies, including the extreme policies for each objective and two compromise policies, we can effectively demonstrate the core argument of this study. The primary goal is not to exhaustively analyze every possible policy but rather to illustrate how different robustness metrics can lead to the identification of different policies as the most robust. This approach allows us to clearly show that the choice of robustness metric significantly influences which policy is considered most robust, thereby proving the concept without needing to analyze the entire policy set.

The directions of the objectives are carefully considered in each robustness analysis—whether a lower or higher value is preferred is directly integrated into the calculations. Specifically, for objectives such as Egypt Irrigation, Egypt Low HAD, and Sudan Irrigation, lower values are preferred, which means that better outcomes are achieved when the values are minimized. In contrast, for Ethiopia Hydropower, higher values are desired, as greater hydropower output represents a better outcome.

This difference in direction is reflected in the mathematical formulation used in the robustness calculations. Take, for example, the Minimax Regret analysis: when calculating the regret score for a policy concerning Ethiopia Hydropower (where higher values are better), the regret is determined by comparing the maximum possible hydropower output across all policies with the actual output achieved by the policy in question. The regret is calculated as the difference between these two values. A regret score of zero, in this context, indicates that the policy achieved the best possible outcome, leaving no room for regret. Conversely, for objectives where lower values are preferred, such as Egypt Irrigation, the calculation is reversed. Here, the regret is calculated by comparing the minimum possible value with the value achieved by the policy. Again, a regret score of zero indicates that the policy achieved the optimal outcome, minimizing regret.

This approach applies consistently across all robustness analyses. In the Undesirable Deviations analysis, deviations are penalized differently based on whether lower or higher values are preferred. For lower-is-better objectives, deviations above the median are penalized, while for higher-is-better objectives like Ethiopia Hydropower, deviations below the median are penalized.

In the Percentile-Based Skewness analysis, the skewness score is adjusted for the direction of preference. For lower-is-better objectives, the score is inverted to favor outcomes clustered towards lower values, while for higher-is-better objectives, positive skewness is preferred.

In the Mean-Variance analysis, the robustness score is also adjusted based on the objective's direction. For lower-is-better objectives, the score is calculated by multiplying the mean by the standard deviation, favoring low and consistent outcomes. For higher-is-better objectives, the score is calculated by dividing the mean by the standard deviation, rewarding high and stable outcomes.

Minimax Regret

The minimax regret scores provide a measure of the worst-case regret that could occur for each policy across the different outcomes. The regret values represent the difference between the best possible outcome and the actual outcome achieved by each policy under the worst-case scenario. The goal in minimax regret analysis is to minimize this worst-case regret.

Table 4.4: Minimax Regret Scores of Policies on Each Objectives.

Policy	Egypt Irr.	Egypt Low HAD	Sudan Irr.	Ethiopia Hydropower
Best Egypt HAD	2.79	0.03	4.54	4.23
Best Egypt Irr.	1.45	0.75	4.52	4.47
Best Ethiopia Hydropower	12.79	0.86	2.44	0
Best Sudan Irr.	4.99	0.72	0	3.42
Compromise: Absolute	3.8	0.52	4.03	1.5
Compromise: Percentile	4.38	0.48	3.41	1.71

The following table provides a clear indication of which policies perform best across the different objectives, considering the worst-case regret scenario for each. The analysis shows that:

- Egypt Irrigation: The "Best Egypt Irrigation" policy performs best with the lowest regret score of 1.45, but it still has a non-zero regret, indicating that even in its optimal objective, it does not eliminate regret.
- Egypt Low HAD: The "Best Egypt HAD" policy has the lowest regret score of 0.03, reflecting its strong performance in this objective. However, like Egypt Irrigation, it is not entirely free of regret.
- Sudan Irrigation: The "Best Sudan Irrigation" policy has a perfect score of 0, meaning it completely minimizes regret for this objective.
- Ethiopia Hydropower: The "Best Ethiopia Hydropower" policy also achieves a perfect score of 0, indicating that it fully meets the objective without any regret.

While we would intuitively expect the extreme policies to perform best in their respective objectives, the presence of residual regret scores in both the "Best Egypt Irrigation" and "Best Egypt HAD" policies suggests that these policies, while optimal in their target objectives, do not eliminate regret. This residual regret implies that there may be competing factors or trade-offs within the same objective, which prevent these policies from being perfect. For instance, the "Best Egypt Irrigation" policy, although focused on minimizing irrigation deficits, still leaves room for improvement when compared to other scenarios that may have performed slightly better under certain conditions.

Interestingly, the Sudan Irrigation and Ethiopia Hydropower objectives have policies ("Best Sudan Irrigation" and "Best Ethiopia Hydropower") that achieve a regret score of 0, indicating that these objectives can be fully optimized without any competing trade-offs. This suggests that the goals for Sudan Irrigation and Ethiopia Hydropower are more straightforward and less influenced by conflicting factors, making it easier for these policies to perform without regret.

Another point of reflection is the performance of the compromise policies. Although these policies are designed to balance across multiple objectives, their regret scores indicate that they perform reasonably well but do not excel in any single objective. This reflects the inherent trade-off in compromise policies: they aim to balance competing interests but may not fully satisfy any objective as well as the extreme policies. Finally, the fact that even the best policies for Egypt Low HAD and Egypt Irrigation still show residual regret suggests while these policies are optimized for one aspect of the objective, other dimensions (perhaps related to environmental or social factors) are still not fully addressed, leaving room for regret.

Undesirable Deviations

The Undesirable Deviations analysis evaluates how much a policy deviates from a target value, which in this case is the median (q50) of each outcome across all scenarios, following the reference target from Kwakkel et al. (2016). This method penalizes deviations from this target, but the nature of the penalty depends on the direction of the objective—whether higher or lower values are preferred.

Table 4.5: Undesirable Deviations Scores of Policies on Each Objectives.

Policy	Egypt Irrigation Deficit	Egypt Low HAD	Sudan Irrigation Deficit	Ethiopia Hydropower
Best Egypt HAD	91631.99	971.63	2695.7	12704.9
Best Egypt Irr.	94634.69	486.27	2325.8	9019.21
Best Ethiopia Hydropower	146080.6	6.61	1083.89	15133.7
Best Sudan Irr.	115002.6	438.52	13.1	10867.1
Compromise: Absolute	108564.5	1019.17	4095.77	15196.5
Compromise: Percentile	112331.8	1059.16	2483.09	16698.4

The following table provides a clear indication of which policies perform best across the different objectives. The analysis shows that:

- Egypt Irrigation Deficit: The “Best Egypt HAD” policy shows the least undesirable deviations, making it the best performer for this objective.
- Egypt Low HAD: The “Best Ethiopia Hydropower” policy has the lowest deviations, which is counterintuitive since this policy is not focused on this objective.
- Sudan Irrigation Deficit: The “Best Sudan Irrigation” policy performs the best, aligning with expectations.
- Ethiopia Hydropower: The “Best Egypt Irrigation” policy surprisingly shows the lowest deviations, despite not being directly targeted at maximizing hydropower output.

The results reveal several interesting and somewhat unexpected findings. First, although we would typically expect the extreme policies to minimize undesirable deviations in their respective objectives, this is not always the case. For instance, the "Best Ethiopia Hydropower" policy, which we might expect to perform well in the Ethiopia Hydropower objective, incurs higher deviations compared to the "Best Egypt Irrigation" policy. Similarly, for Egypt Low HAD, the "Best Ethiopia Hydropower" policy achieves the lowest deviation, despite its primary focus being on a different objective.

These unexpected outcomes suggest that the policies optimized for a specific objective do not necessarily align perfectly with the median-based target (q50), leading to higher deviations. This may be due to the inherent variability in outcomes across different scenarios, where the median value may not fully represent the policy’s strength in its intended objective.

This observation raises the possibility that selecting a different target than q50 might yield more insightful or favorable results. For example, if the target were set based on a different percentile (such as the 75^{th} or 25^{th}), or even an average value, it could potentially reduce the apparent deviations and better reflect the policy’s true performance relative to its intended goal.

The high scores across the board indicate that there is significant variability and deviation from the median target values, reflecting the challenges in achieving consistently optimal outcomes across all scenarios. This variability shows the complexity of the problem space, where achieving a balance between different objectives is difficult, particularly when the policies are tailored to extreme outcomes rather than compromise solutions.

The compromise policies, both “Absolute” and “Percentile” generally perform better than the extreme policies in some objectives but worse in others. This is expected as they are designed to balance performance across multiple objectives rather than excelling in any single one.

Percentile Based Skewness

The Percentile-Based Skewness analysis offers insights into how the distribution of outcomes skews in relation to the desired objectives, where we consider both the direction of preference (whether a higher or lower value is better) and the shape of the distribution. The skewness score indicates whether outcomes are skewed towards more favorable or less favorable results compared to the median (q50), with a positive skewness indicating that outcomes are clustered towards higher values, and a negative skewness indicating clustering towards lower values.

Table 4.6: Percentile Based Skewness Scores of Policies on Each Objectives.

Policy	Egypt Irrigation Deficit	Egypt Low HAD	Sudan Irrigation Deficit	Ethiopia Hydropower
Best Egypt HAD	-0.32	-1	0.18	0.21
Best Egypt Irr.	-0.32	-0.2	0.14	0.27
Best Ethiopia Hydropower	-0.14	0.86	-0.33	-0.01
Best Sudan Irr.	-0.27	-0.14	-0.99	0.28
Compromise: Absolute	-0.32	-0.81	-0.52	0.04
Compromise: Percentile	-0.28	-0.87	-0.71	0.01

The analysis shows that:

- Egypt Irrigation Deficit: The policy “Best Egypt Irr” has a skewness score of -0.32, indicating that outcomes tend to be better (lower) than the median. Similarly, “Best Egypt HAD”, and “Compromise: Absolute” also shows a skewness score of -0.32, reflecting a similar distribution. The negative skewness here makes sense since the objective is to minimize the irrigation deficit. Both policies show a tendency towards achieving lower (better) outcomes than the median.
- Egypt Low HAD: The policy "Best Egypt HAD" achieves a skewness score of -1, which is the most negative across all policies and objectives. This indicates that this policy strongly favors outcomes that are better (lower) than the median, aligning well with the minimization goal.
- “Best Sudan Irr” achieves a skewness score of -0.99, indicating a strong tendency towards better (lower) outcomes than the median, which aligns with the objective of minimizing irrigation deficits.
- The objective is to maximize hydropower, and “Best Sudan Irr” shows the most favorable skewness score (0.28), indicating a tendency towards higher (better) outcomes than the median. “Best Ethiopia Hydropower” shows a slightly negative skewness score (-0.01), suggesting that this policy doesn’t significantly outperform the median and might even perform slightly worse, which is counterintuitive given that it’s optimized for this very objective.

The most notable counterintuitive result is seen in the Ethiopia Hydropower objective. The policy specifically optimized for maximizing hydropower (“Best Ethiopia Hydropower”) does not exhibit the highest positive skewness, and instead, “Best Sudan Irr” performs better according to this metric. This might be due to the method of calculation, where the skewness is influenced by the range and distribution of outcomes.

Best Ethiopia Hydropower, for the Ethiopia Hydropower objective has a slightly negative skewness score of -0.01, this indicates that the outcomes are more evenly distributed around the median or even slightly skewed towards lower values, which is not ideal for a maximization objective. It implies that the policy might not be achieving high outputs as frequently as expected. If the "Best Ethiopia Hydropower" policy produces more consistent but not necessarily high outcomes, it could lead to a skewness score that doesn't reflect high performance relative to the 90^{th} percentile.

The compromise policies show varied performance. For instance, "Compromise: Percentile" generally exhibits negative skewness across most objectives, which is less favorable for objectives where higher values are desired, such as Ethiopia Hydropower. On the other hand, it performs reasonably well (though not optimally) for minimization objectives. The skewness scores for compromise policies indicate that they tend to achieve outcomes that are more balanced but not necessarily extreme or highly favorable, which aligns with their design as compromise solutions rather than objective-maximizing strategies.

Mean-Variance

The mean-variance analysis provides a measure of robustness by considering both the average performance of a policy and the variability of that performance across different scenarios. The goal is to identify policies that not only perform well on average but also do so consistently, with lower variability being preferable for minimization objectives and higher consistency (low variability relative to high mean performance) being preferable for maximization objectives.

Table 4.7: Mean-Variance Scores of Policies on Each Objectives.

Policy	Egypt Irrigation Deficit	Egypt Low HAD	Sudan Irrigation Deficit	Ethiopia Hydropower
Best Egypt HAD	26	0.09	2.86	4.27
Best Egypt Irr.	25.19	0.15	2.65	4.51
Best Ethiopia Hydropower	87.33	0.22	0.45	6.18
Best Sudan Irr.	40.78	0.15	0	4.28
Compromise: Absolute	33.51	0.13	0.91	5.45
Compromise: Percentile	37.3	0.12	0.41	5.35

This analysis shows that:

- Egypt Irrigation Deficit: "Best Egypt Irr" policy scores the best in this objective. This is the expected outcome since this policy is specifically optimized for minimizing the Egypt Irrigation Deficit. The low mean-variance score indicates that this policy performs consistently well across scenarios, maintaining low irrigation deficits with relatively low variability.
- Egypt Low HAD: The scores are very low across all policies, with the "Best Egypt HAD" policy scoring 0.09. As expected, the policy optimized for Egypt Low HAD performs best for this objective. The very low mean-variance score suggests that this policy achieves low HAD levels consistently, with minimal variability across scenarios.
- Sudan Irrigation Deficit: The "Best Sudan Irr" policy achieves a score of 0, indicating it perfectly minimizes the deficit with no variability across scenarios. This makes it the ideal policy for this objective, as expected.
- Ethiopia Hydropower: The "Best Ethiopia Hydropower" policy scores 6.18, which is the highest among all policies for this objective. This score reflects a high average hydropower output combined with low variability, making it the best policy for maximizing hydropower production.

The score for Best Ethiopia Hydropower is the highest because, for the Ethiopia Hydropower objective, the goal is to maximize the outcome rather than minimize it. In the mean-variance analysis, when higher values are preferred (as in the case of maximizing hydropower output), the robustness score is calculated by dividing the mean by the standard deviation. A higher score indicates a better policy because it reflects a high average performance (mean) with relatively low variability (standard deviation).

This is the opposite of minimization objectives, where lower scores are better because they indicate low outcomes with low variability. Thus, the highest score for Best Ethiopia Hydropower signifies that this policy achieves the best balance of high hydropower output with stable performance across scenarios.

Both compromise policies (Absolute and Percentile) do not perform best in any of the individual objectives. Their scores are generally higher than the extreme policies but still reflect a balance between mean performance and variability. For example, "Compromise: Absolute" has a score of 33.51 for Egypt Irrigation Deficit, which is higher than the optimized policy but still represents a reasonable balance between different objectives.

What is noticed with this robustness metrics, is all extreme policies performs as expected, meaning the extreme policies designed to optimize specific objectives generally perform best for those objectives. This is intuitive as they are tailored to minimize or maximize the respective outcomes.

Comparing Robustness Metrics

In this chapter, we take a close look at the different robustness metrics used in this study to assess how well policies perform across various scenarios. Our goal is to see how each metric affects the ranking of policies and to understand the trade-offs that come with each approach. By doing this, we aim to shed light on how different metrics can lead to different conclusions about which policies are the most robust.

We start by providing a clear summary of the rankings for each policy, based on the robustness metrics we've chosen—Minimax Regret, Undesirable Deviations, Percentile-Based Skewness, and Mean Variance. In these rankings, a score of 1 represents the best performance, while a score of 6 represents the worst.

Table 4.8: Relative Robustness Score Rankings based on Minimax Regret Metrics.

Policy	Egypt Irr.	Egypt Low HAD	Sudan Irr.	Ethiopia Hydropower
Best Egypt HAD	2	1	6	5
Best Egypt Irr.	1	5	5	6
Best Ethiopia Hydropower	6	6	2	1
Best Sudan Irr.	5	4	1	4
Compromise: Absolute	3	3	4	2
Compromise: Percentile	4	2	3	3

Table 4.9: Relative Robustness Score Rankings based on Undesirable Deviation Metrics.

Policy	Egypt Irr.	Egypt Low HAD	Sudan Irr.	Ethiopia Hydropower
Best Egypt HAD	1	5	5	3
Best Egypt Irr.	2	4	4	1
Best Ethiopia Hydropower	6	1	3	2
Best Sudan Irr.	5	3	1	4
Compromise: Absolute	3	6	6	5
Compromise: Percentile	4	2	2	6

The comparison of the rankings across different robustness metrics reveals that the identification of the "best" policy is far from straightforward. The rankings show considerable variation depending on the robustness metric applied, highlighting the complexity and trade-offs inherent in each approach.

It is evident that each robustness metric can lead to different conclusions about which policy is the most robust. For instance, using the Minimax Regret metric, compromise policies like "Compromise:

Table 4.10: Relative Robustness Score Rankings based on Percentile-Based Skewness Metrics.

Policy	Egypt Irr.	Egypt Low HAD	Sudan Irr.	Ethiopia Hydropower
Best Egypt HAD	1	1	5	5
Best Egypt Irr.	1	4	4	2
Best Ethiopia Hydropower	4	3	6	6
Best Sudan Irr.	3	5	1	1
Compromise: Absolute	1	3	3	3
Compromise: Percentile	2	2	2	4

Table 4.11: Relative Robustness Score Rankings based on Mean-Variance Regret Metrics.

Policy	Egypt Irr.	Egypt Low HAD	Sudan Irr.	Ethiopia Hydropower
Best Egypt HAD	2	1	5	5
Best Egypt Irr.	1	4	4	4
Best Ethiopia Hydropower	6	6	3	1
Best Sudan Irr.	5	5	1	3
Compromise: Absolute	3	3	2	2
Compromise: Percentile	4	2	2	3

Absolute" and "Compromise: Percentile" tend to perform well across the board, suggesting that these policies strike a balance across various objectives. However, when considering Undesirable Deviations, "Best Egypt Irr" emerges as the top performer, particularly for the "Ethiopia Hydropower" objective, which is not its primary focus.

Interestingly, the Percentile-Based Skewness metric reveals a surprising outcome where extreme policies, which are typically expected to excel in their respective objectives, sometimes rank poorly. For example, "Best Ethiopia Hydropower" which should theoretically perform best for the "Ethiopia Hydropower" objective, ranks worst for that metric under Percentile-Based Skewness.

To allow for better comparison and a clearer overview of the performance across different robustness metrics, we next aggregate the scores. This aggregation aims to provide a consolidated ranking, helping to identify which policies consistently perform well—or poorly—across the various robustness metrics. By summing the relative rankings, we can determine which policies are the most robust overall according to each metric. This approach enables us to better understand the trade-offs and see which policies offer the best balance across different robustness considerations.

To aggregate the scores, we summed the relative rankings each policy received across all robustness metrics. For each policy, the rankings from Minimax Regret, Undesirable Deviations, Percentile-Based Skewness, and Mean-Variance were added together. Importantly, since the rankings are assigned on a scale from 1 to 6—where 1 indicates the best performance and 6 the worst—the policy with the lowest total score is considered the most robust overall. We'll present these aggregated rankings in a parallel coordinate plot to visually depict how each policy performs across the different robustness metrics. This visualization allows us to clearly see the trade-offs between policies and identify which ones consistently rank well across multiple metrics, making it easier to pinpoint the most balanced and robust policies.

The parallel coordinate plot and the aggregated rankings provide a clear illustration of how different robustness metrics yield distinct "best" policies. Again, each robustness metric focuses on different

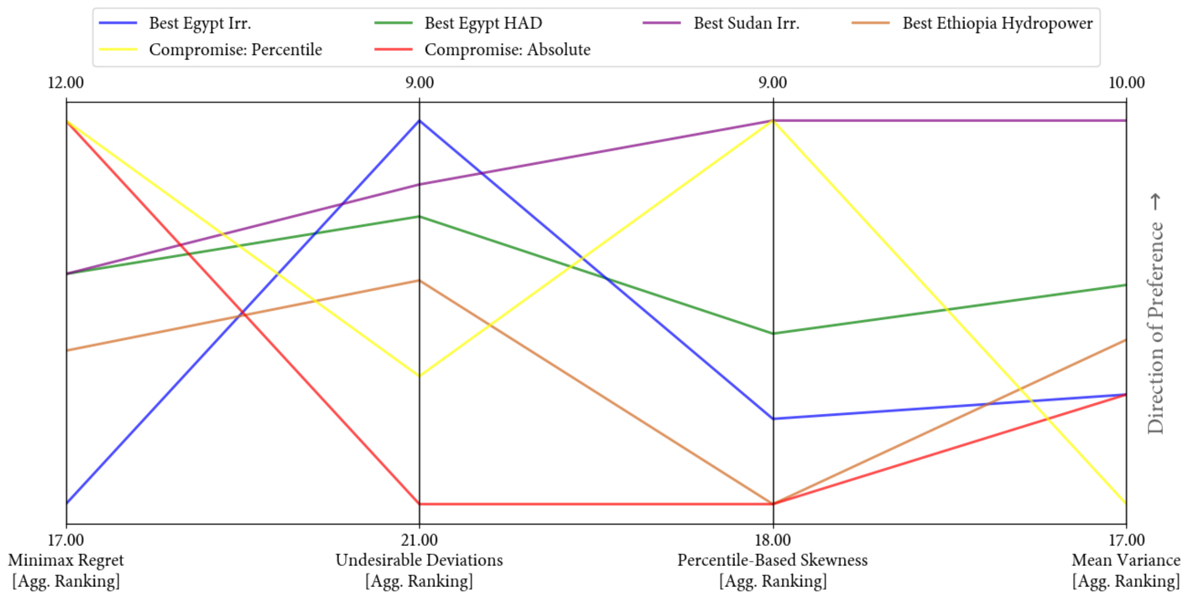

Figure 4.3: Parallel coordinates plot of Aggregated Robustness Ranking Comparison.

aspects of performance, leading to varying conclusions about which policy is the most robust overall. The best policies for each metrics are:

- Minimax Regret aims to minimize the worst-case scenario. "Compromise: Absolute" and "Compromise: Percentile" is identified as the most robust policies, using this metric.
- Undesirable Deviations penalizes large deviations from the median outcome, finds "Best Egypt Irr" to be the most robust policy.
- Percentile-Based Skewness rewards policies with favorable outcome distributions, surprisingly highlights "Best Sudan Irr" and "Compromise: Percentile" as the best options.
- Mean Variance metric seeks a balance between mean performance and variability, favors "Best Sudan Irr".

From our findings, there are several interesting insights. One of them is how "Compromise Absolute" and "Compromise: Percentile" policies perform well under the Minimax Regret metric despite not being tailored to excel in any single objective. This suggests that these compromise solutions effectively balance trade-offs, minimizing the worst-case scenario across multiple objectives. Minimax Regret metric values policies that avoid the worst possible outcomes, which explains why compromise solutions come out on top.

We also see that there's two policies identified to be the best for the Percentile Based Skewness, namely "Best Sudan Irr" and "Compromise: Percentile". This metric rewards policies that have a favorable outcome distribution—those that are skewed toward better outcomes—rather than those that simply perform well on average. This can lead to unexpected results, where policies that are extreme in their focus may perform poorly under this metric because their outcome distributions are less favorable or more variable.

Undesirable Deviations penalizes large deviations from a target, which is why "Best Egypt Irr" is seen as robust—it performs consistently close to the median outcome, avoiding large swings that could be penalized by this metric. Policies that have more variability, like those designed to excel in a specific objective, may be seen as less robust under this metric because they are more likely to have larger deviations from the median.

Mean Variance seeks a balance between average performance and variability. Policies like "Best Sudan Irr" perform well here because they maintain a good balance between achieving solid average results and keeping variability low. This metric is less forgiving of policies with high variance, which may explain why more extreme policies do not perform as well.

"Best Ethiopia Hydropower" stands out in some metrics as not performing as expected. This could be due to its extreme focus on a single objective, which leads to higher variability and less favorable outcome distributions in the eyes of metrics that value consistency or favorable skewness. On the other hand, "Best Egypt HAD" is relatively stable across most metrics but never emerges as the top policy in any metric.

We should also reflect on the method of aggregating rankings by summing them. This provides a straightforward way to compare policies across multiple metrics, but it also has limitations. This approach assumes that all metrics are equally important and that they can be directly compared, which might not always be the case. Furthermore, this method may obscure nuances in how policies perform under different metrics, particularly if a policy scores well in one metric but poorly in another.

Comparing Robustness Metrics and Risk-Aversion Level

It is evident that the robustness metrics embody varying levels of risk aversion, as discussed in Chapter 3.4.2, and this is clearly reflected in the results. Percentile-Based Skewness, positioned at the lower end of the risk aversion spectrum, calculates the skewness of outcomes by comparing extreme values (such as the 90^{th} and 10^{th} percentiles) with the median (50^{th} percentile). These metric rewards policies where the distribution of outcomes is skewed toward better results—those that exhibit favorable high-end outcomes, even if they also carry the possibility of less favorable outcomes. "Best Sudan Irr" and "Compromise: Percentile" were identified as the top-performing policies under this metric. These policies are more tolerant of risk, as they leverage the potential for high rewards, even at the cost of some risk. The focus on extremes allows for greater risk-taking, which might be more appropriate in scenarios where decision-makers are willing to accept variability for the chance of better-than-average results.

On the opposite end of the spectrum, Minimax Regret represents the highest level of risk aversion. This metric is designed to minimize the worst possible outcomes, specifically the maximum regret a decision-maker might experience if a chosen policy performs poorly compared to the best-performing policy in any given scenario. The Minimax Regret results favored "Compromise: Absolute" and "Compromise: Percentile" which suggests that these policies are effective at avoiding the worst-case scenarios across multiple objectives. This highly conservative approach prioritizes stability and predictability, often at the expense of higher potential gains.

The other metrics, such as Undesirable Deviations and Mean Variance, fall somewhere in between these two extremes in terms of risk aversion. Undesirable Deviations penalizes policies that deviate significantly from a target value, which in our analysis was set as the median outcome (q50). For instance, "Best Egypt Irr" was identified as the most robust under this metric, as it consistently performed close to the median, avoiding large deviations. Decision-makers have the flexibility to choose this target and adjusting it could yield different results. If a more aggressive target (like the 75^{th} percentile) were chosen, policies aiming for higher-than-average performance might be rewarded more, while those that focus on stability around the median might be penalized.

Mean Variance seeks to balance average performance with the variability of outcomes, offering a moderate level of risk aversion. It penalizes policies with high variability, even if their average performance is strong. In our analysis, "Best Sudan Irr" was favored under this metric because it maintained a good balance between achieving solid average results and keeping variability low. This approach is less forgiving of policies with high variance.

4.3. Physical System Implications

The previous robustness analysis identified the best policies for each metric. We now analyze the physical system implications of these top-performing policies. The top robust policies analysed include:

- Minimax Regret: "Compromise: Absolute" and "Compromise: Percentile"
- Undesirable Deviations: "Best Egypt Irr"
- Percentile-Based Skewness: "Best Sudan Irr" and "Compromise: Percentile"
- Mean Variance: "Best Sudan Irr"

We will assess how these policies affect key physical parameters such as reservoir water levels, irrigation deficits, and hydropower generation.

4.3.1. Compromise: Absolute Policy

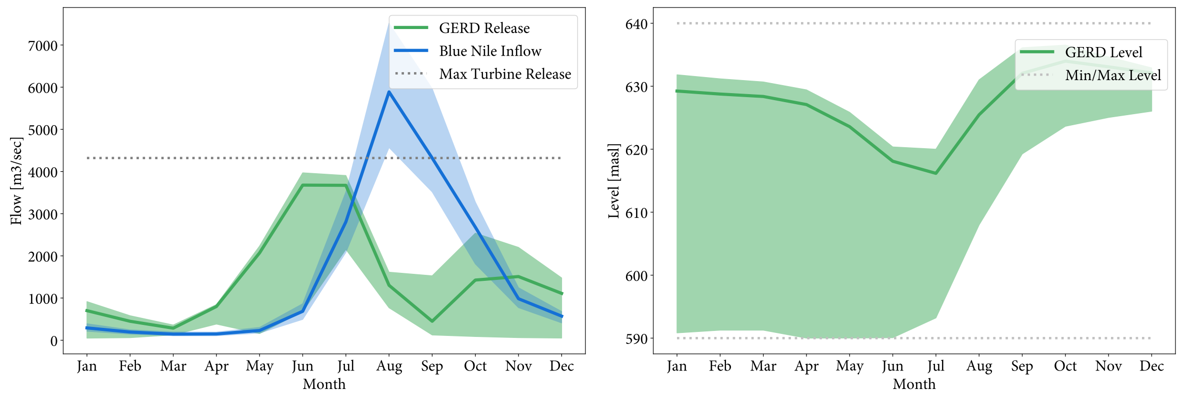

Figure 4.4: GERD Performance under Compromise: Absolute Policy.

Under the "Compromise: Absolute Threshold" policy, the GERD demonstrates effective management, as it is able to align releases with the inflow from the Blue Nile. The GERD releases are generally lower than the inflow during peak months, suggesting that the dam is effectively storing excess water that would be critical for maintaining operational flexibility. The GERD level graph supports this observation, showing expected seasonal fluctuations with water levels decreasing during dry periods and recovering during wet seasons. These fluctuations remain well within the dam's operational range.

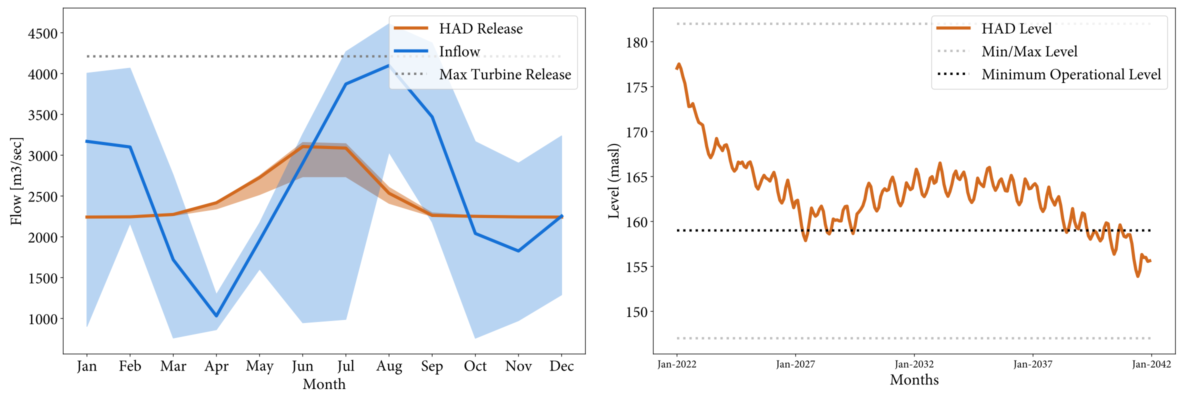

Figure 4.5: HAD Performance under Compromise: Absolute Policy.

While the HAD's releases under this policy track closely with the inflows, the water levels show a steady decline over time. This indicates that the dam is consistently releasing more water than it is receiving, which could potentially lead to long-term sustainability issues, especially during prolonged dry periods. The HAD levels consistently approach and even goes beyond the minimum operational level, which could be a red flag for future water security.

For the Gezira irrigation system, the policy shows that irrigation demands are mostly met during peak flow periods. However, there is significant variability in the received flow, particularly during non-peak months, which could indicate difficulties in maintaining consistent irrigation supply. Looking at Egypt's overall performance under this policy, the demanded flow is generally met during key periods, but there are noticeable discrepancies during the dry season when the received flow falls short of the demand.

4.3.2. Compromise: Percentile Policy

The GERD under the Compromise: Percentile Threshold policy demonstrates a more conservative approach compared to the Absolute Threshold. The releases from GERD are closely aligned with the

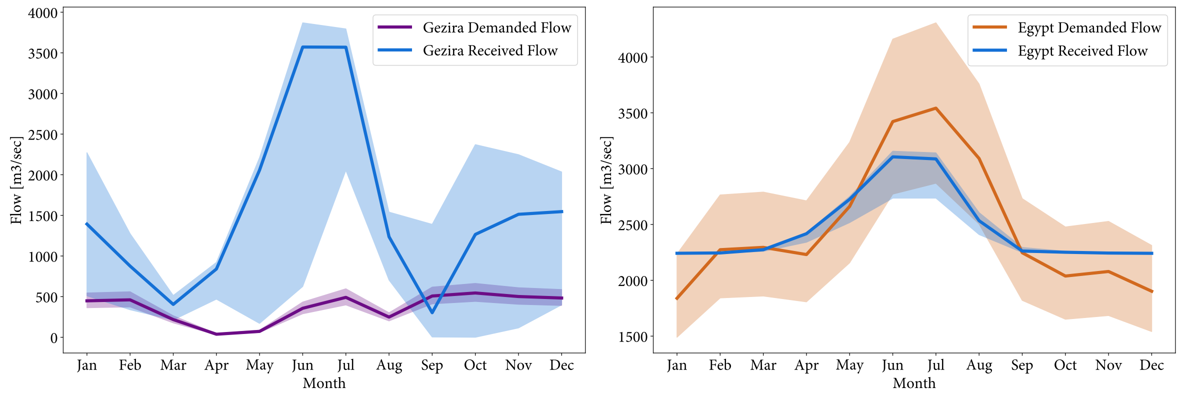

Figure 4.6: Gezira and Egypt Water Levels Performance under Compromise: Absolute Policy.

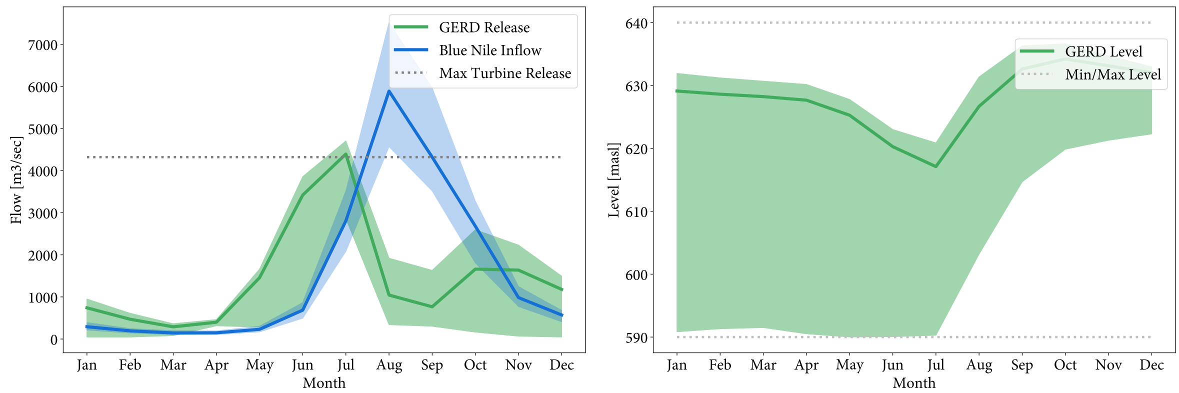

Figure 4.7: GERD Performance under Compromise: Percentile Policy.

Blue Nile inflow, particularly during peak flow months. The GERD release remains lower than the inflow during most months, indicating successful water storage, particularly during the high inflow months of July and August. However, the releases still ensure downstream demands are met without over-releasing water. The GERD level shows significant fluctuations, but it maintains a stable storage level over time.

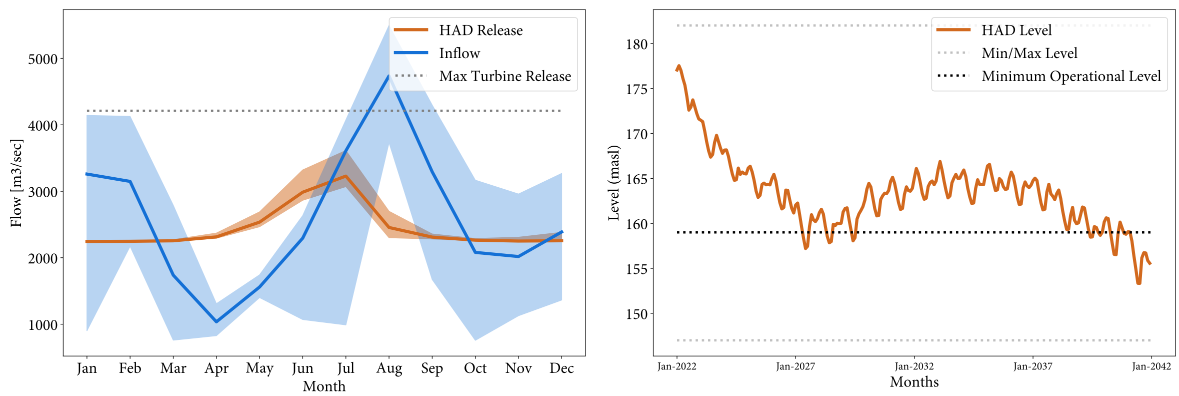

Figure 4.8: HAD Performance under Compromise: Percentile Policy.

The HAD releases are generally aligned with the inflow, but with noticeable fluctuations. These fluctuations are more pronounced during off-peak months, suggesting that the dam may struggle with maintaining a steady release when inflows are low. This is further supported by the gradual decline in the HAD level over time, which is consistent with a pattern of releasing more water than is received.

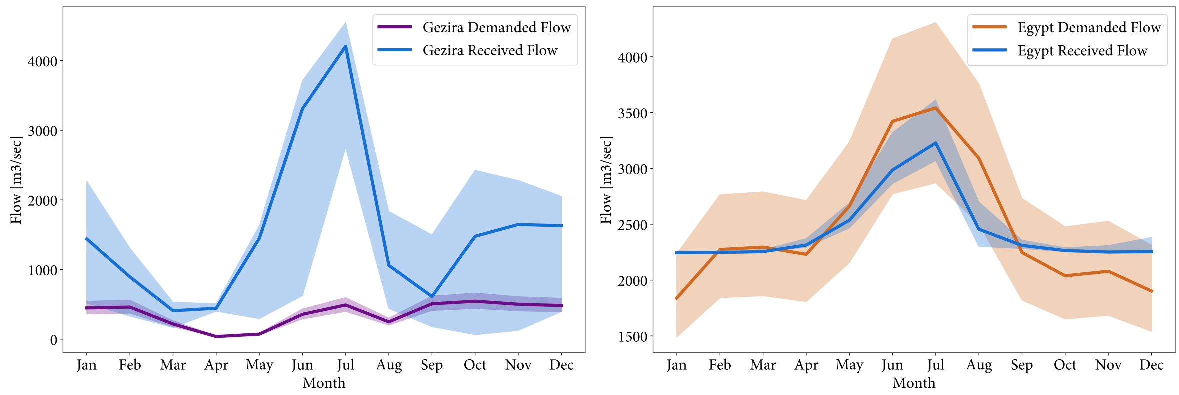

Figure 4.9: Gezira and Egypt Water Level Performance under Best Egypt Irrigation Policy.

The Gezira system's performance under this policy is marked by significant variability in the received flow compared to the demanded flow. During peak months, the demands are largely met, but during off-peak periods, there is a substantial shortfall. This inconsistency could lead to challenges in meeting irrigation demands consistently throughout the year. For Egypt overall, the demanded flow is largely met during critical months under the Compromise: Percentile Threshold policy. However, similar to the Gezira system, there is noticeable variability and occasional shortfalls during off-peak months.

4.3.3. Best Egypt Irrigation

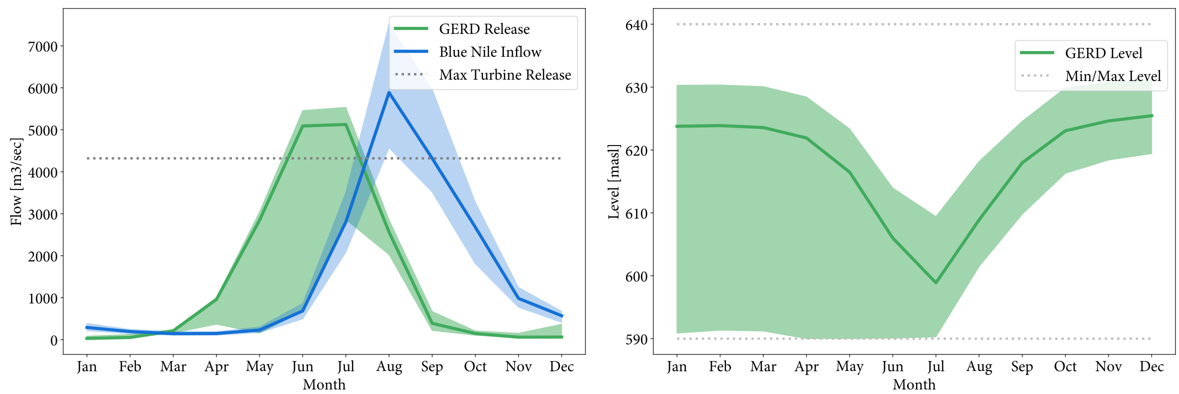

Figure 4.10: GERD Performance under Best Sudan Irrigation Policy.

Under the Best Egypt Irrigation, the GERD appears to perform effectively, particularly in managing its releases in alignment with the Blue Nile inflow. The GERD releases are generally lower than the inflow, indicating that the dam is successfully storing water during peak inflow months. The GERD level graph further supports this, showing significant seasonal fluctuations where water levels decrease during the dry season and rise during the wet season. Importantly, these fluctuations are well within the operational range, suggesting that the dam is managing its storage efficiently, maintaining sufficient reserves while also providing necessary releases downstream.

However, the High Aswan Dam (HAD) displays a more concerning trend. Although the HAD's releases track closely with inflows, particularly during peak months, the overall level of the dam shows a gradual decline over time. This decline suggests that the dam is consistently releasing more water than it is receiving, potentially leading to long-term issues in maintaining sufficient water levels. This trend, if continued, could jeopardize Egypt's water security, particularly during dry years when inflows are already reduced.

The Gezira irrigation system's performance is relatively stable, with irrigation demands largely met during peak flow periods. However, there is noticeable variability in the received flow, indicating

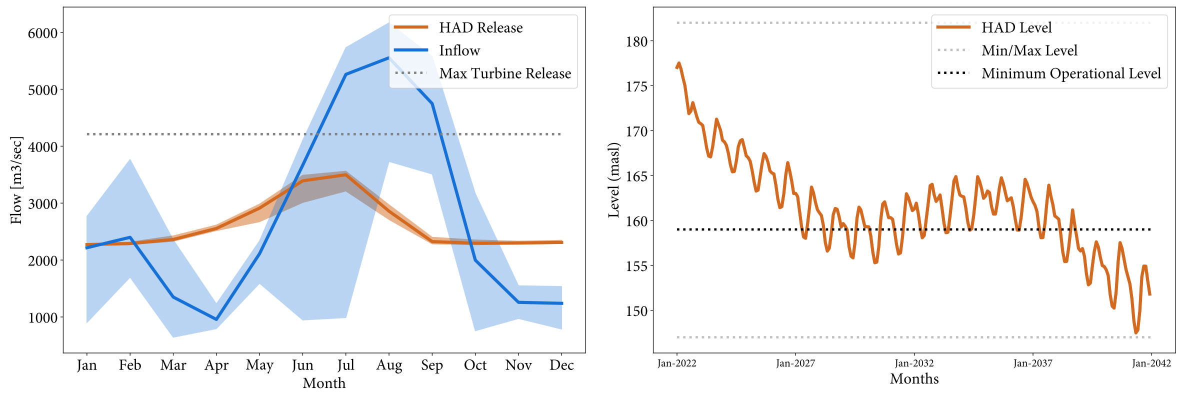

Figure 4.11: HAD Performance under Best Egypt Irrigation Policy.

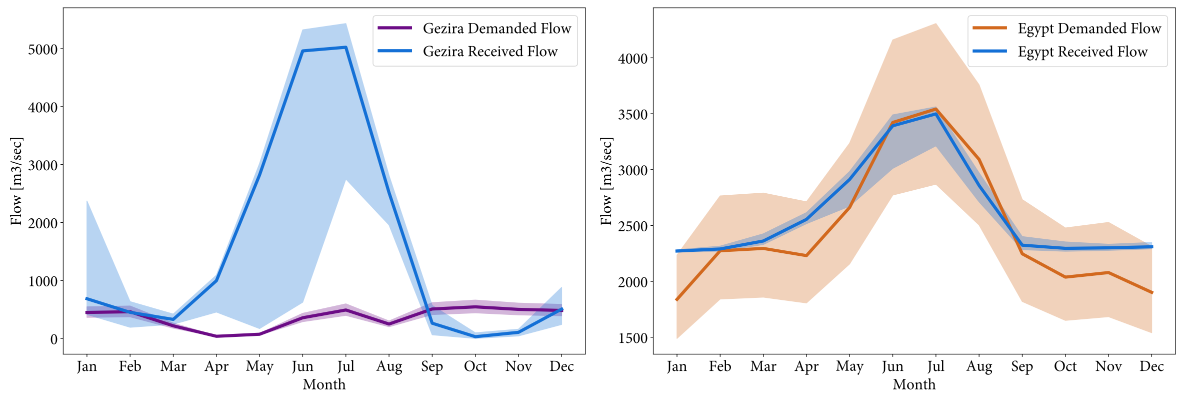

Figure 4.12: Gezira and Egypt Water Level Performance under Best Egypt Irrigation Policy.

challenges in consistently meeting irrigation demands, especially during off-peak periods. Similarly, Egypt's overall water demand is mostly met during crucial months, but there are instances, particularly during the dry season, where the received flow falls short of the demand. This shortfall could negatively impact agricultural productivity and water availability downstream.

4.3.4. Best Sudan Irrigation Policy

The Best Sudan Irrigation Policy demonstrates effective management of the GERD, particularly in aligning water releases with Sudan's irrigation needs, as expected.

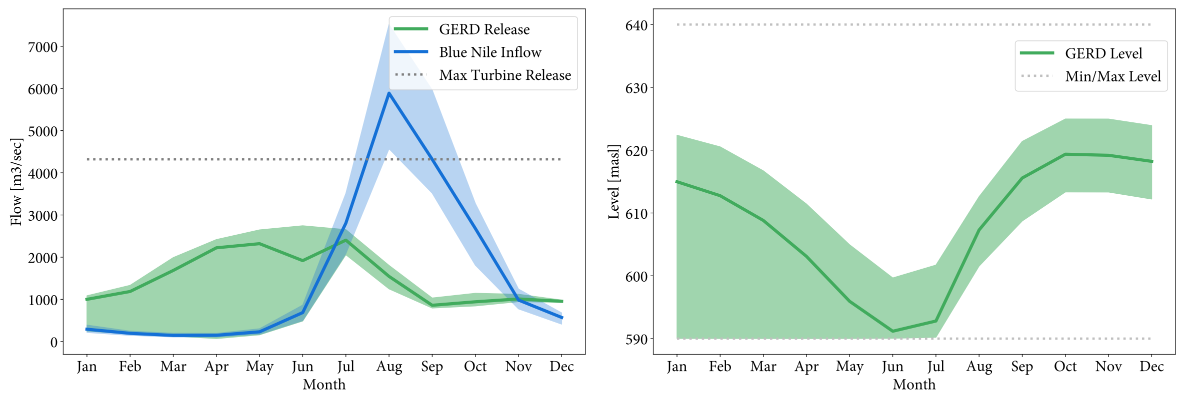

Figure 4.13: GERD Performance under Best Sudan Irrigation Policy.

The GERD's release is carefully controlled and remains below the inflow levels during peak periods, suggesting that the reservoir is prioritizing water retention during times of high inflow. This strategy ensures that the GERD operates well within its turbine capacity. The seasonal fluctuations in the GERD's water level are as expected, with increases during the wet season and decreases during the dry season. Importantly, the levels remain within safe operational limits.

However, while the GERD performance is satisfactory, the policy raises concerns with the HAD operation.

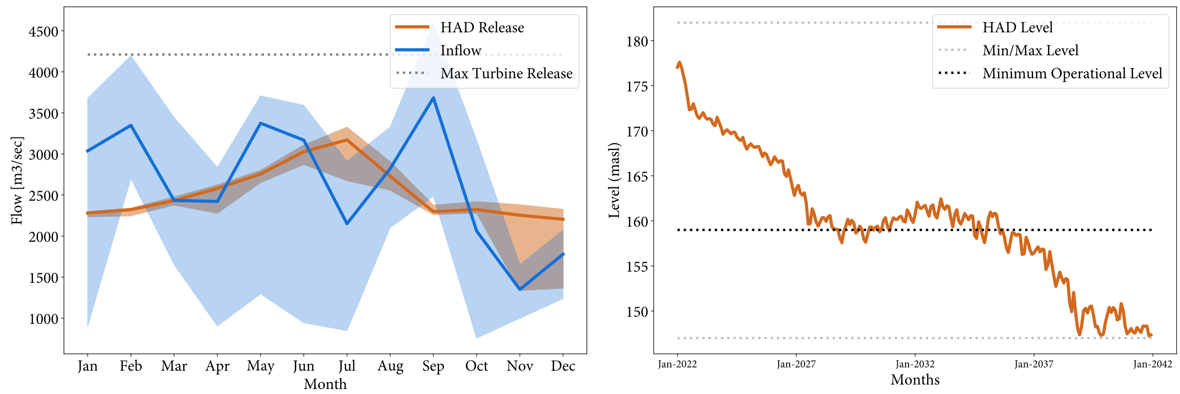

Figure 4.14: HAD Performance under Best Sudan Irrigation Policy.

The HAD's water levels show a steady and concerning decline over time, which suggests that more water is being released than is being stored. This trend poses a serious risk to Egypt's long-term water security. The discrepancy between inflow and release during peak periods at the HAD further suggests potential inefficiencies in water resource management. The HAD's water levels show a steady and concerning decline over time, which suggests that more water is being released than is being stored. This trend poses a serious risk to Egypt's long-term water security. The discrepancy between inflow and release during peak periods at the HAD further suggests potential inefficiencies in water resource management.

For Gezira, while irrigation demands are met especially during peak flow periods, the significant variability in the received flow suggests inconsistencies that could impact agricultural productivity, particularly during off-peak times. Likewise, with Egypt, while the water demands are mostly met, there are periods where the received flow falls short especially during dry season that could have effects on water availability downstream.

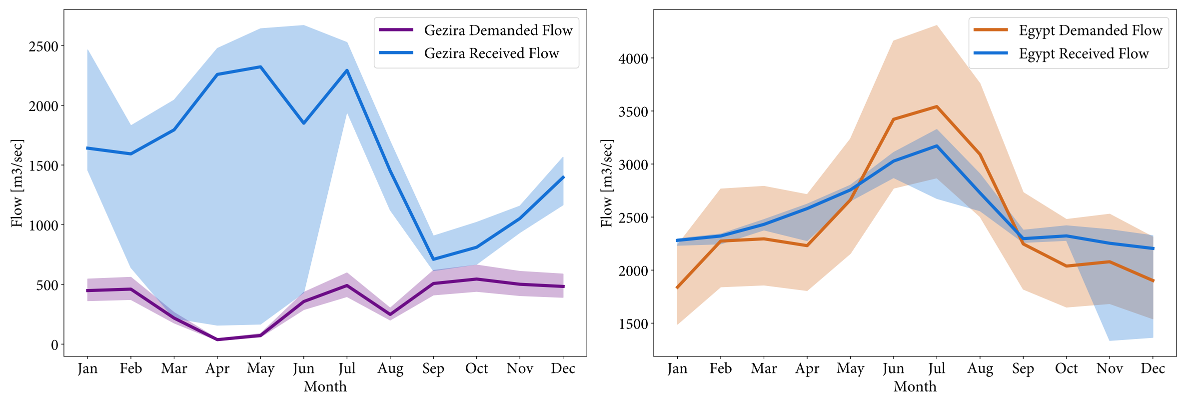

Figure 4.15: Gezira and Egypt Water Level Performance under Best Sudan Irrigation Policy.

5 Discussion

5.1. Robustness Metrics

In this discussion, we reflect on the robustness analysis conducted using four distinct metrics—Minimax Regret, Percentile-Based Skewness, Mean-Variance, and Undesirable Deviations—each embodying varying levels of risk aversion and providing different lenses through which to evaluate policy performance. Our findings confirm that the choice of robustness metric has on identifying the most robust policies, ultimately leading to different "best" solutions depending on the metric employed.

The robustness metrics that we used offers each differing definition of what it means for a policy to be "robust", directly reflected in the way scores are calculated.

1. Minimax Regret focuses on minimizing the worst-case scenario by comparing the maximum regret a policy might incur against the best possible outcome across scenarios. This metric is inherently conservative, representing a high level of risk aversion. It aims to identify policies that avoid the most unfavorable outcomes.
2. Percentile-Based Skewness examines the distribution of outcomes relative to the median, favoring policies that skew towards better results. This metric operates at the lower end of the risk aversion spectrum, as it rewards policies that achieve high-end outcomes, even if they involve some risk.
3. Mean-Variance seeks to balance the average performance of a policy with the variability of that performance across scenarios. It represents a moderate level of risk aversion, penalizing policies that, despite having strong average performance, shows high variability.
4. Undesirable Deviations penalizes policies that deviate significantly from a target value, typically the median outcome. This metric is less extreme in its risk aversion compared to Minimax Regret, yet it still favors policies that perform consistently close to a desired target, minimizing large difference to its target.

The analysis indeed reveals that the choice of robustness metric significantly influences which policies are considered the most robust. This means that it is important for decision-makers to carefully consider their priorities and risk tolerance when selecting a metric and should take a multi-metric approach into selecting which policy is ultimately the most robust.

These findings confirm the presence of inter-utility of robustness, as previously discussed in Chapter 3.4.1 that no single policy will emerge as the most robust across all scenarios, thus it is inherently multi-dimensional and depends on the specific context and the decision-maker's preferences.

It should also be noted that some robustness metrics allows decision makers to choose their target, like Undesirable Deviations, that we chose. The selection of the target value, penalizing policies that deviates significantly from this, most definitely affects the outcome of the robustness score calculation. If a more aggressive target (like the 75^{th} percentile) were chosen, policies aiming for higher-than-average performance might be rewarded more, while those that focus on stability around the median might be penalized. This also applies in other metric that lets decision makers set their own target such as Hurwicz Optimism-Pessimism Rule.

Clearly, for decision making in uncertainty, the choice of the robustness metrics and the eventual policies that is chosen will have tangible implication on the physical system. For instance, the Compromise: Absolute policy, identified as robust by the Minimax Regret metric, effectively manages GERD's water storage but leads to a concerning decline in HAD water levels, potentially threatening Egypt's long-term water security. In contrast, the Best Sudan Irrigation policy, favored by both Percentile-Based Skewness and Mean Variance metrics, shows effective water management for Sudan but similarly results in unsustainable water release patterns from the HAD, raising similar concerns for Egypt.

Ultimately, this analysis reinforces the concept that robustness is inherently multi-dimensional and context dependent. The choice of metric should be aligned with the specific goals and risk preferences of the decision-making body. Additionally, it emphasizes the need for a multi-metric approach in multi-objective optimization to capture the full spectrum of potential outcomes and ensure a well-rounded decision-making process. The findings provide valuable insights for the scientific community, suggesting that the use of a variety of robustness metrics can offer a more comprehensive understanding of policy impacts, particularly in complex systems like water resource management.

5.2. Limitations

5.2.1. Conceptual Limitations

The conceptual framework of this study presents several important limitations that could affect the robustness and relevance of the findings. One major limitation is the assumption of equal weights for all objectives in the analysis. This assumption does not necessarily reflect the diverse priorities, preferences, and needs of different stakeholders within the Nile River Basin, which could vary significantly between and within countries. By assuming equal importance across objectives, the analysis might overlook the fact that some stakeholders might prioritize certain outcomes over others, such as agricultural water use versus hydropower generation. This could lead to policy recommendations that are not fully aligned with the actual needs and priorities of those affected by water management decisions in the region.

The study's scope is also a notable limitation, as it focuses only on a portion of the Nile River Basin, without fully encompassing the entire region that spans 11 countries. This narrower focus means that many important stakeholder activities and interests are not incorporated into the analysis. For instance, water diversion practices and the impact of upstream activities in other parts of the basin were only superficially modeled, potentially missing critical interactions that could influence the system's overall dynamics. A more comprehensive scope that includes these broader interactions and activities would likely enhance the accuracy and relevance of the system modeling.

Moreover, the study aggregates objectives within countries from a utilitarian perspective, which can mask the diverse interests and needs within each nation. This approach may fail to adequately represent specific groups, such as local communities and ecosystems that depend on water regimes. For example, the regulated release regime of the Grand Ethiopian Renaissance Dam (GERD) could adversely impact Sudan's flood-recession agriculture, a critical livelihood for many in the region. The study's aggregation approach does not fully account for these localized impacts, potentially leading to recommendations that might be beneficial on a national scale but harmful to certain communities or ecosystems.

Another conceptual limitation is the lack of consideration for political stability and potential conflicts within the Nile Basin countries. Political instability and conflicts can severely disrupt water management agreements and cooperation efforts, affecting the successful implementation of proposed policies. The analysis assumes a stable political environment, which may not be realistic in a region characterized by complex geopolitical tensions and frequent disputes over water rights. By not incorporating these factors, the study may overestimate the feasibility and sustainability of the proposed water management strategies.

5.2.2. Methodological Limitations

The study presents several methodological limitations that could potentially influence the robustness and accuracy of the findings. One of the primary limitations is the reliance on specific robustness metrics and modeling approaches, which may introduce biases and restrict the full exploration of variability and uncertainty within the system. The choice of robustness metrics, such as Minimax Regret, Percentile-Based Skewness, Mean-Variance, and Undesirable Deviations, inherently shapes the outcomes and may not capture all dimensions of robustness. Each metric defines robustness differently, which can skew the results towards certain policy preferences that may not fully align with the overall system goals.

A key methodological constraint arises from the use of Latin Hypercube Sampling (LHS) to select the best robust solutions from extreme best solutions and their compromise solutions. While LHS is a useful technique for reducing computational load by efficiently sampling the input space, it also limits the exploration of the solution space. By focusing on specific policies (extreme policies and compromise policies), potentially more robust solutions may be overlooked, particularly those that fall outside the preselected sampling range.

For calculating the score using Undesirable Deviations, policies that score closer to the q50 median are generally considered more desirable because they indicate consistent performance across scenarios. However, this approach may inadvertently penalize policies that significantly outperform the median, labeling them as having "undesirable deviations." This is particularly problematic for objectives where exceeding the target is inherently beneficial, such as in maximizing hydropower output. A policy that significantly surpasses the median might be more advantageous, yet the current metric framework could misclassify it as less robust simply due to its divergence from the median.

The method of aggregating scores also presents a limitation. While the aggregation provides a means to synthesize multiple metrics into a single ranking, it assumes that all metrics are equally important and directly comparable. This assumption may not hold true in all contexts, leading to potential inaccuracies in the overall assessment of policy robustness. Moreover, this method could obscure the nuances in policy performance under different metrics, particularly if a policy performs exceptionally well in one metric but poorly in another. The resulting aggregate score may not fully reflect the true robustness of a policy, thereby affecting the final recommendations.

6 Conclusion

6.1. Addressing Research Questions

Main RQ: What is the consequence of applying multiple robustness metrics within many-objective optimization models to address water allocation issues under deep uncertainty in the Nile River Basin?

The consequence of applying multiple robustness metrics within many-objective optimization models in the Nile River Basin is that it reveals the multidimensionality and complexity inherent in robust decision-making under deep uncertainty. Different robustness metrics, each with its own definition and emphasis, lead to the identification of different "best" policies, highlighting the fact that no single policy can be considered the most robust across all scenarios. This emphasizes the importance of a multi metric approach in policy evaluation, as it allows for a more comprehensive understanding of the trade-offs and potential outcomes associated with different policies. By considering various robustness metrics, decision-makers can better aling their choices with specific goals, risk tolerances, and stakeholder preferences, thereby improving the overall resilience and effectiveness of water management strategies in the basin.

SQ1: What are the trade-offs based on the Pareto-optimal policy alternatives of the optimal reservoir control in the Nile River Basin?

Prominent trade-offs exist between Ethiopia's hydropower maximization and the deficit minimization objectives of Egypt and Sudan. The Pareto-optimal policy alternatives for optimal reservoir control in the Nile River Basin illustrate these conflicts. Policies that prioritize Ethiopia's hydropower generation tend to result in undesirable outcomes for Egypt's irrigation and low flow objectives, compromising Egypt's water security. Conversely, policies that optimize Egypt's objectives lead to significantly reduced hydropower outcomes for Ethiopia and poor irrigation outcomes for Sudan. These trade-offs show the inherent conflict in balancing the diverse water needs and objectives of the Nile Basin countries, where optimizing one country's benefit will result in significant detriments to others.

SQ2: How do different robustness metrics influence the selection of optimal policy alternatives in the Nile River Basin?

Different robustness metrics significantly influence the selection of optimal policy alternatives by emphasizing different aspects of policy performance and risk tolerance.

- Minimax Regret prioritizes policies that minimize the worst-case outcome, leading to the selection of compromise policies like "Compromise: Absolute" and "Compromise: Percentile" which balance various objectives without excelling in any single one.
- Percentile-Based Skewness rewards policies with outcomes skewed towards better-than-median results, favoring policies like "Best Sudan Irrigation" which might not have been selected under other metrics.

- Mean-Variance focuses on policies that achieve a good balance between mean performance and consistency, potentially favoring policies that might be more stable but less extreme.
- Undesirable Deviations penalizes policies that deviate significantly from a target, favoring policies like “Best Egypt Irrigation” that perform consistently close to the median.

These differences in metric application shows that the choice of robustness metric can lead to varying conclusions about which policy is the most robust, thus influencing the final decision-making process.

SQ3: What are the implications of using different robustness metrics for stakeholder decision-making in the Nile River Basin?

The implications of using different robustness metrics for stakeholder decision-making in the Nile River Basin include the need for stakeholders to clearly define their priorities and risk tolerance levels before selecting a robustness metric. The choice of metric will directly impact which policies are deemed most robust and, consequently, which policies are recommended for implementation. For instance, a metric like Minimax Regret may be more suitable for stakeholders who are highly risk-averse and concerned with avoiding the worst possible outcomes, while Percentile-Based Skewness might appeal to those willing to accept more risk for potentially higher rewards.

6.2. Recommendations

To enhance water management in the Nile River Basin, we recommend that policies should be evaluated using multiple robustness metrics. This multi-metric approach ensures comprehensive understanding and resilient decision-making across different scenarios and uncertainties. Strengthening cooperative management among Nile Basin countries is crucial, potentially through a central authority or improved joint decision-making frameworks, to balance diverse stakeholder interests and mitigate conflicts. Engaging stakeholders, including local communities and governments, ensures that various perspectives are considered, leading to more equitable policies. Additionally, implementing adaptive management practices allows for flexibility and adjustments based on real-time data and changing conditions, effectively managing uncertainties and improving long-term sustainability.

6.3. Future Research

Future research should expand the geographic scope to include the entire Nile Basin, incorporating the White Nile and Atbara rivers for a more comprehensive understanding of the basin’s hydrology and stakeholder dynamics. This broader scope would allow for a more accurate representation of the complex interactions and water management challenges across the region. Additionally, exploring the long-term impacts of climate change on the Nile River Basin will be crucial for developing adaptive water management policies that can respond to changing environmental conditions.

Further exploration of robustness metrics and their implications for policy evaluation is also necessary. This includes the development of new metrics that address specific regional challenges, such as those related to economic performance indicators like Net Present Value (NPV), cost, and economic efficiency. Incorporating these metrics into the analysis can provide a more comprehensive evaluation of policies by balancing economic considerations with other objectives. This approach would enable decision-makers to better understand the trade-offs between cost-effectiveness and other critical factors, leading to more informed and balanced policy choices.

Moreover, future research should aim to create a unifying framework that systematically compares the results of various robustness metrics. Such a framework would help identify commonalities and differences among metrics, particularly those with similar underlying methodologies. For instance, as demonstrated in studies like McPhail et al. (2018), metrics that assess performance stability or minimize regret could often converge in their policy recommendations. By establishing a unified approach, researchers could streamline the decision-making process, offering a more cohesive understanding of which policies are most robust across different scenarios and metrics.

Finally, while this study focused on six extreme policies, future research should test a broader range of possible policies to identify more robust solutions. Addressing these recommendations and research directions will enhance the robustness and effectiveness of water management policies in the Nile

River Basin, ensuring sustainable and equitable use of this vital resource. By integrating a wider array of metrics and adopting a comprehensive, basin-wide approach, future research can provide more nuanced and effective strategies for managing the Nile's water resources in the face of ongoing and future challenges.

Bibliography

Sarah Alexander, Guang Yang, Girmachew Addisu, and Paul Block. Forecast-informed reservoir operations to guide hydropower and agriculture allocations in the blue nile basin, ethiopia. *International Journal of Water Resources Development*, 37(2):208–233, 2021.

JR Allan, N Levin, KR Jones, S Abdullah, J Hongoh, V Hermoso, and S Kark. Navigating the complexities of coordinated conservation along the river nile. *Science advances*, 5(4):eaau7668, 2019.

Hana Attia and Mona Saleh. The political deadlock on the grand ethiopian renaissance dam. 2021.

Mohamed Abdel Aty. Managing risks on egypt water resources security: Climate change and grand ethiopian renaissance dam (gerd) as challenging aspects. *Water Security Under Climate Change*, pages 313–330, 2022.

Ellen Banzhaf, Henrique Ll Roig, and Fabio Bakker. Characterising the phenomenon of water scarcity in the fast growing federal district of brasilia. In *2009 Joint Urban Remote Sensing Event*, pages 1–7. IEEE, 2009.

Eva HY Beh, Feifei Zheng, Graeme C Dandy, Holger R Maier, and Zoran Kapelan. Robust optimization of water infrastructure planning under deep uncertainty using metamodels. *Environmental Modelling & Software*, 93:92–105, 2017.

Yakov Ben-Haim. *Info-gap decision theory: decisions under severe uncertainty*. Elsevier, 2006.

Nathan Bonham, Joseph Kasprzyk, and Edith Zagona. post-mordm: Mapping policies to synthesize optimization and robustness results for decision-maker compromise. *Environmental Modelling & Software*, 157:105491, 2022.

Casey Brown, Frederick Boltz, Sarah Freeman, Jacqueline Tront, and Diego Rodriguez. Resilience by design: A deep uncertainty approach for water systems in a changing world. *Water security*, 9:100051, 2020.

Judith Bunbury, John P Cooper, Richard Hoath, Salima Ikram, Christine Johnston, and Thomas Schneider. The egyptian nile: Human transformation of an ancient river. *River culture—Life as a dance to the rhythm of the waters*, pages 43–47, 2023.

William S Butcher. Stochastic dynamic programming for optimum reservoir operation 1. *JAWRA Journal of the American Water Resources Association*, 7(1):115–123, 1971.

Jared L Cohon and David H Marks. Multiobjective screening models and water resource investment. *Water Resources Research*, 9(4):826–836, 1973.

Jared L Cohon and David H Marks. A review and evaluation of multiobjective programing techniques. *Water resources research*, 11(2):208–220, 1975.

Ruth Dittrich, Anita Wreford, and Dominic Moran. A survey of decision-making approaches for climate change adaptation: Are robust methods the way forward? *Ecological Economics*, 122:79–89, 2016.

Petra Döll and Patricia Romero-Lankao. How to embrace uncertainty in participatory climate change risk management—a roadmap. *Earth's Future*, 5(1):18–36, 2017.

Hamdy Elsayed, Slobodan Djordjević, Dragan A Savić, Ioannis Tsoukalas, and Christos Makropoulos. The nile water-food-energy nexus under uncertainty: Impacts of the grand ethiopian renaissance dam. *Journal of Water Resources Planning and Management*, 146(11):04020085, 2020.

Augustine O Esogbue and Janusz Kacprzyk. Fuzzy dynamic programming. In *Fuzzy sets in decision analysis, operations research and statistics*, pages 281–307. Springer, 1998.

Hyung-Il Eum and Young-Oh Kim. The value of updating ensemble streamflow prediction in reservoir operations. *Hydrological processes*, 24(20):2888–2899, 2010.

Alexei A Gaivoronski, Giovanni M Sechi, and Paola Zuddas. Balancing cost-risk in management optimization of water resource systems under uncertainty. *Physics and Chemistry of the Earth, Parts A/B/C*, 42:98–107, 2012.

Robel T Geressu and Julien J Harou. Screening reservoir systems by considering the efficient trade-offs—informing infrastructure investment decisions on the blue nile. *Environmental Research Letters*, 10 (12):125008, 2015.

Matteo Giuliani and Andrea Castelletti. Is robustness really robust? how different definitions of robustness impact decision-making under climate change. *Climatic Change*, 135:409–424, 2016.

Matteo Giuliani, Jonathan D Herman, Andrea Castelletti, and P Reed. Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management. *Water resources research*, 50(4):3355–3377, 2014.

Matteo Giuliani, Francesca Pianosi, and Andrea Castelletti. Making the most of data: An information selection and assessment framework to improve water systems operations. *Water Resources Research*, 51(11):9073–9093, 2015.

Matteo Giuliani, Andrea Castelletti, Francesca Pianosi, Emanuele Mason, and Patrick M Reed. Curses, tradeoffs, and scalable management: Advancing evolutionary multiobjective direct policy search to improve water reservoir operations. *Journal of Water Resources Planning and Management*, 142(2): 04015050, 2016.

Matteo Giuliani, Marta Zaniolo, Paul Block, and Andrea Castelletti. Data-driven control of water reservoirs using an emulator of the climate system. *IFAC-PapersOnLine*, 53(2):16531–16536, 2020.

Marjolijn Haasnoot, Hans Middelkoop, Ermond Van Beek, and WPA Van Deursen. A method to develop sustainable water management strategies for an uncertain future. *Sustainable Development*, 19(6): 369–381, 2011.

Antonia Hadjimichael, Julianne Quinn, Erin Wilson, Patrick Reed, Leon Basdekas, David Yates, and Michelle Garrison. Defining robustness, vulnerabilities, and consequential scenarios for diverse stakeholder interests in institutionally complex river basins. *Earth's Future*, 8(7):e2020EF001503, 2020.

Stéphane Hallegatte, Ankur Shah, Casey Brown, Robert Lempert, and Stuart Gill. Investment decision making under deep uncertainty–application to climate change. *World Bank Policy Research Working Paper*, (6193), 2012.

Caner Hamarat, Jan H Kwakkel, Erik Pruyt, and Erwin T Loonen. An exploratory approach for adaptive policymaking by using multi-objective robust optimization. *Simulation Modelling Practice and Theory*, 46:25–39, 2014.

Tsuyoshi Hashimoto, Jery R Stedinger, and Daniel P Loucks. Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. *Water resources research*, 18(1):14–20, 1982.

Jon C Helton and Freddie Joe Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1):23–69, 2003.

Jonathan D Herman, Harrison B Zeff, Patrick M Reed, and Gregory W Characklis. Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty. *Water Resources Research*, 50(10):7692–7713, 2014.

Jonathan D Herman, Patrick M Reed, Harrison B Zeff, and Gregory W Characklis. How should robustness be defined for water systems planning under change? *Journal of Water Resources Planning and Management*, 141(10):04015012, 2015.

Ronald A Howard. Information value theory. *IEEE Transactions on systems science and cybernetics*, 2(1): 22–26, 1966.

Ronald A Howard and James E Matheson. Influence diagrams. *Decision Analysis*, 2(3):127–143, 2005.

Xin Huang, Bin Xu, Ping-an Zhong, Hongyi Yao, Hao Yue, Feilin Zhu, Qingwen Lu, Yu Sun, Ran Mo, Zhen Li, et al. Robust multiobjective reservoir operation and risk decision-making model for real-time flood control coping with forecast uncertainty. *Journal of Hydrology*, 605:127334, 2022.

Leonid Hurwicz. What has happened to the theory of games. *The American Economic Review*, 43(2): 398–405, 1953.

Marc Jeuland and Dale Whittington. Water resources planning under climate change: Assessing the robustness of real options for the blue nile. *Water Resources Research*, 50(3):2086–2107, 2014.

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

Daniel Kendie. Egypt and the hydro-politics of the blue nile river. *Northeast African Studies*, 6(1/2): 141–169, 1999.

Gi Joo Kim, Young-Oh Kim, and Patrick M Reed. Improving the robustness of reservoir operations with stochastic dynamic programming. *Journal of Water Resources Planning and Management*, 147(7): 04021030, 2021.

Young-Oh Kim and Richard N Palmer. Value of seasonal flow forecasts in bayesian stochastic programming. *Journal of Water Resources Planning and Management*, 123(6):327–335, 1997.

Young-Oh Kim, Hyung-Il Eum, Eun-Goo Lee, and Ick Hwan Ko. Optimizing operational policies of a korean multireservoir system using sampling stochastic dynamic programming with ensemble streamflow prediction. *Journal of Water Resources Planning and Management*, 133(1):4–14, 2007.

Joshua B Kollat and Patrick M Reed. Comparison of multi-objective evolutionary algorithms for long-term monitoring design. In *Impacts of Global Climate Change*, pages 1–11. 2005.

Demetris Koutsoyiannis and Athanasia Economou. Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems. *Water Resources Research*, 39(6), 2003.

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

Jan H Kwakkel, Warren E Walker, and Vincent AWJ Marchau. Classifying and communicating uncertainties in model-based policy analysis. *International journal of technology, policy and management*, 10(4):299–315, 2010.

Jan H Kwakkel, Marjolijn Haasnoot, and Warren E Walker. Comparing robust decision-making and dynamic adaptive policy pathways for model-based decision support under deep uncertainty. *Environmental Modelling & Software*, 86:168–183, 2016.

Robert J Lempert. Embedding (some) benefit-cost concepts into decision support processes with deep uncertainty. *Journal of Benefit-Cost Analysis*, 5(3):487–514, 2014.

Robert J Lempert and Myles T Collins. Managing the risk of uncertain threshold responses: comparison of robust, optimum, and precautionary approaches. *Risk Analysis: An International Journal*, 27(4): 1009–1026, 2007.

Robert J Lempert, Steven W Popper, and Steven C Bankes. Shaping the next one hundred years: new methods for quantitative. *Long-Term Policy Analysis*, 208, 2003.

C Lorenz, A Gilbert, and P Vellinga. Sustainable management of transboundary river basins: a line of reasoning. *Regional Environmental Change*, 2:38–53, 2001.

Hector Macian-Sorribes and Manuel Pulido-Velazquez. Inferring efficient operating rules in multireservoir water resource systems: A review. *Wiley Interdisciplinary Reviews: Water*, 7(1):e1400, 2020.

Holger R Maier, Zoran Kapelan, Joseph Kasprzyk, Joshua Kollat, L Shawn Matott, Maria C Cunha, Graeme C Dandy, Matthew S Gibbs, Edward Keedwell, Angela Marchi, et al. Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. *Environmental Modelling & Software*, 62:271–299, 2014.

Evgenii S Matrosov, Ashley M Woods, and Julien J Harou. Robust decision making and info-gap decision theory for water resource system planning. *Journal of hydrology*, 494:43–58, 2013.

Matthew P McCartney and Michael Menker Girma. Evaluating the downstream implications of planned water resource development in the ethiopian portion of the blue nile river. *Water International*, 37(4): 362–379, 2012.

Cameron McPhail, Holger R Maier, Jan H Kwakkel, Matteo Giuliani, Andrea Castelletti, and Seth Westra. Robustness metrics: How are they calculated, when should they be used and why do they give different results? *Earth's Future*, 6(2):169–191, 2018.

Paul CD Milly, Julio Betancourt, Malin Falkenmark, Robert M Hirsch, Zbigniew W Kundzewicz, Dennis P Lettenmaier, and Ronald J Stouffer. Stationarity is dead: Whither water management? *Science*, 319 (5863):573–574, 2008.

Sherif Mohyeldeen. The dam that broke open an ethiopia-egypt dispute. *Malcolm H. Kerr Carnegie Middle East Center*, 2021.

Richard K Paisley and Taylor W Henshaw. Transboundary governance of the nile river basin: Past, present and future. *Environmental Development*, 7:59–71, 2013.

Limeng Pan, Mashor Housh, Pan Liu, Ximing Cai, and Xin Chen. Robust stochastic optimization for reservoir operation. *Water Resources Research*, 51(1):409–429, 2015.

Fiona L Paton, Graeme C Dandy, and Holger R Maier. Integrated framework for assessing urban water supply security of systems with non-traditional sources under climate change. *Environmental Modelling & Software*, 60:302–319, 2014.

Mario VF Pereira and Leontina MVG Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical programming*, 52:359–375, 1991.

JD Quinn, Antonia Hadjimichael, PM Reed, and Scott Steinschneider. Can exploratory modeling of water scarcity vulnerabilities and robustness be scenario neutral? *Earth's Future*, 8(11):e2020EF001650, 2020.

Julianne D Quinn, Patrick M Reed, M Giuliani, and Andrea Castelletti. Rival framings: A framework for discovering how problem formulation uncertainties shape risk management trade-offs in water resources systems. *Water Resources Research*, 53(8):7208–7233, 2017.

Saman Razavi and Hoshin V Gupta. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. theory. *Water Resources Research*, 52(1):423–439, 2016.

Martin Reuss. Is it time to resurrect the harvard water program? *Journal of Water Resources Planning and Management*, 129(5):357–360, 2003.

Charles Revelle, Erhard Joeres, and William Kirby. The linear decision rule in reservoir management and design: 1, development of the stochastic model. *Water Resources Research*, 5(4):767–777, 1969.

Mohsen Saadat and Keyvan Asghari. Feasibility improved stochastic dynamic programming for optimization of reservoir operation. *Water Resources Management*, 33(10):3485–3498, 2019.

Md Sahabuddin and Imran Khan. Multi-criteria decision analysis methods for energy sector's sustainability assessment: Robustness analysis through criteria weight change. *Sustainable Energy Technologies and Assessments*, 47:101380, 2021.

Jazmin Zatarain Salazar, Patrick M Reed, Jonathan D Herman, Matteo Giuliani, and Andrea Castelletti. A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control. *Advances in water resources*, 92:172–185, 2016.

Jazmin Zatarain Salazar, Patrick M Reed, Julianne D Quinn, Matteo Giuliani, and Andrea Castelletti. Balancing exploration, uncertainty and computational demands in many objective reservoir optimization. *Advances in water resources*, 109:196–210, 2017.

Jazmin Zatarain Salazar, Jan H Kwakkel, and Mark Witvliet. Evaluating the choice of radial basis functions in multiobjective optimal control applications. *Environmental Modelling & Software*, 171: 105889, 2024.

Salman MA Salman. The gerd and the revival of the egyptian–sudanese dispute over the nile waters. *Ethiopian Yearbook of International Law 2017*, pages 79–110, 2018.

Yasin Sari. Exploring trade-offs in reservoir operations through many objective optimisation: Case of nile river basin. 2022.

Leonard J Savage. The theory of statistical decision. *Journal of the American Statistical association*, 46(253): 55–67, 1951.

Mohamed S Siam and Elfatih AB Eltahir. Climate change enhances interannual variability of the nile river flow. *Nature Climate Change*, 7(5):350–354, 2017.

Herbert A Simon. Dynamic programming under uncertainty with a quadratic criterion function. *Econometrica, Journal of the Econometric Society*, pages 74–81, 1956.

Rodolfo Soncini-Sessa, Enrico Weber, and Andrea Castelletti. *Integrated and participatory water resources management-theory*. Elsevier, 2007.

Jery R Stedinger, Beth A Faber, and Jonathan R Lamontagne. Developments in stochastic dynamic programming for reservoir operation optimization. In *World environmental and water resources congress 2013: Showcasing the future*, pages 1266–1278, 2013.

Ashok Swain. Ethiopia, the sudan, and egypt: the nile river dispute. *The Journal of Modern African Studies*, 35(4):675–694, 1997.

Tebikachew Betru Tariku and Thian Yew Gan. Regional climate change impact on extreme precipitation and temperature of the nile river basin. *Climate Dynamics*, 51(9):3487–3506, 2018.

J Alberto Tejada-Guibert, Sharon A Johnson, and Jery R Stedinger. The value of hydrologic information in stochastic dynamic programming models of a multireservoir system. *Water resources research*, 31 (10):2571–2579, 1995.

Amaury Tilmant and R Kelman. A stochastic approach to analyze trade-offs and risks associated with large-scale water resources systems. *Water resources research*, 43(6), 2007.

Anwar Hassen Tsega. The geopolitics of water negotiations succeeding the gerd project in the nile river basin: the case of ethiopia, egypt, and sudan. *Humanitarian and Social Research Center (INSAMER). INSAMER, Istanbul*, 2017.

John Tsitsiklis and Benjamin Van Roy. Analysis of temporal-diffference learning with function approximation. *Advances in neural information processing systems*, 9, 1996.

Sean WD Turner, James C Bennett, David E Robertson, and Stefano Galelli. Complex relationship between seasonal streamflow forecast skill and value in reservoir operations. *Hydrology and Earth System Sciences*, 21(9):4841–4859, 2017.

Olli Varis, Cecilia Tortajada, and Asit K Biswas. *Management of transboundary rivers and lakes*. Springer, 2008.

Vlasios Voudouris, Ken'ichi Matsumoto, John Sedgwick, Robert Rigby, Dimitrios Stasinopoulos, and Michael Jefferson. Exploring the production of natural gas through the lenses of the aceges model. *Energy Policy*, 64:124–133, 2014.

Abraham Wald and Jacob Wolfowitz. Bayes solutions of sequential decision problems. *The Annals of Mathematical Statistics*, pages 82–99, 1950.

Warren E Walker, Marjolijn Haasnoot, and Jan H Kwakkel. Adapt or perish: A review of planning approaches for adaptation under deep uncertainty. *Sustainability*, 5(3):955–979, 2013.

Kevin G Wheeler, Mohammed Basheer, Zelalem T Mekonnen, Sami O Eltoum, Azeb Mersha, Gamal M Abdo, Edith A Zagona, Jim W Hall, and Simon J Dadson. Cooperative filling approaches for the grand ethiopian renaissance dam. *Water International*, 41(4):611–634, 2016.

Kevin G Wheeler, Jim W Hall, Gamal M Abdo, Simon J Dadson, Joseph R Kasprzyk, Rebecca Smith, and Edith A Zagona. Exploring cooperative transboundary river management strategies for the eastern nile basin. *Water Resources Research*, 54(11):9224–9254, 2018.

Kevin G Wheeler, Marc Jeuland, Jim W Hall, Edith Zagona, and Dale Whittington. Understanding and managing new risks on the nile with the grand ethiopian renaissance dam. *Nature communications*, 11 (1):1–9, 2020.

Lam Yeh. Optimal control of a finite dam: average-cost case. *Journal of applied probability*, 22(2):480–484, 1985.

Muluneh Yitayew and Assefa M Melesse. Critical water resources issues in the nile river basin. *Nile River basin: hydrology, climate and water use*, pages 401–416, 2011.

Benjamin F Zaitchik, Belay Simane, Shahid Habib, Martha C Anderson, Mutlu Ozdogan, and Jeremy D Foltz. Building climate resilience in the blue nile/abay highlands: A role for earth system sciences. *International journal of environmental research and public health*, 9(2):435–461, 2012.

Christiane Zarfl, Alexander E Lumsdon, Jürgen Berlekamp, Laura Tydecks, and Klement Tockner. A global boom in hydropower dam construction. *Aquatic Sciences*, 77:161–170, 2015.

Mark Zeitoun, Ana Elisa Cascão, Jeroen Warner, Naho Mirumachi, Nathanial Matthews, Filippo Menga, and Rebecca Farnum. Transboundary water interaction iii: contest and compliance. *International Environmental Agreements: Politics, Law and Economics*, 17:271–294, 2017.

Tongtiegang Zhao, Dawen Yang, Ximing Cai, Jianshi Zhao, and Hao Wang. Identifying effective forecast horizon for real-time reservoir operation under a limited inflow forecast. *Water Resources Research*, 48 (1), 2012.

A

Appendix: Results

A.1. Converge Analysis

While the progress curve on the left indicates there is still potential for improvement, the hypervolume graph on the right appears to be stabilizing. Considering the time and computational limits of the project, we decided that the convergence status is adequate with 50,000 NFEs.

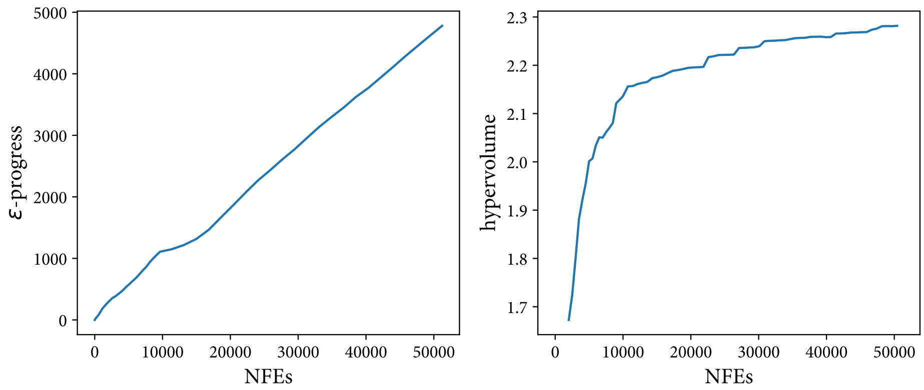

Figure A.1: Convergence Analysis Result.

A.2. Physical System Implications

In this section we present physical system changes imposed by policies that were not presented in the study.

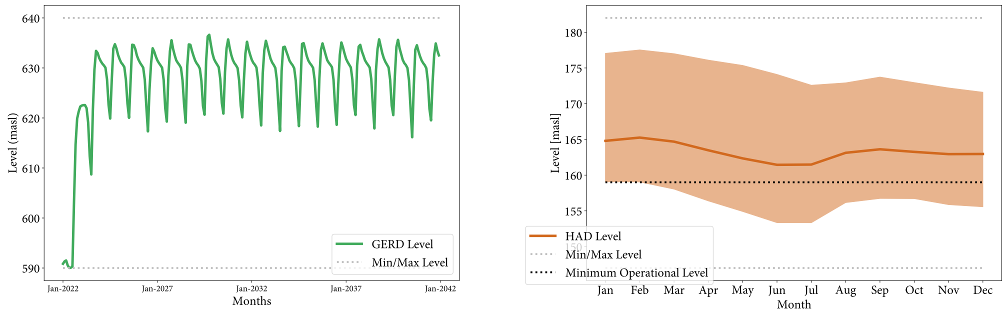

Figure A.2: GERD and HAD Level under Compromise: Percentile Policy

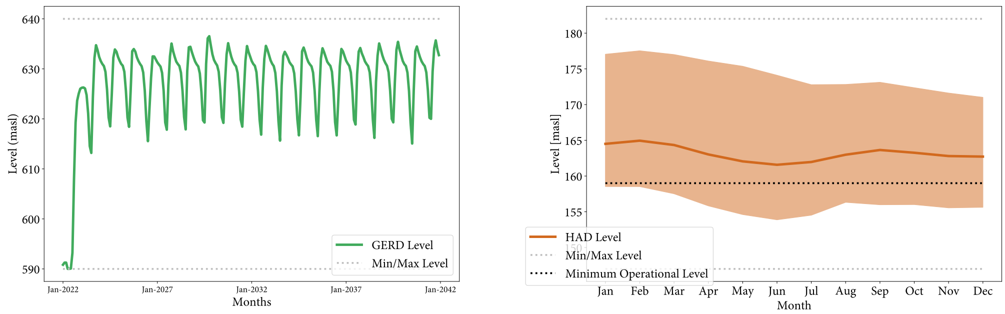

Figure A.3: GERD and HAD Level under Compromise: Absolute Policy

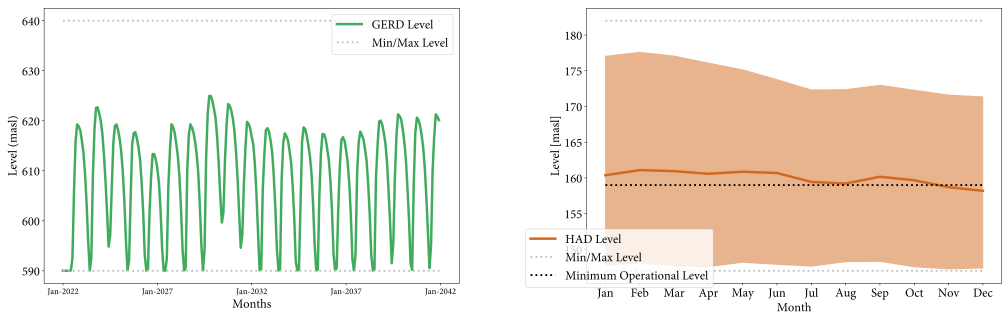

Figure A.4: GERD and HAD Level under Best Sudan Irrigation Policy

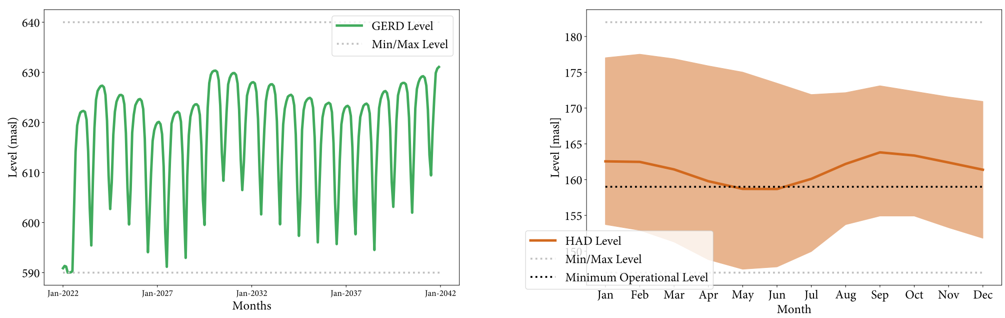

Figure A.5: GERD and HAD Level under Best Egypt Irrigation Policy