Author:
Dobson, Barney

Title:
Uncertainty in modelling and optimising operations of reservoir systems

Uncertainty in modelling and optimising operations of reservoir systems

By

Barnaby Alexander Dobson

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Engineering

Department of Civil Engineering

University of Bristol

October 2018

Word count: 45405

ABSTRACT

Reservoir operation optimisation aims to determine release and transfer decisions that maximise water management objectives such as the reliability of water supply, the hydropower production or the mitigation of downstream floods. This thesis studies two key issues in reservoir operation optimisation. Firstly, despite being an active field of research, the state of uptake of optimisation techniques by practitioners is largely unknown. Secondly, there are sources of uncertainty in the simulation models that underpin optimisation and the impact of these uncertainties on operation optimisation results has not yet been considered. We present a literature review that classifies different optimisation techniques based on what types of problem they are applicable to rather than the mathematical workings behind them, as previous reviews have done. This review is contrasted with a practitioner survey that reaches water managers and consultants around the world. We find that practitioners do not typically use operation optimisation tools, instead following decision-making procedures that are more informal than the formulaic operating policies presented in research. The survey suggests that a key reason for hesitation in the uptake of optimisation techniques is the limited fidelity of simulation models that underpin optimisation results. We discuss sources of uncertainty in these models and find that no work has yet considered the impact of structural uncertainty (i.e. arising from how interrelationships within the system model are defined) or contextual uncertainty (i.e. around definition of the model boundaries) on reservoir operation optimisation. Consequently, we formulate 'rival framings' of a real-world reservoir operation problem, each making different assumptions about structural/contextual uncertainties affecting the model of the system. We then test how the estimated performance of optimised decisions changes when evaluated under different framings. We find that contextual uncertainty in particular has a significant impact on estimates of performance. Finally, we investigate the applicability of 'robust optimisation', i.e. an approach where operations are directly optimised under multiple model formulations at once. In our case study, robust optimisation is effective because it produces a set of solutions that have greater robustness than would be achievable using conventional optimisation.

PLAIN LANGUAGE SUMMARY

There are more than 33,000 large reservoirs in the world that have been constructed to supply water for domestic, industrial, energy generation, flood control and agricultural needs. However, many of these reservoirs do not come close to delivering the benefits that were envisaged during their design. One explanation for this is that the reservoirs are not managed and operated as effectively as they could be. Therefore, it is common for researchers to create computational models of reservoir systems that water managers could use to improve the operation and management of dams. These models enable the reservoir operators to anticipate how different decisions would affect the objectives that they aim to fulfil. In research it is also common to apply mathematical techniques (optimisation) to determine which decisions will be the most effective, however this is not yet common in practice. In this thesis our first aim is to identify barriers to uptake of optimisation in practice. Therefore, we surveyed water managers about how they operate their reservoir systems. We found that a key reason that they do not use optimisation is that they do not trust the models that represent their reservoir systems.

Consequently, the remainder of the work in this thesis focused on the assumptions required to create models of reservoir systems. To test these assumptions, we created a model of a reservoir system in the UK that is operated by two water companies. We find that the most important modelling assumption is how much the two companies coordinate their decisions. If the companies do not represent each other in their simulation models, then the outputs of optimisation (for example, estimates about the cost of pumping) will be very variable (with differences of up to £200/day in our case study, about 25% of the total cost). Another important modelling assumption we identify is around how to represent the reservoir inflows statistically (where small changes can impact estimates of the amount of water to be supplied by external sources by 2 megalitres/day). We also present a technique to help ensure that operators will not expose themselves to the vulnerabilities that arise from making different model assumptions. We hope that the work presented in this thesis will give reservoir operators greater confidence that the outputs of optimisation will beneficial when applied to real systems and not just in the models used to demonstrate them on.

DEDICATION AND ACKNOWLEDGEMENTS

This work was funded as part of the WISE CDT under a grant from the Engineering and Physical Sciences Research Council (EPSRC), grant number EP/L016214/1.

This thesis has been made possible by many people. However, none have devoted more than Francesca Pianosi, whose supervision and patience has helped me to create this piece of work and improve myself as a researcher. I would also not be in this position without Thorsten Wagener, who originally encouraged my interest in research and has also provided excellent supervision for this thesis. I greatly appreciate Patrick Reed for hosting me in a research visit and teaching me about the rigour required to make good research. I am very grateful to Chris Hutton and Wessex Water for the data, time and effort they devoted to this work, and to the individuals and teams that participated in our survey. Wouter Knoben, thank you for being a solid friend throughout this and for entertaining my most bizarre research ideas and answering the endless simple questions that make the day-to-day easier. Colleagues and friends in Bristol, Cornell and beyond who have helped me to both generate good ideas and provided many fun times are too numerous to mention, but thank you for your support and kindness.

Finally, thank you to my family and to Collette Taylor for their unwavering belief and encouragement throughout these years.

AUTHOR’S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed: …………………………………………**Date**: …………………………………………….

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Context

Water resources is a broad term encompassing any source of water that is, or is potentially, useful [*Nature.com*, 2018]. If water is withdrawn from the environment it is most commonly used for agriculture (e.g. irrigation), industry (e.g. coolant) or municipalities (e.g. drinking) [*FAO*, 2015b]. However, the FAO additionally considers the environmental services provided by water to be one of its primary uses. Water resources can also refer to non-consumptive uses such as hydropower generation, which accounted for 16% of global electricity production in 2015 [*IEA*, 2017]. The management of water resources focuses broadly on ensuring that the right amount of water is available at the right time for its uses. In addition, water resources management typically encompasses managing too much water (i.e. flooding) and managing water quality. One of the key challenges in water resources management is the temporal variability in water supply [*J W Hall et al.*, 2014], with droughts being the most visible example of this.

Reservoirs are a water resources infrastructure that enable the 'banking' of excess water to use at a later time, thus reducing the variability in supply from connected water sources. A water supply with low variability is significantly and positively correlated with a nation's per capita GDP [*Brown and Lall*, 2006] and the creation of reservoirs has historically been the most common way to reduce supply variability. In the past 60 years the water security afforded by reservoirs has encouraged the widespread construction of dams across the world, as visualised in Figure 1.1. Recent estimates place the total global storage capacity of reservoirs between 7,000 and 8,000 km^3 [*Lehner et al.*, 2011]. This capacity is twice the annual global water withdrawal of around 3,700 km^3 [*FAO*, 2015a].

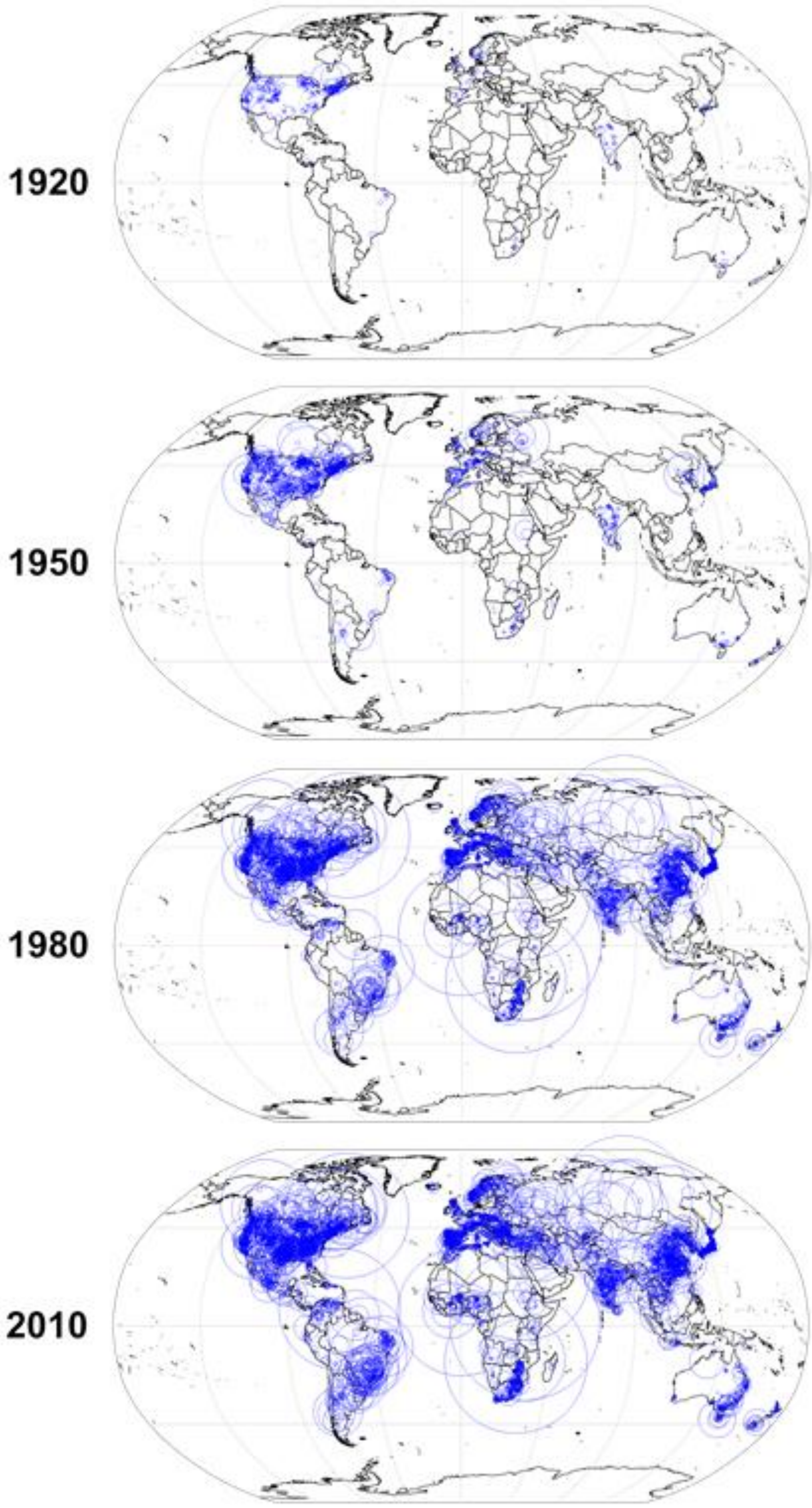


Figure 1.1: Distribution and size of dams in the Global Reservoirs and Dams (GRanD) database [*Lehner et al.*, 2011] at different stages in the past 100 years. Each blue circle represent an individual dam and the circle size is proportional to the capacity of the reservoir.

Dams are among the most massive man-made structures in the world, and the reservoirs they create have far-reaching impacts both upstream and downstream. About 50% of global river volume is presently being moderately or severely impacted by dams [*Grill et al.*, 2015]. These impacts are most commonly environmental, such as habitat destruction and sediment obstruction, or social, such as displacement of communities and inundation of productive lands [*Kraljevic et al.*, 2013]. Besides these environmental and social costs, evidence also suggest that many dams do not achieve the economic returns projected during their design [*WCD*, 2000]. The World Commission on Dams (WCD) identified that this failure is the result of a number of reasons. The most common causes for failure differ depending on the purpose of the dam. For hydropower dams they are underestimation of the costs of construction, mitigation and operation (for example, *de Sousa Júnior and Reid* [2010] describes how a large hydropower dam under construction is likely to become an economic burden due to over-optimistic cost estimates). For supply dams they are overestimation of future water demands (for example, *McCulloch* [2006] describes how a water supply dam was constructed for an expected increase in demand that never happened). For irrigation dams they are institutional failures that prevent the delivery of benefit (for example, *Van Wicklin* [2018] describes how many of the intended recipients of irrigation from a dam failed to receive their water due to an inability to resolve disputes about the land around planned canal routes). Interestingly, two out of three of these causes are related to estimation errors, highlighting the high level of uncertainty that affects our predictions of water resources systems.

In spite of these problems, dam building is undergoing a resurgence, primarily in developing or less developed countries. The total number of hydropower dams is expected to increase from around 2,000 to 6,000 by 2030 [*Ansar et al.*, 2014; *Zarfl et al.*, 2014]. These hydropower dams alone are projected to increase the total river volume that is moderately or severely impacted by dams to 93% [*Grill et al.*, 2015]. However, as *Muller et al.* [2015] suggests, it is unlikely that the negative impacts of dams will prevent their construction since their benefits are simply too essential for developing and less developed countries. These countries typically have rapidly growing populations [*Jahan*, 2015] – meaning dams will be required to both close the development gap and supply more people. In cases where there is no alternative but to build a dam, we should aim to minimize impacts where possible and, at the very least, ensure that the projected returns are realized.

In the UK, it is expected that significant investments will be needed to maintain a resilient water supply under the effects of climate change and population growth [*NIC*, 2018]. The National Infrastructure Commission (NIC) suggests that an increase in available water of over 25% (4000 megalitres/day) is required over the next 30 years to maintain the current level of resilience to droughts. The NIC proposes that this increase should be met in equal parts by reducing leakage, reducing demand, and creating new infrastructure, including new reservoirs and a national water transfer network. Such a network would enable transferral of water from wetter regions (e.g. the North of England) to drier ones (e.g. the South-East), increasing the effective storage of the system and providing greater operational flexibility [*OFWAT*, 2015; *Young*, 2016].

1.2 Problem Analysis

Reservoir systems are typically complex to manage for a range of reasons:

- Interconnectivity of reservoirs (for example, as would occur as a result of the NIC's proposed national water network above) gives flexibility around which water resource to use but makes operating the system less straightforward.

- The presence of multiple stakeholders with conflicting needs (for example, flood protection requires a low water level in a reservoir while hydropower production needs a high water level to maximise hydraulic head) makes it hard to find a course of operations that will satisfy all needs.
- Many inputs that force reservoir systems are highly variable in time (in particular the demand of water and the inflow into reservoirs), this variability makes operation a challenging problem.
- Finally, reservoir systems contain factors that are uncertain beyond the variability described above. For example, at the time of the dam's operation design, future demands are not known.

As reservoir systems increase in complexity it becomes more difficult to identify sensible choices for operational procedures [*Moss et al.*, 2016]. At some level of complexity, the only recourse for determining effective operations is likely to be the application of sophisticated modelling and optimisation techniques [*Yakowitz*, 1982; *Yeh*, 1985; *Hiew et al.*, 1989; *Labadie*, 2004; *Rani and Moreira*, 2009]. In cases such as the uncertainties described in the final point above, the focus should shift towards creating operating rules that are 'robust' against future uncertainty and perform acceptably in many situations [*Lempert and Collins*, 2007].

Much of the early literature in reservoir operation optimisation focused on creating operational strategies that could help overcome the difficulties of time variability in system forcing and interconnectivity in systems [*Maass et al.*, 1962]. The difficulty of multiple stakeholders and thus multiple objectives was raised by *Haimes and Hall* [1977] and has been extensively studied in reservoir operation optimisation since. While the number of studies on the optimisation of reservoir operations continues to increase every year, as shown in Figure 1.2, dams that do not deliver their planned benefits and exceed their predicted damages continue to be built [*Kraljevic et al.*, 2013]. This suggests a disconnect between research and practice, which has also been pointed out by several researchers in the field, such as *Yeh* [1985], *Simonovic* [1992], *Labadie* [2004] or *Brown et al.* [2015]. These authors have proposed many reasons for this disconnect, including: the lack of involvement of practitioners in the development of reservoir simulation and optimisation models; a lack of suitable data; the focus of researchers on over-simplified reservoir systems; the existence of institutional constraints that prevent innovation in water resource management practice; and the lack of accessible, credible and user-friendly software that implements reservoir simulation and optimisation methods. Although the opinions of experienced academics who have worked closely with practitioners is surely helpful to provide a good starting point, it is an indirect source of information. In the water resources optimisation literature the only direct survey of practitioners that has been published to date is *Rogers and Fiering* [1986]. One key finding of that survey was that the uncertainty present in simulation models contributed to a significant lack of trust in the results of the optimisation process.

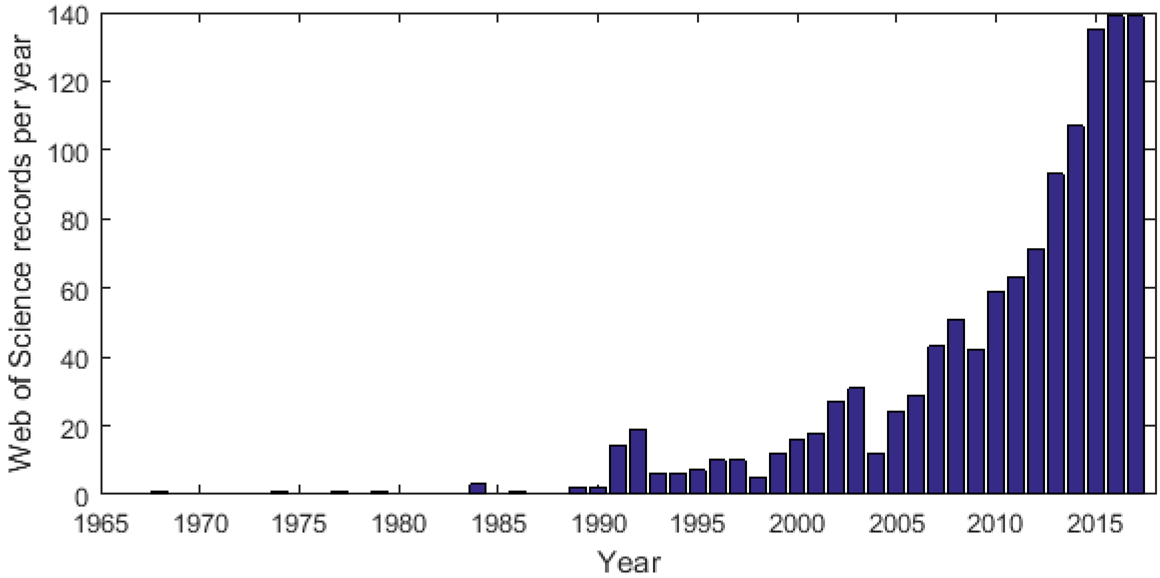


Figure 1.2: The records by publication year from a web of science search for the topic “reservoir operation optimisation”. The graph contains results from categories: water resources, engineering civil and environmental science up to the year 2017.

All water resources system models are subject to some forms of uncertainty. The key question in the context of reservoir operation optimisation is how valid a set of results will be in the face of these uncertainties [*Loucks*, 1992]. This thesis will focus specifically on contextual uncertainty, i.e. the uncertainty around the definition of the system boundaries when developing the model, and structural uncertainty, i.e. the uncertainty arising from the definition of the interrelationships between the model variables. Structural uncertainty in particular has received much attention in related fields. For example, in hydrological modelling it has been shown to significantly impact streamflow predictions [*Clark et al.*, 2008; *Fenicia et al.*, 2011; *Fowler et al.*, 2016]. Yet, of all the references pictured in Figure 1.2 (and in a far more broad search of ”water resource* optim*” that returned almost 12,000 papers), the sub-searches for topics “struct* uncert*” and “context* uncert*” did not return any result focusing on the contextual or structural uncertainty of water resources system models, but only 5 results focusing on the structural uncertainty of climate/hydrological/environmental models linked to the water resources system model. Thus, it seems that no work in reservoir operation optimisation, nor more broadly in operations research or water resources optimisation, has studied the impact of structural or contextual uncertainty on optimisation results.

1.3 Research Objectives

The goal of this thesis is to identify the concerns of potential users of reservoir operation optimisation tools and make progress towards mitigating them while improving our understanding of water resources systems. Based on the identified concerns, we focus on one in particular since it has received little to no attention in research: the impact of structural and contextual uncertainty in reservoir operation simulation models on the performance of operations. These naturally provide two overarching research questions:

- What are the barriers to uptake of reservoir operation optimisation techniques in practice and how can they be overcome?
- Are optimisation results being undermined by model uncertainties and what does this teach us about water resources systems?

We address these questions by pursuing the following research objectives:

- Review and classify the existing reservoir operation optimisation literature in a way that focuses on the more practical aspects of the optimisation methods, such as when they are applicable, rather than their mathematical properties.
- Survey water managers to determine how practice reflects what has been developed in research and to identify barriers to the uptake of reservoir operation optimisation methods.
- Develop a methodology to quantify the impact of uncertainties on estimates of performance of optimised operations and demonstrate this methodology. To

demonstrate our methodology, we create a new model of a real-world reservoir system.

- Develop a methodology to optimise operations that perform effectively over a wide range of the uncertainties present in the above model.
- Create a measure of performance under uncertainty that can scale to multiple objectives because multiple objectives are common in reservoir operation optimisation problems.

1.4 Thesis outline

The research chapters (2 to 5) are all linked in that they address the research objectives and questions proposed above and follow logically from each other as shown in Figure 1.3.

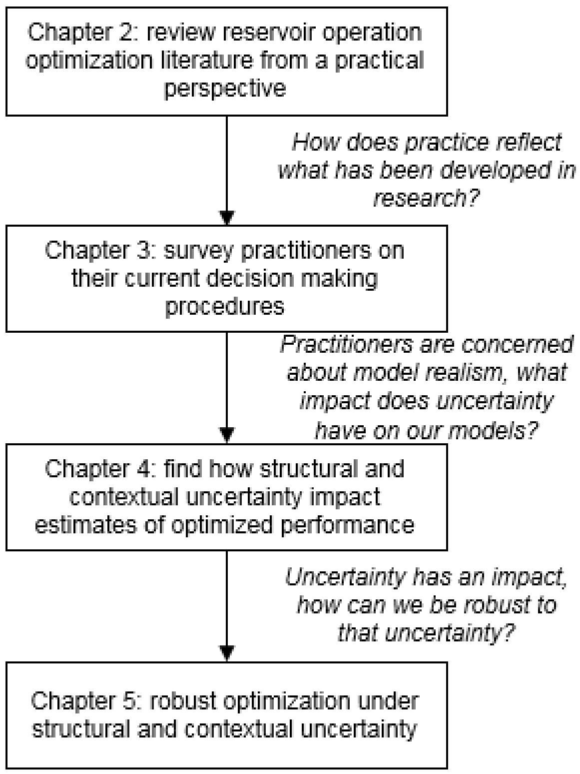


Figure 1.3: The logical progression of the work performed in this thesis. Motivating questions that link chapters are shown in italics while chapter descriptions are provided in boxes.

Chapter 2 focuses on identifying how operation optimisation is framed in research. We review the reservoir operation optimisation literature to make it more accessible to those who wish to use the tools, rather than those who wish to develop the tools. We achieved this by changing the focus to *what* different methods achieve and *when* they are applicable, rather than *how* they work.

Chapter 3 introduces a survey of water resources companies and consultants in the UK, Australia, South Africa and South Korea, to determine their current decision-making procedures, their usage and views on modelling tools, and the potential for and barrier to uptake of optimisation tools. We also provide a discussion on how to bring research and practice's interpretations of reservoir operation together.

Chapter 4 addresses a key result from the survey: that practitioners are concerned about uncertainty in the simulation models that underlie optimisation results. We therefore demonstrate a methodology to quantify how structural and contextual uncertainty can impact the estimates of optimised performance in a real-world case study of a two-reservoir pumped storage system. This work also discusses the different uncertainties present in water resources systems and how to isolate their impact such that specific uncertainties can be studied.

Chapter 5 considers how to define, compare and achieve robustness in water resources operations. We introduce a novel method for measuring robustness which provides a simple but powerful interpretation of the uncertainties present in a multi-objective reservoir operation problem. Through application to the same case study system of Chapter 4, robust optimisation under structural and contextual uncertainty is demonstrated to be an effective technique to attain robustness.

Chapter 6 concludes by summarising the previous chapters and discussing how to improve knowledge transfer from research to practice in reservoir operation optimisation.

CHAPTER 2: A REVIEW AND CLASSIFICATION OF RESERVOIR OPERATION OPTIMISATION METHODS[1]

2.1 Introduction

A recent estimate places the total global storage capacity of reservoirs and dams between 7,000-8,000km^3 [*Lehner et al.*, 2011]. While dam construction has slowed in countries with a Human Development Index (HDI) above 0.7 (the UN Development Programme's threshold for a 'high' level of development [*Jahan*, 2015]), it is likely to continue at a considerable rate in countries with an HDI below 0.7. The latter countries contain around half of the human population and have the highest projected growth rates, with a total population increase of 18-27% by 2030 [*UN*, 2015], as shown in Figure 2.1. Currently, the per capita water storage of low HDI (<0.7) countries is around one third of high HDI (>0.7) countries. Besides building new dams to fulfil irrigation and water supply needs, hydropower dams are also expected to triple worldwide (from 2,000 to nearly 6,000) by 2030, under growing electricity demand [*Ansar et al.*, 2014; *Zarfl et al.*, 2014].

[1] This Chapter is has been accepted conditional on minor revisions as a review article in 'Advances in Water Resources'

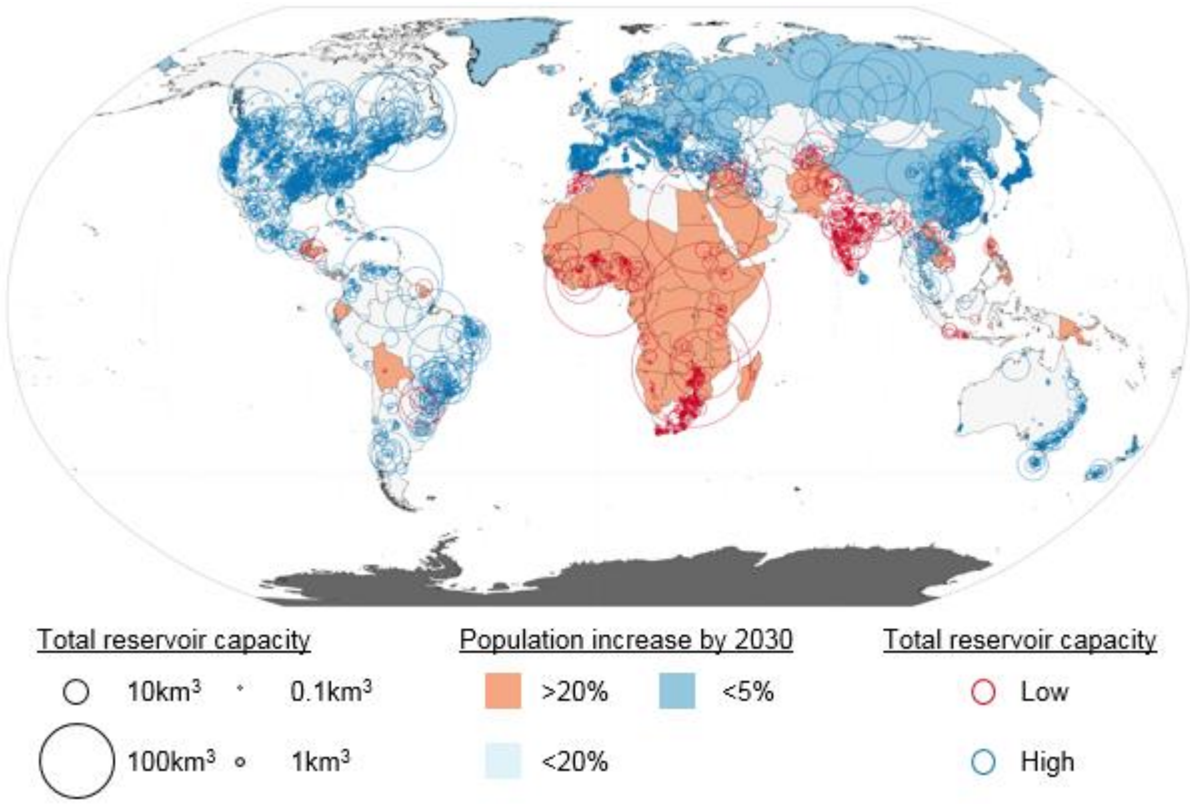


Figure 2.1: Map of the reservoirs listed in the GRanD database [*Lehner et al.*, 2011], centres of circles are a dam's location, the size is proportional to capacity and the colour indicates HDI. Countries are coloured by projected population growth by 2030.

Despite their importance and the level of planning and resources required to construct a dam, it is common for reservoirs not to achieve the goals envisaged in their design, in terms of both economic returns and mitigation of negative impacts [*WCD*, 2000; *Labadie*, 2004]. Dams are most commonly criticised for causing social and environmental damage, such as the displacement of communities or the obstruction of sediments [*Graf*, 1999; *Ouyang et al.*, 2010; *Tockner et al.*, 2011; *Liermann et al.*, 2012; *Scudder*, 2012], which may not be sufficiently understood beforehand. Therefore, damages are underestimated and poorly mitigated by actions recommended in social and environmental impact assessments [*Nakayama*, 1998; *Fearnside*, 2016]. Nonetheless, countries with unsatisfactory water resource infrastructures continue dam construction in a drive to increase quality of life [*Muller et al.*, 2015]. In countries with adequate water security, negative impacts have occasionally resulted in the decommissioning of dams [*Allan*, 2003; *Bellmore et al.*, 2016; *Ho et al.*, 2017] and in substantial legal regulation for water resources management. Regulations include changing investment preferences towards efficiency gains that can be achieved through, for example, water transfer or network projects [*Brown et al.*, 2015]. The coordinated operation of these linked systems allows greater scope for increased efficiency than could be achieved by uncoordinated operation of individual sources, with efficiency gains that could be applied towards environmental impact mitigation [*Poff and Schmidt*, 2016].

Optimisation of reservoir operations is therefore more relevant than ever, both as a complement to the efficient design of new dams and for the revision of operations in existing ones. Here, we would define reservoir operation as the determination of how much water to abstract from sources (e.g. rivers), to transfer between reservoirs, and to release from reservoirs to points of consumption (e.g. for irrigation, domestic or industrial consumption) or use (e.g., hydropower production). Reservoir operation is a challenging decision-making problem because it requires finding a balance between decisions conflicting in time (for example, whether to accept a cost in the short-term in order to avoid a larger, but more uncertain, cost in the mid-term) and across uses (for example, between irrigation, hydropower and municipal supply).

Research has demonstrated that the use of mathematical models to simulate and optimise reservoir operation can significantly enhance the performance of existing reservoirs, as well as enable efficient design of new dams or their repurposing/expansion. Traditionally, dam design has been based on Rippl's method, an approach that aims to find the smallest reservoir capacity that can ensure releasing the target water demand through a worst-case drought [*Rippl*, 1883; *Hazen*, 1914; *Loucks et al.*, 2005]. Drawbacks of this approach include the difficulty in applying it to systems that go beyond the simple single-reservoir and single-purpose case, for example coordinated reservoir networks or multiple purpose reservoirs [*Maass et al.*, 1962]. More flexible approaches that can accommodate these drawbacks have been proposed for many years [*Vogel and Stedinger*, 1987; 1988; *Douglas et al.*, 2002; *Celeste*, 2016] and are now widely adopted in scientific research [*Loucks et al.*, 2005]. These design methods simulate the reservoir system against long time series of reservoir inflows and iterate the simulation until finding the minimum reservoir capacity that meets the target objectives under a variety of hydrological conditions. As such, they require an explicit formulation (and preferably nested optimisation) of the reservoir operating policy that will be used to make release decisions in the various simulated circumstances.

Reservoir operation optimisation is a mature and yet very active research area (see Figure 1.2) and a number of reviews of the available optimisation methods have been carried out over time [*Yeh*, 1985; *Labadie*, 2004; *Castelletti et al.*, 2008; *Rani and Moreira*, 2009; *Ahmad et al.*, 2014]. While these reviews may differ in the emphasis given to a particular group of methods or another, they all share the same fundamental approach to classifying and presenting methods, which is based on the mathematical properties of the optimisation algorithms. However, we believe that an alternative approach to classifying methods is possible and useful, particularly for new and non-specialised users, by focusing on the argument of the optimisation problem. In order to better understand this point, we note that there are four elements to an optimisation problem: (1) the objective(s), i.e. the variable(s) to be minimised/maximised, such as the average water supply, or hydropower production, level of flood protection, etc.; (2) the argument of the optimisation problem, i.e. the decision variable(s) whose optimal choice would deliver the minimum/maximum objective value(s); (3) the constraints, i.e. the set of equations that link the decision variables to the objectives; and (4) the optimisation method, i.e. the algorithm used to determine the values of the decision variables that optimise the objectives while respecting all the constraints. We have presented these elements in the order in which they should be defined in practice. Indeed, when formulating an optimisation problem the optimisation method should be the last element to be chosen, and yet previous reviews in this field focus on this element as the key to present and compare literature contributions. We propose instead that the highest level of classification should be the argument, which determines the 'output' of the optimisation task (which type of variable is being optimised, i.e. a sequence of release/transfer decisions or an operating policy, as further explained in the following sections). This changes the focus to the practical aspects that make an optimisation approach more or less suitable for the problem at hand (*what* type of solution they deliver and *when* they are useful), rather than the mathematical properties of the solution algorithm (*how* the methods achieve those solutions).

This work hence offers a new review of the scientific literature on reservoir operation optimisation where optimisation methods and applications are presented according to the type of argument to the optimisation problem instead of the underlying mathematics in use. Indeed we will show that the same type of algorithm (for instance, a genetic algorithm) can be used to solve reservoir optimisation problems with very different arguments (e.g. deriving the optimal sequence of short-term decisions vs determining the long-term optimal operating policy); while an optimisation problem with the same type of argument (and hence solution) can be solved by using very different algorithms (e.g. a genetic algorithm vs a nonlinear programming one). We complement the review with a terminology disambiguation table to help the reader navigate both our review and the wider literature, where terms are sometimes used with different meanings by different authors. By focusing less on the mathematical properties of solution algorithms in favour of an argument-based classification of the optimisation methods, we are also able to draw a comparison between them, discuss important practical factors such as the different assumptions required by each method, and ultimately provide guidelines towards selecting a suitable method for the decision-making problem at hand. We hope the work presented in this Chapter will make the reservoir operation optimisation literature accessible to a wider audience besides the academic community already active in water systems analysis and optimisation.

In the remainder of this Chapter, we present our classification system by argument and review optimisation methods and applications accordingly. Beforehand, however, in the following section, we briefly define the two other elements of the optimisation problem formulation described previously: the objectives and constraints.

2.2 Objectives and constraints

In the optimisation literature, an 'objective' is a scalar variable that summarises the system performance over a temporal period. The optimisation method aims at either minimizing or maximizing the objective; throughout this review we will assume that objectives must be minimized, i.e. they represent either costs or benefits changed in sign. Objectives that are commonly used to evaluate the performance of reservoir systems capture the system's reliability, resilience or vulnerability [*Hashimoto et al.*, 1982; *Loucks et al.*, 2005; *Kasprzyk et al.*, 2013]. Reliability objectives measure the frequency of occurrence of a specified failure event (for example, failure to supply adequate amounts of water to a demand node), resilience objectives measure the recovery time from a failure event, and vulnerability objectives measure the severity of the failure's consequences. The choice and definition of objectives can vary greatly depending on the specific reservoir system under study, the availability of data, etc. and as a general rule should reflect as much as possible the reservoir operator's targets and preferences. However, two factors in the formulation of the objective impact the applicability of operation optimisation methods. The first is the presence of non-linear components in the objective's mathematical definition, which may prevent the application of some methods that assume linearity, as will be summarised in Section 2.4.4. The second is the so called 'time-separability', i.e. the fact that the objective is defined by temporal aggregation (for instance, averaging) of 'step costs' (or 'step benefits') that only depend on system variables at one time-step [*Barro and King*, 1982]. An example of a time non-separable objective is the profit from selling water in a water market, where the price at each time step is dependent on water sales at previous time steps.

Another critical aspect that may strongly influence the applicability of reservoir optimisation methods is the number of objectives that the operation aims to minimize. In fact, reservoirs are typically expected to serve multiple purposes. For example, nearly half of all large dams included in the World Register of Dams [*ICoLD*, 2003] have multiple uses - most commonly irrigation, hydropower, water supply and flood control. A possible approach to handle multiple objectives is to make them commensurable by appointing them an economic value so that they can be summed up into a single objective that expresses the total net benefit (or cost) over the simulation period [*Maass et al.*, 1962]. However, this technique may not sufficiently compensate for non-economic indicators and does not express the available trade-offs between objectives to the decision maker, as described in detail in *Kasprzyk et al.* [2013]. An increasingly preferable alternative is to solve a multi-objective optimisation problem, which returns a set of Pareto-optimal (or Pareto efficient) solutions, instead of one optimal solution. Pareto-optimal solutions are characterised by the property that an improvement in one objective is unattainable without a deterioration in at least one other objective [*Cohon and Marks*, 1975]. The choice of the 'best' solution within the set of Pareto-optimal ones is not considered as part of the optimisation process because it involves a subjective evaluation of what acceptable trade-offs between the objectives should be. However, in order to assist the decision maker in such evaluation and choice, the set of Pareto optimal solutions can be displayed in the objective space (this representation is called the ‘Pareto front’) to reveal and quantify those trade-offs. The benefit of visualizing the Pareto front lies in enabling the decision maker to view the impact of their decisions in the context of all objectives rather than a single, prior weighted objective. In selecting one solution within the Pareto front the decision maker implicitly selects a posterior set of weights. For the sake of simplicity, in the next section we will first introduce optimisation methods with reference to the single-objective case, and in Section 2.4.2 we will discuss their ability in handling multi-objective optimisation problems, in particular when the number of objectives increases above 3 – the so called ‘many objective’ problems [*Fleming et al.*, 2005].

The ’constraints’ of an optimisation problem are all the equations that are needed to compute the objective(s) from the decision variables. In a reservoir operation optimisation problem, this link is established via a simulation model of the reservoir system, which is run over the simulation period for given initial condition and trajectory of forcing inputs. Conservation of mass, in the form of a water balance equations, typically forms the foundation for the physics of reservoir simulation models, while nodes and links are the basis for the dynamics [*Ford and Fulkerson*, 1962]. For a more detailed description of typical reservoir system simulation equations, we refer the reader to *Rani and Moreira* [2009], *Matrosov et al.* [2011], *Mo et al.* [2013] and *Seifollahi-Aghmiuni et al.* [2016], and to *Coelho and Andrade-Campos* [2014] and *Wurbs* [2005] for examples of reservoir simulation software. This mathematical description is typically complemented with several hard and soft constraints on individual variables. Hard constraints are those constraints that cannot be violated under any circumstance and typically represent physical limits, such as non-negativity of storage and flow variables. Less commonly used hard constraints include equations to impose conservation of energy and wave travel times. Soft constraints, instead, are those constraints that should not be violated but that are not physically impossible to break [*Mayne et al.*, 2000], for example a minimum environmental flow requirement downstream of a reservoir. Soft constraints may be included in the optimisation problem either as additional objectives or as hard constraints. Treating soft constraints as objectives allows exploring the trade-off between breaking the soft constraints and preventing a greater cost elsewhere. The downside is the increase in the number of objectives, which may increase the difficulty of solving the multi-objective optimisation problem. Depending on the case study application, a balance can be found between the ease of optimisation (which would suggest using hard constraints) and completeness of information delivered to decision makers (which would suggest using objectives). Some interesting examples of swapping constraints and objectives include *Sigvaldson* [1976], which uses channel capacity as an objective rather than a constraint, as most commonly treated, and *Koutsoyiannis and Economou* [2003], which uses deficit as a constraint rather than an objective.

2.3 Classification of methods by argument

This section presents our classification system of reservoir operation optimisation methods, which focuses on a higher-level understanding of the decision variables to which they are applied (the argument of the optimisation problem). For each method, we will review applications in the literature, and provide a short description of the most commonly used optimisation algorithms, with reference for further reading on more mathematical details. The classification system is summarised in Figure 2.2, while further details about the adopted terminology are given in Figure 2.3.

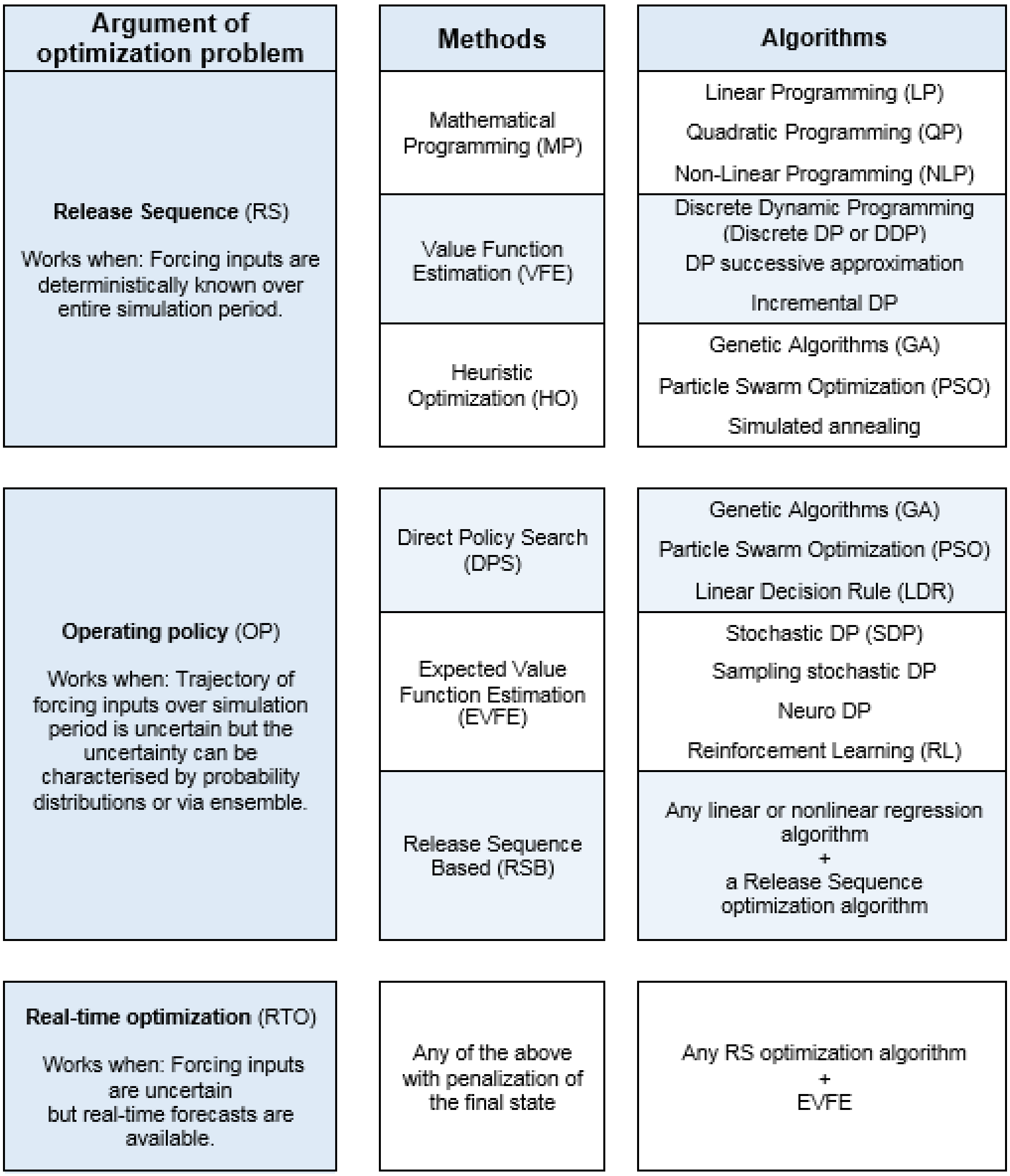


Figure 2.2: Our proposed classification system of reservoir operation optimisation methods based on the argument of the optimisation problem. The list of algorithms is not intended to be exhaustive, but it covers the literature applications reviewed in this Chapter.

Definition used in this paper	Alternative definitions in the literature	Examples of alternate definitions in the literature	Justification for our choice
Release Sequence (RS)	Operating policy	[*Afshar*, 2012]	Operating policy is much more frequently used in the context as described in this paper.
	Network flow optimization	[*Wurbs*, 1993]	RS is more specific to reservoir systems. Also, "network flow optimization" commonly refers to linear systems only.
Operating Policy (OP)	Reservoir (*or* Operating, *or* Release) rule	[*Oliveira and Loucks*, 1997]	OP avoids confusion with a *rule curve* (a format to present a specific and simplified OP dependent only on storage and time of year).
	Hedging curve (*or* function, *or* rule)	[*Draper and Lund*, 2004; *You and Cai*, 2008]	OP avoids implying that the only purpose of operation is hedging (i.e. to accept small deficits in the current time-step to avoid larger deficits in the future)
Real-Time Optimization (RTO)	Model Predictive Control (MPC)	[*Raso et al.*, 2014]	MPC is used because the (decision) *model* is being used in *predictive* mode for the purpose of *control* (of the system). We prefer RTO to MPC because RTO also indicates *what* is being done (an optimization in real-time) and not only on *how* (by using model predictions).
Value function	Cost-to-go function	[*Castelletti et al.*, 2007]	
Direct Policy Search (DPS)	Parameterization-Simulation-Optimization (PSO)	[*Koutsoyiannis and Economou*, 2003]	This is a vague term that could be applicable to a range of model building processes.
Release Sequence based	Implicit Stochastic Optimization (ISO)	[*Labadie*, 2004]	We decided to avoid ISO because we would like not to emphasize the implicit/explicit divide (see discussion in Section 3.6) and because DPS also considers forcing input variability implicitly.

Figure 2.3: Disambiguation table aimed at clarifying the terminology used in this paper and commonly found in the literature, where different terms can have similar but subtly different meanings, or the same terms are used by different authors to refer to substantially different concepts. The examples are for illustrative purposes only and are by no means exhaustive.

In our classification we distinguish three main types of argument:

- Release Sequences (RS) optimisation methods. Optimisation aims at finding the sequence of release decisions over a prescribed time period (Figure 2.4a) that minimises operational objectives under a given scenario of forcing inputs, for example a given time series of reservoir inflows and water demand. RS optimisation can be used to directly inform operational decisions if the underlying assumption that forcing inputs can be deterministically predicted is valid. The larger the deviations from the assumed deterministic scenario, the less effective the 'optimal' RS will actually be when applied in reality. Since forcing inputs are typically very uncertain and the mismatch between predictions and actual trajectories very large, 'optimal' RS are rarely implemented in practice. More commonly, RS optimisation is an intermediate step within a more complex optimisation process of the other two types below (discussed in Section 2.3.3). Another possible use of RS optimisation is in what-if studies, for example to determine a reference baseline for comparison with other optimisation solutions or to assess the upper bound of system performance – the maximum that could be achieved with the existing infrastructure under the "ideal" assumption of perfect foresight of all future inflows (for an example see *Castelletti et al.* [2012b]).

- Operating policy (OP) optimisation methods. Optimisation aims at finding the optimal operating policy, i.e. a function that can be used to determine the release conditional on the state of the reservoir system in the current time-step (Figure 2.4b). In other words, optimisation returns a strategy (the OP) for making release decisions, rather than the release decisions themselves. At each time-step, the optimal OP should return the decisions that will perform best over the expected trajectories of forcing inputs that may occur from that time-step onwards. The assumption here is not that the future forcing inputs trajectory is deterministically known (as with RS optimisation) but only that the trajectories (historical or synthetic time-series) or distributions assumed in the OP optimisation are representative of actual conditions. The state variables that OPs depend on typically include reservoir storage and time of year. They may also include other variables, e.g. current inflow [*Oliveira and Loucks*, 1997], depending on the characteristics of the study site, the reservoir system equations and the chosen optimisation method.

- Real Time Optimisation (RTO) methods. RTO uses an optimised RS over a rolling time horizon for which real-time forecasts of forcing inputs are available. The first release in the RS is implemented, and then at the next time step the optimisation process is performed again with updated forecasts, as displayed in Figure 2.4c. RTO is ideal if real time computing resources and accurate input forecasts are available.

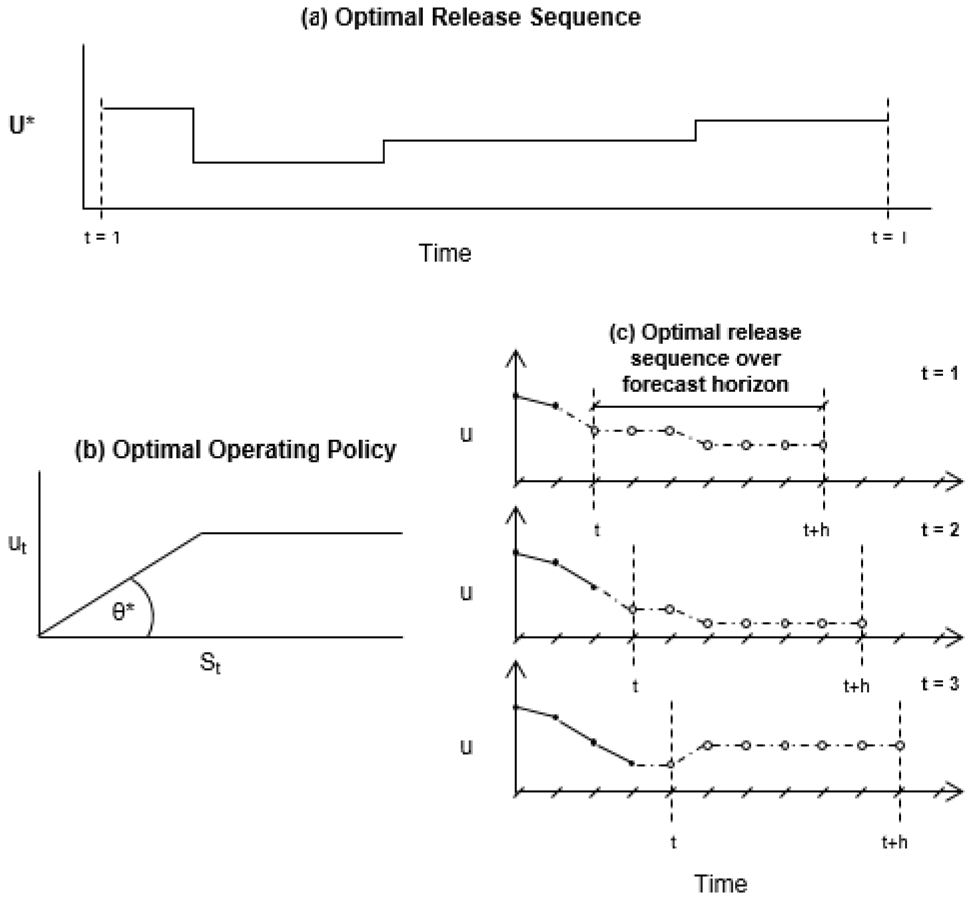


Figure 2.4: Schematic examples of the three possible outputs of reservoir operation optimisation, depending on the argument of the optimisation problem. (a) An optimal Release Sequence, i.e. a sequence of release decisions (**U***, in the notation of Section 2.3.1) over time. (b) An optimal Operating Policy, i.e. a function that returns a release decision for a given time step (u_t) depending on the system state (e.g. storage, S_t) at that time (t). (c)

Schematic illustrating the working principle of Real Time Optimisation. Here, the Release Sequence is re-optimised at every time-step over a rolling horizon (from current time t to t+h) for which input forcing forecasts are available, but only the first release decision of the sequence is actually implemented.

For each of the three above cases, our classification system (Figure 2.2) distinguishes optimisation methods based on their key working principles, i.e. essential mathematical properties of the optimisation problem formulation. For each method, the optimisation problem can be solved using different algorithms, as shown in the last layer of our classification system. While there are certainly differences between algorithms under the same method, they do not significantly affect the broader type of reservoir system to which the overarching method is applicable (with the exception of algorithms for Mathematical Programming, as further discussed in Section 2.3.1). Therefore, in the following sections we will focus on the description of the different methods and only provide references for further details on the specific algorithms. These descriptions form the basis for our discussion in Section 2.4, where we will compare the applicability of the various methods to different types of reservoir systems (for example, presence of multiple reservoirs or multiple objectives, linearity or non-linearity of the reservoir simulation model, etc.) and give practical guidelines towards selecting an appropriate method for a given system.

2.3.1 Methods for Release Sequence (RS) optimisation

The first case identified by our classification system is that of Release Sequence (RS) optimisation (see Figure 2.3 for disambiguation of terminology). A RS is a sequence of reservoir release decisions over a prescribed time period. Thus, each release in the sequence is a variable in the optimisation problem. An optimal RS is the release sequence for which an objective is minimized (under a given deterministic scenario of the system forcing inputs, e.g. reservoir inflows and demands), i.e.:

$$\boldsymbol{U}^* = arg \min_{\boldsymbol{U}} J \tag{2.1}$$

where **U** is a matrix containing all releases over the simulation period for all the reservoirs in the system under study (i.e. a RS), J is the aggregated objective associated with these releases, and **U*** is the optimal RS. Since the optimisation argument is the matrix **U**, the solution space to be explored in the optimisation quickly grows with the length of the simulation period and the number of reservoirs. The large search space is a characteristic difficultly of RS optimisation. The three most commonly used methods for RS optimisation are summarized below.

Mathematical programming (MP). We classify as MP any method that exploits the mathematical properties of the optimisation problem (for example, linearity and convexity of the constraints and objective) to efficiently find an optimal RS. As such, MP is most effective where speed is important and simplifications to fit the required assumptions (e.g. linearizing non-linear components) are acceptable. MP employs a broad range of algorithms, distinguishable primarily by the level of non-linearity allowed in the objective and constraint definitions. Linear and quadratic programming algorithms (LP, QP) require that all constraints be described by linear equations and that the objectives be either linear (LP, e.g. applied to the RS optimisation by *Hiew et al.* [1989] and *Terlaky* [2013]) or quadratic (QP, e.g. *Mariño and Loaiciga* [1985]). While these assumptions are strong, the advantage of LP and QP is that they can quickly find global optima even for large RS optimisation problems. However, as the linearity assumptions become less acceptable and non-linear equations are needed for a more realistic representation of the reservoir system, non-linear programming (NLP) is required. Sequential linear programming (e.g. *Martin* [1983], *Grygier and Stedinger* [1985]) and sequential quadratic programming (e.g. *Boggs and Tolle* [1995]) have been most commonly applied to RS optimisation, however other NLP algorithms exist and continue to be developed (see, for example, *Bazaraa et al.* [2013] for a recent collection of available algorithms). The disadvantage of NLP algorithms is that they cannot guarantee reaching a globally optimal solution in usable computation time for many problems [*Bazaraa et al.*, 2013].

Value function estimation (VFE). This method exploits the dynamic nature of the optimisation problem by breaking it into a sequence of easier to solve sub-problems, each relevant to one time-step in the simulation period. The key idea is to define a value function that, for each time-step, represents the cost it takes to transition from the state at that time-step (t) to the state at the final time-step of the simulation period (T) *if only optimal decisions are made*, i.e. via the optimal RS from t to T [*W A Hall and Buras*, 1961]. The value function can be derived by solving the recursive Bellman equation of dynamic programming [*Bellman*, 1956], which has been extensively used for reservoir operation optimisation for a long time – the first review of its application dating back to *Yakowitz* [1982].

There are two primary strengths to the VFE method. Firstly, it does not impose any limitation on the level of non-linearity of the objective or constraints. Secondly, the solution time only increases linearly with the length T of the simulation period (in contrast to the other RS optimisation methods, which increase polynomially or worse) so that it can be applied to find optimal RS that are very long. The drawback is that, since at each time-step the numerical resolution of the Bellman equation requires the evaluation of all possible combinations of state variables (e.g. storages) and decision variables (e.g. releases), the solution time scales exponentially with the number of states and decisions. This problem was named by Bellman as the *curse of dimensionality* [*Bellman*, 1956] and it severely limits the applicability of this method to large reservoir systems. A second drawback is that, since the value function is only defined at discrete points, interpolation between point evaluations is required. The first weakness compounds the second: the curse of dimensionality pushes towards using a coarser resolution and this makes the interpolation less accurate. Several variants of the discrete DP algorithm have been proposed to mitigate the problem in the context of reservoir operation optimisation, for example incremental dynamic programming [*W A Hall et al.*, 1967] and dynamic programming successive approximation [*Shim et al.*, 2002], however none of these have established as standard practice. Another very important limitation of the VFE method, which no technical advances will overcome, is that the very definition of a value function requires a time-separable objective (as discussed in Section 2.2), making the method incompatible with common performance metrics such as resilience metrics [*Hashimoto et al.*, 1982].

Heuristic optimisation (HO). This term covers a wide range of algorithms that can use very different working principles, but have as a common trait the fact that they attempt to find an approximate solution to a problem (in our case, highly non-linear and/or with large number of reservoirs) for which classic methods (MP and VFE in our case) are not applicable. Given such variety of HO algorithms, we do not discuss the entire spectrum of options but highlight that the two most common methods currently in use for RS optimisation are genetic algorithms (GA) [*Wardlaw and Sharif*, 1999; *Hınçal et al.*, 2010] and particle swarm optimisation (PSO) [*Kumar and Reddy*, 2007; *Noory et al.*, 2012]. However, numerous other algorithms have been tried in the context of RS optimisation, such as honey bees mating [*Haddad et al.*, 2006], ant colony optimisation [*Kumar and Reddy*, 2006], simulated annealing [*Georgiou et al.*, 2006] and many more [*Garousi-Nejad et al.*, 2016]. To the best of the authors' knowledge, HO was first applied to the RS problem by *Wardlaw and Sharif* [1999]. Given that no single algorithm dominates in all cases, newer algorithms use a combination of optimisation search strategies blended from different algorithms [*Reed et al.*, 2013], which are selected in an adaptive manner throughout the optimisation process. An example that appears to be very successful is the Borg algorithm [*Hadka and Reed*, 2013].

The advantage of HO is that it can be equally applied to linear or non-linear constraints and objectives, as well as to either time-separable or non-separable objectives. Hence it can be applied to problems where complex decisions are investigated (for example, planning drought revenue loss insurance as in *Herman et al.* [2014]). Since HO covers a large variety of algorithms it is difficult to make generic statements about its weaknesses, which may vary from one algorithm to another. However, one general comment is that, as the size of the RS increases (either due to many decisions per time-step or a long simulation period, or both) the solution time can become prohibitively long.

2.3.2 Methods for Operating Policy (OP) optimisation

An Operating Policy (OP) is a function that takes the current state of the system and returns a release decision, or set of release decisions, to be implemented in the current time-step. At a minimum, the system state vector (i.e. the independent variables used as inputs to the OP) should include the reservoir storages at the current time-step; in a more sophisticated OP it may also include additional information such as time of year (useful for reservoir systems with strong seasonal behaviour), reservoir inflows at the current or previous time-step [*Tejada-Guibert et al.*, 1995], or other information like flows at upstream locations in the reservoir network [*Giuliani et al.*, 2015]. In the following we denote an OP as

$$\boldsymbol{U}_t = m(\boldsymbol{\theta}, \boldsymbol{X}_t) \quad (2.2)$$

where $\mathbf{U}_t$ is the vector of all release decisions to be made at time t, $\mathbf{X}_t$ is the vector of relevant state variables (such as storages, reservoir inflows, etc.) at time t, and $\boldsymbol{\theta}$ is a set of parameters to be determined as part of the OP optimisation task. The OP optimisation problem can be described by

$$\mathrm{m}^* = arg \min_{\mathrm{m}} J \tag{2.3}$$

We classify OP optimisation methods into three categories below:

Release sequence based (RSB). The first step for these methods is to solve a RS optimisation problem and thus obtain an optimal RS (**U***) and associated optimal states (**X***). The OP (the function m and its parameters $\boldsymbol{\theta}$ in Equation (2.2)) is then derived as the result of a regression between the time series of state variables (**X***) and the optimal RS (**U***). In other words, the OP is a “generalization” of the optimal RS it originates from. It follows that a better OP is obtained when the optimal states (**X***) cover the state space as widely as possible. This in turn is more likely to be achieved if the RS is optimised over a long simulation period. Resultantly, in most cases HO will not be applicable for the RS optimisation step, while either MP or VFE will need to be employed, hence imposing constraints on the objective formulation and model structure, as discussed in Section 2.3.1. As for the second step, many sophisticated regression techniques have been demonstrated, most commonly artificial neural networks, fuzzy logic and decision trees [*Celeste and Billib*, 2009; *Celeste et al.*, 2009; *Kumar et al.*, 2012]. Since each regression algorithm has different benefits and drawbacks (discussed in referenced papers), it is unlikely that a single algorithm is preferable for all possible reservoir systems [*Labadie*, 2004].

A limitation of the RSB method is that it provides an OP that is only ‘optimal’ to the accuracy of the regression, i.e. it is actually sub-optimal even under the deterministic scenario used in the RS optimisation step. Furthermore, and possibly more importantly, the very RSB optimisation problem is somehow ill-posed. In fact, the ultimate goal of reservoir optimisation is to find the OP that minimizes the management objective, and not the distance from an "optimal trajectory" (**X***) that most likely will never occur (because it is based on a deterministic scenario of uncertain input forcing). Directly minimizing the objective function is precisely the key idea of the DPS method described in the next section. The RSB method thus appears to be an unnecessarily indirect way to achieve (in a sub-optimal way), what DPS can achieve more directly. The authors would note that, due to the efficacy of the DPS method, in recent years the RSB method has been seldom used and in the future is likely to be replaced by DPS.

Direct policy search (DPS). These methods aim to directly derive the OP by directly finding the parameterization $\boldsymbol{\theta}$ of a pre-selected function (m) that minimizes the objective under a deterministic time-series of forcing inputs, in a special case of equation (2.3):

$$\boldsymbol{\theta}^* = arg \min_{\boldsymbol{\theta}} J \qquad (2.4)$$

The DPS approach can be linked back to early works by *Maass et al.* [1962] and *Revelle et al.* [1969] on the Linear Decision Rule (LDR). In fact, the OP of a single-purpose, single-reservoir and single-demand node reservoir system can be expressed by the LDR:

$$m = s_t - \theta_t \qquad (2.5)$$

where θ_t is the LDR parameter, essentially the target storage for time-step *t*, to be optimised. If the objective and all constraints are linear, then the optimisation problem can be formulated as a linear program and the set of optimal parameters (one per time-step) obtained by MP. The LDR approach can be expanded to include a non-linear OP, i.e. a non-linear version of Equation (2.5) by using first order Taylor series approximation of non-linear objectives [*Shih and ReVelle*, 1994; 1995; *Pan et al.*, 2015].

HO algorithms can be used to extend the applicability of DPS to non-linear objectives and constraints, as well as any function form for the OP beyond the linear case. All the HO algorithms described previously (Section 2.3.1) are in principle suitable to solve Equation (2.4). Indeed both GA [*Oliveira and Loucks,* 1997; *Koutsoyiannis and Economou,* 2003; *Ahmed and Sarma*, 2005; *Chang et al.*, 2005; *Momtahen and Dariane*, 2007] and PSO [*Ostadrahimi et al.*, 2011] have been tested for this purpose. The additional benefits of HO algorithms such as no limits on using time non-separable objectives [*Giuliani et al.*, 2014], spontaneous multi-objective formulation and scalability to many-objective problems will be further discussed in Section 2.4.2.

As for the choice of the OP form, many options beyond the simple linear curve in Equation (2.5) have been proposed. The OP may be represented by, for example, a piecewise linear function (Oliveira and Loucks, 1997), as depicted in Figure 2.4b. Given that non-linear and piecewise constraints are handled by HO algorithms, it is possible to introduce variable policy structures within the same reservoir system, for example to operate some reservoirs based on their inflows and some others based on their storage levels [*Ashbolt et al.*, 2016]. Universal approximating functions can also be used, for example Artificial Neural Networks (ANN) [*Pianosi et al.*, 2011] or Radial Basis Functions (RBF), which according to *Giuliani et al.* [2015b] can outperform ANNs in many different aspects. In all these cases, the parameter vector $\boldsymbol{\theta}$ contains the weights and biases of the ANN or RBF. The advantage of such universal approximating functions is that they do not a priori constrain the OP to any specific structure, and that they scale efficiently with the number of input arguments of the approximating function [*Barron*, 1993], i.e. in our case the number of independent variables used as inputs to the OP. The drawback is that the resulting OP is a black-box that is difficult to interpret and therefore possibly more difficult to communicate to decision-makers. A possible solution to this problem is to optimise the ‘rule curves’ (for definition see OP disambiguation in Figure 2.3) in use by the current operators, for example *Chang et al.* [2005]. However, rule curves are limited by their flexibility and typically only depend on storage and time of year. An alternative is to create OPs that have highly flexible structures but are easy to visualize, such as in the form of a decision tree [*Herman and Giuliani*, 2018].

Expected value function estimation (EVFE). The Expected Value Function Estimation (EVFE) method extends the VFE approach discussed for RS optimisation to the case of OP optimisation, i.e. optimisation under uncertain forcing inputs. Typically, forcing inputs are regarded as stochastic variables described by probability distributions, and the OP is obtained by the minimization of the expected value of the value function. The solution algorithm, called Stochastic Dynamic Programming (SDP), follows similar steps as the discrete DP algorithm but with one more layer of discretization for the forcing input variables. The value function is thus evaluated against all possible combinations of the forcing inputs and the sample mean is used to approximate its expected value for each discretised state. Another possible approach, although much less common, is to describe forcing inputs by membership sets (rather than probability distributions) and search for the OP that minimizes the maximum possible value function [*Nardini et al.*, 1992].

The SDP algorithm has been widely used for reservoir operation (for reviews see *Yakowitz* [1982] or *Nandalal and Bogardi* [2007]). However, its applicability is subject to the same limitations as deterministic DP, i.e. the need for time-separable objectives, limited scalability to multi-objective problems, and the curse of dimensionality (as discussed in Section 2.3.1). The latter problem is even more severe here given that each state-decision combination must be evaluated against each combination of forcing input variables. To partially mitigate the computing burden, several variants of the SDP algorithm have been proposed, including the

Neuro-Dynamic Programming (NDP) algorithm [*Castelletti et al.*, 2007], which uses a neural network to interpolate the value function evaluations, allowing for a coarser state discretization grid.

Another limitation of the SDP approach is that each forcing input must be characterized by an independent probability density function, which might be an overly simplistic approach for input processes (e.g. inflows), which typically exhibit complex temporal and spatial structures [*Carrillo et al.*, 2011]. On the other hand, including temporal and spatial correlations among probability distributions would increase the number of variables required for discretization, up to a point that the problem becomes computationally intractable. An SDP variant that aims at overcoming the issue is Sampling SDP (SSDP), which uses a large number of sample inflow sequences in place of inflow probability distributions (see *Kelman et al.* [1990] for one of the earlier works on this, and *Stedinger et al.* [2013] for a more recent review). More recently, Reinforcement Learning (RL) algorithms have been demonstrated as viable options for EVFE optimisation using time-series of inflows rather than distributions (e.g. *Castelletti et al.* [2010], *Castelletti et al.* [2013] or *Dariane and Moradi* [2016]).

2.3.3 Real-Time Optimisation (RTO)

When forcing inputs of the reservoir system are uncertain but a forecasting system is in place, Real-Time Optimisation (RTO) is an interesting alternative to the OP approach. Differently from OP optimisation, where the optimisation task is concentrated in one effort, in RTO the optimisation is repeated each time an operational decision needs to be taken. This allows for exploiting forecasts (for example inflow or demand forecasts) that are available from a continuously updated forecasting system. The optimisation problem is typically formulated as an RS optimisation with forcing inputs set equal to their (deterministic) forecasts. Although the RS so obtained provides release decisions over the entire forecast horizon, only the decisions for the current time-step are implemented, and at the next time step the optimisation process is performed again with updated forecasts (Figure 2.4c). Long-term costs beyond the forecast horizon are accounted for by including a term that penalizes 'unfavourable' final states into the objective function, i.e.

$$\boldsymbol{U}^{*} = arg \min_{\boldsymbol{U}} \left[J_{[t,t+h-1]} + p(\boldsymbol{X}_{t+h}) \right] \tag{2.6}$$

where p represents the penalisation function, h is the length of the forecast horizon, **U** is the RS over the period [t, t+h-1], $J_{[t,t+h-1]}$ is the cost associated with implementing **U** over this forecasted period and $\mathbf{X}_{t+h}$ represents the system state at the end of the forecasted period. For example, for a supply reservoir the penalisation function would help finding a balance between maximising supply reliability over the forecast horizon and not leaving the storage depleted at the end of the period. The RTO problem in Equation (2.6) can be solved by any of the RS optimisation methods discussed in Section 2.3.1, provided they can accommodate any non-linearity associated with the penalization function. As a result, RS research in the context of RTO typically focuses on non-linear optimisation over short horizons, a case in which MP is mainly outperformed by the other methods.

A key issue in the application of RTO is the adequate definition of the penalisation function. Different approaches have been demonstrated, from using deviations from seasonal 'target storages' (as given, for example, by the reservoir's filling curves, e.g. *Ficchì et al.* [2015]), to linking the penalization function to the solution of an optimisation problem where 'off-line' forecasts (e.g. seasonal averages) are used in place of 'posterior' (real-time) forecasts (e.g. *Galelli et al.* [2014]). In principal, the value function from EVFE would be suitable, as demonstrated in *Pianosi and Soncini-Sessa* [2009]. However, as discussed, EVFE is not applicable to systems that contain many reservoirs or require many decisions to be made each time step.

As the deviations from the assumed deterministic forecast increase, the RTO release becomes less effective when applied in reality. Therefore, the benefits of using RTO are highly dependent on the quality of the real-time forecasts. If these forecasts are not significantly better than ‘off-line forecasts’ (e.g. seasonal averages) then using an OP will be equivalent to RTO (but at lower implementation costs, as the optimisation effort of an OP is done once and for all before the operation starts). This is the primary reason why RTO has only recently received significant attention, as a result of increasingly accurate forecasting systems [*Anghileri et al.*, 2016]. A proven way to increase RTO performance in the presence of inaccurate forecasts is by explicitly taking into account forecast uncertainty in the optimisation problem. This has been mainly implemented using two approaches. The first is to explicitly characterise forecast’s uncertainty by probability distributions and solve the resulting stochastic optimisation problem by EVFE. The initial illustration of the idea (although with an extremely simplified flow forecasting approach) dates back to *Bras et al.* [1983] (more recent applications include *Pianosi and Soncini-Sessa* [2009] and *Zhao et al.* [2011]). The second is to optimise an RS against an ensemble of inflow forecasts [*Zhao et al.*, 2011; *Raso et al.*, 2014; *Ficchì et al.*, 2015]. Interestingly, all these authors have found that including forecast uncertainty consistently outperforms any single deterministic ‘worst-case’ or ‘most likely forecast’ RTO approach.

Finally, an interesting question for RTO is the impact of the forecast horizon length (or ‘lead-time’) on RTO performance. For example, *Zhao et al.* [2012] investigated how forecast horizon length and forecast uncertainty trade off against each other, aiming to find the ‘effective forecast horizon’ for which the forecast provides the most valuable information for operators. If the forecast horizon is short, the optimised decision is highly sensitive to the horizon length; as the horizon length increases, the decision becomes increasingly sensitive to forecast uncertainty. Interestingly, in *Zhao et al.* [2012] the inclusion of ensemble forecasts improved performance but had no effect on determining the effective forecast horizon. Seasonal forecasts (between a month and a year) with some skill are becoming widely available for water resources operators, although the value of seasonal forecasts to improve operation by RTO has proved limited thus far [*Celeste et al.*, 2008; *Anghileri et al.*, 2016]. We would expect this to become an increasingly active area of research as the skill of these forecasts improves (or perhaps as characterization of their uncertainties becomes more accurate).

2.4 Comparison and choice of reservoir optimisation methods

In the previous sections we have briefly reviewed operation optimisation methods individually. In this section we will discuss some concepts and properties that are relevant across methods and can be useful for the comparison and choice of the most adequate method for the problem at hand. We start by discussing how OP optimisation methods could be further classified based on the approach they use to handle uncertainty in forcing inputs (typically reservoir inflows but possibly also other input variables/parameters like water demand or energy price), i.e. 'implicit' or 'explicit'. Such distinction is useful both in mathematical and in practical terms. Following this discussion, we debate the concept of 'optimality' within a practice-oriented research field such as reservoir operation optimisation and compare the extent to which different methods can be regarded as 'optimal' given the uncertainties that the reservoir modelling and optimisation process is subject to. Finally, we compare the ability of different methods to scale from single-objective to multi-objective problems, which, as anticipated in Section 2.2, is a very important feature in the context of reservoir operation optimisation. These topics and the advantages and disadvantages anticipated in Sections 2.3 are then brought together into a set of practical guidelines towards appropriate method selection.

2.4.1 Implicit versus explicit treatment of forcing inputs variability

In the context of OP optimisation, a typical distinction is made between two 'classes' of optimisation methods based on the way they handle the variability in forcing inputs (the distinction is used, for example, in the review by *Labadie* [2004]). According to this distinction, RSB and DPS methods belong to the same class as they account for the variability in forcing inputs 'implicitly' by using a long and diverse time series (or multiple time-series) to force the simulation model (typically a time-series of historical observation or a synthetic time-series generated by a model). In contrast, the EVFE method represents variability of forcing inputs 'explicitly', because it uses probability distributions. Authors who have adopted the implicit-explicit divide seem to suggest that an explicitly approach is preferable because it is more rigorous. However, it should be noted that forcing input probability distributions are also subject to simplifying assumptions, such as simplification or omission of spatial and temporal correlations. Furthermore, probability distributions are estimated from historical data and therefore can also be affected by scarcity or poor quality of the data [*Koutsoyiannis*, 2000; *Chatfield*, 2013]. So, in our opinion the preference for explicit characterisation of uncertainty is often not strongly motivated, except for the simplest case of a single-input reservoir system where a complete characterization of inflow uncertainty by probability distribution is often possible. More generally, we would argue that classifying OP optimisation algorithms based on the implicit-explicit divide is mathematically elegant but much less salient from a user's perspective.

2.4.2 Scaling methods to multi-objective optimisation

As anticipated in Section 2.2, reservoir operation is typically a multi-objective optimisation problem and as such it creates a set of Pareto-optimal solutions, each realising a different trade-off between the multiple objectives (i.e. one point on the so-called Pareto front), instead of a single solution. Multi-objective optimisation problems can be approached through two distinct methods [*Cohon and Marks*, 1975]. *A priori* techniques find Pareto-optimal solutions one at a time by repeatedly solving single-objective optimisation problems with different combinations of the multiple objectives (where each combination implicitly defines a ‘prior’ weighting of the objectives, hence the term of ‘a priori’ techniques). *A posteriori* techniques, instead, obtain a complete set of Pareto-optimal solutions in a single optimisation run. The main drawback of a priori techniques is that the number of optimisations required to approximate the Pareto front (at given resolution) increases factorially with the number of objectives [*Reed and Kollat*, 2013]. Therefore a priori methods become quickly unsuitable with growing number of objectives, while a posteriori techniques scale much more efficiently.

In the context of the reservoir operation optimisation methods presented so far, we note that the algorithms for Mathematical Programming (MP) and Value Function Estimation (VFE) (and, by extension, for EVFE and for RSB using MP or VFE, see Figure 2.2) are inherently single-objective. *A priori* techniques are therefore the only available way to handle multi-objective problems if these approaches are used.

On the contrary, most Heuristic Optimisation (HO) algorithms, and in particular population-based algorithms such as GA and PSO, can equally handle single or multi-objective optimisation problems and therefore provide an entire set of Pareto solutions in a single optimisation run in the multi-objective case [*Sharif and Wardlaw*, 2000]. They thus constitute an a posteriori approach, and make HO (whether it is used to directly obtain a RS or within DPS) particularly efficient when the number of objectives is large [*Reed et al.*, 2013]. A review and comparison of many state-of-art population-based approaches for MO optimisation for DPS is provided by *Salazar et al.* [2016].

Finally, one exception to the distinction delineated above is the Reinforcement Learning (RL) algorithm introduced by *Castelletti et al.* [2013], which is the first (and to the authors' knowledge, only) *a posteriori* algorithm for multi-objective EVFE optimisation. This method is also applied and compared with DPS in *Dariane and Moradi* [2016], where it is found that DPS outperforms RL, although it is not possible to conclude whether the result would hold in a water system of lower dimensionality in which the limitations of EVFE are less prohibitive.

2.4.3 Optimality and modelling assumptions

The optimisation algorithms reviewed in the previous sections provide different degrees of confidence with regard to the optimality or sub-optimality of their solutions. For example, a correctly executed MP algorithm provides an optimal solution of the optimisation problem, VFE or EVFE provide approximately optimal solutions (i.e. accurate to the resolution of the interpolation), while HO algorithms give no guarantee of optimality and simply return the best solution that could be found in a given number of iterations. However, it is important to highlight that such 'optimality' statements are only valid within the given problem formulation. If the problem formulation is not 'correct', i.e. the underlying assumptions (for example, linear reservoir equations or a single, time-separable objective) provide an oversimplified representation of the system behaviour, then the 'optimal' solution will perform sub-optimally when applied in the real world. This is an important factor to be considered when selecting an optimisation method, as we will further discuss in the following section.

Interestingly, the 'optimal' solution subject to the above caveats may perform worse than a heuristic solution obtained with a more realistic (and hence more mathematically challenging) problem formulation [*Momtahen and Dariane*, 2007]. Therefore, a trade-off curve may exist between optimality and the degree of realism of the system model. A theoretical depiction of this trade-off curve is shown by the red line in Figure 2.5: point A represents a highly-simplified problem formulation that can be associated with an 'optimal' solution, which degrades significantly to A' when implemented in real life, while point B represents an extremely sophisticated representation, for example including a very detailed hydraulic model, that may be intractable by current optimisation algorithms and can therefore only be associated with a heuristic solution.

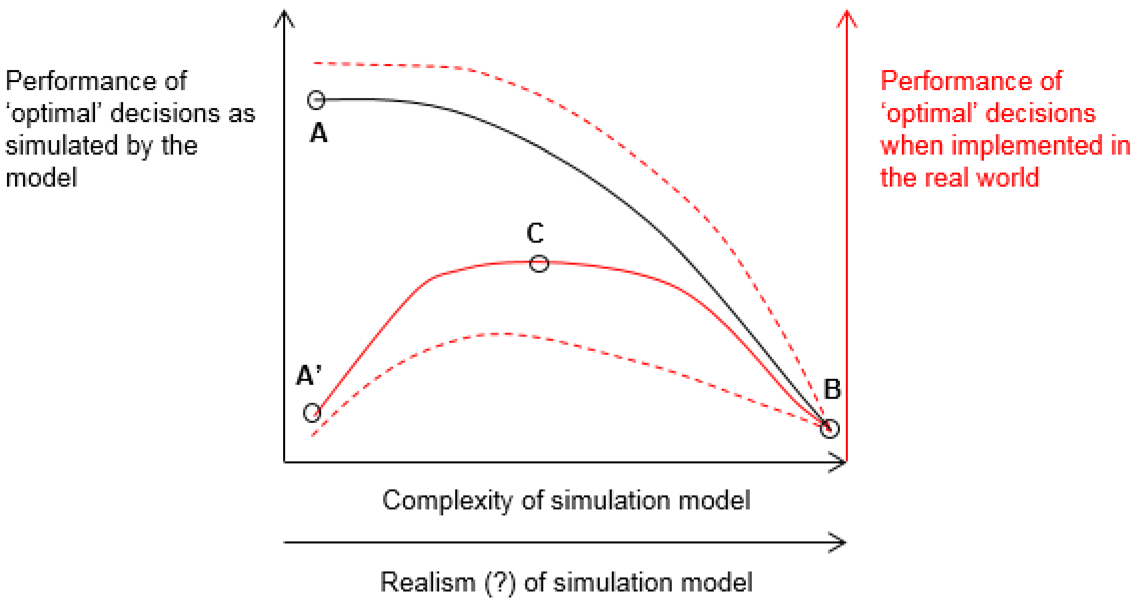


Figure 2.5: A theoretical representation of the trade-off between performance (to be maximized) simulated in the optimisation process (solid black curve) and the real performance when this solution is implemented in practice (solid red curve), for increasing complexity (and presumably realism) of the simulation model. We make the assumptions that more realistic models will be more complex and that, as the realism increases, uncertainty bands around the actual performance (dashed red curves) will reduce.

In the reviewed literature, it is common to find application case studies where the modelling of the reservoir system is over-simplified in order to make the optimisation problem more easily tractable, or even just to facilitate the explanation of a newly proposed optimisation method. In other words, most literature focuses on advancing and demonstrating methods lying on the black line in Figure 2.5, a practice that does not foster faith about the applicability and performance of reservoir optimisation methods in the real world (the red line). Although we believe that a trade-off (point C in Figure 2.5) should exist, we expect that this trade-off cannot be found while ignoring the 'red line'. Since any model is a simplification of reality, one can never precisely identify this line, rather the uncertainty around it must be characterized (i.e. the dashed lines in Figure 2.5). *Herman et al.* [2014] provide a framework to do this for a single point on Figure 2.5's X-axis, by evaluating different solutions under a range of plausible scenarios rather than simply the 'most likely'. We believe that expanding this approach by introducing plausible simulations into the optimisation process could be a way towards further exploring the X-axis in Figure 2.5 and the trade-off curve. In Chapter 4 we build on these ideas and discuss the concept of uncertainty in reservoir operation models.

2.4.4 Practical guidelines towards selecting reservoir operation optimisation methods

The literature review presented in the previous sections was aimed to provide practical information about the advantages and limitations of reservoir operation optimisation methods. Another contribution of this review is to identify a set of guidelines for the selection of the most appropriate optimisation method for a given reservoir system. Our advice is summarised in the comparison table presented in Figure 2.6. The table can be used to narrow down the number of suitable methods (horizontal axis) for a reservoir system of given characteristics (vertical axis).

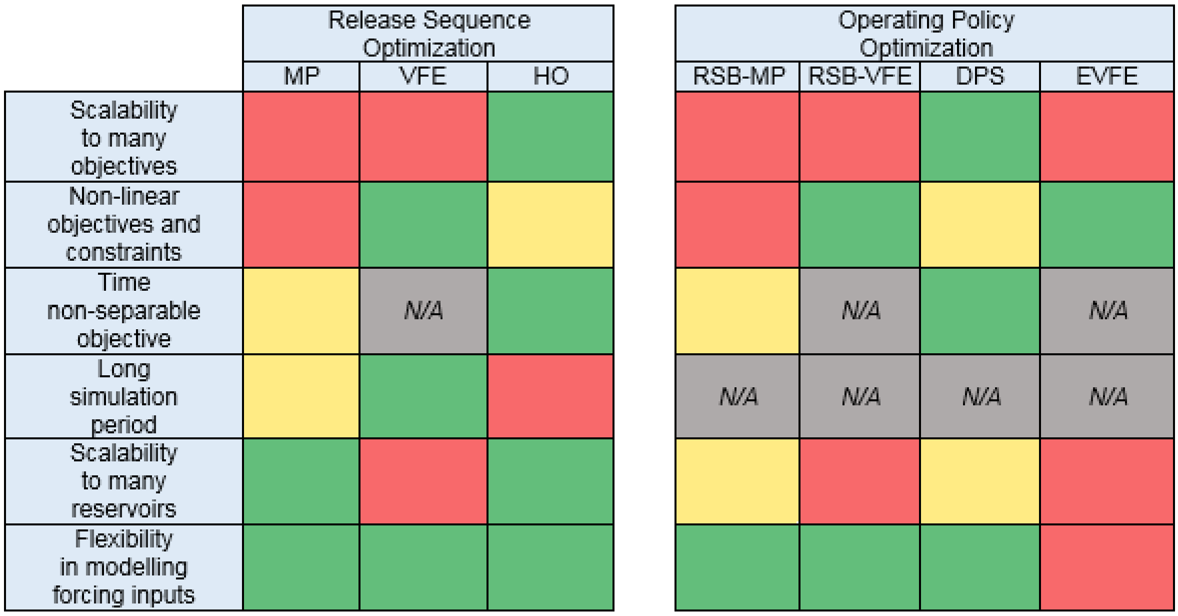


Figure 2.6: A summary of the suitability of different reservoir optimisation methods to different characteristics of the reservoir system and decision-making problem. Green indicates the highest suitability, yellow indicates medium suitability and red lowest suitability (based on the authors' review of the literature). N/A stands for "Not Applicable", other acronyms as in Figure 2.2. RSB-MP and RSB-VFE refer to Release Sequence Based optimisation of an OP where Mathematical Programming or Value Function Estimation are used in the RS optimisation step. RTO is not included because it can use methods from both release sequence optimisation and operating policy optimisation and so the same considerations apply.

For RS optimisation, the choice is least obvious and highly dependent on the system characteristics. The greatest advantage of MP is its speed, scalability and the guarantee of analytical optimality (albeit under the caveat that the problem must be simplified to fit the required assumptions on objectives and constraints). VFE can solve non-linear (and thus more realistic) optimisation problems over long time periods (of the order of T=10,000 time-steps, e.g. decades if the time-step is daily or centuries if it is weekly) but is limited to small reservoir networks (up to 3-4 reservoirs at current computing power) and the solution's optimality is subject to the interpolation accuracy of the value function. HO is a more flexible method that can handle multiple objectives efficiently and allows for time non-separable objectives, but it is limited to short lengths of the simulation-optimisation period.

For OP optimisation, we suggest that DPS is the most widely applicable method, even if it cannot provide any assessment of the optimality or accuracy of the solution. Still, we do not think this is a major issue for practical purposes, given the difficulty in evaluating whether simplifying assumptions required by other methods are satisfied for the problem at hand, and to what extent. On the other hand, EVFE may still be preferred in those situations where it is computationally feasible (i.e. relatively small reservoir networks) and when one can reasonably presume that its underlying assumptions (in particular, time-separability of the objectives) are acceptable.

For Real-Time optimisation, no specific method has been clearly established yet (although VFE/EVFE has possibly been employed more frequently in the literature) but we would expect that more research will be carried out in this context given the increasing availability and advances in real-time monitoring and forecasting systems.

2.5 Conclusion

In this Chapter, we have reviewed the ever-growing body of literature in the field of reservoir operation optimisation, based on a novel classification system that uses the argument of the optimisation problem as the main criterion to classify methods. Our classification system shows that while the use of different arguments leads to substantially different problem formulations and types of solution (an optimal release sequence versus an optimal operating policy), the algorithms used for solving the optimisation problem are to some extent interchangeable. We hope this way of introducing the literature enables to shift the focus from the mathematical properties of solution algorithms, which we expect to be less accessible to users, to the more obvious and tangible properties of the reservoir operation problem. A comparison between different types of optimisation algorithm and some guidance as to what types of system they are more likely to be applicable has been provided to further improve the accessibility of this literature.

CHAPTER 3: ARE RESERVOIR OPERATION OPTIMISATION METHODS USED IN PRACTICE? RESPONSES OF A PRACTITIONER SURVEY

3.1 Introduction

Extensive scientific literature exists on the study of how reservoir operation could be designed or improved using mathematical optimisation methods. Figure 1.2 shows that the field has been increasingly active in recent years and in Chapter 2 we have shown the variety of approaches that exist. Despite such increasing research effort, the state of uptake of these optimisation methods outside academia has been little investigated. *Rogers and Fiering* [1986] gathered evidence in the form of interviews from representatives in four federal agencies in charge of dam operation in the US. They found that, at the time, the uptake of optimisation methods in those agencies was negligible. They attributed this in part to a lack of trust in the validity of results due to the uncertainty in the underlying simulation models, and in part due to optimisation producing very different optimal operating policies that appear to perform equally well. They thus suggested the field should focus more on sensitivity and uncertainty analysis rather than optimisation. Several other studies spanning the last 30 years [*Yeh*, 1985; *Simonovic*, 1992; *Labadie*, 2004; *Brown et al.*, 2015] have also confirmed that the uptake remains limited and proposed reasons for this. These reasons include: the lack of involvement of practitioners in the development of reservoir simulation and optimisation models, a lack of suitable data, the focus of researchers on over-simplified reservoir systems, the existence of institutional constraints that prevent innovation in water resource management practice, and the lack of accessible, credible and user-friendly software that implements reservoir simulation and optimisation methods. To the authors' knowledge, however, since *Rogers and Fiering*

[1986] no other survey of practitioners has been attempted on the topic, while the number of scientific publications on the topic of reservoir operation optimisation has kept increasing.

We thus present and discuss a survey-based study that we carried out on a range of water supply decision-makers and experts. The survey covers four countries: UK, South Africa, South Korea and Australia. We have interviewed practitioners at water companies (UK and South Korea) and also consultants working for these companies (Australia, South Africa). We have also included a UK consultant to determine whether differences in responses are geographically determined or due to the differences between consultants and supply companies.

We started our survey by focusing on the UK water industry, which we expect to provide useful insights regarding the perceived gap between scientists and practitioners, since we assume that the UK water industry should be in principle quite open to innovation. A range of factors contribute to this assumption: water utilities are private companies, which should push them to seek for efficiency gains; yet the industry is also partially regulated and companies are explicitly obliged to seek low cost solutions for the sake of their customers; companies have a certain degree of flexibility in the operation of their reservoirs as over 60% of water treatment works can be supplied by multiple sources, so they should particularly benefit from the use of formal and quantitative approaches to decision-making; and finally, technical staff of water companies are typically highly-skilled. Additionally, high population densities, particularly in the South-East, makes the UK a relatively water stressed country, having the 63rd smallest renewable water resources per capita worldwide [*FAO*, 2008].

We then expanded our survey by interviewing practitioners from South Africa, South Korea and Australia. We chose these countries because they all experience some degree of water stress while facing a range of different operational problems. South Africa is an arid country that faces a large amount of 'unaccounted for water' (water that has been illegitimately removed from the water network), and a small amount of financial resources to leverage (compared to the other surveyed countries). South Korea has a very high population density and a strong imbalance between rainy and dry seasons, with 85% of the rainfall occurring in six months. Australia is the driest continent on Earth and yet has a very high water consumption, mainly due to water intensive industries such as agriculture and mining.

We chose to focus our survey on water supply reservoirs with the aim of gaining comparable responses and because water supply is the target of much of the operation optimisation literature. Our conclusions may thus not be as applicable to reservoirs used mainly for other purposes such as hydropower reservoirs, where the uptake of optimisation methods is often presumed to be higher [*Brown et al.*, 2015] (even if to our knowledge there is no published evidence to support this presumption). We would anecdotally note that, in scoping the out which sector our survey should be focused on, we also reached two hydropower companies in the UK and they said that they do not use reservoir operation optimisation methods because of the financial uncertainty in electricity prices.

3.2 Survey methodology

Before carrying out the survey it was important to define a coherent target group and determine a set of questions and terminology appropriate for that group. Therefore, we first performed pilot interviews with water resource managers in 2 UK water supply companies. Pilot interviews were conducted via telephone and lasted around 40 minutes each. They were loosely guided by some core questions, but mainly consisted of a free discussion aimed at scoping the company's operational procedures and difficulties, and understanding the terminology in use. We found that the UK water supply industry is quite homogeneous in that water companies are mostly of similar size and are subject to the same set of rather stringent regulations. Thanks to this homogeneity, we found it possible to design a questionnaire that could be meaningfully answered by all the targeted water companies. The addition of the remaining countries provides a useful set of case studies for a number of reasons, including: (i) a variety of hydrological conditions, ranging from relatively water rich to near constant drought; (ii) a variety of institutional frameworks, with a fully privatised industry, a fully public industry and publicly owned assets that take most operational decisions guided by private consultants; and (iii) a large potential for increasing coordination of water supply sources since many systems involve multiple connected reservoirs as well as other surface and groundwater sources.

The questionnaire was created using the methodology set out in *Scheuren* [2004], i.e. select questions only relevant to the objective of the investigation, pre-test to check the clarity of questions (which we performed on the pilot study companies) and add explanations where terminology might be ambiguous (for example, in the definition of a rule curve). We selected the format of self-administered questionnaire via the internet to enable recipients easy response in an environment without time pressure and to avoid introducing 'interviewer effects' into the results (i.e. subconsciously guiding the interviewee towards certain responses) [*Opdenakker*, 2006]. Given that the relatively low number of water supply companies (10 water companies were approached representing 94% of UK reservoir storage capacity used for water supply) would have not allowed a statistical analysis of the responses, we allowed respondents to both select from multiple answers for each question or write their own answer, so to maximise the amount of information gained through the questionnaire.

The questionnaire was split into 4 sections:

- Decision-making and rule curve procedures in normal conditions. We asked 2 questions to characterise the decision-making procedures followed in 'normal operation' conditions. Although the definition of 'normal conditions' (as opposed to 'extreme conditions') varies from company to company, all companies are required by UK regulation to define a set of triggers (such as exceeding reservoir storage thresholds) to discriminate between the two situations. Since pilot interviews suggested that the decision-making process in extreme conditions is notably different than in normal conditions, we prepared two different sets of questions (this and the following section) for each situation.

- Decision-making procedures in extreme conditions.
- Use of software. We asked 4 questions about the use of software to assist operation, and in particular software for water resources system simulation, which, based on our pilot interviews, is the type of software mainly in use.
- Future challenges. We also wanted to find out what respondents considered to be the largest challenge they expect their water supply industry to face in the next 10 years.

Due to the UK specific questions and language that featured in the questionnaire, we administered the survey as a structured interview, [*Turner and Jeffrey*, 2015], via telephone or Skype to non-UK participants. The questions were read out and the responses recorded. While this is not ideal due to the previously mentioned 'interviewer effects', it was necessary to avoid confusion over terminology (which is often UK specific). We found that one benefit of the structured interview format was that respondents could explain the reasoning behind their selection – this added depth provided us with the helpful quotes shown throughout Section 3.4 that helped guide our discussion (it is much better to use the words of respondents to explain the results).

3.3 Survey results

The questionnaire was sent to the 10 largest UK water companies and we received responses from 7 (representing 67% of the national storage capacity used for water supply). The structured interview was administered to 1 UK consulting company, 2 Australian consulting companies (representing around 30% of capacity used for water supply), 1 South Korean water supply company (representing 65% of capacity used for water supply) and 3 South African consulting companies (representing around 60% of capacity used for water supply). The answers are summarised in Figure 3.1 below. Similar types of surveys have been able to attract similar sample sizes of water managers, for example in a study about the perception of uncertainty in water resources systems *Höllermann and Evers* [2017] received responses from 12, while *Whateley et al.* [2015] received responses from 8 in a study on the uptake of seasonal forecasts.

Decision-making and rule curve procedure in normal conditions

Q1A: How are abstraction/release/transfer decisions made?

	A	SA	SK	U1	U2	∑
Expertise/Calculation/Experience	2	1		7	1	11
Software simulation and iteration	2	3	1	4	1	11
Rule curves (informally used)	1	2		6	1	10
Rule curves (formally used)	1	2	1	1		5
Real-time optimization				1		1

Q1B: What are your reservations with this decision-making process?

	A	SA	SK	U1	U2	∑
Makes knowledge transfer difficult		1		2	1	4
Overly risky	1	1		1		3
Overly conservative		1		1		2
Consumes too much resources				1		1
Lacks transparency						0

Decision-making procedure in extreme conditions

Q2A: How are abstraction/release/ transfer decisions made in extreme conditions?

	A	SA	SK	U1	U2	∑
Expertise/Calculation/Experience	2	1	1	5	1	10
Involvement of additional staff	2	3		4	1	10
Rule curves		1	1	6		8
Software simulation and iteration	2	3	1	2		8
Drought plan		1	1	5		7
Real-time optimization						0

Q2B: What informs the trigger storage levels that define extreme conditions?

	A	SA	SK	U1	U2	∑
Software simulation and iteration	2	3	1	5	1	12
Expertise/Calculation/Experience	2	1		5	1	9
Optimization of software simulation	1			3		4

Use of software to aid operation

Q3A What is the aim of optimization software?

	A	SA	SK	U1	U2	∑
Source-supply allocation	1	2	1	6		10
Rule curve creation	2	2	1	4	1	10
Decision making during droughts		1	1	5		7
Planning for extreme conditions		1	1	4		6
Identifying trade-offs		2	1	1		4

Q3C Key features in optimization software for you

	A	SA	SK	U1	U2	∑
Effective visualization functions	1	1		5	1	8
Access to source code	2	1		1		4
Ability to interact with other software				3		3
Cost		1		2		3
Availability of a GUI				2	1	3

Q3B What are the limitations of optimization software?

	A	SA	SK	U1	U2	∑
Oversimplified physical processes	1	1		3		5
Fit for purpose		1		2		3
Difficult to use		1		1		2
Too slow	1				1	2
N/A - Not in use				1		1

Q3D: Do you use optimization software and if not why?

	A	SA	SK	U1	U2	∑
In use already or under development	2		1	5		8
Not in use because:						
lack of trained staff able to use it			1		1	2
too inflexible problem formulation		1			1	2
unnecessary - unstressed system				1		1
unnecessary - simulation is cheap		1				1
sub optimality in provided solutions						0

Future challenges

Q4: What do you expect to be the biggest challenge to meeting your water supply over the next 10 years?

	A	SA	SK	U1	U2	∑
Climate and hydrological change	1			2		3
Insufficient or legacy assets			1	2		3
Increasing limits to abstraction				2	1	3
Lack of water sources to exploit	1	1				2
Institutional instability		2				2
Lack of experts with essential skills		1			1	2
Simultaneous unexpected failures				1		1

Figure 3.1: A selection of the results from the survey. 14 companies responded but were allowed to select multiple responses to questions for all Questions except 5e. Results are presented in columns footed by: A - Australia, SA - South Africa, SK - South Korea, U1 - UK water companies, U2 - UK water consultants, ∑ - total across all responses.

Starting from Question 1a, we see that in normal conditions the decision-making process relies heavily on expertise/calculation/experience, informally applied rule curves and on software simulation. It relies on formal application of rule curves to some extent and does not rely on real-time optimisation at all (we should note that the one respondent who answered that they use real-time optimisation later gave an answer to Question 3a such that we presume they incorporate real-time information informally in their decision-making Chapter 2.3.3 process rather than by means of real time optimisation methods such as those described in). Responses to Questions 2a and 2b suggest that the same holds true also during extreme conditions. We do see a noticeable difference between the responses by practitioners at water companies and by consultants, in fact consultants appear to lean more often towards software simulation and iteration. From the answers to Question 1b we also see that 10 out of 13 respondents have reservations about the current decision-making process – albeit for quite diverse reasons.

Interestingly, when asked to define the purpose of operation optimisation tools (Question 3a) the answer least frequently selected by practitioners at UK water companies was "creation of rule curves", which is the closest to how the scientific community assumes optimisation results would be implemented in practice. Instead, most respondents defined operation optimisation tools as "source-supply allocation", i.e. tools that provide the optimal allocation of water fluxes across the network nodes for a given time step (while reservoir optimisation literature typically focuses on optimal allocation over time). A possible reason for this emphasis on spatial optimisation, instead of temporal optimisation, is that the software simulation tools currently in use in the UK industry include basic source-supply solvers. In contrast, consultants selected rule curve creation most frequently, followed by source-supply allocation solvers. Although it may be surprising to see that "identifying trade-offs between users" was the least selected option, we would note that we have focused our survey on managers of water supply reservoirs, which may explain the lower interest in trade-off analysis.

We also asked whether they use optimisation software and what are the perceived limitations (Question 3b/d). Again, we would note the large amount water supply companies who suggested that optimisation software is in use are likely referring to their operation simulation software – based on the compulsory and publicly available water resource management plans, no UK water supply company uses operation optimisation beyond the linear programming source to supply solvers incorporated into the simulation software. For several respondents the biggest concern with their software tools is the lack of adequate representation of physical processes (Question 3b). Contrastingly only two respondents voiced concerns over calculation speed, and none were concerned about solution optimality (Question 3b), which are two typical focuses of reservoir optimisation research, as we will further explain in our review of the scientific literature (although we note that none of the software currently in use implements the more time-consuming algorithms discussed in the review). A slight difference between consultants and supply companies arises here insofar that consultants appear to be more critical of optimisation software; respondents from water companies were far less likely to list problems in Questions 3b/d, this difference is presumably due to the greater exposure of consultants to simulation software, thus increasing the likelihood of criticisms.

3.3 Discussion

The overall picture painted by the survey results reveals a substantial difference in the approach to reservoir operation of practitioners and researchers. In the literature described in Chapter 2, reservoir operation is usually defined as a very structured, necessarily automated, process. In contrast, practice frames operation as an informal decision-making problem drawing on expertise, existing rule curve information and, increasingly, use of simulation software for addressing "what-if" questions. Indeed, the question of "what optimisation algorithms are used?" (the question we originally aimed to ask) had essentially no meaning to most respondents. They either told us it didn't matter what algorithm was used, or were not clear about what decision variables would actually be optimised and how the optimisation problem would be formulated. Besides the evidence presented in the results section, we can also support this with some quotes from our interviews to practitioners. For example, a (South African) consultant who worked on creating rule curves for reservoir operators, said:

"We find that the rule curves we produce are either followed rigidly or not at all; we would prefer that they are incorporated with a wider understanding of the water resources system in question"

The fact that a consultant who creates rule curves wants them to be followed informally suggests that this informal decision-making process is no accident.

For the reader's own interest, we take this opportunity to mention that, of our entire survey, only a single (Australian) respondent applied optimisation in a manner that approached that later described in the literature. They used a genetic algorithm to optimise rule curves in their simulation model. These optimised rule curves were used to represent the water company (their client) so that they could simulate operation under out-of-record inflow scenarios. It is important to note that it was the results from these Monte Carlo simulations (in the form of an assessment of the system's sensitivity to droughts), and not the optimised rule curves that were provided to their client, again reinforcing that, even in situations where optimisation is used, the decision-making process in practice still appears to be informal. The mixture of responses to question 3a highlights that there is uncertainty around the actual aim of operation optimisation. Indeed, the fact that the only instance of optimisation being used was not actually aimed at directly improving decisions (as is most commonly the case in the literature) highlights that the purpose of operation optimisation tools among practice is not necessarily clear.

However, the survey seems to reveal space for increased uptake of optimisation methods. For example, knowledge transfer within the company was most frequently selected as a significant issue in Question 1b, which could be connected to the high reliance of the decision-making process on personal experience and could be addressed through the adoption of a more structured approach. Operation optimisation could also help to address other problems with the decision-making process (e.g. risk vs conservatism, Question 1b) and some of the expected biggest challenges for the future (e.g. climate change, Question 4) since many of these are commonly studied in research. Based on how many of these problems can be addressed by optimisation tools, and based on the generally positive responses on the value of existing simulation software, we conclude that the lack of uptake of operation optimisation methods does not appear to originate from an actual (or perceived) lack of value of those tools for practitioners' needs. It stems rather from a lack of awareness of their availability and value, again reinforcing our conclusion about the lack of awareness on the purpose of operation optimisation.

Despite this, even if clarity can be provided, which our literature review aimed to do, there is still the disconnect of formal and informal decision-making between academia and practice. We may look to the responses to Question 3b for an explanation of this hesitance to rely on automated decision-making; any model is limited in its ability to capture the true complexity of water resources systems. These concerns around the realism of the modelling will naturally encourage optimisation results not to be taken at face value considering the scepticism around the simulation models that were used in their creation. This result corroborates the result presented in *Rogers and Fiering* [1986] suggesting that practitioners had reservations about optimisation that stem from uncertainty in the simulation models. We support this suggestion with two quotes, from a South African and Australian consultant:

> *"The human elements of our system are so enormously complex that anything as formal as optimisation is unlikely to be of benefit"*

> *"Optimised results are inherently optimistic due to the assumption that the system is working perfectly; this results in decisions that are overly risky"*

Thus, as our results match *Rogers and Fiering* [1986], so must our recommendations – for a greater focus on uncertainty analysis in the simulation models that underpin optimisation.

3.4 Conclusion

We have presented the results of a survey of UK and global water supply companies aimed at assessing current practitioners' decision-making procedures and awareness of reservoir operation optimisation methods as well as barriers to their uptake in practice. We found that such awareness among the interviewed operators and consultants is very limited and that reservoir management decisions are still largely based on expert judgement with little assistance from formal methods. However, most of the difficulties in current practice and future challenges that reservoir operators identify could benefit from a more consistent application of optimisation methods, which suggests that the potential for increasing the uptake is large. We would urge researchers to be explicit about how the optimisation methods they develop should be used in practice and give due consideration to the uncertainties in the simulation models that underpin their results.

Based on the analysis of the scientific literature in Chapter 2 and our practitioner survey in this chapter, we identify two particularly promising areas for future research aimed at improving the uptake of methods by practice. The first is understanding and addressing the impact of informal decision-making by practitioners on the solutions derived from automated decision-making optimisation processes, based on the disparity between the decision-making process in practice and that assumed in research. The second area is demonstrating the real-world effectiveness of the outputs from optimisation problems, based the physical realism concerns of practitioners. We see this as key to address uptake; any other technical gains are meaningless if a decision maker does not believe that the benefits are transferrable to the real world. We hope to make progress in this by the uncertainty and robustness analyses provided in the following chapters.

CHAPTER 4: HOW IMPORTANT ARE MODEL STRUCTURAL AND CONTEXTUAL UNCERTAINTIES WHEN ESTIMATING THE OPTIMISED PERFORMANCE OF RESERVOIR SYSTEMS?[2]

4.1 Introduction

Models and model-based optimisation are widely used to support operations in water management. Within the broad area of water management, this study will focus on the optimisation of reservoir operations [*Yakowitz*, 1982; *Yeh*, 1985; *Hiew et al.*, 1989; *Labadie*, 2004; *Rani and Moreira*, 2009]. Reservoir operation optimisation typically refers to identifying the operational decisions (for example, operating policies that make reservoir release decisions based on such system conditions as reservoir storage or time of year) that achieve optimal values of certain objectives (for example, reliability of water supply), as discussed in greater detail in Chapter 2. Objectives are evaluated using a numerical model that simulates the interaction between decisions and forcing inputs (for example, demands and reservoir inflows) over time. The optimisation process consists of the (usually) iterative improvement of objectives achieved by altering the operational policies. Optimisation can be particularly beneficial in systems of interconnected reservoirs, where even a relatively small increase in system complexity can make the definition of effective operating policies far from trivial [*Moss et al.*, 2016]. Another difficulty in making operational decisions is the need for balancing multiple conflicting objectives, which 30% of large dams face worldwide [*ICoLD*, 2003].

[2] This Chapter has been accepted as: "How important are model structural and contextual uncertainties when estimating the optimized performance of water resource systems?" Water Resources Research (2019).

When multiple objectives exist, the aim of optimisation is not to find a single optimal operating policy, but rather to characterise the possible trade-offs within a set of candidate policies [*Cohon and Marks*, 1975; *Haimes and Hall*, 1977; *Guariso et al.*, 1986; *Kasprzyk et al.*, 2013; *Reed and Kollat*, 2013].

Although reservoir operation optimisation methods have been extensively studied, little is known about the uptake of these methods by water management practitioners. The first attempt to survey the uptake of these tools [*Rogers and Fiering*, 1986] revealed that the uncertainty present in the underpinning simulation models contributed to a significant lack of trust in the end results of the optimisation process. A second, more recent, survey presented in Chapter 3 corroborated this scepticism and suggested that practitioners tend to prefer using simulation models via 'what-if' analyses rather than formal optimisation tools. In a survey on the perception of uncertainty by water managers, *Höllermann and Evers* [2017] found that the uncertainty around boundary conditions, which is an example of what we will later define as contextual uncertainty, was the most commonly listed source of uncertainty for practitioners. In climate change impacts studies, *Mahmoud et al.* [2009] found that stakeholders did not trust the study results if they were not convinced by the system conceptualization underlying the simulation models used. Because of these more recent studies, we believe that *Rogers and Fiering* [1986]'s finding may still be valid today; that practitioners are sceptical about results from optimisation due to the model uncertainties that unavoidably affect the simulation model used during optimisation. In order to build trust that optimisation results will remain valid when applied in reality, we believe that it is essential to understand whether an operational solution will meet its estimated level of performance in the face of these uncertainties, and contextual uncertainty in particular.

A common conceptual classification of uncertainties affecting simulation models distinguishes between *aleatory* and *epistemic* uncertainty [*Walker et al.*, 2003; *Beven et al.*, 2017]. Aleatory uncertainty arises from intrinsic random variability in the system, such as variability in weather conditions. It is typically considered irreducible but can be characterised statistically. Epistemic uncertainty instead can be defined as the uncertainty that is attributable to a lack of historical observations [*Beven et al.*, 2017] which results in a lack of understanding about the system, its properties and its expected behaviour [*Walker et al.*, 2003]. Examples are the uncertainty in the projected magnitude of a flood event with return period exceeding the length of historical time series, or the uncertainty in the subsurface properties of a catchment, which are typically not observable. Epistemic uncertainty is in principle reducible, even if this is difficult to do in practice. Below we discuss how aleatory and epistemic uncertainties affect WRS simulation models, and we review the techniques that have been used to address them within optimisation studies.

Practically unavoidable in WRSs is the variability in hydrological forcing, such as inflows into reservoirs, which was the main focus of the earliest water management studies [*Maass et al.*, 1962]. A common practice in the field is to assume that inflows are aleatory and stationary processes, although the validity of the stationarity assumption is highly debated [*Milly et al.*, 2008; *Montanari and Koutsoyiannis*, 2014] and represent them by fitting a statistical model to the historic data (see for example *Matalas* [1967] for an early application and *Vogel* [2017] for a recent review of the available techniques). Reservoir operation is then stochastically optimised under this statistical model, for example via Stochastic Dynamic Programming (e.g. *Stedinger et al.* [1984], *Nardini et al.* [1992] or *Castelletti et al.* [2012a]) or by generating a synthetic sequence of forcing for which the operations are deterministically optimised [*Koutsoyiannis and Economou*, 2003]. The more densely sampled the statistical model (either by using a high-resolution discretization grid for Stochastic Dynamic Programming or by generating a long sequence of synthetic forcing), the longer the optimisation will take. Hence, it is good practice to keep the sample size limited during the optimisation process and then re-evaluate the optimised operations over an expanded sample, so to ‘validate’ their estimated performance [*Kasprzyk et al.*, 2013]. Similar considerations apply to other system variables that can be regarded as aleatory uncertainties, such as water demand and evaporation from reservoir surfaces, which are often modelled using similar statistical models to inflows [*Donkor et al.*, 2014].

As for epistemic uncertainties, we distinguish four types: parametric, objective, structural and contextual. Parameters are constant values in a model, typically identified through measurement or calibration [*Walker et al.*, 2003]. The measurement and calibration processes not being exact, it results in a certain amount of uncertainty in parameter values, which is in principle reducible by further measurement and testing. However, in WRS simulation and particularly for long-term evaluation of WRS performance, it is typically necessary to use *conceptual* parameters that do not relate to specific physical quantities but instead encapsulate and simplify complex and diverse phenomena. Examples are trend parameters that summarise long-term changes in water demand or in inflow statistics as a consequence of climate change. In recent years, several studies have investigated the robustness of solutions to parametric epistemic uncertainties, either by including sampling of the uncertain parameter space in the re-evaluation of optimised operating policies [*Kasprzyk et al.*, 2012; *Herman et al.*, 2014] or by directly incorporating the sampling of epistemically uncertain parameters into the optimisation process [*Trindade et al.*, 2017; *Watson and Kasprzyk*, 2017].

Another source of epistemic uncertainty is the choice and formulation of *model outcomes* [*Walker et al.*, 2003] such as, in the case of WRS optimisation, the metrics of system performance (or *objectives* hereafter). It is well known that similar formulations (for example, vulnerability vs reliability) of the same objective (for example, water supply) may yield different suggested operations [*Hashimoto et al.*, 1982]. A further difficulty is that the decision-makers themselves may not be aware of their true preferences until they are able to visualise operating policies and their respective objectives in the context of the trade-offs available to them [*Kasprzyk et al.*, 2013]. Unlike parametric uncertainty, which can typically be characterised by random sampling within a defined range of plausible parameter values, the geometry of the objective space is not so clearly defined and thus cannot be characterised in the same way. Recently, *Quinn et al.* [2017] presented a method to investigate the effects of competing formulations of uncertain objectives on multi-objective optimisation results. This method creates different framings of the WRS management problem using different objective formulations – it considers each framing of the system as a unique simulation model and a single sample in the space of uncertain objectives. By application to a hydropower reservoir system in Vietnam, *Quinn et al.* [2017] found that the choice of objective has a significant impact on how effective an operating policy would be considered. *Watson and Kasprzyk* [2017] provide an approach that samples epistemically uncertain parameter spaces and different objective formulations to incorporate both objective and parametric uncertainty into the search process.

Another source of uncertainty, and a key focus of this study, is model structural uncertainty. We use the definition of *Walker et al.* [2003], who suggests that model structural uncertainty is uncertainty about *"the behaviour of the system and the interrelationships among its elements"*. Examples in WRS management might include the type of statistical model used to describe aleatory variables or the omission of processes that are poorly understood or unsupported by data, such as pump failures. The effects of structural uncertainty on the prediction of environmental or socio-economic variables has been relatively well studied, for example in hydrological [*Clark et al.*, 2008], water quality [*Beck*, 1987], ecological [*Ayala et al.*, 2014] and water distribution system [*Hutton and Kapelan*, 2015] modelling. However, to the best of the authors' knowledge, it has not yet been considered in any detail for its impact on optimised solutions of WRS management problems. Instead, the structural choices underlying simulation models used in this field often seem to be guided by a lack of data or knowledge, or by the need to make a certain optimisation method applicable (computationally tractable), rather than their appropriateness [*Giuliani et al.*, 2015b].

Finally, we list a source of uncertainty that is rarely considered in WRS modelling: contextual uncertainty. *Walker et al.* [2003] defines it as the uncertainty about *"the boundaries of the system to be modelled"*. Since few WRS exist in isolation, a certain degree of contextual uncertainty is unavoidable, just like in the modelling of any open system [*Dooge*, 1973]. Typical examples of contextual uncertainties in WRS modelling include aggregating demand nodes beyond the chosen system boundary, and assuming cooperation between multiple operators in the same system. We focus on this last element specifically because it is common for optimisation studies to assume that if multiple infrastructures are present in the same system their operations are perfectly coordinated, while in reality there often are different operators that either do not coordinate their decisions or do so through ad hoc discussions rather than formal rules that can be represented within a simulation model [*Giuliani et al.*, 2015a]. Central to this point, a growing number of studies have demonstrated that cooperation in water systems is a critically important factor in improving operational decisions [*Tilmant and Kinzelbach*, 2012; *Anghileri et al.*, 2013; *Giuliani and Castelletti*, 2013; *Marques and Tilmant*, 2013; *Wu et al.*, 2016].

Our study makes three contributions. Firstly, we introduce and assess a workflow to measure the impact of model structural and contextual uncertainties on the estimated performance of WRS management solutions obtained by optimisation. The workflow enables modellers to assess whether optimisation results are robust to uncertainty in their underlying simulation models. It builds on the 'rival framings' framework by [*Quinn et al.*, 2017] and is expanded to address the relevance of structural and contextual modelling choices in estimating the performance of the solutions of a multi-objective optimisation problem. Secondly, our study demonstrates the value of this workflow in a specific case study of a two-reservoir pumped storage system. In this application we answer the question "What is the extent to which the performances of optimised reservoir operating policies change upon re-evaluation under a simulation model that makes different modelling choices?" or, more simply: "How worse/better off can performances be when optimised under a modelling choice that turns out to be incorrect?". As we later discuss, the conclusions we draw from this case study application are, in varying degrees, generalizable to other types of WRS optimisation problems. Thirdly, we test the importance of aleatory uncertainty in the re-evaluation of optimised operating policies since *Mortazavi et al.* [2012] has suggested this can severely impact the validity of optimisation results. Recent studies have used expanded sampling strategies to account for this uncertainty for validation of optimisation results [*Kasprzyk et al.*, 2013; *Herman et al.*, 2014; *Quinn et al.*, 2017], we present a simple approach to justify the chosen sample size for re-evaluation.

4.2 Methodology

Modelling choices are often made under significant uncertainty, as discussed in the previous section. To study the impact of this uncertainty, we use a workflow built on the 'Rival Framings' framework introduced by *Quinn et al.* [2017]. In this workflow, each rival framing is a plausible hypothesis for formulating the WRS management problem. In *Quinn et al.* [2017], each framing uses different formulations of the objectives with no changes in the underlying simulation model. In our study, each framing makes different choices about some elements of the simulation model structure and the context. Figure 4.1 presents this workflow.

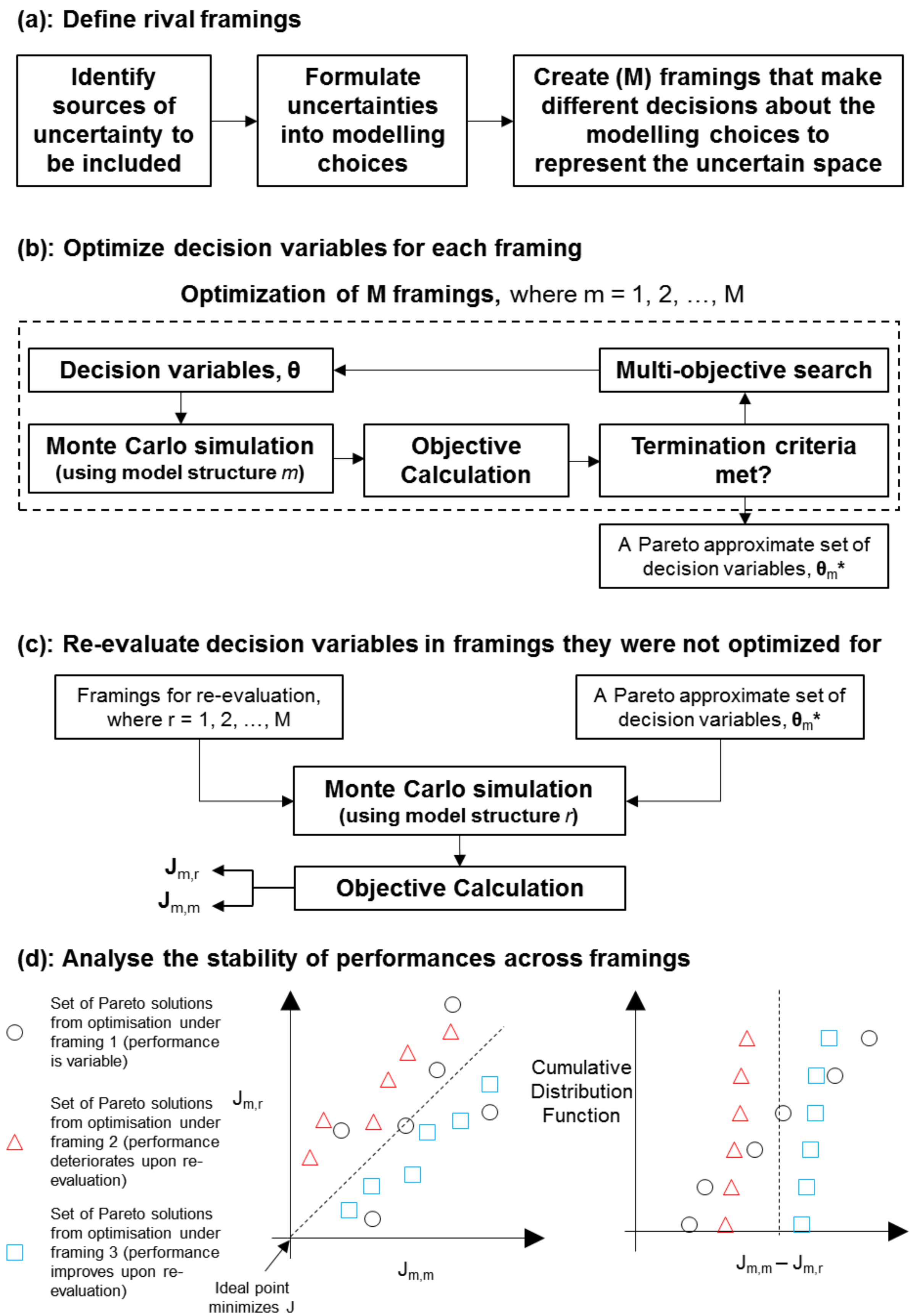
The first step is to define the rival framings, as depicted in Figure 4.1a. During this step, uncertainties are identified and characterised. For example, in *Quinn et al.* [2017] the sources of uncertainty are the objectives, which are characterised by a set of different options for their formulation. In our case, the sources of uncertainty are some of the assumptions underlying the model structure (for example, whether to include pump failures) and context (for example, whether the two water companies that manage the two reservoirs in the WRS coordinate their operations). Each framing will then consist of a unique set of modelling choices relating to these uncertain assumptions. Ideally, the range of considered framings should fully represent the uncertain space under investigation. In our case, this means that they should capture the uncertain assumptions in the modelling process that either the decision-maker(s) or the modeller(s) are sceptical about or wish to study their exposure to, in line with the second and sixth principles of best practice in collaborative modelling [*Langsdale et al.*, 2013]: *"all stakeholder representatives participate early and often to ensure that all their relevant interests are included"* and *"the model addresses questions that are important to the decision makers and stakeholders"*. This step can also be mapped into the 'identify uncertainties' stage in the XLRM framework presented by *Lempert et al.* [2003] and demonstrated in *Lempert et al.* [2006] and *Kasprzyk et al.* [2013]. The remaining stages of that framework – identify decision 'levers' (L), map actions to outcomes (R) and define performance metrics (M) – should then be followed to formulate a relevant simulation model of the system and thus create an appropriate management problem. As suggested by *Mahmoud et al.* [2009], it is important to interact with the decision maker(s) throughout the modelling process since their trust in the model outcomes increases with their trust in the underlying system conceptualization.

Next, as depicted in Figure 4.1b, decision variables are optimised under each framing, which results in a set of (approximate) Pareto optimal solutions (hereafter, a set of Pareto solutions) for each framing. In our case study, the decision variables are not the operational decisions (reservoir releases and pumped inflows) directly, but rather the parameters defining the operating policies that will be used to compute those decisions based on the WRS state (this distinction is described in more detail in Chapter 2, we discuss the specific formulation in the case study section and Appendix A1). A set of Pareto solutions are those whose performance in any objective can only be improved with a corresponding reduction in performance in one or more of the remaining objectives. In order to account for aleatory uncertainties (e.g. in our problem the streamflow, demand and potential evaporation time series) we use Monte Carlo simulation for the calculation of the objective function values against a range of possible realisations of those uncertainties. For the multi-objective search, we use the Borg multi-objective evolutionary algorithm (MOEA) since it has been shown to perform very effectively for multi-objective reservoir operation problems [*Salazar et al.*, 2016]. However, any optimisation tool capable of robustly solving stochastic, multi-objective formulations could be used here.

The key step in the Rival Framings workflow is the use of an independent re-evaluation of the optimised solutions under different candidate framings (Figure 4.1c). This is effectively testing how stable the estimated performances and trade-offs are to the assumptions made in the framing used for optimisation. As in the optimisation step, we use Monte Carlo simulation to estimate the objective values under aleatory uncertainty. Because the aim of this step is to show the stability of estimated performances, it is important that the approximation error from the

Monte Carlo simulation be small enough to enable meaningful comparisons between different sets of simulations. Previous studies that performed a re-evaluation step to validate results of stochastic optimisation have often used a larger sample size than the one used for optimisation, so to reduce approximation error in the re-evaluation (for example, *Kasprzyk et al.* [2013], *Herman et al.* [2014], *Giuliani et al.* [2015b], or *Quinn et al.* [2017]). Here, we propose linking the re-evaluation sample size with the decision-maker's sensitivity to differences in objective values. For example, if the decision-maker is only sensitive to differences in cost greater than 10 £/day, we should choose a sample size such that the approximation error in the objective calculation is, at a maximum, 10 £/day. Further increasing the sample size would be unnecessary, given that the decision-maker would not discriminate between solutions with cost differences lower than 10 £/day. In the experimental setup section, we will provide a simple technique to implement this idea and derive an adequate sample size for given value of the decision-maker's sensitivity.

Finally, the results of the re-evaluation step can be visualised through an objective-objective plot (shown on the left in Figure 4.1d), where the performances in the framing used for optimisation are plotted against the performances in the framing used for re-evaluation [*Quinn et al.*, 2017]. If the points lie along the bisector (the x=y line), there is no difference in performance between the two framings. If instead the points deviate from the bisector, then the choice of the framing impacts performance estimates. The larger the deviations from the bisector, the less robust the solutions are to the modelling choices underpinning the different framings. Given that it is the deviations from the x=y line that are critical to assessing robustness, we propose a simpler visualisation (shown on the right in Figure 4.1d), which shows the cumulative distribution function (CDF) of these deviations. In this plot, if the performances in the framing used for optimisation are similar to the performances in the re-evaluation framing, the CDF would follow the x=0 line. Again, such a result would mean the modelling choices that distinguish the two framings have minimal impact on optimised performances. If instead the CDF lies to the left of the x=0 line (as in the case of the red triangles in Figure 4.1d), then the framing used for optimisation is estimating lower objective values (i.e. more desirable, if we assume all objectives are to be minimized) than in the framing used for re-evaluation – i.e. solutions perform worse upon re-evaluation with respect to what suggested by optimisation. If the CDF lies to the right (blue squares) then solutions perform better upon re-evaluation – i.e. the estimates of performance produced by optimisation are conservative and likely to be exceeded if the assumptions underpinning the optimisation model are wrong. Obviously, the latter situation is preferable than the former, although we would suggest that both outcomes are not satisfying as they reveal a significant amount of uncertainty. Similarly, we would expect decision-makers to be most concerned by sets of Pareto solutions that exhibit the widest variation in performance, such as the black circles in Figure 4.1d, since pinning down the expected performance of a given solution under uncertainty will be difficult. Overall, these CDF plots provide decision makers with an indication of how stable individual solutions, and whole sets of Pareto solutions, are likely to be under different sources of uncertainty. This both enables them to select solutions from a set of Pareto solutions with characterised robustness and directs them towards sources of uncertainty whose monitoring and reduction will reduce the variability in performance estimates the most.



(a): Define rival framings
Identify sources of uncertainty to be included
Formulate uncertainties into modelling choices
Create (M) framings that make different decisions about the modelling choices to represent the uncertain space
(b): Optimize decision variables for each framing
Optimization of M framings, where m = 1, 2, …, M
Decision variables, θ
Multi-objective search
Monte Carlo simulation (using model structure m)
Objective Calculation
Termination criteria met?
A Pareto approximate set of decision variables, θm*
(c): Re-evaluate decision variables in framings they were not optimized for
Framings for re-evaluation, where r = 1, 2, …, M
A Pareto approximate set of decision variables, θm*
Monte Carlo simulation (using model structure r)
Objective Calculation
Jm,r
Jm,m
(d): Analyse the stability of performances across framings
Set of Pareto solutions from optimisation under framing 1 (performance is variable)
Set of Pareto solutions from optimisation under framing 2 (performance deteriorates upon re-evaluation)
Set of Pareto solutions from optimisation under framing 3 (performance improves upon re-evaluation)
Jm,r
Jm,m
Ideal point minimizes J
Cumulative Distribution Function
Jm,m – Jm,r

Figure 4.1: The rival framings workflow used in this study to estimate the impact of model structure and contextual uncertainties on the performance of optimized WRSs management solutions.

4.3 Case study and experimental setup

4.3.1 Description of case study and simulation model

Figure 4.2 depicts the water system used in this study to demonstrate our workflow. It is a simplified version of a two-reservoir system in the South West of the UK (labelled as S1 and S2 in Figure 4.2), with a pumped storage element to provide a supply of water in dry weather (to S1). The system is partly shared between two different water companies, reservoir S1 being the system element used by both companies. This reservoir is used by Company 1 to support downstream abstraction during low river flows for around 400,000 people (demand node D1 in Figure 4.2). While company 2 uses it to complement releases from S2 in supplying around 150,000 people (D2). The two reservoirs are moderately sized (relative to other large UK reservoirs, that have an average of 1377 Ml, [*EA*, 2017]) with storage capacities in the order of 20,000 megalitres (Ml) (S1) and 5000 Ml (S2). Both reservoirs must make environmental compensation releases, in addition S1 is occasionally required to deliver larger releases for downstream fisheries. Besides ensuring a reliable supply to D1 and D2, and delivering environmental compensation releases, the reservoirs' operation also aims at minimising pumping costs.

The two companies that operate the system liaise regularly, particularly regarding pumped storage operation, which is constrained by control curves and has operated in eight years since 1995. In simulation exercises for strategic and long-term planning, the two companies do not jointly model the system, instead making agreed conservative assumptions about each other's operation (described in the Section 4.3.2). Decision procedures are negotiated, and individual decisions are made cooperatively in either emergency situations or as periods of dry weather extend, informed by wider system considerations such as regional demand. Hence, for simulation purposes, the definition of system boundaries and the degree of cooperation assumed in the model of this system is a good example of contextual uncertainty.

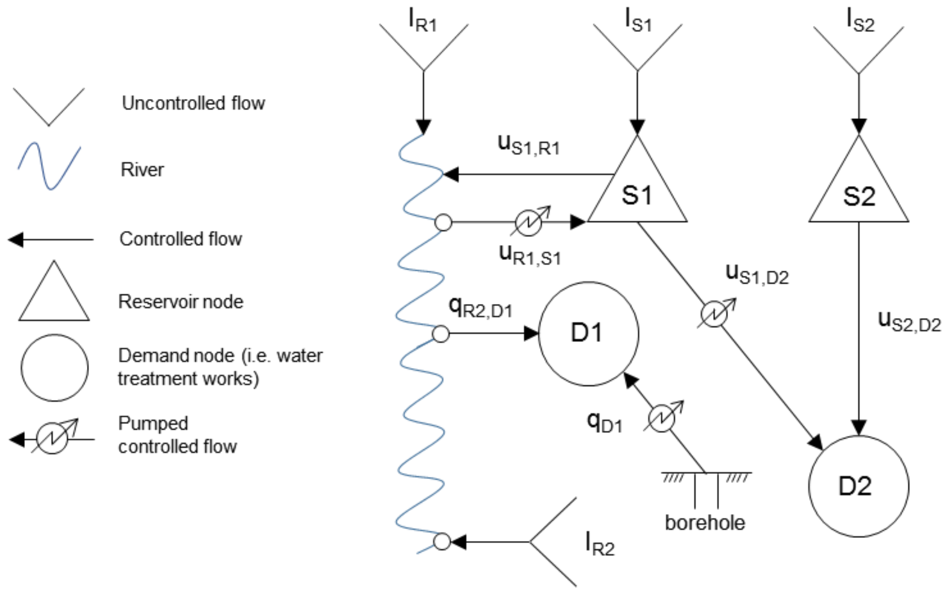


Figure 4.2: A schematic showing the main components and flows of the two-reservoir system used as a case study. Two companies operate half of the system each, with one reservoir (S1) as a shared resource. Company 1 takes the release and abstraction decisions $u_{R1,S1}$, $u_{S1,R1}$ and q_{D1} with the aim of supplying D1, while Company 2 makes the release decisions $u_{S1,D2}$ and $u_{S2,D2}$ to supply D2. Reservoir inflows are described by I_{S1} and I_{S2}, and river streamflows by I_{R1} and I_{R2}.

In our simulation model, we use 'operating policies' to represent the decision-making behaviour of the reservoir operators, as described in more detail in Chapter 2. Operating policies are parameterised functions that take the system's state variables as inputs (for example, reservoir storages and inflows) and return operational decisions as outputs (i.e. the three releases denoted by $u_{S1,R1}$, $u_{S1,D2}$ and $u_{S2,D2}$, and the pumped flux $u_{R1,S1}$ in Figure 4.2). As further explained in the next section, the choice of whether to use one operating policy to produce all operational decisions at once, or a separate policy for each reservoir, depends on the assumed degree of cooperation in the model. In either case, the parameters of the operating policies are the decision variables in our optimisation problem. This approach to reservoir operation optimisation is common in the water resources systems literature [*Guariso et al.*, 1986; *Oliveira and Loucks*, 1997; *Koutsoyiannis and Economou*, 2003; *Quinn et al.*, 2017], and is a specific instance of the more general 'direct policy search' approach to dynamic systems optimisation [*Rosenstein and Barto*, 2001]. For the parameterised functions, it is quite common to use universal approximators such as artificial neural networks [*Pianosi et al.*, 2011] or radial basis function networks [*Giuliani et al.*, 2015b] – we use the latter here. The reservoirs' mass balance equations and model forcing inputs (i.e. reservoir inflows, evaporation losses and water treatment work demands) are resolved at a daily time-step. A mathematical description of the simulation model, including the operating policies, can be found in Appendix A1. The model forcing inputs are generated synthetically using statistical models trained on historical data. We describe this process in detail in Appendix A2. The model is coded in the C programming language and parallelised with the Open MPI library, which enables highly efficient simulation runs.

4.3.2 Definition of modelling choices and framings

The model structure and contextual uncertainties considered in our rival framings methodology are summarized in Figure 4.3. These uncertainties lead to different possible modelling choices for generating inflows, demand and evaporation, and for representing pump failures, fisheries releases and the level of cooperation between the two companies. In this study, we consider each modelling choice as a binary option. For the inflow, demand, evaporation and fisheries releases, the binary choice is between a more sophisticated or less sophisticated representation of the process. For the pump breaks and cooperation between companies, the choice is between including them in the model or not. While we recognise that limiting our study to such binary choices may reduce some of the nuances in our interpretation, we believe it is useful to assess how important the choice is before investigating it in detail: if the exclusion of a process has no impact on the results, it is unlikely that the specific formulation of that process will matter. Then, we combine these binary choices to formulate 8 progressively more complex framings. In the following paragraphs, we briefly describe these sources of uncertainty and associated modelling choices, while further mathematical details are provided in Appendix A2.

Rival framing	Modelling choices					
	Pump breaks occur	Inflow model	Demand model	Potential evaporation model	Fisheries allowable period	Company cooperation
1:NoBrk-AR1-Sep-UC	×	AR(1)	AR(1)	AR(1)	September	×
2:Brk-AR1-Sep-UC	✓	AR(1)	AR(1)	AR(1)	September	×
3:Brk-AR2-Sep-UC	✓	AR(2)	AR(2)	AR(2)	September	×
4:Brk-AR2-All-UC	✓	AR(2)	AR(2)	AR(2)	All year	×
5:NoBrk-AR1-Sep-C	×	AR(1)	AR(1)	AR(1)	September	✓
6:Brk-AR1-Sep-C	✓	AR(1)	AR(1)	AR(1)	September	✓
7:Brk-AR2-Sep-C	✓	AR(2)	AR(2)	AR(2)	September	✓
8:Brk-AR2-All-C	✓	AR(2)	AR(2)	AR(2)	All year	✓

Figure 4.3: The set of rival framings used in this study. Each row indicates a different framing and the modelling choices associated with it. AR(1)/AR(2) describes the type of autoregressive statistical model used to generate the input (with lag of 1 day or 2 respectively). Fisheries releases may occur in September or year-round (except Spring). Company cooperation indicates that all objectives (and decisions) in the system are optimised together while non-cooperation indicates that each company's objectives (and decisions) are optimised separately.

Pump breaks. Based on communication with operators at the two water companies, there is significant uncertainty associated with pump failures resulting from pump breaks. They occur infrequently, for unique reasons and in unique ways making them an epistemic uncertainty. Based on the authors' experience pump failures account for some of the largest operational failures that water suppliers face and are a key cause of the practitioners' scepticism about the validity of simulation models, which typically neglect them. Uncertainty in modelling pump failures exists around both the choice of the statistical distribution of occurrence, severity and duration of failures (structural uncertainty) and the choice of the parameters of those distributions (parametric uncertainty). For the purpose of this study, we limit the choice in our framings to whether to include pump failures or not. While this may simplify the aforementioned uncertainties, it will at least indicate whether including pump failures in the WRS model significantly affects the optimisation results or not. When pump breaks are included, we chose to represent both their duration and the duration between consecutive breaks by a Poisson distribution because this is commonly used for characterising failure frequency in systems with electronic components [*Weiss*, 1956].

Inflow, demand and evaporation. The characterisation of reservoir inflow uncertainty dates to early works in the field [*Maass et al.*, 1962] and has been an active field of research since [*Fiering and Bund*, 1971; *Hirsch*, 1979; *Salas et al.*, 2005; *Rajagopalan et al.*, 2010; *Herman et al.*, 2016]. Here we assume that inflows are stationary and thus, using the definitions set out in our introduction, they can be fully represented through statistical models (see *Vogel* [2017] for a discussion of the non-stationary case). This representation introduces a source of aleatory uncertainty but also of parametric uncertainty (associated with the parameters of the statistical model) and structural uncertainty (associated with the overall form of the statistical model). The latter is particularly significant when the quality and quantity of historic data that can be used to fit the statistical model is low, as in our case. Here we consider two possible model structures to generate inflows: both are periodic autocorrelated (AR) models [*Salas and Obeysekera*, 1982] but with different lags (of 1 and 2 days). Further details on the AR models and their calibration from historical data are given in Appendix A2. As for demand and evaporation, we note that many techniques that address inflow uncertainty may broadly be applied to these variables too, as demonstrated by the similarity between the methods described in [*Donkor et al.*, 2014] for statistical modelling of water demand and those described in the references above for statistical modelling of inflows. Hence, we also use lag-1 and lag-2 AR models for evaporation and demand (see Appendix A2).

Fisheries releases. Reservoir S1 is occasionally required by the UK Environment Agency to make a large release over a few days to support downstream fisheries. Predicting when such request may occur is difficult because it depends on the decision made by an external stakeholder, the Environment Agency, who acts according to conditions and demands (for example, downstream water quality and pressure from the fisheries' owners) that occur outside the boundaries of the WRS under study. This problem of how to represent forcing inputs driven by the behaviour of external stakeholders is rather common in WRSs modelling and makes the fisheries releases a good example of contextual uncertainties. In their simulation exercises for long-term planning, Company 1 assumes that fisheries releases may only occur over a period in September, since this is the most common time of year for them to occur. However, the historic data shows that the releases have also occurred at many other times of the year, apart from the Spring period. In this study we thus characterise this uncertainty by starting the fisheries releases on a random day inside a feasible time window, which is either limited to September (as in the Company's assumption) or expanded to the entire year except Spring (as in the historic data). Once the starting date has been randomly selected, the overall volume of water released is fixed (the historic data shows that the total volume released each year is relatively constant) and distributed over a period of random duration between 3 and 15 days (the historic data determines these limits).

Representation of company cooperation. The last modelling choice we consider is how to represent cooperation between the two companies when making release or pumping decisions. We include this choice because previous studies (e.g. *Anghileri et al.* [2013], *Giuliani and Castelletti* [2013])) demonstrated that model assumptions about coordination between connected reservoirs can dramatically impact the performance of optimised reservoir operations. We note that the current situation is that the two companies coordinate their operations but do so on a case-by-case basis accounting for conditions in the wider water resources system of which the two reservoirs are part. For these reasons it is difficult to formalise their current coordination of operations into a set of mathematical equations. On the other hand, in the simulation models that the two companies use for long-term planning, these elements of cooperation are represented by a set of conservative assumptions about the other company's operations, which is a precaution deemed acceptable as the system is typically in surplus. In order to capture the uncertainty around the assumed level of cooperation in modelling this system, we thus define two extreme scenarios: a 'non-cooperative' modelling scenario and a 'cooperative' one. In the 'cooperative' modelling scenario the model simulates the entire two-reservoir system as one WRS (as depicted in Figure 4.2), and the optimisation produces one operating policy that returns all the release and pumping decisions simultaneously. In the 'non-cooperative' modelling scenario, instead, two separate operating policies are produced, each controlling the company's own decisions independently from the other. The two separate simulation models used to optimise the two policies are shown in Figure 4.4. The model of Company 1, shown on the left, makes the conservative assumption that Company 2 will always draw as much water as possible from reservoir S1, i.e. $u_{S1,D2}$ is equal to the maximum licensed. Conversely, the model of Company 2, on the right, assumes that they will always be able to draw their licensed volume from reservoir S1. We note that whilst these two scenarios capture the two possible extremes of modelling the system, the actual operation varies somewhere between, depending on the nature of the situation.

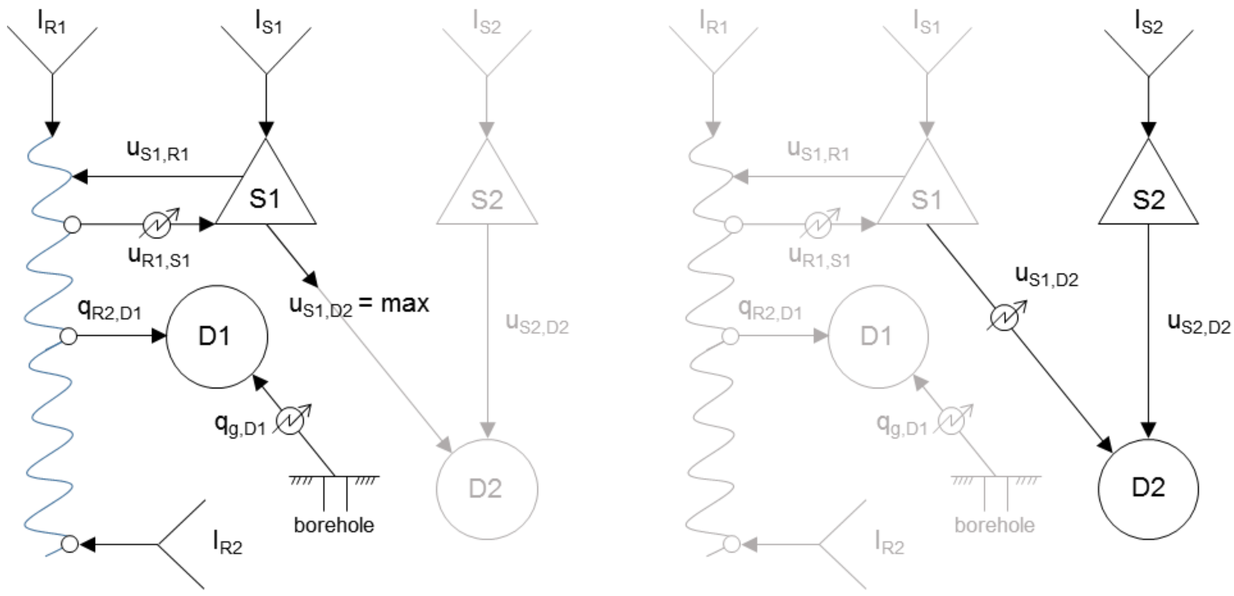


Figure 4.4: Schematic of the two separate simulation models used in the 'non-cooperative' modelling scenario.

4.3.3 Optimisation and re-evaluation

Both the optimisation and re-evaluation steps of our methodology (Figure 4.1 (b) and (c)) require calculation of the objectives via Monte Carlo simulation. We have already defined the operator's objectives broadly: each company aims to reduce both deficits in supply and the cost of pumping. This gives four objectives: the average daily deficit in supply for company 1 (J_{D1}) and for company 2 (J_{D2}); and the average daily pumping cost for company 1 ($J_{£1}$) and company 2 ($J_{£2}$). The mathematical details of the objectives formulation are given in Appendix 1. *Quinn et al.* [2017] has demonstrated that the objective formulation is important and that, for example, the optimal operations can be different even just for a change in the temporal statistic used to aggregate costs, e.g. the mean, the worst case or another distribution quantile. However, this is not the focus of our study and so we will not explore the impact of using different objective function formulations. For each objective we will take simulated daily costs and average them across both the simulation period and the Monte Carlo realisations, as shown in Figure 4.5 and summarised in equation (4.1):

$$J_i = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{T}\sum_{t=1}^{T} g_{t,i}^{k}\left(\theta, p_t^k, I_t^k\right) \tag{4.1}$$

where i is the objective index (running from 1 to 4 in our case), k is the index of the Monte Carlo realisations, t is the time index (day); g are the daily costs of operation (i.e. supply deficits at the two demand nodes and daily pumping costs for the two companies), θ are the searched for parameters of the operating policy, p and I are uncorrelated and autocorrelated forcing respectively, T is the length (days) of the simulation period and K is the number of runs in the Monte Carlo simulation. In the optimisation step, we aim to find the set of θ vectors that is the set of Pareto solutions between the 4 objectives. We use the Borg MOEA [*Hadka and Reed,* 2013] to solve the optimisation problem since it has been shown to be highly effective for reservoir operation optimisation problems based on direct policy search [*Salazar et al.*, 2016].

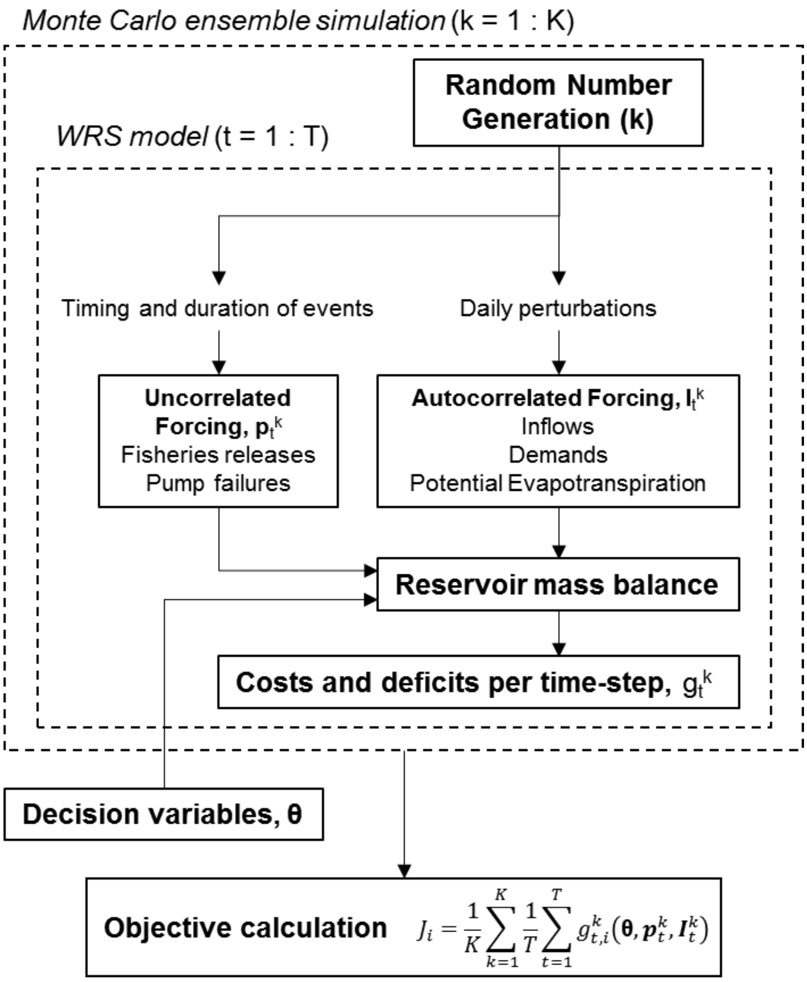


Figure 4.5: The Monte Carlo simulation process behind evaluating the objectives associated with a given operating policy. k is the index for the Monte Carlo ensemble member, and t is the index for the time-step.

As anticipated above, the Monte Carlo simulation requires specification of the ensemble size K and of the simulation length T – together the ‘sample size’. A given sample size will have an associated approximation error in the objective values. As the sample size tends to infinity, the approximation error should reduce to zero. Thus, for a given operating policy, the approximation error at a given sample size can be quantified as the difference between an objective value at that sample size and the objective value at an infinite sample. Given the high computational efficiency of our simulation code, we approximate the objective value at an infinite sample by the objective value at a very large sample size ($4*10^7$ days, coming from K

= 400 and T = 100,000). We can then monitor the trajectory of approximation errors at other sample sizes. An example for the water deficit objective of company 2 is given in Figure 4.6.

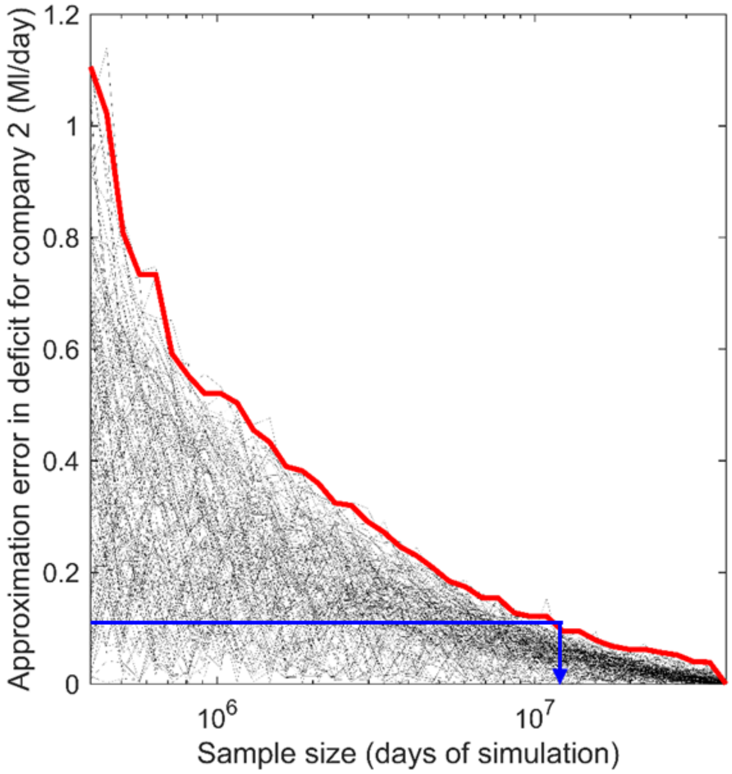


Figure 4.6: (Black lines) approximation error of 132 random policies evaluated over simulation periods of independent and increasingly large sample sizes. (Red line) the 99th percentile of these errors at each given sample size. (Blue line) an example of how to start with the sensitivity of the decision maker for this objective (0.11Ml/d) and determine an appropriate sample size ($1.2*10^7$ days). Approximation errors are defined as the absolute difference between the objective function value at a given sample size and the value at the largest sample size possible (in this case, $4*10^7$ days).

During the optimisation step, smaller sample sizes will result in less accurate objective evaluations (as is clear from Figure 4.6) but speedier computation, thus enabling more function evaluations during optimisation. A vast literature on 'noisy optimisation' indeed shows that optimisation algorithms can often find good solutions in spite of approximate objective values [*Fitzpatrick and Grefenstette*, 1988; *Miller and Goldberg*, 1996; *Smalley et al.*, 2000; *Yun et* *al.*, 2010]. *Salazar et al.* [2017] provide an extensive discussion and computational experiments to demonstrate the complex trade-off between objective approximation and optimisation efficacy. Additionally, using the Borg MOEA requires specification of an 'epsilon' value, which is the minimum difference in objective value that must be exceeded for the optimiser to consider one solution to outperform another. *Kasprzyk et al.* [2012] show that this epsilon value should reflect the minimized likelihood of one solution being selected over another due to approximation error for the given sample size.

During re-evaluation, we must ensure that validation results have approximation errors smaller than the decision-maker's 'sensitivity', i.e. their ability to discriminate between different solutions. As exemplified by the blue line in Figure 4.6, we can start from the pre-specified sensitivity of the decision maker and calculate the sample size that would guarantee approximation errors below that sensitivity. With this approach and assuming sensitivities of 0.17 Ml/day for J_{D1}, 0.11Ml/day for J_{D2}, £9/day for $J_{£1}$ and £1.3/day for $J_{£2}$, we obtain here a required sample size of K = 400, T = 30,000, i.e. $1.2*10^7$ days (the plots equivalent to the 99th percentile in Figure 4.6 for all objectives and all framings are given in Appendix A3, Figure A3.1). Because our simulation model is computationally highly efficient, for the optimisation step we simply use the same sample size as in the re-evaluation step. Consistent with the interpretation of epsilon values given by *Kasprzyk et al.* [2012], we set epsilons to the decision maker sensitivities given above.

With epsilon values specified, the last tuning parameter to be specified for running Borg MOEA is the maximum number of function evaluations. To make this choice, we repeat the optimisation process multiple times with different seeds to determine an appropriate number of function evaluations to produce a stable hypervolume. A hypervolume indicates the volume of objective space that is captured by a set of Pareto solutions, as described in *Knowles and*

Corne [2002] and shown in Chapter 5, Figure 5.2. An example plot, for framing 8, can be found in Appendix A3, Figure A3.2. This allows us to conclude that an optimisation process with 10^5 iterations should be more than satisfactory.

Finally, for framings that include no cooperation between companies, we need to implement the following small adjustments to the workflow, to account for the specifics of our case study:

- *In the optimisation step*. Each company has its own model and its own operating policy that is completely independent from the other company's model and policy. This results in two separate optimisations for a single framing, one for each company (i.e. separate optimisations of the two models depicted in Figure 4.4). Hence there is no trade-off between Company 2's objectives and Company 1's, given that every operational solution for Company 1 is compatible with every operational solution for Company 2 (and vice versa). We note that, while this may not be true in reality, it is the result of the assumptions made under the non-cooperation modelling choice.
- *In the re-evaluation step*. The policies developed under the cooperative modelling choice use as inputs the state variables from the entire system (i.e. storage at both reservoirs, demands, and all uncontrolled flows). Under the non-cooperation choice, instead, the system is de-coupled, thus it would be impossible to simulate a policy developed in the cooperative case under a non-cooperative assumption since certain inputs to that policy are simply not represented in the two de-coupled models. Consequently, in the re-evaluation step we can only re-evaluate policies in the cooperative framings. Note that non-cooperative policies can instead be simulated in the cooperative framings since they control different release variables and both policies' input data are represented in the coupled, cooperative model.

4.4 Results

As anticipated in the Methodology section, our key result is a comparison of the performances of a set of Pareto solutions as estimated in the optimisation step and in a re-evaluation step where different model framings are used. We show the results in full, for each of the four objectives and each combination of every set of Pareto solutions re-evaluated in every framing, in Figure A3.3 of Appendix 3. The majority of the subplots in this figure show large deviations from the x=0 line, which means that the different modelling choices made in the 8 framings have a large impact on estimated objective values. For simplicity of illustration, in Figure 4.7 we focus on the results for a specific objective (the mean deficit in the water supply for Company 2, J_{D2}) and re-evaluation on a specific framing (8).

Figure 4.7 shows the cumulative distribution function (CDF) of the differences between the J_{D2} estimates under the framings used for optimisation (from 1 to 8) and under framing 8. The variability in estimated performance for the non-cooperative framings (1, 2, 3, 4 - red lines) is noticeably larger than for the cooperative ones (5 ,6, 7, 8 - blue lines) – indicated by the larger spread over the x-axis. We also see a clear difference between the framings that use an AR(1) model for generating synthetic forcing during optimisation (framings 1, 2, 5, 6 - lines with circles) and the ones that use an AR(2) model (3, 4, 7, 8 – no circles). In fact, the CDFs of the AR(1) framings are more commonly positive, meaning that the objective values typically improve when re-evaluated using an AR(2) model, i.e. that using the AR(1) synthetic generator provides a conservative estimate of performances. The differences attributable to other modelling choices (i.e. pump failures and fisheries releases) are far smaller. While the details of these results are specific to objective J_{D2} and framing 8, the conclusions of which modelling choices make a significant difference are similar across all cases, as we show in the expanded results plots in Appendix A3, Figure A3.3.

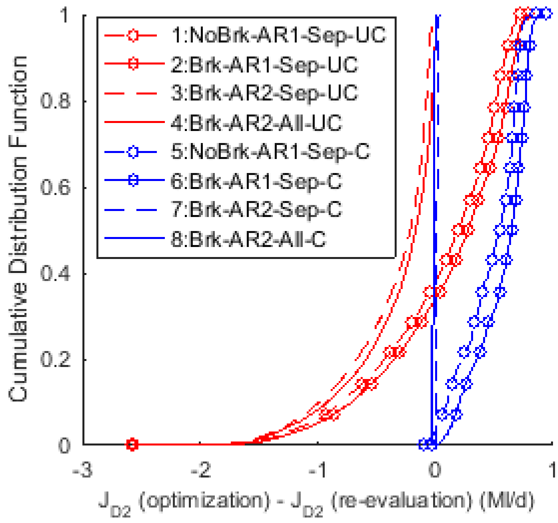


Figure 4.7: A cumulative distribution function of the differences between the J_{D2} objective values estimated in the optimisation step (under framings 1 to 8) and the J_{D2} values re-evaluated under framing 8. Blue lines indicate non-cooperative framings and red lines indicate cooperative framings. Results are obtained with a sample size of $1.2*10^7$ days for both optimisation and re-evaluation.

To offer a more detailed interpretation of the impact of contextual uncertainty, we further analyse two of the eight sets of Pareto solutions shown in Figure 4.7: framing 4 (i.e. a non-cooperative framing) and 8 (i.e. a cooperative framing). From each of these two sets, we extract the subset of solutions that form the set of Pareto solutions between the objectives deficit in water supply for company 2 (J_{D2}) and pumping costs for company 1 ($J_{£1}$) . The objective values of these subsets are shown in Figure 4.8 as red points (framing 4) and black points (framing 8). From this figure we see that there is an area in which the red points are higher than the black points, i.e. an area where solutions optimised under framing 8 perform systematically better than those optimised under framing 4 in terms of water deficit (J_{D2}), while being equal in terms of pumping costs ($J_{£1}$ between 300 and 1000 £/d). Thus, it is clear that optimising under framing 4 simply does not allow access to part of the objective trade-off space that could instead be accessed if optimising under framing 8. This effect is due to the non-cooperative assumption made in framing 4, given that it is persistent across all non-cooperative framings: if a set of policies optimised in any non-cooperative framing is evaluated under any cooperative framing, they will always lose access to this section of the objective subspace, which is instead accessible to all sets of policies optimised under any cooperative framing. This is clear when viewing the complete set of results in Appendix A3, Figure A3.4.

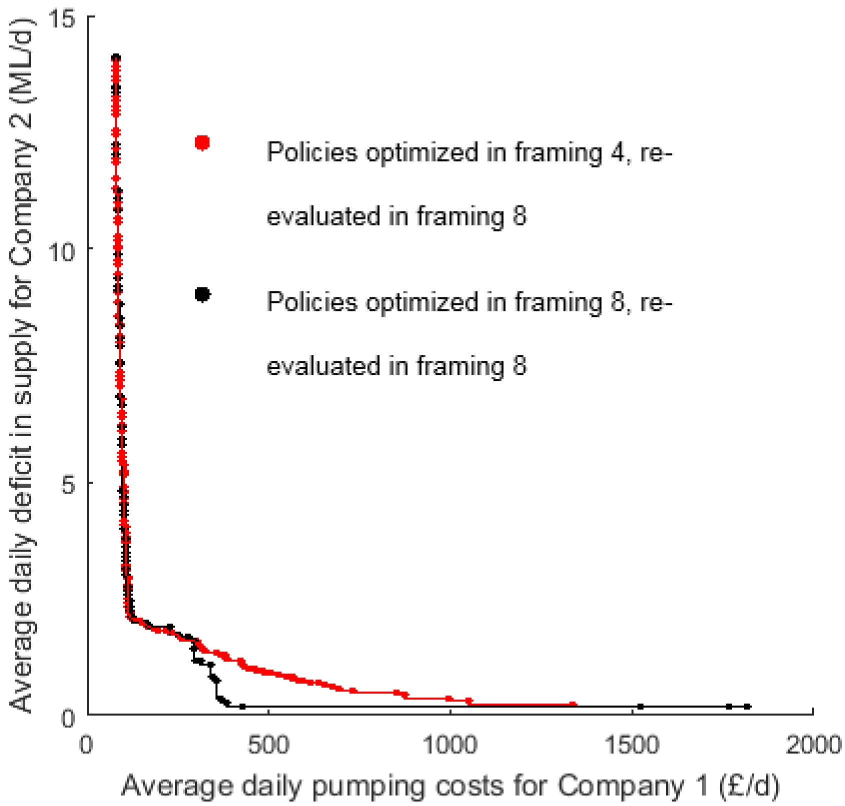


Figure 4.8: (Red points) estimated performances of the set of Pareto solutions optimised under framing 4 and re-evaluated using framing 8. Only policies that lie on the trade-off between the pump costs for company 1 (X-axis) and deficit for company 2 (Y-axis) are shown for clarity. (Black points) the same but with the Pareto set of operating policies optimised in framing 8.

The results in Figure 4.7 were obtained by using a very large sample size ($1.2*10^7$ days) in both optimisation and re-evaluation. Using such a large sample size to validate optimised solutions is not common as the majority of studies in this field use validation periods of few years or decades – with some exceptions such as *Kasprzyk et al.* [2013], *Herman et al.* [2014], [*Giuliani et al.*, 2015b] or *Quinn et al.* [2017]. Hence, we thought it helpful to repeat our analysis using a smaller sample size of 10^4 days, which corresponds to a more typical simulation length of 30 years. We present a sample of the results so obtained in Figure 4.9, again for objective J_{D2} and re-evaluation in framing 8. From this figure, it seems that all sources of uncertainty are influential. However, a closer inspection of the results shows that even the estimates of performance for framing 8 vary from optimisation to re-evaluation (i.e. the CDF deviates from the $x = 0$ line). Since the only difference from optimisation and re-evaluation here is the Monte Carlo sample used for simulation, we can attribute the differences in estimated performance to approximation errors from the small sample size. Therefore, we expect that Figure 4.9 shows the combined influence of both modelling choices and approximation errors. Inspecting these results in their entirety (Figure A3.5 in Appendix A3), shows that there is no discernible pattern to the influence of different uncertainties, signifying that the role of aleatory uncertainty is possibly as large as that of modelling choices.

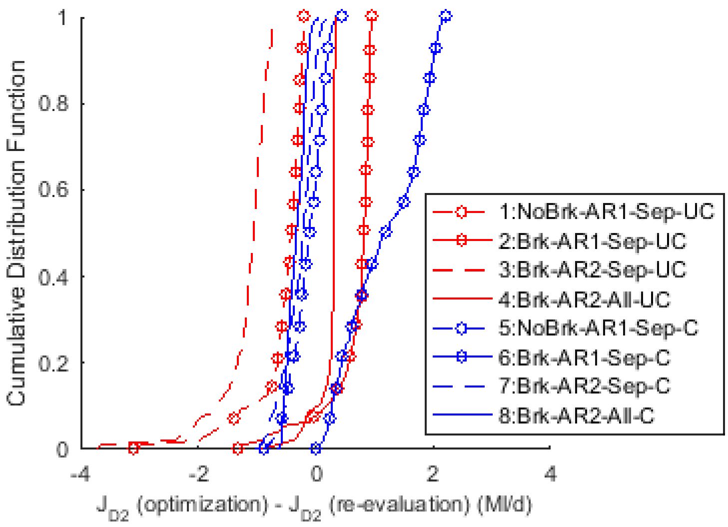


Figure 4.9: The same as in the left panel but with an optimisation and re-evaluation sample size of $1.1*10^4$ days (i.e. 30 years).

4.5 Discussion

The key aim of this study was to measure the impact of model structural and contextual uncertainties on the estimated performance of water management decisions obtained by optimisation. Our key result is that, in our case study, the assumption about the level of cooperation between water companies has a greater impact on estimating objective values than any other modelling choice. Our estimates of performance vary largely with this assumption, as shown in Figure 4.7, and if we model either company separately from the other, the benefit of optimisation is hindered by the inaccessible trade-off space, as shown in Figure 4.8. This confirms what other studies have found, i.e. that cooperation in decision-making is greatly beneficial [*Tilmant and Kinzelbach*, 2012; *Anghileri et al.*, 2013; *Giuliani and Castelletti*, 2013; *Marques and Tilmant*, 2013; *Wu et al.*, 2016]. It is important to note that this occurs even though the assumptions made in the non-cooperative scenario about each company's operations are compatible and seemingly conservative. In company 2's simulation model, it is assumed that they can take as much water as they want from the other company's reservoir (S1), while in company 1's model it is assumed that company 2 will take as much as they can from S1. These assumptions expose company 2 to an over-reliance on S1 and results in company 1 over-abstracting from the river in their pumped storage operations, caused by their conservative assumptions about company 2. In turn, this two-reservoir system that we have modelled is part of a larger inter-connected water resources system. Therefore, it is possible that even the most robust results we present here are themselves subject to a similar amount of uncertainty if one considered the larger system. In general, few water systems exist in isolation and thus these contextual uncertainties likely impact the results of many optimisation studies. *Beven and Alcock* [2012] have suggested that the choice of system boundaries is crucial in making predictions about environmental systems, here we have shown that the same is also true for a coupled human-environmental system.

Another conclusion from our results is that the objective values are quite sensitive to the choice of the autoregressive synthetic generators (Figure 4.7). In Appendix A3, Figure A3.3, the complete set of CDF plots show that the performance estimates of policies optimised under framings 5 and 6 (AR1, cooperative) nearly always improve significantly when re-evaluating in framings 7 and 8 (AR2, cooperative). This implies that the AR(1) generator leads to more conservative estimates. That including just one additional autocorrelation lag term impacts the operational performance corroborates the conclusions found in *Tejada-Guibert et al.* [1995]: that small changes in the statistical characterisation of input forcing can have large operational impacts.

An encouraging result is the seeming lack of importance of the choice of including pump failures, a factor that is often mentioned by practitioners as critically missing in simulation models. For example, in Figure 4.7 we see only a small translational shift in the CDF between framings 1 (no breaks, non-cooperative) and 2 (breaks, non-cooperative), and between framings 5 (no breaks, cooperative) and 6 (breaks, cooperative). This indicates that, while pump breaks do reduce performance, the choice of the optimal policy is unlikely to change. This is because, in the event of a pump failure, there is little that can be done in terms of the operational decisions available in the model that will resolve the failure. We expect this result to be generalisable since it would require a level of redundancy not usually present in water resources systems to alter the conclusion. The authors hope that more studies will include this rarely considered source of uncertainty on the basis that it may help to build a case that excluding asset failures in a simulation model is not a reason to reject optimised operational policies.

In Figure 4.9, we show that approximation errors can lead to falsely attributing differences in objective values to (in our case) structural/contextual uncertainty. Despite this, the use of an expanded sample for validation of optimisation results is the exception and not the rule in this field (for examples of expanded validation samples see *Kasprzyk et al.* [2013], *Herman et al.* [2014], [*Giuliani et al.*, 2015b] or *Quinn et al.* [2017]), and seeing studies where a simulation period of 20-30 years (or even less) is used to demonstrate the efficacy of a new optimisation algorithm is not unusual. For our case study, Figure 4.6 shows that this simulation length is far too short to produce accurate objective estimates, and Figure 4.9 (and the expanded results shown in Appendix A3, Figure A3.5) show that, with a 30-year re-evaluation period, there are seemingly many significant differences between framings, which could be misattributed to structural/contextual uncertainty if we were not aware of how much approximation error was present in the objective value estimates.

4.6 Conclusion

In this Chapter we reviewed different types of uncertainty and different approaches for addressing it in water resources operation problems. We found that no previous study had formally investigated the impacts of model structural uncertainty nor of contextual uncertainty on optimisation results. We formulated different 'rival framings' of a water resources system based on alternative modelling choices for the most uncertain assumptions about the system functioning. We obtained a set of Pareto solutions (where solutions are operating policies) via multi-objective optimisation under each framing and then re-evaluated those policies against all other framings. This enabled us to test how robust the objective value estimates are against the various framings. Our results revealed four key findings, three about the impact of uncertainty and one around the importance of validation:

- Cooperation between operators is often assumed in water resources models. We find that this assumption and thus the definition of the system boundaries (i.e. the model's context) had the most significant impact on estimated objective values and trade-offs.
- The model structural uncertainty that had the most significant impact was around the level of temporal persistence (auto-correlation) in the forcing inputs. Our results suggest that even minor differences in the statistical formulation of forcing generators can significantly impact performance estimates. One implication of this result is that the common simplification of using an AR(1) model for generating forcing inputs (as is often done to reduce the problem's dimensionality for stochastic dynamic programming) may not always be a suitable assumption.
- Other modelling choices, such as whether to introduce pump failures or not, had much less impact. This result is encouraging since it shows that simplifying assumptions in simulation models do not always affect optimisation results significantly and hence simulation-optimisation models can be operationally useful even if they are not a perfect picture of the real-world.

- Recreating results by re-evaluation on a shorter (30 year) simulation period produced dramatically different conclusions – this shows that insufficiently accounting for aleatory uncertainty (i.e. intrinsic random variability) can lead to misleading results.

Our findings highlight why it is important to consider structural and contextual uncertainties in water resources system optimisation. Re-evaluation under uncertainty enables decision-makers test how ‘optimal’ their optimisation results really are, and thus identify the modelling choices that merit careful consideration. It can also identify simplifying assumptions that, although seemingly unrealistic, can be acceptable for the purpose of operation optimisation.

CHAPTER 5: MEASURING AND ACHIEVING ROBUSTNESS IN MULTI-PURPOSE RESERVOIR SYSTEMS UNDER STRUCTURAL AND CONTEXTUAL UNCERTAINTY

5.1 Introduction

Chapter 3 found that water managers are concerned by the uncertainties present in the simulation models that underpin optimisation solutions, these results are in line with the only other survey of this type [*Rogers and Fiering*, 1986]. Chapter 4 evaluates optimised *operating policies*, a specific type of optimisation solution (defined in Chapter 2), under uncertainty and shows that their performance can change significantly depending on the assumptions made about the model structure (i.e. how the interrelationships within the system model are defined) or context (i.e. how the model boundaries are defined). Because water managers are concerned about uncertainty (Chapter 3) and because structural and contextual uncertainty has been shown to significantly affect the performance of optimised operating policies (Chapter 4), we believe that there is likely to be a need for methodologies to identify operating policies that are robust to structural and contextual uncertainty. Below we discuss several formal definitions of robustness and how to create robust solutions. This chapter focuses specifically on operating policies, however the discussion and methodology presented here would be equally applicable to other types of optimisation solutions. These solutions could be from operation optimisation (for example, as presented in Chapter 2, i.e. release sequences or real-time optimisation) or elsewhere in water resources research (for example, the size of a dam).

We first answer the question, what is robustness? Robustness in water resources systems is most commonly described as a solution's ability to 'perform under uncertainty'. Many studies use the term but do not provide a more formal definition than this [*Bankes et al.*, 2001; *Lempert*, 2002; *Hine and Hall*, 2010; *Kasprzyk et al.*, 2013; *Walker et al.*, 2013]. Based on this description, defining a metric of robustness requires specifying definitions for: a) the uncertainty, and b) the performance to be guaranteed under uncertainty.

We first discuss the uncertainty. As established in Chapter 4, there are both aleatory uncertainties (arising due to intrinsic variability) and epistemic uncertainties (arising due to lack of data/understanding). We extend these definitions. Aleatory uncertainty has sufficient data/understanding to be characterised with a probability distribution (or statistical model) that is agreed upon by experts. In contrast, epistemic uncertainty lacks sufficient data/understanding to be characterised with an agreed upon probability distribution, this matches *Walker et al.* [2013]'s definition of 'deep uncertainty'. We follow *Bankes* [2002] and say that the uncertainty that is the subject of robustness is epistemic (or deep, as put by Bankes), because, if probability distributions are available for a source of uncertainty then statistical modelling and analysis should be sufficient to create a stochastic optimal solution (for example, if we had probability distributions of streamflow moments under climate change, we could create an operating policy that optimises our objectives conditional on these distributions). We also note, to avoid confusion, that robustness has commonly been defined as the worst-case performance under aleatory uncertainty in the past, for example see *Daniels and Kouvelis* [1995].

We now provide a discussion around what should be guaranteed under uncertainty must be provided to categorise metrics of robustness. One such discussion is given by *Gabrel et al.* [2014], who proposes three possible definitions of a robust solution:

- Delivering stable objective values under uncertainty, for example, the operating policy's cost should not vary too much when evaluated under different scenarios that cover the uncertain space. This is equivalent to *Kwakkel et al.* [2016]'s definition of 'statistical robustness' for an environmental decision-making problem. *Herman et al.* [2015] and *Kwakkel et al.* [2016] also specify a category 'satisficing'. Robust satisficing requires a solution's objective value to exceed a pre-specified threshold under different uncertain scenarios. We believe this is covered by 'delivering stable objective values' because it could be re-formulated as a piece-wise objective value.
- Delivering a stable distance to optimality under uncertainty, for example, for each uncertain scenario, the operating policy must cost within £X of the optimal operating policy for that scenario. This is equivalent to *Herman et al.* [2015]'s definition of 'regret' in a water resources management problem.
- Deliver a feasible solution under uncertainty, for example, the operating policy must not break any constraints for X% of scenarios in the uncertain space. To the author's knowledge, this definition is not used in water resources or environmental decision-making – possibly because the feasibility of typical solutions is unaffected by uncertainty. We interpret the constraints here to be of the type described as 'hard constraints' in Chapter 2, i.e. constraints that describe physical impossibilities rather than 'soft constraints' that express preferences of decision makers.

Whichever definition of robustness is ultimately selected, the creation or identification of robust solutions is required. This is commonly achieved through one of two philosophies, which we will refer to as: (1) robustness via re-evaluation, or (2) robust optimisation (RO).

In robustness via re-evaluation, a set of candidate solutions is first selected and then they are evaluated against a range of uncertain scenarios. This reveals the robust ones and highlights which uncertain scenarios commonly cause solutions to 'fail' (i.e. not perform). Candidate solutions may be specified a-priori by the decision maker, this is quite common in planning and design (e.g. *Matrosov et al.* [2013] or *Borgomeo et al.* [2018]) but is not the focus of this chapter. In reservoir operation literature, candidate solutions are commonly obtained by optimisation, which is becoming increasingly widespread due to improvements in computing power and progress in the development of optimisation algorithms. Typically, this optimisation step will include aleatory uncertainty, thus making it 'stochastic optimisation' by the definitions of uncertainty in robustness we have set out previously. If multiple objectives are present, which is also common in water resources systems (discussed in Chapter 2), optimisation can result in potentially thousands of candidate solutions where each solution satisfies a different trade-off between the objectives in question – termed a set of (approximate) Pareto optimal solutions (hereafter, a set of Pareto solutions) [*Reed et al.*, 2013]. *Kasprzyk et al.* [2013] demonstrate all of these steps in a framework they term 'multi-objective robust decision making', where the solutions they optimise are policies that specify volumes of water to be transferred and the sources of uncertainty they re-evaluate under are parametric that, for example, describe those demand growth rate.

In contrast, RO builds epistemic uncertainty into the optimisation problem, and thus directly optimise the robustness of a solution(s) [*Beyer and Sendhoff*, 2007; *Bertsimas et al.*, 2011]. Robust optimisation has been applied to a range of environmental decision-making problems. For example, *Hamarat et al.* [2014] performs robust optimisation of EU strategic decisions in the energy sector (such as the percentage of renewables and amount of subsidies) under uncertainties such as economic growth or electrification rate. More specific to water resources, *Trindade et al.* [2017] optimises management decisions (including triggers for water transfers and water restrictions) using objective values that are aggregated across an ensemble of scenarios that contain both aleatory uncertainty (about streamflows) and epistemic uncertainty (about parameters that, for example, determine demand growth and transfer costs). To the author's knowledge, structural and contextual epistemic uncertainties have not been considered in RO so far, despite our previous work (Chapter 4) that has shown that they significantly impact estimates of performance for optimised operating policies.

Robust optimisation has been so far a less common method than robustness via re-evaluation, despite its intuitive advantage of directly searching for robustness rather than selecting for robustness from a pre-specified set of solutions. This may be due to a range of reasons. Firstly, the epistemic uncertain space that is included during the optimisation process for robust optimisation is typically very large. For example, the uncertainties discussed in Chapter 4 change the simulation model itself and significantly alter performance estimates. This may prohibit the optimisation algorithm from being able to effectively optimise the objectives, or the operating policy from having sufficient flexibility to perform well in the face of so much uncertainty. Secondly, being robust to all sources of uncertainty may prevent the solution provided by RO from performing effectively in any specific scenario. Thirdly, each iteration of the optimisation algorithm in RO requires running simulations against a very large ensemble of scenarios, thus robust optimisation has been prohibited by limits in computing power until recent years.

In our experience working with water companies (building the model presented in Chapter 4) and based on the survey results (presented in Chapter 3) the decision-making process is rarely automated. Because of this, practitioners aren't necessarily interested in picking an individual solution to 'replace' the decision maker – rather the value of optimised operations is in scoping their available options and exploring the system model. Yet, all metrics of robustness to the author's knowledge define it for individual solutions only. This disconnect is magnified in the multi-objective case where thousands of solutions may be present – focusing on individual robust solutions in a multi-objective context seems to defeat the entire point of having a set of Pareto solutions. *Trindade et al.* [2017] solves this to some extent in comparing the robustness of two set of Pareto solutions (one attained using RO, the other using robustness via re-evaluation), by measuring each solution's robustness for each objective and then plotting the cumulative frequency of a solution's robustness in the two sets of Pareto solutions. However, in this case the robustness metric is still measured on a solution-per-solution basis and the comparison is still made on an objective-by-objective basis (although this last point is not a problem in Trindade's case specifically since one set of solutions vastly outperforms the other on all objectives).

The novel contribution of this work is robust optimisation and robustness via re-evaluation on a case study to:

- Determine whether robust optimisation is able to overcome the potential barriers listed above for structural and contextual uncertainty (i.e. is able to identify an optimum in a large uncertain space, to do this without compromising performance in any specific scenario). To do this we compare robust optimisation with robustness via re-evaluation.
- Develop and demonstrate a metric to compare the robustness of sets of Pareto solutions that does not require an objective-by-objective comparison (because this defeats the point of having a set of Pareto solutions) and is not measured on a solution-per-solution basis (because practitioners are not necessarily interested in picking individual solutions).

5.2 Methodology

In this section we first present our methodology for robustness via re-evaluation and robust optimisation under structural and contextual uncertainty. This methodology builds upon the 'Rival Framings Framework' that was introduced by *Quinn et al.* [2017] and expanded to include structural and contextual uncertainty in Chapter 4. We then discuss how to measure the robustness of a set of solutions, instead of a single solution, introducing a new metric that we term 'hypervolume-regret', and demonstrating it on a benchmark problem.

5.2.1 Robust optimisation and robustness via re-evaluation under structural and contextual uncertainty

In the introduction we presented two philosophies for attaining robust solutions: robustness via re-evaluation takes solutions and evaluates them under uncertain scenarios to identify the most robust, while robust optimisation directly incorporates uncertain scenarios into the objective values and thus optimises for robustness. In this work we demonstrate both philosophies. We expect robust optimisation should be more effective at producing robust solutions than robustness via re-evaluation, we compare the two to see if this is true. Below we present a workflow for robustness via re-evaluation under structural and contextual uncertainty. We then build robust optimisation into this workflow. For clarity, we present this workflow for the single-objective case here and will expand it to the multi-objective case in the following section. This workflow is depicted in Figure 5.1 and explained thereafter.

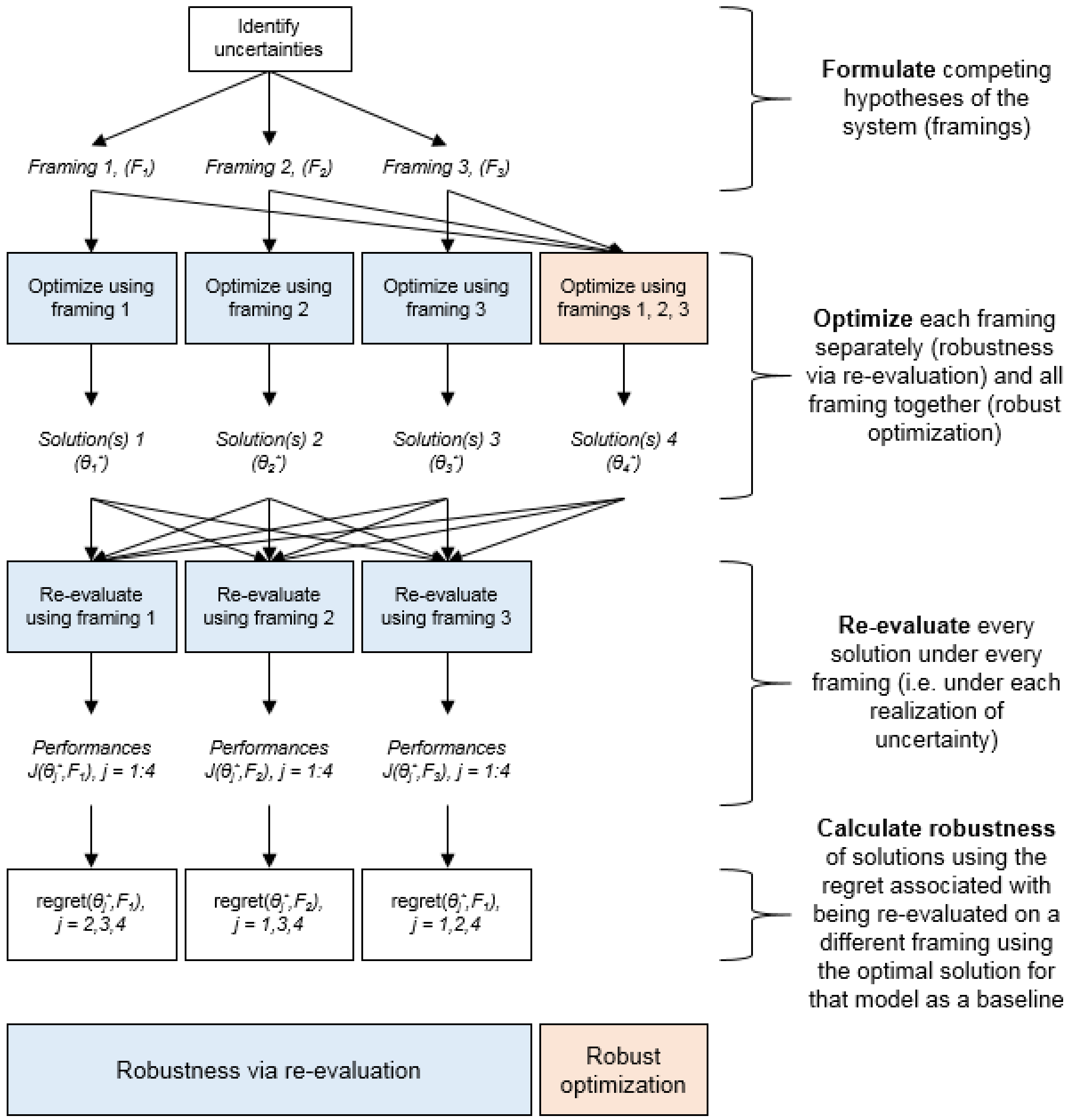


Figure 5.1: Our methodology to compare the robustness of solution(s) from robust optimisation (RO) and robustness via re-evaluation. In this example uncertainties are formulated into 3 plausible framings – each with its own optimised solution(s) under robustness via re-evaluation (blue boxes). We also show a 4th solution(s), which has been optimised for all 3 models simultaneously for robust optimisation (orange boxes)). 'Processes' are shown in boxes, while 'outputs' are shown in italics.

The 'Rival Framings Framework' developed by *Quinn et al.* [2017] and adapted for structural and contextual uncertainty in Chapter 4. The first three steps of the framework with an additional 'measure robustness' step provides a basis for achieving robustness via re-evaluation:

- Formulate uncertainties into competing hypotheses (framings) of the system. In our case framings are distinguished by the set of assumptions made about structural and contextual uncertainty. As described in Chapter 4, these framings should form scenarios that cover the uncertain space of interest.
- Perform optimisation under every framing to find the optimal solution for each, as described by equation (5.1),

$$\min_{\theta} J(\theta, F) \tag{5.1}$$

, where the J is the performance of a solution, θ, when evaluated on the framing used for optimisation, F. Possible optimisation algorithms are reviewed in Chapter 2.

- Re-evaluate each solution under every framing, this will result in objective values for every solution on every framing.
- These objective values are then used to measure robustness. In this study we define robustness as the distance between a given solution's performance in an uncertain scenario and the optimal solution in that scenario (regret hereafter).

We choose the regret metric because, as described in *J W Hall et al.* [2012], it helps to focus the decision-maker's attention to situations where the decision makes a difference – it essentially has the benefit of putting robustness into context, thus making it a useful metric for comparison. We build on *Savage* [1954]'s formulation of regret, as described in equation (5.2),

$$regret(\theta, F_i) = J(\theta, F_i) - J(\theta_i^*, F_i) \tag{5.2}$$

, where J is the objective value of a solution, θ, evaluated in a framing, F. F_i is the framing used for re-evaluation and so θ_i^* is the optimal solution in that framing. θ is the solution for which regret is being measured. This considers single objective formulation, which we expand to multiple objectives in the following section.

The workflow presented above is expanded to include robust optimisation in which the epistemic uncertainty is directly incorporated into the optimisation stage. The robust optimisation is an alternative to the second step above that optimises to find the optimal solution for all framings at once rather than for any specific framing. To achieve this, the objective value across all framings is aggregated into a robustness metric and optimised. As demonstrated in *Kwakkel et al.* [2016], in principle there are many different robustness metrics that could be used to aggregate the objective values across the framings (for example, worst case or signal-to-noise ratio) – however for simplicity we aggregate using the mean objective value over all framings. We do not directly optimise for regret (as would be preferable since this is our overall goal) since the optimal solution for each framing would not be available during optimisation (i.e. robust optimisation should be possible without having to individually optimise for every framing). The robust optimisation problem is described in equation (5.3),

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} J(\theta, F_i) \tag{5.3}$$

, where N describes the number of framings, and J is the objective values of solution θ in framing, F. J can be vectorised (**J**) to change the equation into a multi-objective optimisation problem, as discussed in more detail in Chapter 2.4.2.

5.2.2 Measuring the robustness of a set of Pareto solutions by a new hypervolume-regret metric.

As described in the introduction, a robustness metric measures the (single objective) performance of a solution under uncertainty, for example see *Kwakkel et al.* [2016] for a comparison of many possible metrics. However, we perform multi-objective optimisation which results in a set of Pareto solutions. Thus, we require a robustness metric that does not focus on single objectives, and thus will 'miss' the multi-objective nature of the problem, but instead considers multiple objectives simultaneously. Because decision makers are not necessarily interested in picking individual solutions, the metric should also measure the robustness of sets of solutions rather than individual solutions.

To the author's knowledge, the robustness of sets of Pareto solutions has only been compared by *Trindade et al.* [2017]. Trindade calculates robustness objective-by-objective and on a solution-per-solution basis, thus not meeting the requirements we set out in the previous paragraph. However, there is an active literature that studies different methods of comparing sets of solutions yielded by different multi-objective optimisation algorithms [*Auger et al.*, 2009; *Reed and Kollat*, 2013; *Li et al.*, 2014]. These studies generally acknowledge that a hypervolume-based indicator is the most effective metric to measure and compare the performance of a set of solutions in a multiple-objective space. Hypervolume is defined as the N-dimensional volume (where N is the number of objectives) dominated by the objective values of the set of solutions. We illustrate the concept by presenting a two-dimensional (N=2) example in Figure 5.2a.

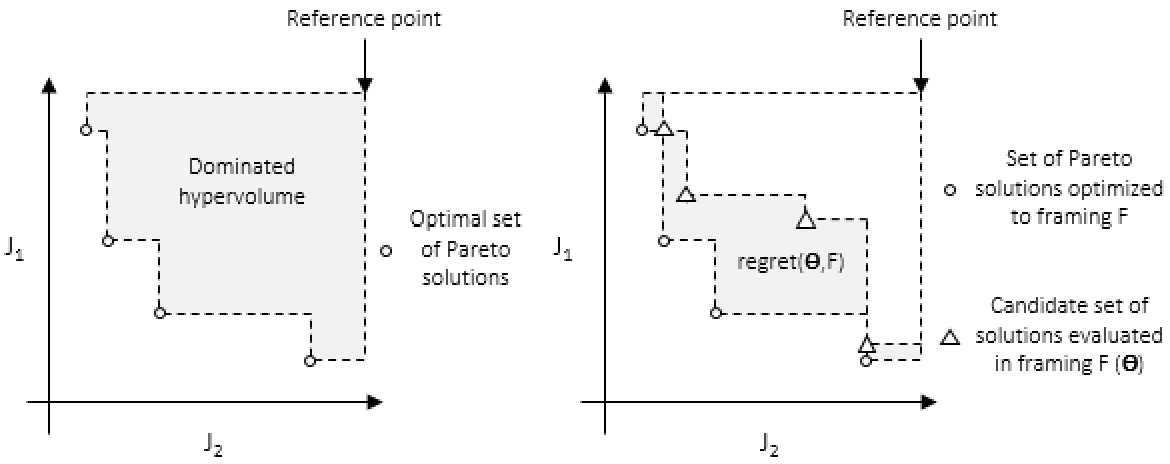


Figure 5.2: (a, left) the hypervolume, shaded in grey, dominated by a set of Pareto solutions for two objectives: J_1 and J_2. (b, right) the hypervolume-regret, shaded in grey, of a candidate set of solutions (**Θ**) relative to the set of Pareto solutions optimised to framing F. In both plots, objectives are to be minimized.

Hypervolume-based indicators are generally acknowledged to have some weaknesses. Firstly, the slow computational speed of calculating the metric has, thus far, limited hypervolume indicators for use as an objective during optimisation [*Reed et al.*, 2013], but has not prevented its application for comparative purposes (for example, in comparing the performance of optimisation algorithms) in which fewer evaluations of the metric are typically required. Secondly, the choice of reference point can lead to inconsistent evaluations of solution sets (i.e. the metric might prefer a different solution set depending on the reference point) [*Auger et al.*, 2009; *Cao et al.*, 2015]. We have repeated our analysis with multiple reference points to ensure this does not influence results. Finally, interpreting the metric is not obvious, owing to it being a N-dimensional volume where each objective is normalized [*Reed et al.*, 2013]. This makes the metric most effective when contextualised, i.e. when used to compare different sets of solutions (via differences in hypervolume) instead of evaluating the performance of a single set of solutions (via the hypervolume value itself), explaining its popularity in the comparison of different MOEAs.

A robustness metric that uses hypervolume will be able consider a set of solutions with multiple objectives simultaneously. If it is used to compare sets of solutions, it should not face any of the difficulties presented above. As discussed in the previous section, we will be using the regret metric of robustness for comparison. Thus, we formulate the 'hypervolume-regret' metric that measures the loss in hypervolume for a candidate set of solutions relative to the optimal set of solutions in a given framing, as demonstrated in Figure 5.2b and described in equation (5.4),

$$HV.regret(\boldsymbol{\theta}, F_i) = HV\big(\boldsymbol{J}(\boldsymbol{\theta}, F_i)\big) - HV\big(\boldsymbol{J}(\boldsymbol{\theta}_i^*, F_i)\big) \tag{5.4}$$

, where **J** is the multiple objective values of every solution in the set of solutions, **Θ**, evaluated in a framing, F. i is the framing used for re-evaluation and so $\mathbf{\Theta}_i^*$ is the set of Pareto solutions optimised in that framing. **Θ** is the solution for which regret is being measured. To update our workflow in Figure 5.1, simply replace 'regret' with 'HV.regret' and use solutions (**Θ**) with their associated objective values (**J**).

5.2.2.1 Demonstration of hypervolume-regret

In this section we present an example of hypervolume-regret, to test its usefulness as a tool for comparison. It is presented alongside Cumulative Distribution Functions (CDFs), used in [*Trindade et al.*, 2017], which is the only example of comparing the robustness of sets of solutions that we could identify. We also present other methods of comparing sets of solutions: Parallel Coordinates Plots (PCPs), used in *Quinn et al.* [2017], 2D and 3D 'slices' of the set of Pareto solutions, used in Chapter 4, Figure 4.8, (2D) and *Zeff et al.* [2014] (2D and 3D).

The example is performed on the DTLZ2 function (introduced by *Deb et al.* [2002]), which is commonly used as a benchmark for multi-objective optimisation (specifically minimization, since all objectives are to be minimized). This function has a user specified number of decision variables and objectives. We use 3 objectives and 4 decision variables. By the function's design, with 3 objectives and 4 variables, any solution that sets the final 2 decision variables equal to 0.5 will be on the Pareto front. We compare 3 sets of solutions:

- The set of Pareto solutions from 10,000 sets of randomly sampled parameters (denoted 'random sampling'), we include this to help contextualise the performance of the following Borg MOEA to improve interpretability. We would expect this set to perform worst.
- A set of solutions acquired through use of the Borg MOEA [*Hadka and Reed*, 2013] with 10,000 iterations (denoted 'Borg MOEA'). We describe this minimization problem in equation 5.5:

$$\min_{\boldsymbol{x}}[J_1(\boldsymbol{x}), J_2(\boldsymbol{x}), J_3(\boldsymbol{x})] \tag{5.5}$$

- A set of 10,000 solutions that have the final 2 decision variables equal to 0.5, which we would expect to perform best (denoted 'exact').

The results are presented below in Figure 5.3 and discussed thereafter.

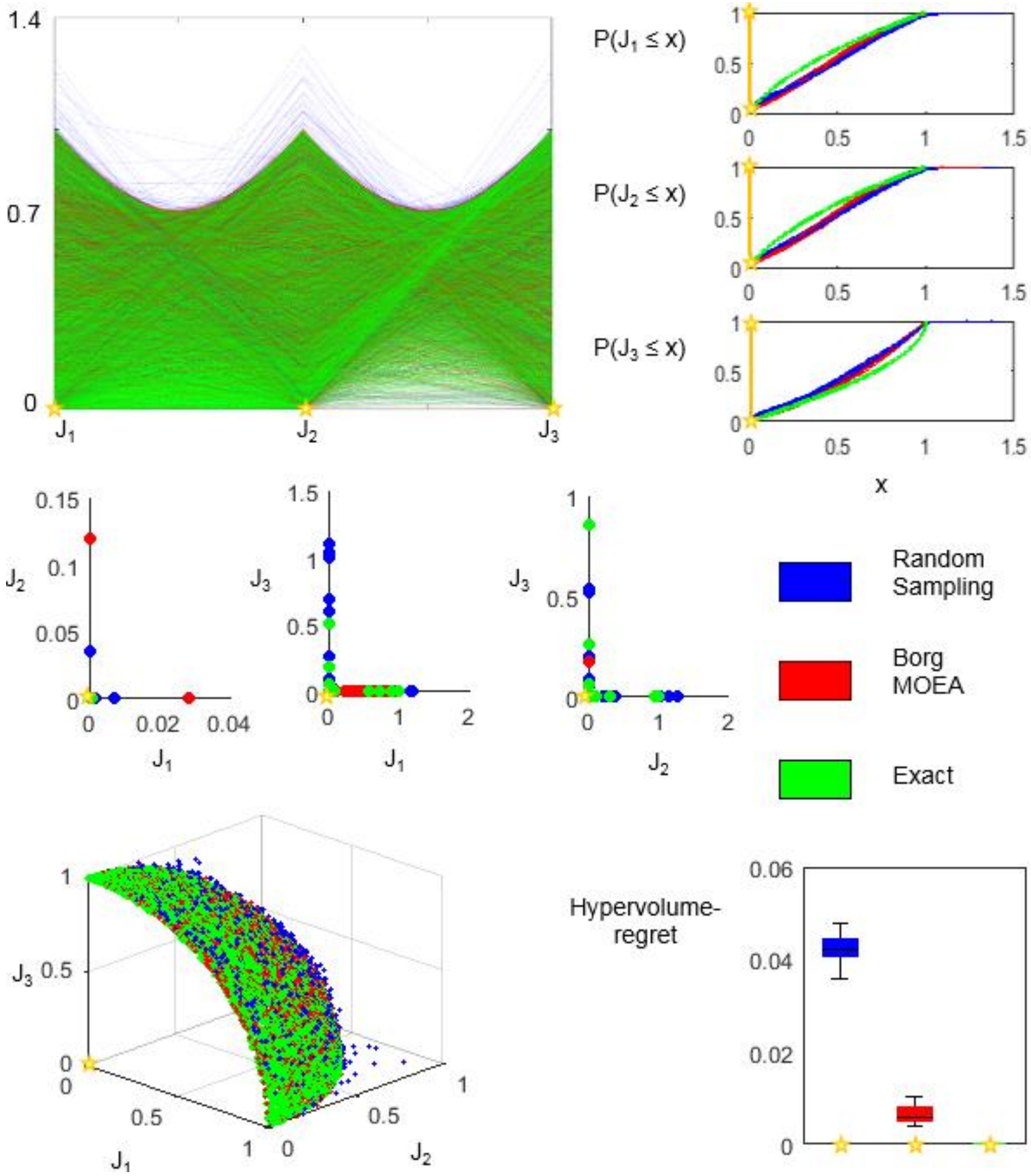


Figure 5.3: (Top row, left) A parallel coordinates plot of the 3 sets of Pareto solutions (blue: produced by random sampling; red: produced by the Borg MOEA; green: analytical solutions). The point at which a line intersects each Y-axis indicates the value delivered by the solution for the objective labelled on that axis(Top row, right) cumulative distribution function of each of the 3 objectives for each set of Pareto solutions. (Middle row, left) The three plots each show a pairwise comparison between two of the three objectives. Only solutions that form the set of Pareto solutions between the two objectives in a given set of solutions are included for clarity. These plots can be considered '2-objective slices' of the 3-objective space. (Bottom row, left) A 3D

scatterplot showing each of the three sets of Pareto solutions in the full 3-objective space. (Bottom row, right) The hypervolume-regret of all sets of solutions, relative to the 'Exact' set of solutions. The experiment was repeated 20 times to account for the variability in the random sampling and optimisation process – the quantiles and extremes of hypervolume-regret from these repetitions is shown using boxplots. On all plots except the CDFs, the 'ideal' points are indicated by yellow stars. The 'ideal' CDF is indicated by the yellow line between two yellow stars.

The 'exact' (green) objectives are very dense on the PCPs (top left), indicating that they densely cover the objective space. There are also a few 'random' (blue) solutions that have higher objectives than most other solutions, indicating there are some poor solutions in that set.

The CDFs (top right) distinguish between the 'exact' and 'random'/'Borg' solutions because the green line is noticeably separate from the red and blue lines. Interestingly the 'exact' set of solutions is further away from the 'ideal' point (red star) on the CDF of objectives for J_3 (even though we expect it to be the 'best' set of solutions).

The pairwise plots (middle row) suggest that no trade-off exhibits between any two objectives. This is because trade-offs in the DTLZ2 function exist only in the Nth dimension, where 'N' is the number of objectives. This behaviour of the DTLZ2 demonstrates why 'slicing' a Pareto front to make comparisons may not be sufficient to compare the robustness of sets of solutions.

These trade-offs become apparent in the 3D scatterplot (bottom left). Although, besides the 'exact' solutions densely covering the trade-off front, it is hard to tell the difference between the 'random' and 'Borg' solutions in this plot (with the exception of a few 'random' solutions that are notably worse, as with the PCPs).

The hypervolume-regret metric clearly highlights the differences between the sets of solutions, showing us the expected result, i.e. that 'exact' outperforms 'Borg', which outperforms 'random'. We also believe that both the use of a relative metric, and the comparison with random sampling alleviates problems of interpreting results: for example, 10^{-2} of 4-dimensional hypervolume would be meaningless on its own but is meaningful when viewed in the context of regret and the 'random sampling' set of solutions. Thus, we will adopt hypervolume-regret to demonstrate the robustness of sets of solutions but will include a comparison with random sampling.

5.3 Experimental setup for the case study application

The methodology described in Section 5.2.1, using the hypervolume-regret metric described and exemplified in Section 5.2.2, is demonstrated on the case study introduced in Chapter 4. This is a multi-reservoir, multi-objective water resources system in the South West of the UK. The operators control releases and abstractions with an aim to minimize pumping costs and water supply deficits. Although the system is multi-operator we assume that decisions are made cooperatively between the two companies, although the impact of doing otherwise is shown in Chapter 4. The decision-making behaviour of operators is represented by 'operating policies'. In our study, these policies are universal approximation functions that control decisions in the system. The optimisation problem aims to minimize objectives by iteratively adjusting the operating policies. This is achieved by altering their parameters. The simulation and operation formulations are described in Chapter 4, in this study they are unchanged.

We do, however, consider different structural and contextual uncertainties, summarizing the problem framings that cover the uncertain space in Figure 5.4 and described in the following paragraph. We do not consider pump failures or the fisheries releases because Chapter 4 showed them to not impact estimates of performance. We do not consider cooperation/non-cooperation, as in Chapter 4, because creating an operating policy that could be robustly optimised to this uncertainty is not possible due to the non-cooperative framing being controlled with two separate policies.

Framing	Inflow, Demand, Evaporation	'Storage threshold'
1: AR0-Regular (Reg)	AR(0)	Regular
2: AR2-Reg	AR(2)	Regular
3: AR2-Low	AR(2)	Low

Figure 5.4: The framings formulated for this study. Each row is a different framing. The Inflow, Demand, Evaporation column denotes whether an AR(0) or AR(2) model is used to generate these inputs. The 'storage threshold' is the storage level at which the water company decides that storage should not drop below.

Auto-correlation in forcing. Chapter 4 has already demonstrated that the amount of auto-correlation in forcing is a structural uncertainty that has operational significance. We consider both an AR(0) and an AR(2) forcing model.

Minimum target storage. The companies that operate this system suggest that operation optimisation would only be possible provided that it does not draw down the reservoirs' storage below an acceptable threshold. Because this threshold is typically dependent on a range of factors that exist beyond the modelled system, we consider it an excellent example of contextual uncertainty. We include a modelling choice of a 'regular' threshold that fits the historic 5th percentile of daily storages, and a more optimistic 'low' threshold that is equal to half of the 'regular' threshold. We present these thresholds against the historic storage in Figure 5.5. In the model, when storage drops below a threshold, decisions are only permitted to be made at the minimum required flow.

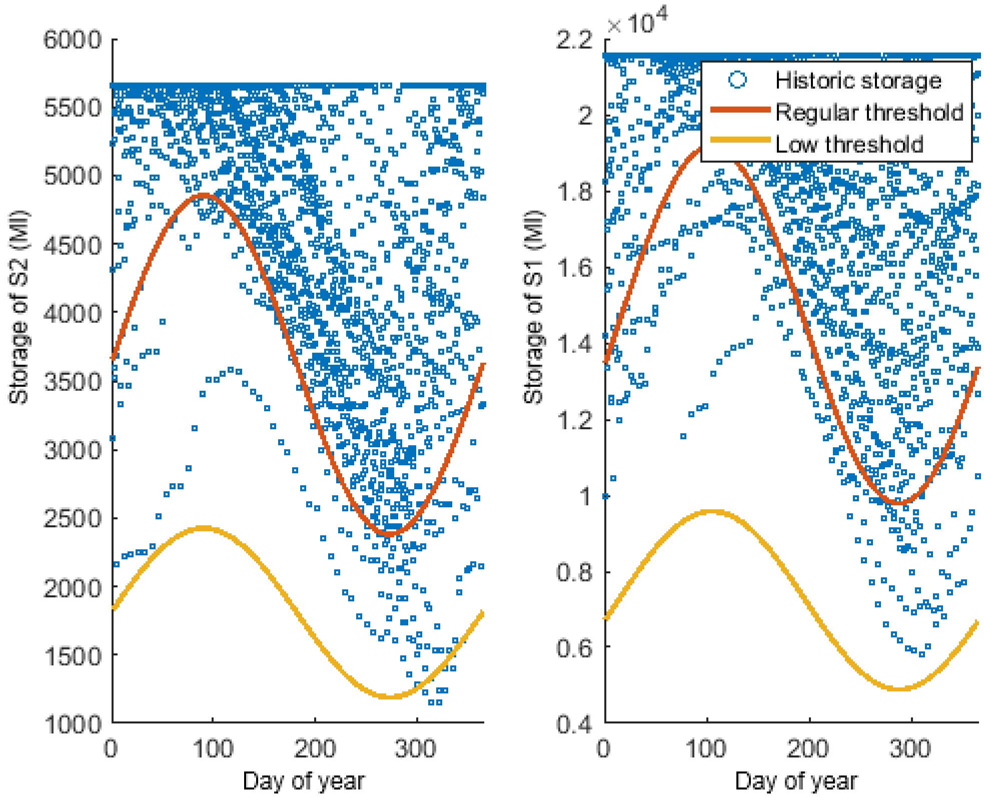


Figure 5.5: Weekly historic storage of reservoir S2 between 1975-2016 (left). Weekly historic storage of reservoir S1 between 1979-2016 (right). The regular threshold (red) fits the 5th percentile of storage on a given day, while the low threshold (yellow) is half the regular one.

Since the framings are different from Chapter 4, we re-assess the optimisation-simulation parameters; i.e. the determination of sample size and the of the number of function evaluations to be used in optimisation. The results of these are not shown since they are indistinguishable from those presented in Chapter 4, resulting in a sample size of $1.2*10^7$ days and an optimisation process with 10^5 iterations. We note here that the sample size used in robust optimisation (i.e. the aggregated ensemble containing all three framings) is the same as the sample size of the other framings, thus, it uses a sample size of $4*10^6$ days for each of the three framings.

As anticipated in Section 5.2.2.1, we will also compare the robustness via re-evaluation and robust optimisation results with random sampling. A set of solutions from random sampling uses the same number of model evaluations as the optimisation process, but it selects the parameters of the policies randomly rather than with the Borg MOEA guided search. This contextualises the regret of the optimised solutions by placing it next to the regret associated with not using a guided search (instead simply choosing random points).

5.4 Results

Figure 5.6 presents the hypervolume-regret (Y-axis), introduced in Section 5.2.2, for five sets of Pareto solutions (columns) when evaluated in each framing considered in our study (panels). The first three sets of Pareto solutions are solutions that were optimised to each framing separately (AR0-Regular, AR2-Regular and AR2-Low) and then re-evaluated against the other framings, according to the 'robustness via re-evaluation' philosophy presented in the introduction and described in Section 5.2.1. The fourth set of Pareto solutions was obtained by including all framings in the optimisation process, falling under the 'robust optimisation' philosophy. The fifth set of solutions was created by randomly sampling the parameter space of the operating policies – without optimisation - and selecting the Pareto dominant solutions in the sample. These sets of Pareto solutions were created using the respective framing for re-evaluation, i.e. the Pareto dominance of a set of solutions was checked by looking at the performance estimates obtained in framing AR0-Regular (top row, fifth column), AR2-Regular (middle row, fifth column) and AR2-Low (bottom row, fifth column).

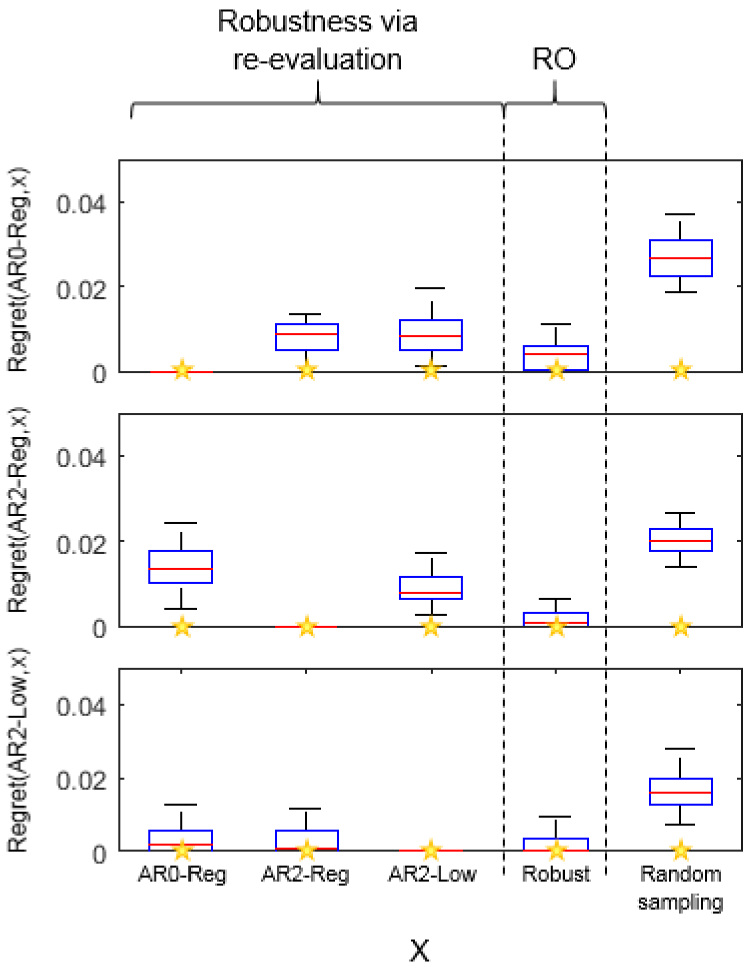


Figure 5.6: Hypervolume-regret plots from the re-evaluation stage. Each panel presents the results with a different framing used for re-evaluation. The Y-axis is the value of the hypervolume-regret metric. Regret is displayed in a boxplot with the extremes, median and 25th and 75th percentiles marked over 21 repetitions of the experiment to account for intrinsic variability in the optimisation process. The X-axis indicates [1-3] the framing used for optimisation in robustness via re-evaluation, [4] that robust optimisation (RO) was used (i.e. optimised over all framings), [5] that random sampling was used in place of a guided search algorithm. Yellow stars indicate the 'ideal' point for each box.

In Figure 5.6 we see that robust optimisation (fourth column) has the lowest regret under all framings. Regret is particularly low when re-evaluating in the AR2-Regular and AR2-Low framings (middle and bottom panel) and is only slightly increased when evaluated in the AR0-Regular framing (top panel). Of the solutions obtained by Robustness via re-evaluation, the ones that were optimised under the AR0-Regular and AR2-Regular framings (first and second columns) have a larger regret when evaluated in each other's framings (middle and top panels), but a relatively low regret when evaluated in the AR2-Low framing (bottom row). The set of Pareto solutions optimised in the AR2-Low framing (third column) has high regret in both the AR2-Regular and AR0-Low framings (top and middle rows). Not surprisingly, the set of solutions obtained by random sampling (fifth column) has the highest regret in all framings. This is the expected result given that random sampling should perform worse than optimisation because it makes no attempt to improve solutions based on their objective values.

The differences between sets of Pareto solutions are, in many cases, quite large. Robust optimisation (fourth column) identifies zero-regret sets of Pareto solutions in 52%, 48% and 24% of experiment repetitions for the AR2-Low (bottom row), AR2-Regular (middle row) and AR0-Regular (top row) framings respectively. Meanwhile, the robustness via re-evaluation sets of Pareto solutions (first three columns) only achieve zero-regret for the AR2-Low framing (bottom row) and the framing that is used for optimisation (by definition).

By further analysing the regret of the solutions obtained by 'robustness via re-evaluation', we can gain insights about which sources of uncertainty cause regret and which do not. We can see that the choice of the level of autocorrelation in forcing inputs is very important. For example, the set of Pareto solutions optimised under the AR0-Regular framing has a high regret when re-evaluated in AR2-Regular (middle row, first column) and almost as high as the regret of the solutions obtained by random sampling (middle row, fifth column). In other words, all the gains delivered by optimisation are lost if the choice of the AR0 model proves incorrect. We see that contextual uncertainty (around the threshold of acceptable storage) significantly influences regret because the set of Pareto solutions optimised in the AR2-Low framing performs worse when evaluated in the AR2-Regular and AR0-Regular framings (third column, top and middle rows). We also see that all sets of Pareto solutions that are obtained by optimisation (i.e. the first four columns) have low regret when re-evaluated in the AR2-Low framing (bottom row) – implying that, for this case study and set of framings, optimising under the less optimistic (regular) assumptions about the storage threshold does not appear to have a downside.

As anticipated in Section 5.2.2.1, we believe that hypervolume regret is the most effective way to compare the robustness of sets of Pareto solutions. However, we also presented alternative visualisation techniques that can be, or have been, used to compare the robustness of sets of Pareto solutions on an objective-by-objective basis. Because the hypervolume regret metric aggregates across the objectives in a set of Pareto solutions, we may miss interpretation that could have been gained by not aggregating. Therefore, in Figure 5.7 we present the CDFs of the objective values for each set of Pareto solutions. Given that all the objectives are to be minimised, we would ideally like CDFs to be as close as possible to the X=0 axis (shown with the yellow stars). We see that the closest sets of Pareto solutions to the X=0 axis are the ones optimised in AR2-Regular (red lines) and using robust optimisation (black lines) for objective J_{D2} (top row), and the ones optimised in AR0-Regular (blue lines) and AR2-Low (green lines) are closest for the other objectives J_{D1}, $J_{£2}$, $J_{£1}$ (bottom three rows). These observations are the same in every framing used for re-evaluation (i.e. in every column). Thus, from Figure 5.7 we see clear differences between the different framings and might expect that the sets of Pareto solutions optimised to AR0-Regular and AR2-Low are more robust, this conclusion is different from what was drawn from Figure 5.6.

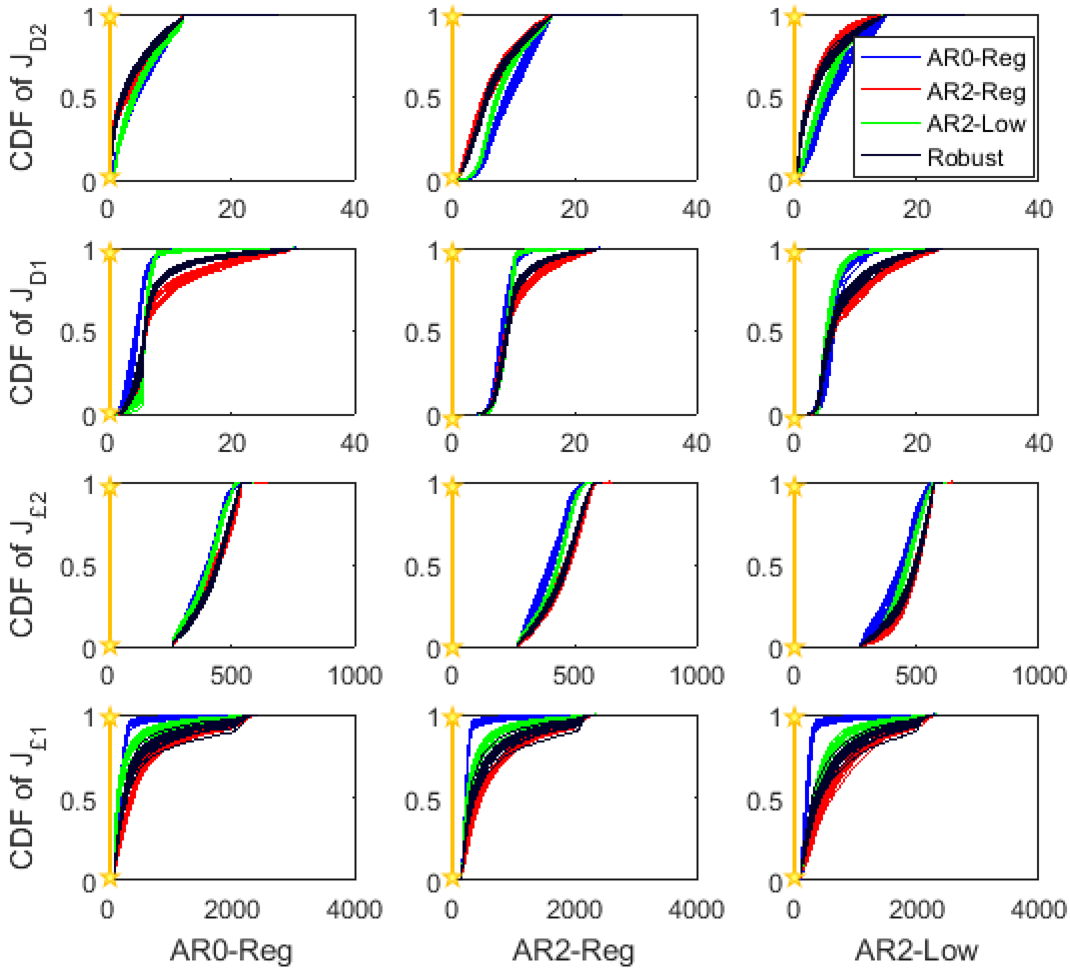


Figure 5.7: CDFs of the performances of sets of solutions, where the Y-axis is the cumulative probability of a given X-axis objective value occurring within a given set of solutions. Each row indicates a different objective. Each column indicates a different framing used for re-evaluation. Coloured lines (red, green, blue) represent operating policies created using different framings for optimisation under robustness via re-evaluation. Black lines represent the operating policies created from the robust optimisation. Each line is a different set of solutions, there are multiple lines because we show the results for multiple repetitions of the experiment to account for variability in the optimisation process. Filled circles indicate the location of the 'most robust' solution within a given set of solutions, where robustness is measured by signal-to-noise ratio. The 'ideal' CDF is indicated by the yellow line between two yellow stars.

For comparison with the plots presented in Section 5.2.2.1, we also include the PCP plots and 2D 'slices' in Appendix B, Figures B.1 and B.2 respectively. The PCP plots show the same observations as Figure 5.7 above (i.e. that the AR2-Regular and Robust perform better on J_{D2} while AR0-Regular and AR2-Low perform better on the other objectives). The 2D ‘slices’ indicate that there are many trade-offs between different objectives but do not distinguish between different sets of Pareto solutions.

Finally, in Figure 5.8 we visualise our sets of Pareto solutions via the cumulative frequency plots presented in *Trindade et al.* [2017] and discussed in the introduction and Section 5.2.2.1. To the authors knowledge, these are the only plots that have been used so far to compare the robustness of sets of Pareto solutions. However, as with the CDFs shown in Figure 5.7, these are still calculated objective-by-objective. In the left column of this figure, for each objective (rows) and each set of Pareto solutions (an individual line), the robustness of each solution (Y-axis) is ranked-ordered (X-axis) and plotted as a cumulative frequency. The robustness of a solution for an objective is calculated as the signal-to-noise ratio (normalized across all framings used for evaluation, as used in *Kwakkel et al.* [2016]. We define the signal-to-noise as the mean multiplied by the standard deviation of the normalized objective values of a solution over all framings. Unlike Trindade, this robustness metric should be minimized, therefore more robust solutions will have lower values on the Y-axis. As with Figure 5.7, the sets of Pareto solutions optimised in the AR0-Regular framing (blue lines) perform best, i.e. are closest to Y=0, for objectives J_{D1}, $J_{£2}$ and $J_{£1}$ (bottom 3 rows). Differently from Figure 5.7, the set of Pareto solutions optimised in AR2-Low (green lines) performs more similarly to the robust optimisation (black lines) rather than the AR0-Regular (blue lines) sets of Pareto solutions in J_{D1}, $J_{£2}$ and $J_{£1}$, while the AR2-Regular (red lines) set of Pareto solutions performs worst in every objective. These differences in interpretation result from the skew that occurs by plotting the absolute cumulative frequency rather than the distribution that is relative to the number of solutions in a set of solutions. Therefore, we normalize the X-axis by dividing the rank in the cumulative frequency by the total number of solutions in a given set of Pareto solutions. We show these normalized cumulative frequencies in the right column of Figure 5.8, where the interpretations are consistent with Figure 5.7 (i.e. the sets of Pareto solutions optimised in AR2-Regular and robust are closest to ideal in J_{D2} and the sets of Pareto solutions optimised in AR0-Regular and AR2-Low are closest to ideal for J_{D1}, $J_{\pounds 2}$, $J_{\pounds 1}$).

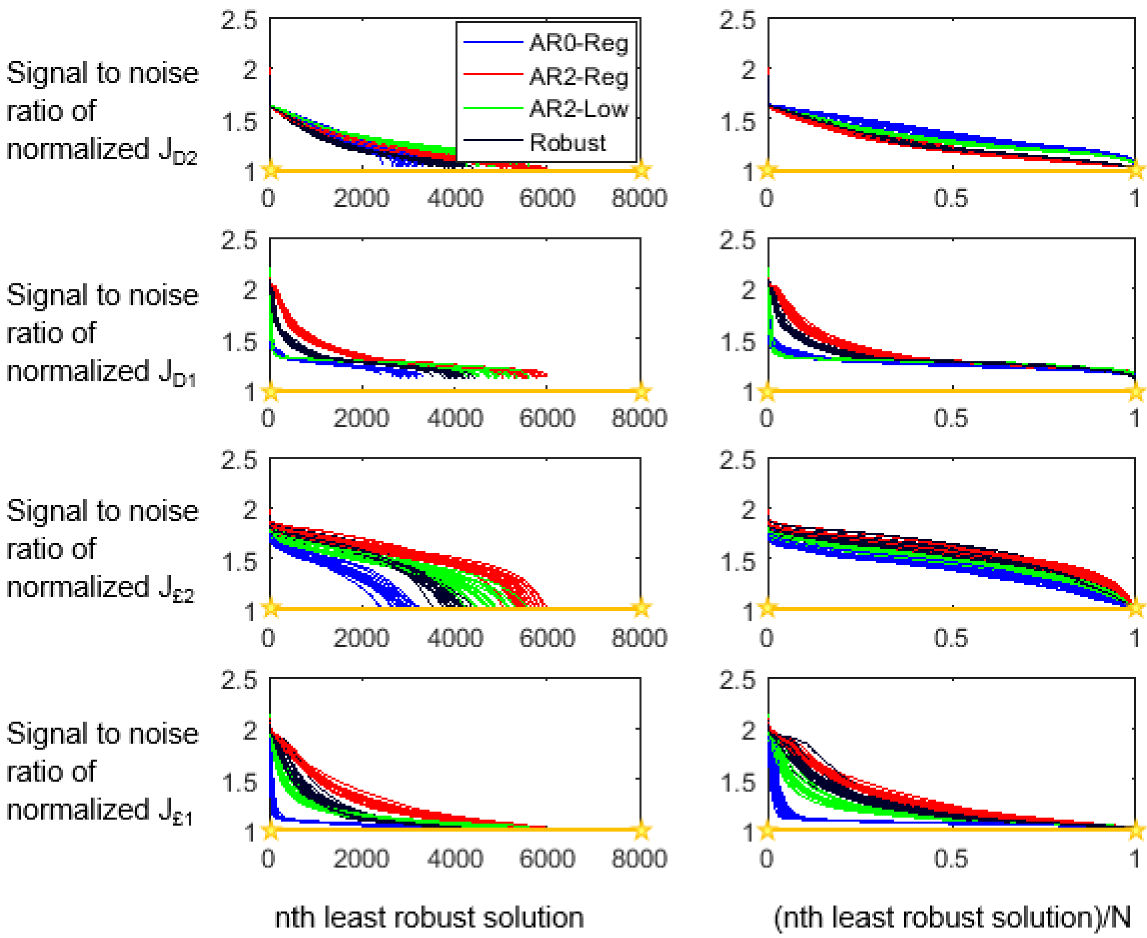


Figure 5.8: (left column) Each line indicates the rank-ordered robustness of each operating policy in a given set of Pareto solutions. An operating policy's robustness is calculated by the signal to noise ratio for a given normalized objective across all framings. Each colour (red, blue, green) represents a different framing used for optimisation in robustness via re-evaluation. Black lines represent the robust optimisation policies. The multiple lines of the same colour represent repetitions of the experiment to account for variability in the optimisation process. These plots are based off plots from *Trindade et al.* [2017], although it should be noted that the signal to noise metric is to be minimized (rather than maximised as in the metric for robustness in their study) so points

following the Y=0 line are ideal, is indicated by the yellow lines between yellow stars. (right column) The same but normalized to 'N', where N is the number of solutions in a set of solutions.

5.5 Discussion

A key aim of this study was to determine whether robust optimisation is possible and effective for structural and contextual uncertainty. In Figure 5.6 we find that robust optimisation can create a set of solutions that satisfies the objectives under both structural and contextual uncertainties. In the introduction we described why one might have detracted from using robust optimisation: there is too much uncertainty to find good solutions, aiming to be robust to all uncertainty would prevent effectiveness in any specific scenarios, and computational cost (which we ignore in the following). We find that, in our case study, the first point is not true. The low regret of the robust optimisation set of Pareto solutions shown in Figure 5.6 makes it clear that the function describing the operating policies is sufficiently flexible to accommodate a robust solution and the uncertainty does not prohibit the Borg MOEA from optimising the more complex objectives. The second point is true to some extent because the robust optimisation has some regret when evaluated in the AR0-Regular framing (although it has less regret than either AR2 framing). However, in the other cases, when evaluated in the AR2-Regular and AR2-Low framings, robust optimisation has very low regret. Thus, in this case our results would suggest a more detailed study of the flow regime to identify which statistical models of the inflows are more likely.

From carrying out robustness via re-evaluation we have revealed some interesting insights about the system. For example, using the regular acceptable storage threshold during optimisation does not appear to have any associated regret while being more optimistic and selecting a lower threshold does. Meanwhile neither AR0 or AR2 are low regret assumptions since the sets of Pareto solutions optimised to the AR0-Regular and AR2-Regular framings both have regret when evaluated in each other's framings. The hypervolume-regret metric is equivalent to the hypervolume of objective space that cannot be accessed by optimisation due to the differences between two framings. Thus, it demonstrates how assumptions about different uncertainties constrain the multi-objective space. This provides a unique interpretation about the impacts of the uncertainties in question on the water resources simulation model. These results can direct modellers towards identifying which assumptions matter and thus which uncertainties are beneficial to reduce.

In the introduction we stated that the part of the value of optimisation was in *scoping available options and exploring the system model*. The above observations about the system model are not possible to draw from the CDFs shown in Figures 5.8 and 5.9 (or other plots shown in Appendix B). Although the CDFs detected that uncertainties had an impact, their interpretation was different from the hypervolume-regret interpretation. The CDFs could not identify that choosing a specific AR model during optimisation would result in regret, or that being less optimistic about the acceptable storage would not. Nor could they identify that robust optimisation was resulting in a low-regret set of solutions. This supports the value in using our proposed hypervolume-regret metric.

Although outside the scope of this study, it is important to note that robustness often concerns not 3 uncertain scenarios as we have used, but 1000's of uncertain scenarios. Each scenario realizing a different instance of parametric uncertainty. *Watson and Kasprzyk* [2017] and *Trindade et al.* [2017] demonstrate robust optimisation for a sample of 1000's of uncertain scenarios that contain parametric and aleatory uncertainty, however we see no reason this sample could not also contain different instances of structural/contextual uncertainty. To compare the robustness of sets of Pareto solutions re-evaluated in 1000's of scenarios is certainly possible since modern hypervolume evaluation tools are highly efficient [*Fonseca et* *al.*, 2006]. The next logical step for studies of this type is to combine aleatory and epistemic parametric, structural, contextual and (possibly) objective uncertainties into one study for both robust optimisation and robustness via re-evaluation to determine how important each of these sources of uncertainty are, and whether robust solutions can be identified.

5.6 Conclusion

In this Chapter we have demonstrated robust optimisation for a water resources operations optimisation problem using a hypervolume-based regret metric. This analysis is novel in that it is, to the authors' knowledge, the first to perform robust optimisation of a decision-making problem under structural and contextual model uncertainty. It also appears to be the first to calculate and compare the robustness of sets of solutions (sets of Pareto optimal solutions attained via multi-objective optimisation) as a whole rather than comparing on an objective-by-objective or solution-per-solution basis. We find that comparing the robustness using the hypervolume regret metric provides a unique interpretation of the simulation-optimisation problem that cannot be achieved by analysis of individual robust solutions or objectives, as is typically performed.

CHAPTER 6: SUMMARY AND CONCLUSIONS

6.1 Summary of the research

Chapter 2: A review and classification of reservoir operation optimisation methods

Reservoir operation optimisation aims to determine release and transfer decisions that maximise water management objectives such as the reliability of water supply, the hydropower production, the mitigation of downstream floods, etc. An extensive and growing body of scientific literature exists on advancing and applying mathematical optimisation methods to reservoir operation problems. In this Chapter, we reviewed such literature according to a novel classification system of optimisation approaches, which focuses on the characteristics of the actual operation problem – i.e. what needs to be optimised, or in mathematical terms, the argument of the optimisation problems - rather than the mathematical properties of the optimisation algorithm. This enabled us to further discuss advantages, limitations and the scope of application of the different methods from a more practical perspective. We hence concluded the Chapter with a set of guidelines that should help potential users to match the properties of their system and operation problem with a suitable optimisation method.

Chapter 3: Are reservoir operation optimisation methods used in practice? Responses of a practitioner survey

Despite the extensive scientific literature on mathematical optimisation methods for solving reservoir operation problems, very little is known about the actual uptake of those methods by reservoir operators. In this Chapter, we presented the results of a survey of water resources managers to analyse how reservoir management decisions are made in practice, to determine the level of uptake of reservoir optimisation algorithms and to identify possible barriers to uptake. We reached a range of companies and consultancies in the UK, Australia, South Africa and South Korea. We found that the decision-making process in practice is much more informal and experience-based than is assumed by most scientific reservoir optimisation studies and that practitioners are concerned about the validity of optimisation results due to uncertainty in the underlying simulation models. We conclude that studies aiming at the application of optimisation results to real-world problems are required to build faith in the applicability of optimisation methods to achieve uptake in practice.

Chapter 4: How important are model structural and contextual uncertainties when estimating the optimised performance of reservoir systems?

Uncertainty in simulating water resources systems (WRSs) makes it difficult to assess how effective different water management decisions will be, hence undermining the credibility of simulation and optimisation studies and the uptake of their results. In this Chapter, we identified different sources of uncertainty in WRS models and found that structural uncertainty (i.e. arising from how interrelationships within the system are defined) and contextual uncertainty (i.e. around the definition of the system boundaries) are rarely considered when simulating and optimising WRSs. We proposed a methodology to quantify the effects of structural and contextual uncertainties on the estimated performance of optimised water management decisions, and demonstrated that they have a significant impact on a real-world case study of a pumped-storage system in the UK. To the best of the authors' knowledge this is the first study to consider the impact of these types of uncertainty on optimised operating policies and their simulated performances. Our first key finding is that, of all the considered uncertainties, the assumptions made about context – specifically around the level of cooperation between neighbouring water companies – had the greatest impact on performance estimates. This is important because few WRSs exist in isolation, yet discussion of the effects that a given definition of the system boundaries have on the simulation/optimisation results is uncommon. Our second key finding is that, of the structural uncertainties we analysed, the amount of auto-correlation represented in forcing inputs (reservoir inflows and water demands) had the greatest impact. This implies that seemingly small changes in the statistical characterisation of those inputs can have a significant impact on the estimated performance of optimal operating policies. We also highlighted the significance of adequately considering aleatory uncertainty when validating performance estimates – something that few studies do – and presented a simple technique to justify the sample size used for validation. We believe that, by studying and quantifying the uncertainties present in WRS models more fully, researchers can increase faith in their models and their results, ultimately encouraging their uptake.

Chapter 5: Measuring and achieving robustness in multi-purpose reservoir systems under structural and contextual uncertainty

Chapter 4 showed that ignoring the model structural and contextual uncertainties that unavoidably affect WRS models may lead to optimised solutions not performing at the level anticipated during optimisation. One option to create solutions that are ‘robust’ to these uncertainties is to use ‘robustness via re-evaluation’: optimising to attain a set of candidate solutions that are re-evaluated in uncertain scenarios to identify those that perform effectively under a wide range of scenarios. Robust solutions can also be attained via ‘robust optimisation’: optimising a solution’s performance across a range of uncertain scenarios. In Chapter 5, we applied both approaches to the same two-reservoir system introduced in Chapter 4, to derive a set of operating policies that are robust under structural and contextual uncertainty. This appears to be the first such study in water resources or operations research. Robust optimisation was shown to be effective, exhibited by it producing a set of solutions that are more robust than would be achievable by robustness via re-evaluation. An important element of most water resources systems models is the presence of competing objectives, which means that optimisation yields a set of Pareto optimal solutions rather than a single one. We found that comparing the robustness of sets of solutions (and not individual solutions, as is customary) is essential to both preserve trade-offs between objectives and to understand the impact of the uncertainties on the model predictions.

6.2 Outlook

6.2.1 Promoting reservoir operation optimisation in practice

Our first research question focused on understanding the use and usefulness of reservoir operation optimisation in practice. There are both authors who suggest that reservoir operation optimisation is used in practice [*Loucks et al.*, 1985; *Wurbs*, 1998; *Zagona et al.*, 2001; *Ibanez et al.*, 2014] and those who suggest that it is not used in practice [*Rogers and Fiering*, 1986; *Walski*, 2001; *Harou et al.*, 2009; *Savic et al.*, 2009; *Nicklow et al.*, 2010]. In either case, with the exception of *Rogers and Fiering* [1986], there is a distinct absence of empirical studies. For a practice orientated field, this is a curious state of affairs. Chapter 3 in this thesis has aimed to provide the first structured survey of uptake by practice in 30 years. The results presented here suggest that uptake appears to be low in the surveyed countries (UK, South Africa, South Korea, Australia) and industry (water supply), both in public and private organisations. We would recommend future surveys to target different regions and industries, in particular hydropower, which is assumed to have a higher level of uptake [*Brown et al.*, 2015]. It is also not clear, even among the authors listed above that present a positive view of WRSO research dissemination, *how* the results of these studies shape practice and ultimately the decision-making process.

To understand how the results of studies should shape practice, we believe it will be useful to consider what the practical outputs of research in WRSO are. In Figure 6.1, we highlight typical examples of outputs from WRSO studies that might be expected to be used in practice, and link them to our survey results. Each type of study may provide different outputs, for example the output from a modelling study (top row) might be a bespoke model of an individual water resources system or it might be a generalisable piece of software that can be applied to many systems. Our survey presented in Chapter 3 highlighted that simulation models appear to be the main WRSO output that is used by practice. We discussed in Chapter 3 how simulation is used to answer 'what-if' questions, as opposed to the 'what's best' questions that might be answered by optimisation. Interestingly, even the single case of optimisation being used in practice that we identified (bottom row, right column) was still being applied to a 'what-if' question. In contrast, the implicit aim of WRSO is to create optimal solutions, such as optimal operating policies, which appears to be the least likely research output to be implemented. We believe that it is important to be explicit about what is actually being transferred to practice when investigating the uptake of WRSO.

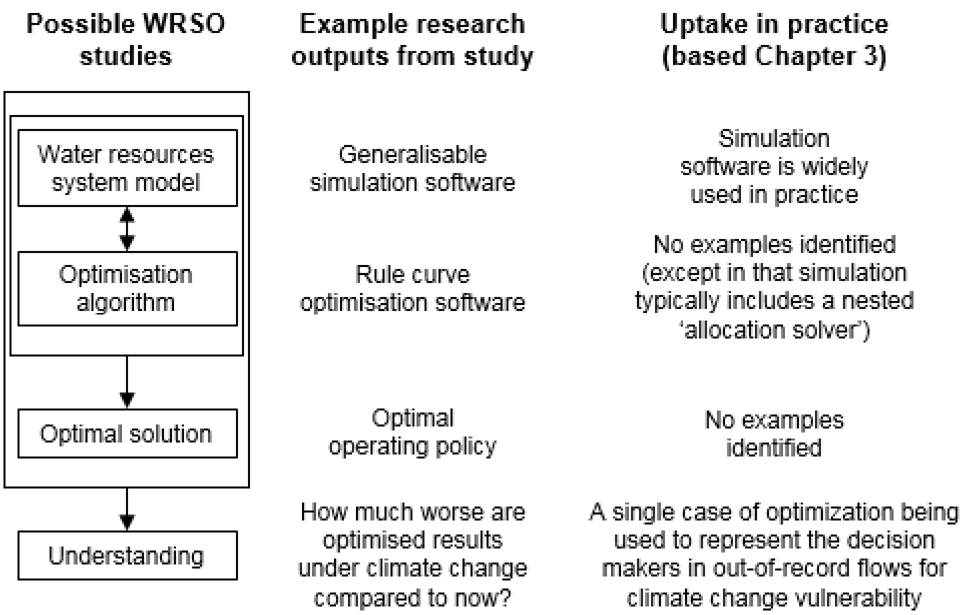


Figure 6.1: Examples of common research outputs of WRSO studies and the extent to which the survey presented in Chapter 3 identified their uptake in practice.

By being specific in this way we believe that future research on the uptake of WRSO outputs can provide the answers to a variety of questions beyond 'was the research output used/not used'. Examples of these questions (adapted from the questions presented in *Greenhalgh et al.* [2004]) are:

- What characteristics of the research output led to its successful/unsuccessful uptake? (For example, was the optimisation software used because it was highly efficient or because it was complemented with good visualization tools?)
- How and why has the success/failure of the uptake changed over time? (For example, why was the operating policy used in normal conditions but ignored during extremes?)
- How did decision makers change their decision-making process when presented with research outputs? (For example, did the understanding provided about the system vulnerabilities lead to investment in recommended backup water sources?)

- Did the water resource system performance improve/worsen after uptake of research outputs? (For example, did the use of the optimised operating policy reduce the amount of money the company was spending on pumping?)
- What external factors have influenced the uptake of research outputs? (For example, did the optimisation software become part of the operation process because the water regulator demanded it or because individuals in organisations championed its use?)

6.2.2 Uncertainty in reservoir operation optimisation

As our survey highlighted, the models used in WRSO must be ‘trustworthy’ before practitioners will consider using optimisation that is underpinned by them. We believe that uncertainty analysis is required to provide some form of ‘evidence base’ that the optimisation results are expected to be effective when applied in reality.

In Chapter 4 we created a methodology to quantify the impact of uncertainty on optimised operating policies. We expect this methodology will help to build evidence that optimal solutions will or will not deliver the benefits estimated by the simulation model when applied to a system of which the model is a necessarily incomplete representation. In Chapter 5 we also proposed a method to directly attain robust solutions and demonstrate their efficacy. These Chapters provide a starting point to creating evidence that WRSO research outputs are effective in the face of different uncertainties. Future work in building this evidence might include a wider range of uncertainties (such as a combination of parametric, structural and contextual) and perform a global sensitivity analysis to understand the impact of different uncertainties simultaneously, rather than the ‘one at a time’ approach used here [*Saltelli et al.*, 2008].

In particular the results from Chapter 4 highlight how the choice of system boundaries in a WRS model can strongly influence the estimates about the performance of optimisation. The importance of this modelling choice raises the question of the very definition of an ‘optimal solution'. In a field where different optimisation algorithms might be shown to outperform each other by a few percentage points [*Pan et al.*, 2015], the drawing of boundaries in our case study model (which in itself is a relatively small reservoir network compared to other water resources systems in the UK) can influence estimates of objective values by up to 25%. Because of this we believe that WRSO must move towards creating national (or even larger) scale water resources system models. However, large scale WRS models will have their own set of difficulties, as they will either require their own simplifications (e.g. linear internal processes) or face prohibitively long simulation times. Researchers must continue study into what types of assumptions will teach us the most about the system and which will give practitioners enough confidence in their models to use optimisation.

In this same line of thinking, boundaries may be extended to include the processes that create the forcing inputs of WRSs and integrate WRS models with climate, land-use and other earth system models [*Kim et al.*, 2016; *Monier et al.*, 2018]. For example, instead of representing inflow with a statistical model, as was done in this thesis, the inflow generators could be coupled to a hydro-climatological model. This coupled model would enable a more representative study of how the water resources system may respond climate change. Integrating WRS models with other earth system models will help to understand the interaction between different processes and potentially reveal vulnerabilities that could not have been identified otherwise. It will also, however, raise many challenges. Different earth system models work at different scales (in space and time) and often there is not agreement around how to best represent certain processes. 'Models of everything', [*Beven and Alcock*, 2012], are a powerful concept but a model will always and unavoidably contain assumptions. Scrutinising these assumptions and getting to know what optimising a model can achieve and what it cannot will help us use models to improve our understanding and put that understanding to good use.

6.3 Concluding remarks

We believe the key to overcoming the barriers to uptake that we have identified in Chapter 3 and discussed in Section 6.2 is primarily focused around the provision of evidence that optimisation will be effective when applied in reality by demonstrating it under uncertainty. This perhaps embodies an 'idealised' view that, provided one can show with reasonable confidence that certain research outputs will provide benefits in practice, then uptake will surely follow. However, many of the references provided throughout this thesis list other barriers to uptake (for example, institutional resistance and lack of expertise). Although these other barriers did not emerge from our survey, we expect, from personal contact with practitioners, that they do exist and may be as relevant to uptake as persuasive evidence of efficacy. Future work may wish to survey practitioners at many levels within an organisation to better understand these barriers. We advocate this route with a pinch of salt however, as research benefits by its removed position from the institutional constraints of companies or governments. Our field might be best served by letting innovation guide necessity and not the reverse.

REFERENCES

Afshar, M. H. (2012), Large scale reservoir operation by Constrained Particle Swarm Optimization algorithms, *Journal of Hydro-environment Research*, *6*(1), 75-87.

Ahmad, A., A. El-Shafie, S. F. M. Razali, and Z. S. Mohamad (2014), Reservoir Optimization in Water Resources: a Review, *Water Resources Management*, *28*(11), 3391-3405.

Ahmed, J. A., and A. K. Sarma (2005), Genetic Algorithm for Optimal Operating Policy of a Multipurpose Reservoir, *Water Resources Management*, *19*(2), 145-161.

Allan, T. (2003), IWRM/IWRAM: a new sanctioned discourse, *Occasional paper*, *50*, 1-27.

Anghileri, D., A. Castelletti, F. Pianosi, R. Soncini-Sessa, and E. Weber (2013), Optimizing Watershed Management by Coordinated Operation of Storing Facilities, *Journal of Water Resources Planning and Management*, *139*(5), 492-500.

Anghileri, D., N. Voisin, A. Castelletti, F. Pianosi, B. Nijssen, and D. P. Lettenmaier (2016), Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments, *Water Resources Research*, *52*(6), 4209-4225.

Ansar, A., B. Flyvbjerg, A. Budzier, and D. Lunn (2014), Should we build more large dams? The actual costs of hydropower megaproject development, *Energy Policy*, *69*, 43-56.

Ashbolt, S. C., S. Maheepala, and B. J. C. Perera (2016), Using Multiobjective Optimization to Find Optimal Operating Rules for Short-Term Planning of Water Grids, *Journal of Water Resources Planning and Management*, *142*(10), 04016033.

Auger, A., J. Bader, D. Brockhoff, and E. Zitzler (2009), Theory of the hypervolume indicator: optimal μ-distributions and the choice of the reference point, paper presented at Proceedings of the tenth ACM SIGEVO workshop on Foundations of genetic algorithms, ACM.

Ayala, A. I., A. Cortés, W. E. Fleenor, and F. J. Rueda (2014), Seasonal scale modeling of river inflows in stratified reservoirs: Structural vs. parametric uncertainty in inflow mixing, *Environmental Modelling & Software*, *60*, 84-98.

Bankes, S. C. (2002), Tools and techniques for developing policies for complex and uncertain systems, *Proceedings of the National Academy of Sciences of the United States of America*, *99 Suppl 3*, 7263-7266.

Bankes, S. C., R. J. Lempert, and S. W. Popper (2001), Computer-assisted reasoning, *Computing in Science & Engineering*, *3*(2), 71-77.

Barro, R. J., and R. King (1982), Time-separable preference and intertemporal-substitution models of business cycles, edited, National Bureau of Economic Research Cambridge, Mass., USA.

Barron, A. R. (1993), Universal approximation bounds for superpositions of a sigmoidal function, *Information Theory, IEEE Transactions on*, *39*(3), 930-945.

Bazaraa, M. S., H. D. Sherali, and C. M. Shetty (2013), *Nonlinear programming: theory and algorithms*, John Wiley & Sons.

Beck, M. B. (1987), Water quality modeling: A review of the analysis of uncertainty, *Water Resources Research*, *23*(8), 1393-1442.

Bellman, R. (1956), Dynamic programming and Lagrange multipliers, *Proceedings of the National Academy of Sciences*, *42*(10), 767-769.

Bellmore, R. J., J. J. Duda, L. S. Craig, S. L. Greene, C. E. Torgersen, M. J. Collins, and K. Vittum (2016), Status and trends of dam removal research in the United States, *Wiley Interdisciplinary Reviews: Water*.

Bertsimas, D., D. B. Brown, and C. Caramanis (2011), Theory and Applications of Robust Optimization, *SIAM Review*, *53*(3), 464-501.

Beven, K. J., and R. E. Alcock (2012), Modelling everything everywhere: a new approach to decision-making for water management under uncertainty, *Freshwater Biology*, *57*, 124-132.

Beven, K. J., et al. (2017), Epistemic uncertainties and natural hazard risk assessment. 1. A review of different natural hazard areas, *Natural Hazards and Earth System Sciences Discussions*, 1-53.

Beyer, H.-G., and B. Sendhoff (2007), Robust optimization – A comprehensive survey, *Computer Methods in Applied Mechanics and Engineering*, *196*(33-34), 3190-3218.

Boggs, P. T., and J. W. Tolle (1995), Sequential Quadratic Programming, *Acta Numerica*, *4*, 1.

Borgomeo, E., M. Mortazavi-Naeini, J. W. Hall, and B. P. Guillod (2018), Risk, Robustness and Water Resources Planning Under Uncertainty, *Earth's Future*, *6*(3), 468-487.

Bras, R. L., R. Buchanan, and K. C. Curry (1983), Real time adaptive closed loop control of reservoirs with the High Aswan Dam as a case study, *Water Resources Research*, *19*(1), 33-52.

Broomhead, D. S., and D. Lowe (1988), Radial basis functions, multi-variable functional interpolation and adaptive networks, Royal Signals and Radar Establishment Malvern (United Kingdom).

Brown, C., and U. Lall (2006), Water and economic development: The role of variability and a framework for resilience, *Natural Resources Forum*, *30*(4), 306-317.

Brown, C., J. R. Lund, X. Cai, P. M. Reed, E. A. Zagona, A. Ostfeld, J. Hall, G. W. Characklis, W. Yu, and L. Brekke (2015), The future of water resources systems analysis: Toward a scientific framework for sustainable water management, *Water Resources Research*, *51*(8), 6110-6124.

Cao, Y., B. J. Smucker, and T. J. Robinson (2015), On using the hypervolume indicator to compare Pareto fronts: Applications to multi-criteria optimal experimental design, *Journal of Statistical Planning and Inference*, *160*, 60-74.

Carrillo, G., P. A. Troch, M. Sivapalan, T. Wagener, C. Harman, and K. Sawicz (2011), Catchment classification: hydrological analysis of catchment behavior through process-based modeling along a climate gradient, *Hydrology and Earth System Sciences*, *15*(11), 3411-3430.

Castelletti, A., F. Pianosi, and R. Soncini-Sessa (2008), Water reservoir control under economic, social and environmental constraints, *Automatica*, *44*(6), 1595-1607.

Castelletti, A., F. Pianosi, and R. Soncini-Sessa (2012a), Stochastic and Robust Control of Water Resource Systems: Concepts, Methods and Applications, 383-401.

Castelletti, A., F. Pianosi, and M. Restelli (2013), A multiobjective reinforcement learning approach to water resources systems operation: Pareto frontier approximation in a single run, *Water Resources Research*, *49*(6), 3476-3486.

Castelletti, A., S. Galelli, M. Restelli, and R. Soncini-Sessa (2010), Tree-based reinforcement learning for optimal water reservoir operation, *Water Resources Research*, *46*(9).

Castelletti, A., F. Pianosi, X. Quach, and R. Soncini-Sessa (2012b), Assessing water reservoirs management and development in Northern Vietnam, *Hydrology and Earth System Sciences*, *16*(1), 189-199.

Castelletti, A., D. de Rigo, A. E. Rizzoli, R. Soncini-Sessa, and E. Weber (2007), Neuro-dynamic programming for designing water reservoir network management policies, *Control Engineering Practice*, *15*(8), 1031-1038.

Celeste, A. B. (2016), Managing spills in reservoir design optimisation models, *Proceedings of the Institution of Civil Engineers - Water Management*, *169*(5), 221-228.

Celeste, A. B., and M. Billib (2009), Evaluation of stochastic reservoir operation optimization models, *Advances in Water Resources*, *32*(9), 1429-1443.

Celeste, A. B., K. Suzuki, and A. Kadota (2008), Integrating Long- and Short-Term Reservoir Operation Models via Stochastic and Deterministic Optimization: Case Study in Japan, *Journal of Water Resources Planning and Management*, *134*(5), 440-448.

Celeste, A. B., W. F. Curi, and R. C. Curi (2009), Implicit Stochastic Optimization for deriving reservoir operating rules in semiarid Brazil, *Pesquisa Operacional*, *29*(1).

Chang, F.-J., L. Chen, and L.-C. Chang (2005), Optimizing the reservoir operating rule curves by genetic algorithms, *Hydrological Processes*, *19*(11), 2277-2289.

Chatfield, C. (2013), *The analysis of time series: an introduction*, CRC press.

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resources Research*, *44*(12).

Coelho, B., and A. Andrade-Campos (2014), Efficiency achievement in water supply systems—A review, *Renewable and Sustainable Energy Reviews*, *30*, 59-84.

Cohon, J. L., and D. H. Marks (1975), A review and evaluation of multiobjective programing techniques, *Water Resources Research*, *11*(2), 208-220.

Daniels, R. L., and P. Kouvelis (1995), Robust Scheduling to Hedge Against Processing Time Uncertainty in Single-Stage Production, *Management Science*, *41*(2), 363-376.

Dariane, A. B., and A. M. Moradi (2016), Comparative analysis of evolving artificial neural network and reinforcement learning in stochastic optimization of multireservoir systems, *Hydrological Sciences Journal*, *61*(6), 1141-1156.

de Sousa Júnior, W. C., and J. Reid (2010), Uncertainties in Amazon hydropower development: Risk scenarios and environmental issues around the Belo Monte dam, *Water Alternatives*, *3*(2).

Deb, K., L. Thiele, M. Laumanns, and E. Zitzler (2002), Scalable multi-objective optimization test problems, paper presented at Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on, IEEE.

Donkor, E. A., T. A. Mazzuchi, R. Soyer, and J. Alan Roberson (2014), Urban Water Demand Forecasting: Review of Methods and Models, *Journal of Water Resources Planning and Management*, *140*(2), 146-159.

Dooge, J. (1973), *Linear theory of hydrologic systems*, Agricultural Research Service, US Department of Agriculture.

Douglas, E. M., R. M. Vogel, and C. N. Kroll (2002), Impact of Streamflow Persistence on Hydrologic Design, *Journal of Hydrologic Engineering*, *7*(3), 220-227.

Draper, A. J., and J. R. Lund (2004), Optimal Hedging and Carryover Storage Value, *Journal of Water Resources Planning and Management*, *130*(1), 83-87.

EA, E. A. (2017), Large Raised Reservoirs (AfA134), edited by E. A. EA, data.gov.uk, Online.

FAO (2008), AQUASTAT online database - Total Renewable Water Resources, edited by U. Food and Agriculture Organization.

FAO (2015a), AQUASTAT online database - Water withdrawal by source, edited by U. Food and Agriculture Organization.

FAO (2015b), AQUASTAT online database - Water withdrawal by sector, edited by U. Food and Agriculture Organization.

Fearnside, P. M. (2016), Tropical dams: To build or not to build?, *Science*, *351*(6272), 456-457.

Fenicia, F., D. Kavetski, and H. H. G. Savenije (2011), Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, *Water Resources Research*, *47*(11).

Ficchì, A., L. Raso, D. Dorchies, F. Pianosi, P. O. Malaterre, P. J. Van Overloop, and M. Jay-Allemand (2015), Optimal Operation of the Multireservoir System in the Seine River Basin Using Deterministic and Ensemble Forecasts, *Journal of Water Resources Planning and Management*, 05015005.

Fiering, M. B., and B. Bund (1971), *Synthetic streamflows*, American Geophysical Union.

Fitzpatrick, J. M., and J. J. Grefenstette (1988), Genetic Algorithms in Noisy Environments, *Machine Learning*, *3*(2/3), 101-120.

Fleming, P. J., R. C. Purshouse, and R. J. Lygoe (2005), Many-Objective Optimization: An Engineering Design Perspective, paper presented at EMO, Springer.

Fonseca, C. M., L. Paquete, and M. Lopez-Ibanez (2006), An Improved Dimension-Sweep Algorithm for the Hypervolume Indicator, paper presented at 2006 IEEE International Conference on Evolutionary Computation, 16-21 July 2006.

Ford, L., and D. R. Fulkerson (1962), *Flows in networks*, Princeton Princeton University Press.

Fowler, K. J. A., M. C. Peel, A. W. Western, L. Zhang, and T. J. Peterson (2016), Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models, *Water Resources Research*, *52*(3), 1820-1846.

Gabrel, V., C. Murat, and A. Thiele (2014), Recent advances in robust optimization: An overview, *European Journal of Operational Research*, *235*(3), 471-483.

Galelli, S., A. Goedbloed, D. Schwanenberg, and P. J. van Overloop (2014), Optimal Real-Time Operation of Multipurpose Urban Reservoirs: Case Study in Singapore, *Journal of Water Resources Planning and Management*, *140*(4), 511-523.

Garousi-Nejad, I., O. B. Haddad, and H. A. Loáiciga (2016), Modified Firefly Algorithm for Solving Multireservoir Operation in Continuous and Discrete Domains, *Journal of Water Resources Planning and Management*, 04016029.

Gentle, J. E. (2009), *Computational statistics*, Springer Science & Business Media.

Georgiou, P. E., D. M. Papamichail, and S. G. Vougioukas (2006), Optimal irrigation reservoir operation and simultaneous multi-crop cultivation area selection using simulated annealing, *Irrigation and Drainage*, *55*(2), 129-144.

Giuliani, M., and A. Castelletti (2013), Assessing the value of cooperation and information exchange in large water resources systems by agent-based optimization, *Water Resources Research*, *49*(7), 3912-3926.

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

Giuliani , M., J. D. Herman, A. Castelletti, and P. Reed (2014), Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management, *Water Resources Research, 50*(4), 3355-3377.

Giuliani, M., A. Castelletti, F. Amigoni, and X. Cai (2015a), Multiagent Systems and Distributed Constraint Reasoning for Regulatory Mechanism Design in Water Management, *Journal of Water Resources Planning and Management*, *141*(4).

Giuliani, M., A. Castelletti, F. Pianosi, E. Mason, and P. M. Reed (2015b), Curses, Tradeoffs, and Scalable Management: Advancing Evolutionary Multiobjective Direct Policy Search to Improve Water Reservoir Operations, *Journal of Water Resources Planning and Management*, 04015050.

Graf, W. L. (1999), Dam nation: A geographic census of American dams and their large-scale hydrologic impacts, *Water Resources Research*, *35*(4), 1305-1311.

Greenhalgh, T., G. Robert, F. Macfarlane, P. Bate, and O. Kyriakidou (2004), Diffusion of innovations in service organizations: systematic review and recommendations, *Milbank Q*, *82*(4), 581-629.

Grill, G., B. Lehner, A. E. Lumsdon, G. K. MacDonald, C. Zarfl, and C. Reidy Liermann (2015), An index-based framework for assessing patterns and trends in river fragmentation and flow regulation by global dams at multiple scales, *Environmental Research Letters*, *10*(1).

Grygier, J. C., and J. R. Stedinger (1985), Algorithms for Optimizing Hydropower System Operation, *Water Resources Research*, *21*(1), 1-10.

Guariso, G., S. Rinaldi, and R. Soncini-Sessa (1986), The Management of Lake Como: A Multiobjective Analysis, *Water Resources Research*, *22*(2), 109-120.

Haddad, O. B., A. Afshar, and M. A. Mariño (2006), Honey-Bees Mating Optimization (HBMO) Algorithm: A New Heuristic Approach for Water Resources Optimization, *Water Resources Management*, *20*(5), 661-680.

Hadka, D., and P. Reed (2013), Borg: an auto-adaptive many-objective evolutionary computing framework, *Evol Comput*, *21*(2), 231-259.

Haimes, Y. Y., and W. A. Hall (1977), Sensitivity, responsivity, stability and irreversibility as multiple objectives in civil systems, *Advances in Water Resources*, *1*(2), 71-81.

Hall, J. W., R. J. Lempert, K. Keller, A. Hackbarth, C. Mijere, and D. J. McInerney (2012), Robust climate policies under uncertainty: a comparison of robust decision making and info-gap methods, *Risk analysis : an official publication of the Society for Risk Analysis*, *32*(10), 1657-1672.

Hall, J. W., D. Grey, D. Garrick, F. Fung, C. Brown, S. J. Dadson, and C. W. Sadoff (2014), Water Security. Coping with the curse of freshwater variability, *Science*, *346*(6208), 429-430.

Hall, W. A., and N. Buras (1961), The dynamic programming approach to water-resources development, *Journal of Geophysical Research*, *66*(2), 517.

Hall, W. A., R. Shepherd, W. Butcher, A. ESOGBUE, and S. Parikh (1967), Optimum operations for planning of a complex water resources system.

Hamarat, C., J. H. Kwakkel, E. Pruyt, and E. T. Loonen (2014), An exploratory approach for adaptive policymaking by using multi-objective robust optimization, *Simulation Modelling Practice and Theory*, *46*, 25-39.

Harou, J. J., M. Pulido-Velazquez, D. E. Rosenberg, J. Medellín-Azuara, J. R. Lund, and R. E. Howitt (2009), Hydro-economic models: Concepts, design, applications, and future prospects, *Journal of Hydrology*, *375*(3-4), 627-643.

Hashimoto, T., J. R. Stedinger, and D. P. Loucks (1982), Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation, *Water Resources Research*, *18*(1), 14-20.

Hazen, A. (1914), Storage to be provided in impounding reservoirs for municipal water supply: American Society of Civil Engineers, v. 77.

Heaton, J. (2008), *Introduction to neural networks with Java*, Heaton Research, Inc.

Herman, J. D., and M. Giuliani (2018), Policy tree optimization for threshold-based water resources management over multiple timescales, *Environmental Modelling & Software, 99*, 39-51.

Herman, J. D., H. B. Zeff, P. M. Reed, and G. W. Characklis (2014), Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty, *Water Resources Research*, *50*(10), 7692-7713.

Herman, J. D., P. M. Reed, H. B. Zeff, and G. W. Characklis (2015), How Should Robustness Be Defined for Water Systems Planning under Change?, *Journal of Water Resources Planning and Management*, *141*(10), 04015012.

Herman, J. D., H. B. Zeff, J. R. Lamontagne, P. M. Reed, and G. W. Characklis (2016), Synthetic Drought Scenario Generation to Support Bottom-Up Water Supply Vulnerability Assessments, *Journal of Water Resources Planning and Management*, *142*(11).

Hiew, K. L., J. W. Labadie, and J. F. Scott (1989), Optimal operational analysis of the Colorado-Big Thompson project, paper presented at Computerized decision support systems for water managers, ASCE.

Hınçal, O., A. B. Altan-Sakarya, and A. Metin Ger (2010), Optimization of Multireservoir Systems by Genetic Algorithm, *Water Resources Management*, *25*(5), 1465-1487.

Hine, D., and J. W. Hall (2010), Information gap analysis of flood model uncertainties and regional frequency analysis, *Water Resources Research*, *46*(1).

Hirsch, R. M. (1979), Synthetic hydrology and water supply reliability, *Water Resources Research*, *15*(6), 1603-1615.

Ho, M., U. Lall, M. Allaire, N. Devineni, H. H. Kwon, I. Pal, D. Raff, and D. Wegner (2017), The future role of dams in the United States of America, *Water Resources Research*.

Höllermann, B., and M. Evers (2017), Perception and handling of uncertainties in water management—A study of practitioners' and scientists' perspectives on uncertainty in their daily decision-making, *Environmental Science & Policy*, *71*, 9-18.

Hutton, C. J., and Z. Kapelan (2015), A probabilistic methodology for quantifying, diagnosing and reducing model structural and predictive errors in short term water demand forecasting, *Environmental Modelling & Software*, *66*, 87-97.

Ibanez, E., T. Magee, M. Clement, G. Brinkman, M. Milligan, and E. Zagona (2014), Enhancing hydropower modeling in variable generation integration studies, *Energy*, *74*, 518-528.

ICoLD (2003), World Register of Dams, edited by I. C. o. L. Dams, Paris.

IEA, I. E. A. (2017), *Key world energy statistics*, International Energy Agency Paris.

Jahan, S. (2015), Human Development Report 2015, *Work for Human Development. New York: United Nations Development Programme*.

Kasprzyk, J. R., P. M. Reed, G. W. Characklis, and B. R. Kirsch (2012), Many-objective de Novo water supply portfolio planning under deep uncertainty, *Environmental Modelling & Software*, *34*, 87-104.

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

Kelman, J., J. R. Stedinger, L. A. Cooper, E. Hsu, and S.-Q. Yuan (1990), Sampling stochastic dynamic programming applied to reservoir operation, *Water Resources Research*, *26*(3), 447-454.

Kim, S. H., M. Hejazi, L. Liu, K. Calvin, L. Clarke, J. Edmonds, P. Kyle, P. Patel, M. Wise, and E. Davies (2016), Balancing global water availability and use at basin scale in an integrated assessment model, *Climatic Change*, *136*(2), 217-231.

Knowles, J., and D. Corne (2002), On metrics for comparing nondominated sets, paper presented at Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on, 12-17 May 2002.

Koutsoyiannis, D. (2000), A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, *36*(6), 1519-1533.

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

Kraljevic, A., J. Meng, and P. Schelle (2013), Seven Sins of Dam Building, *WWF International, Berlin*.

Kumar, A. R. S., M. K. Goyal, C. S. P. Ojha, R. D. Singh, P. K. Swamee, and R. K. Nema (2012), Application of ANN, Fuzzy Logic and Decision Tree Algorithms for the Development of Reservoir Operating Rules, *Water Resources Management*, *27*(3), 911-925.

Kumar , N. D., and J. M. Reddy (2006), Ant Colony Optimization for Multi-Purpose Reservoir Operation, *Water Resources Management*, *20*(6), 879-898.

Kumar , N. D., and J. M. Reddy (2007), Multipurpose Reservoir Operation Using Particle Swarm Optimization, *Journal of Water Resources Planning and Management*, *133*(3), 192-201.

Kwakkel, J. H., S. Eker, and E. Pruyt (2016), How robust is a robust Policy? Comparing alternative robustness metrics for robust decision-making, in *Robustness Analysis in Decision Aiding, Optimization, and Analytics*, edited, pp. 221-237, Springer, Switzerland.

Labadie, J. W. (2004), Optimal Operation of Multireservoir Systems: State-of-the-Art Review, *Journal of Water Resources Planning and Management*, *130*(2), 93-111.

Langsdale, S., A. Beall, E. Bourget, E. Hagen, S. Kudlas, R. Palmer, D. Tate, and W. Werick (2013), Collaborative Modeling for Decision Support in Water Resources: Principles and Best Practices, *JAWRA Journal of the American Water Resources Association*, *49*(3), 629-638.

Lehner, B., et al. (2011), High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, *Frontiers in Ecology and the Environment*, *9*(9), 494-502.

Lempert, R. (2002), A new decision sciences for complex systems, *Proceedings of the National Academy of Sciences of the United States of America*, *99 Suppl 3*, 7309-7313.

Lempert, R., and M. T. Collins (2007), Managing the risk of uncertain threshold responses: comparison of robust, optimum, and precautionary approaches, *Risk analysis : an official publication of the Society for Risk Analysis*, *27*(4), 1009-1026.

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

Lempert, R., D. G. Groves, S. W. Popper, and S. C. Bankes (2006), A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios, *Management Science*, *52*(4), 514-528.

Li, M., S. Yang, and X. Liu (2014), Diversity comparison of Pareto front approximations in many-objective optimization, *IEEE Trans Cybern*, *44*(12), 2568-2584.

Liermann, C. R., C. Nilsson, J. Robertson, and R. Y. Ng (2012), Implications of Dam Obstruction for Global Freshwater Fish Diversity, *BioScience*, *62*(6), 539-548.

Loucks, D. P. (1992), Water Resource Systems Models: Their Role in Planning, *Journal of Water Resources Planning and Management*, *118*(3), 214-223.

Loucks, D. P., J. R. Stedinger, and U. Shamir (1985), Modelling water resource systems: issues and experiences, *Civil Engineering Systems*, *2*(4), 223-231.

Loucks, D. P., E. Van Beek, J. R. Stedinger, J. P. Dijkman, and M. T. Villars (2005), *Water resources systems planning and management: an introduction to methods, models and applications*, Paris: Unesco.

Maass, A., M. M. Hufschmidt, R. Dorfman, H. A. Thomas Jr, S. A. Marglin, and G. M. Fair (1962), *Design of water-resource systems*, Harvard University Press Cambridge, Massachusetts.

Mahmoud, M., et al. (2009), A formal framework for scenario development in support of environmental decision-making, *Environmental Modelling & Software*, *24*(7), 798-808.

Mariño, M. A., and H. A. Loaiciga (1985), Quadratic model for reservoir management: Application to the Central Valley Project, *Water Resources Research*, *21*(5), 631-641.

Marques, G. F., and A. Tilmant (2013), The economic value of coordination in large-scale multireservoir systems: The Parana River case, *Water Resources Research*, *49*(11), 7546-7557.

Martin, Q. W. (1983), Optimal Operation of Multiple Reservoir Systems, *Journal of Water Resources Planning and Management*, *109*(1), 58-74.

Matalas, N. C. (1967), Mathematical assessment of synthetic hydrology, *Water Resources Research*, *3*(4), 937-945.

Matrosov, E. S., J. J. Harou, and D. P. Loucks (2011), A computationally efficient open-source water resource system simulator – Application to London and the Thames Basin, *Environmental Modelling & Software*, *26*(12), 1599-1610.

Matrosov, E. S., A. M. Woods, and J. J. Harou (2013), Robust Decision Making and Info-Gap Decision Theory for water resource system planning, *Journal of Hydrology*, *494*, 43-58.

Mayne, D. Q., J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert (2000), Constrained model predictive control: Stability and optimality, *Automatica*, *36*(6), 789-814.

McCulloch, C. (2006), The Kielder Water Scheme: the last of its kind?, paper presented at In Improvements in reservoir construction, operation and maintenance: Proceedings of the 14th conference of the British Dam Society at the University of Durham from 6 to 9 September 2006, University of Durham Durham.

Miller, B. L., and D. E. Goldberg (1996), Optimal sampling for genetic algorithms, *Urbana*, *51*, 61801.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008), Stationarity Is Dead: Whither Water Management?, *Science*, *319*(5863), 573-574.

Mo, L., P. Lu, C. Wang, and J. Zhou (2013), Short-term hydro generation scheduling of Three Gorges–Gezhouba cascaded hydropower plants using hybrid MACS-ADE approach, *Energy Conversion and Management*, *76*, 260-273.

Momtahen, S., and A. B. Dariane (2007), Direct Search Approaches Using Genetic Algorithms for Optimization of Water Reservoir Operating Policies, *Journal of Water Resources Planning and Management*, *133*(3), 202-209.

Monier, E., et al. (2018), Toward a consistent modeling framework to assess multi-sectoral climate impacts, *Nat Commun*, *9*(1), 660.

Montanari, A., and D. Koutsoyiannis (2014), Modeling and mitigating natural hazards: Stationarity is immortal!, *Water Resources Research*, *50*(12), 9748-9756.

Mortazavi, M., G. Kuczera, and L. Cui (2012), Multiobjective optimization of urban water resources: Moving toward more practical solutions, *Water Resources Research*, *48*(3).

Moss, R., et al. (2016), Understanding Dynamics and Resilience in Complex Interdependent Systems, in *U.S. Global Change Research Program Interagency Group on Integrative Modeling*, edited, Washington, D.C.

Muller, M., A. Biswas, R. Martin-Hurtado, and C. Tortajada (2015), Built infrastructure is essential, *Science*, *349*(6248), 585-586.

Nakayama, M. (1998), Post-project Review of Environmental Impact Assessmentfor Saguling Dam for Involuntary Resettlement, *International Journal of Water Resources Development*, *14*(2), 217-229.

Nandalal, K., and J. J. Bogardi (2007), *Dynamic programming based operation of reservoirs: applicability and limits*, Cambridge University Press.

Nardini, A., C. Piccardi, and R. Soncini-Sessa (1992), On the integration of risk aversion and average-performance optimization in reservoir control, *Water Resources Research*, *28*(2), 487-497.

Nature.com (2018), Definition of "Water Resources", edited.

NIC, N. I. C. (2018), Preparing for a drier future, England's water infrastructure needs.

Nicklow, J., et al. (2010), State of the Art for Genetic Algorithms and Beyond in Water Resources Planning and Management, *Journal of Water Resources Planning and Management*, *136*(4), 412-432.

Noory, H., A. M. Liaghat, M. Parsinejad, and O. B. Haddad (2012), Optimizing Irrigation Water Allocation and Multicrop Planning Using Discrete PSO Algorithm, *Journal of Irrigation and Drainage Engineering*, *138*(5), 437-444.

OFWAT (2015), A study on potential benefits of upstream markets in the water sector in England and Wales, edited.

Oliveira, R., and D. P. Loucks (1997), Operating rules for multireservoir systems, *Water Resources Research*, *33*(4), 839-852.

Opdenakker, R. (2006), Advantages and Disadvantages of Four Interview Techniques in Qualitative Research, *2006*, *7*(4).

Ostadrahimi, L., M. A. Mariño, and A. Afshar (2011), Multi-reservoir Operation Rules: Multi-swarm PSO-based Optimization Approach, *Water Resources Management*, *26*(2), 407-427.

Ouyang, W., F. Hao, K. Song, and X. Zhang (2010), Cascade Dam-Induced Hydrological Disturbance and Environmental Impact in the Upper Stream of the Yellow River, *Water Resources Management*, *25*(3), 913-927.

Pan, L., M. Housh, P. Liu, X. Cai, and X. Chen (2015), Robust stochastic optimization for reservoir operation, *Water Resources Research*, *51*(1), 409-429.

Pianosi, F., and R. Soncini-Sessa (2009), Real-time management of a multipurpose water reservoir with a heteroscedastic inflow model, *Water Resources Research*, *45*(10).

Pianosi, F., X. Q. Thi, and R. Soncini-Sessa (2011), Artificial Neural Networks and Multi Objective Genetic Algorithms for Water Resources Management: An Application to the Hoabinh Reservoir in Vietnam, 10579-10584.

Poff, N. L., and J. C. Schmidt (2016), How dams can go with the flow, *Science*, *353*(6304), 1099-1100.

Quinn, J. D., P. M. Reed, M. Giuliani, and A. Castelletti (2017), 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.

Rajagopalan, B., J. D. Salas, and U. Lall (2010), STOCHASTIC METHODS FOR MODELING PRECIPITATION AND STREAMFLOW, in *Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting*, edited, pp. 17-52, WORLD SCIENTIFIC.

Rani, D., and M. M. Moreira (2009), Simulation–Optimization Modeling: A Survey and Potential Application in Reservoir Systems Operation, *Water Resources Management*, *24*(6), 1107-1138.

Raso, L., D. Schwanenberg, N. C. van de Giesen, and P. J. van Overloop (2014), Short-term optimal operation of water systems using ensemble forecasts, *Advances in Water Resources*, *71*, 200-208.

Reed, P. M., and J. B. Kollat (2013), Visual analytics clarify the scalability and effectiveness of massively parallel many-objective optimization: A groundwater monitoring design example, *Advances in Water Resources, 56*, 1-13.

Reed, P. M., D. Hadka, J. D. Herman, J. R. Kasprzyk, and J. B. Kollat (2013), Evolutionary multiobjective optimization in water resources: The past, present, and future, *Advances in Water Resources*, *51*, 438-456.

Revelle, C., E. Joeres, and W. Kirby (1969), The Linear Decision Rule in Reservoir Management and Design: 1, Development of the Stochastic Model, *Water Resources Research*, *5*(4), 767-777.

Rippl, W. (1883), The Capacity of Storage-Reservoirs for Water-Supply. (Including Plate), *Minutes of the Proceedings, 71*(1883), 270-278.

Robinson, E. L., E. Blyth, D. B. Clark, E. Comyn-Platt, J. Finch, and A. C. Rudd (2016), Climate hydrology and ecology research support system potential evapotranspiration dataset for Great Britain (1961-2015) [CHESS-PE], edited, NERC Environmental Information Data Centre.

Rogers, P. P., and M. B. Fiering (1986), Use of systems analysis in water management, *Water Resources Research*, *22*(9S), 146S-158S.

Rosenstein, M. T., and A. G. Barto (2001), Robot weightlifting by direct policy search, paper presented at International Joint Conference on Artificial Intelligence, Citeseer.

Salas, J. D., and J. T. B. Obeysekera (1982), ARMA Model identification of hydrologic time series, *Water Resources Research*, *18*(4), 1011-1021.

Salas, J. D., C. Fu, A. Cancelliere, D. Dustin, D. Bode, A. Pineda, and E. Vincent (2005), Characterizing the Severity and Risk of Drought in the Poudre River, Colorado, *Journal of Water Resources Planning and Management*, *131*(5), 383-393.

Salazar, J. Z., P. M. Reed, J. D. Herman, M. Giuliani, and A. Castelletti (2016), A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control, *Advances in Water Resources*, *92*, 172-185.

Salazar, J. Z., P. M. Reed, J. D. Quinn, M. Giuliani, and A. Castelletti (2017), Balancing exploration, uncertainty and computational demands in many objective reservoir optimization, *Advances in Water Resources*, *109*, 196-210.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008), *Global sensitivity analysis: the primer*, John Wiley & Sons.

Savage, L. J. (1954), *The Foundations of Statistics*, 86-89 pp., Wiley Publications in Statistics.

Savic, D. A., Z. S. Kapelan, and P. M. R. Jonkergouw (2009), Quo vadis water distribution model calibration?, *Urban Water Journal*, *6*(1), 3-22.

Scheuren, F. (2004), What is a Survey?, American Statistical Association.

Scudder, T. (2012), Resettlement outcomes of large dams, in *Impacts of large dams: A global assessment*, edited, pp. 37-67, Springer.

Seifollahi-Aghmiuni, S., O. Bozorg Haddad, and M. A. Mariño (2016), Generalized Mathematical Simulation Formulation for Reservoir Systems, *Journal of Water Resources Planning and Management*, *142*(4), 04016004.

Sharif, M., and R. Wardlaw (2000), Multireservoir Systems Optimization Using Genetic Algorithms: Case Study, *Journal of Computing in Civil Engineering*, *14*(4), 255-263.

Shih , J. S., and C. ReVelle (1994), Water-Supply Operations during Drought: Continuous Hedging Rule, *Journal of Water Resources Planning and Management*, *120*(5), 613-629.

Shih , J. S., and C. ReVelle (1995), Water supply operations during drought: A discrete hedging rule, *European Journal of Operational Research*, *82*(1), 163-175.

Shim, K.-C., D. G. Fontane, and J. W. Labadie (2002), Spatial Decision Support System for Integrated River Basin Flood Control, *Journal of Water Resources Planning and Management*, *128*(3), 190-201.

Sigvaldson, O. T. (1976), A simulation model for operating a multipurpose multireservoir system, *Water Resources Research*, *12*(2), 263-278.

Simonovic, S. P. (1992), Reservoir Systems Analysis: Closing Gap between Theory and Practice, *Journal of Water Resources Planning and Management*, *118*(3), 262-280.

Smalley, J. B., B. S. Minsker, and D. E. Goldberg (2000), Risk-based in situ bioremediation design using a noisy genetic algorithm, *Water Resources Research*, *36*(10), 3043-3052.

Stedinger, J. R., B. F. Sule, and D. P. Loucks (1984), Stochastic dynamic programming models for reservoir operation optimization, *Water Resources Research*, *20*(11), 1499-1505.

Stedinger, J. R., B. A. Faber, and J. R. Lamontagne (2013), Developments in stochastic dynamic programming for reservoir operation optimization, paper presented at World Environmental and Water Resources Congress 2013: Showcasing the Future.

Tejada-Guibert, J. A., S. A. Johnson, and J. R. Stedinger (1995), The Value of Hydrologic Information in Stochastic Dynamic Programming Models of a Multireservoir System, *Water Resources Research*, *31*(10), 2571-2579.

Terlaky, T. (2013), *Interior point methods of mathematical programming*, Springer Science & Business Media.

Tilmant, A., and W. Kinzelbach (2012), The cost of noncooperation in international river basins, *Water Resources Research*, *48*(1).

Tockner, K., M. Pusch, J. Gessner, and C. Wolter (2011), Domesticated ecosystems and novel communities: challenges for the management of large rivers, *Ecohydrology & Hydrobiology*, *11*(3-4), 167-174.

Trindade, B. C., P. M. Reed, J. D. Herman, H. B. Zeff, and G. W. Characklis (2017), Reducing regional drought vulnerabilities and multi-city robustness conflicts using many-objective optimization under deep uncertainty, *Advances in Water Resources*, *104*, 195-209.

Turner, S. W. D., and P. J. Jeffrey (2015), Industry views on water resources planning methods - prospects for change in England and Wales, *Water and Environment Journal*, *29*(2), 161-168.

UN (2015), World Population Prospects: The 2015 Revision, edited by D. E. a. S. A. United Nations, Population Division, United Nations, Online.

Van Wicklin, W. (2018), *Involuntary resettlement: Comparative perspectives*, Routledge.

Vogel, R. M. (2017), Stochastic watershed models for hydrologic risk management, *Water Security*, *1*, 28-35.

Vogel, R. M., and J. R. Stedinger (1987), Generalized storage-reliability-yield relationships, *Journal of Hydrology*, *89*(3-4), 303-327.

Vogel, R. M., and J. R. Stedinger (1988), The value of stochastic streamflow models in overyear reservoir design applications, *Water Resources Research*, *24*(9), 1483-1490.

Walker, W. E., M. Haasnoot, and J. Kwakkel (2013), Adapt or Perish: A Review of Planning Approaches for Adaptation under Deep Uncertainty, *Sustainability*, *5*(3), 955-979.

Walker, W. E., P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen, and M. P. Krayer von Krauss (2003), Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support, *Integrated Assessment*, *4*(1), 5-17.

Walski, T. M. (2001), The Wrong Paradigm—Why Water Distribution Optimization Doesn't Work, *Journal of Water Resources Planning and Management*, *127*(4), 203-205.

Wardlaw, R., and M. Sharif (1999), Evaluation of Genetic Algorithms for Optimal Reservoir System Operation, *Journal of Water Resources Planning and Management*, *125*(1), 25-33.

Watson, A. A., and J. R. Kasprzyk (2017), Incorporating deeply uncertain factors into the many objective search process, *Environmental Modelling & Software*, *89*, 159-171.

WCD (2000), *Dams and Development: A New Framework for Decision-making: the Report of the World Commission on Dams*, Earthscan.

Weiss, H. K. (1956), Estimation of Reliability Growth in a Complex System with a Poisson-Type Failure, *Operations Research*, *4*(5), 532-545.

Whateley, S., R. N. Palmer, and C. Brown (2015), Seasonal Hydroclimatic Forecasts as Innovations and the Challenges of Adoption by Water Managers, *Journal of Water Resources Planning and Management*, *141*(5).

Wu, X., C. Cheng, Y. Zeng, and J. R. Lund (2016), Centralized versus Distributed Cooperative Operating Rules for Multiple Cascaded Hydropower Reservoirs, *Journal of Water Resources Planning and Management*, *142*(11).

Wurbs, R. A. (1993), Reservoir-System Simulation and Optimization Models, *Journal of Water Resources Planning and Management*, *119*(4), 455-472.

Wurbs, R. A. (1998), Dissemination of Generalized Water Resources Models in the United States, *Water International*, *23*(3), 190-198.

Wurbs, R. A. (2005), *Comparative evaluation of generalized river/reservoir system models*, Texas Water Resources Institute, Texas A&M University.

Yakowitz, S. (1982), Dynamic programming applications in water resources, *Water Resources Research*, *18*(4), 673-696.

Yeh, W. W. G. (1985), Reservoir Management and Operations Models: A State-of-the-Art Review, *Water Resources Research*, *21*(12), 1797-1818.

You, J.-Y., and X. Cai (2008), Hedging rule for reservoir operations: 1. A theoretical analysis, *Water Resources Research*, *44*(1).

Young, E. (2016), Changing course through water trading, edited by S. T. Water.

Yun, R., V. P. Singh, and Z. Dong (2010), Long-Term Stochastic Reservoir Operation Using a Noisy Genetic Algorithm, *Water Resources Management*, *24*(12), 3159-3172.

Zagona, E. A., T. J. Fulp, R. Shane, T. Magee, and H. M. Goranflo (2001), Riverware: A Generalized Tool for Complex Reservoir System Modeling1, *JAWRA Journal of the American Water Resources Association*, *37*(4), 913-929.

Zarfl, C., A. E. Lumsdon, J. Berlekamp, L. Tydecks, and K. Tockner (2014), A global boom in hydropower dam construction, *Aquatic Sciences*, *77*(1), 161-170.

Zeff, H. B., J. R. Kasprzyk, J. D. Herman, P. M. Reed, and G. W. Characklis (2014), Navigating financial and supply reliability tradeoffs in regional drought management portfolios, *Water Resources Research*, *50*(6), 4906-4923.

Zhao, T., X. Cai, and D. Yang (2011), Effect of streamflow forecast uncertainty on real-time reservoir operation, *Advances in Water Resources*, *34*(4), 495-504.

Zhao, T., D. Yang, X. Cai, J. Zhao, and H. Wang (2012), Identifying effective forecast horizon for real-time reservoir operation under a limited inflow forecast, *Water Resources Research*, *48*(1).

APPENDICES

Appendix A: Supporting information and results for Chapter 4

Appendix A1: Simulation and optimisation model

Simulation

The model dynamics are simulated at a daily time-step through the following two mass balance equations

$$S_{S2,t+1} = S_{S2,t} + I_{S2,t} - u_{S2,D2,t} - E_{S2,t} - w_{S2,t} - env_{S2,t} \quad \text{(A1.1)}$$

$$S_{S1,t+1} = S_{S1,t} + I_{S1,t} - u_{S1,D2,t} - u_{S1,R,t} + u_{R,S1,t} - E_{S1,t} - w_{S1,t} - env_{S1,t} \quad \text{(A1.2)}$$

where $S_{k,t}$ is the storage at time t for reservoir k; $I_{k,t}$, $E_{k,t}$, $w_{k,t}$ and $env_{k,t}$ are natural inflow, evaporation, spills and compensation flow (released to meet downstream ecological flow requirements) respectively; and $u_{k,j,t}$ are controlled flows along a link between two nodes (k,j) at time t. These controlled flows include the abstractions from the two reservoirs to the demand nodes and the abstraction from the river that is pumped back into reservoir S1 (see Figure 2 in the main manuscript). Evaporation fluxes are computed by multiplying the reservoir surface areas by the unit evaporation rate. Reservoir surface areas at each time step are calculated from storages using the available storage-elevation curves and the unit evaporation rate (assumed equal for both reservoirs) is taken from *Robinson et al.* [2016]. Spills are calculated by imposing the hard constraint that storages at the following time-step (left hand side of Equation (A1.1-2)) should never exceed the reservoir capacities, hence they are either equal to zero or to the excess volumes generated by all other terms on the right-hand side of Equation (A1.1-2). Environmental compensation flows are equal to prescribed values that are constant over the year (1 Ml/day for S1 and 5 Ml/day for S2) plus occasional fish releases. Controlled fluxes are calculated via a set of operating rules, further explained below, and subject to a range of licensing and operational constraints.

The aim of the system operation is to reliably supply water while reducing pumping costs. This leads to formulating four ‘daily costs’, all to be minimised, shown in equations (A1.3-6)

$$\text{Deficit for Company 1: } g_{1,t} = \max(d_{D1,t} - I_{R2,t} + u_{R,S1,t} - u_{S1,R,t} - q_{D1,t}, 0) \quad \text{(A1.3)}$$

$$\text{Deficit for Company 2: } g_{2,t} = \max(d_{D2,t} - u_{S1,D2,t} - u_{S2,D2,t}, 0) \quad \text{(A1.4)}$$

$$\text{Cost for Company 1: } g_{3,t} = c_{d1} * q_{D1,t} + c_{rs1} * u_{R,S1,t} \quad \text{(A1.5)}$$

$$\text{Cost for Company 2: } g_{4,t} = c_{s1d2} * u_{S1,D2,t} \quad \text{(A1.6)}$$

where $d_{j,t}$ is the demand for water treatment works j at time t, p_{D1} is the minimum flow required after abstraction at point D1, and c_y is the pumping costs associated with a given flow along link y. These daily costs are then translated into four objectives by taking their averages over time and over a Monte Carlo simulation ensemble, as discussed in the experimental setup section of the main manuscript.

Operating policy

Of the variables in equations (A1.1-2), the controlled fluxes $u_{i,j,t}$ are the decision variables that the operators have control over, and which determine the performance of operations. To determine their values, we formulate an ‘operating policy’, i.e. a function that takes system state variables at the current time-step as inputs, and returns the decision variables for that time-step as outputs. For the policy, we use a Radial Basis Function Network (RBFN) in the formulation originally described in *Broomhead and Lowe* [1988]. We visualize a single evaluation of an operating policy (as would occur every time-step) in Figure A1.1.

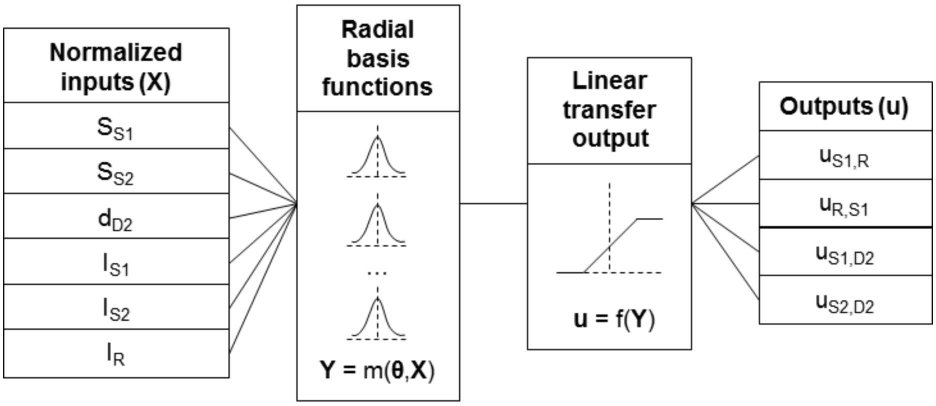


Figure A1.1: A schematic showing a single evaluation of an operating policy for a given time step. This process takes the system states at a given time-step as inputs (normalized between their maximum and minimum possible values) and returns normalized outputs that are then scaled to specify abstractions from reservoirs and rivers in the system.

The process inside the 'Radial basis functions' step is given by equation (A1.7) below:

$$Y_{p,t} = a_p + \sum_{n=1}^{N} b_{n,p} q_n, \quad \text{(A1.7)}$$

$$\text{where } q_n = exp\left(-c_n \sum_{m=1}^{M} (d_{m,n} - X_{m,t})^2\right)$$

where X is the vector of the network inputs (with M=6 in our case), and *a* (output biases), *b* (output weights), *c* (inverse variances) and *d* (centres) are the network parameters, which all together form the parameter vector θ used in Figure A1.1, N is the number hidden nodes in the network, and $Y_{p,t}$ is the network output that, after de-normalisation, becomes $u_{p,t}$. In our application we follow the rule-of-thumb described by *Heaton* [2008] – that the number of hidden nodes should lie between the number of inputs and the number of outputs – and use N=5.

Appendix A2: Synthetic generation of forcing inputs

The Table below lists the stochastic variables that appear in the water resources system model and the model used for their synthetic generation.

Variables	Model used for synthetic generation
Inflows: I_{R1}, I_{R2}, I_{S1}, I_{S2}	Periodic logarithmic autoregressive model
Demands: d_{D2}	Periodic autoregressive model
Unit Evaporation: ue	Periodic autoregressive model
Pump failures	Poisson duration of breaks and between breaks
Fisheries release: p_{fish}	Uniform probability of occurrence (over either September or whole year, depending on the framing) and duration

Periodic autocorrelated variables (inflows, demands or potential evaporation) are generated at each time-step following the equation:

$$X_t = \mu_t \exp(Y_t) \qquad \text{(A2.1)}$$

in the logarithmic case (i.e. reservoir inflows) and

$$X_t = \mu_t + Y_t \qquad \text{(A2.2)}$$

otherwise (demand and evaporation) where X_t represents the autocorrelated variable (I, d or ue), μ_t is the periodic component and Y_t the autocorrelated component.

The periodic component μ_t is given by the equation:

$$\mu_t = b_1 + b_2 \sin(\lambda_1 \pi f_t) + b_3 \cos(\lambda_1 \pi f_t) + \ldots \qquad \text{(A2.3)}$$

$$+ b_{N-1} \sin(\lambda_{(N-1)/2} \pi f_t) + b_N . \cos(\lambda_{(N-1)/2} \pi f_t)$$

where f_t = (t mod P)/P with P = 365 represents the time of the year, the coefficients λ_1, λ_2, … , $\lambda_{(N-1)/2}$ represent the harmonic frequencies characterising the modelled variable and the coefficients b_1, b_2, …, b_N are the amplitudes of those frequencies. In our application we use two harmonics for the inflows (annual and biannual) and hence set N = 5, $\lambda_1 = 2$ and $\lambda_2 = 4$;

two harmonics for the demands (annual and weekly), i.e. N = 5, λ_1=2 and λ_2=2P/7; and one harmonic (annual) for the evaporation, i.e. N = 3, λ_1=2. Once the harmonic frequencies have been set, the coefficients b_1, b_2, …, b_N are found using least-squares fitting of the historic data.

The autocorrelated component Y_t is given by the equation:

$$Y_t = a_0 + a_1.Y_{t-1} + a_2.Y_{t-2} + \ldots + a_L.Y_{t-L} + \varepsilon_t, \quad \text{(A2.4)}$$

where a_0 is the expected value of Y_t and the coefficients a_1, a_2, …, a_L represent the lagged correlations and ε represents the 'innovation' (a normal random variable). The number of lag terms (L) takes the value 1 or 2 depending on the framing. For given L, the coefficients a_1, a_2, …, a_L are determined by a least-squares fitting of historic data using. Since ε is correlated across variables, we transform uncorrelated random normal numbers by the Cholesky decomposition of the variance-covariance matrix of ε (found from the historic data) to create correlated normal variables, as described in [*Gentle*, 2009].

Pump failures are generated using two Poisson distributions, one describing the duration between breaks and one describing the duration of each break. The expected duration between breaks is 300 days for $u_{S1,D2}$ and $u_{R,S1}$, and 800 days for $u_{S2,D2}$. The expected duration of a break is 3 days for $u_{S1,D2}$ and $u_{R,S1}$, and 5 days for $u_{S2,D2}$. We note that, although $u_{S2,D2}$ is not a pumped flow, pump failures represent any failure to use the pipe, thus we suggest that a failure to use this link is possible (albeit less likely). We assume that, as a direct dam release, $u_{S1,R}$ is always possible.

Fisheries releases are assumed to happen at most once a year. The release event must occur in a certain time window that, depending on the framing, either covers the all year (except Spring) or spans over September only. Each day in this window has an equal chance of being the beginning of the fisheries release. Once the first release day has been randomly extracted, the release duration may last from 3 to 14 days (this duration is also randomly selected). The specified volume (900 Ml) is then uniformly released over the duration of the release.

Appendix A3: Supporting results

In this Appendix we include sets of results that either a more detailed or complete version of what is presented in the body of this thesis.

Pre-optimisation

To create Figure A3.1 we evaluated 132 random operating policies over increasingly large sample sizes for simulation and recorded the objective values. We then recorded the absolute difference between an objective value at a given sample size and the objective value at the largest sample size ($1.2*10^7$ days). As established in Section 4.3.3 this is the 'approximation error'. The lines in this plot indicate the 99^{th} percentile (across the 132 random policies) of the approximation error for a given framing. These plots enable us to choose a sample size that limits the approximation error to a specified value (indicated by the dotted lines). There are 8 lines because we use 8 framings.

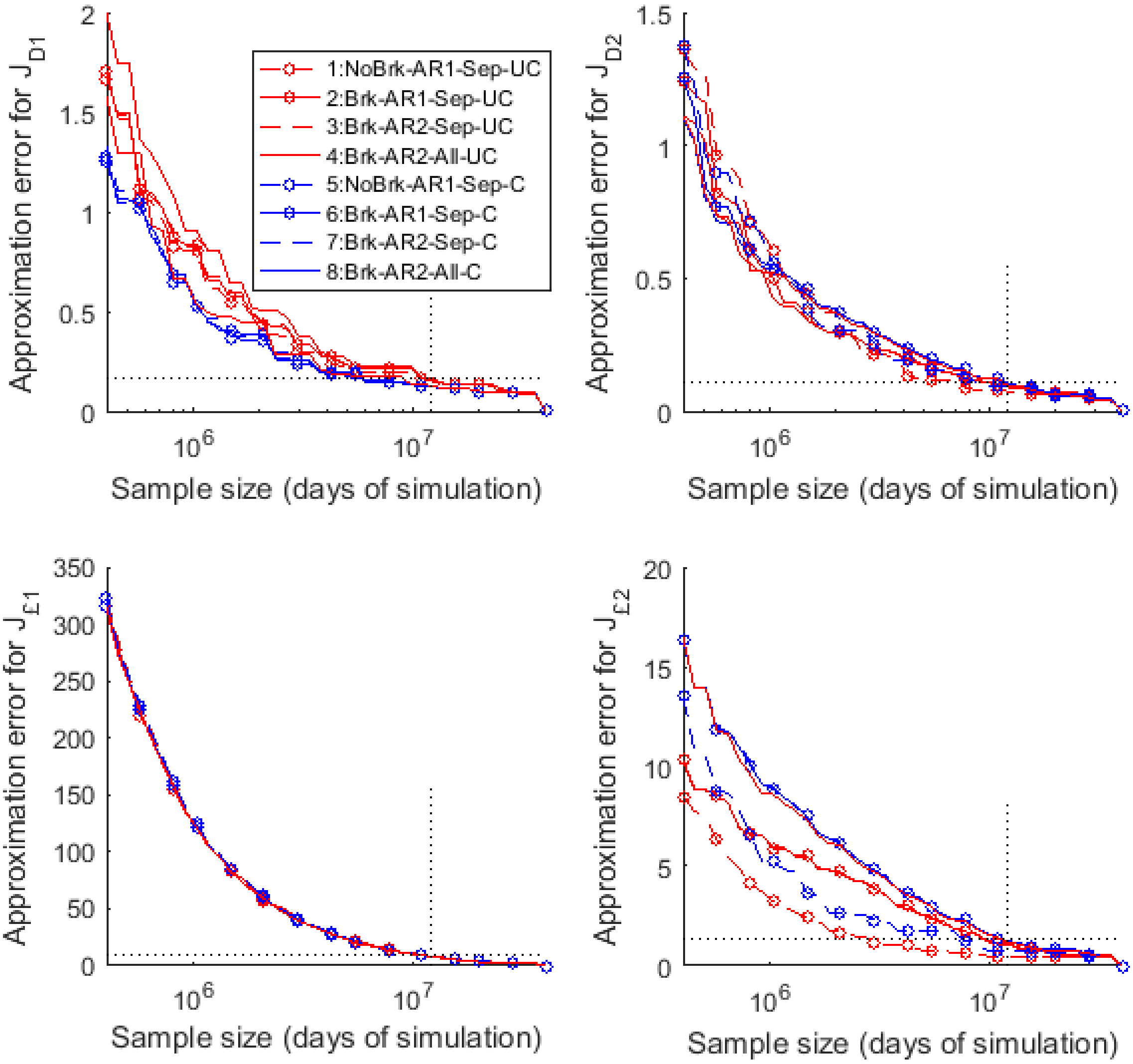


Figure A3.1: Each coloured line marks the trajectory of the 99th percentile of error for a given objective, in a given framing. The black dotted lines mark the select K*T ($1.2*10^7$) and the approximation error which we also define as the epsilon value.

To create Figure A3.2 we performed optimization (in framing 8) 25 times, recording the objective values of the population of operating policies every 100 iterations of the optimisation. These objective values are then normalized between 0 and 1. The hypervolume that is dominated by these objective values is plotted in the plot below. This enables us to determine a suitable number of function evaluations for the optimisation process to ensure that hypervolume will be converged.

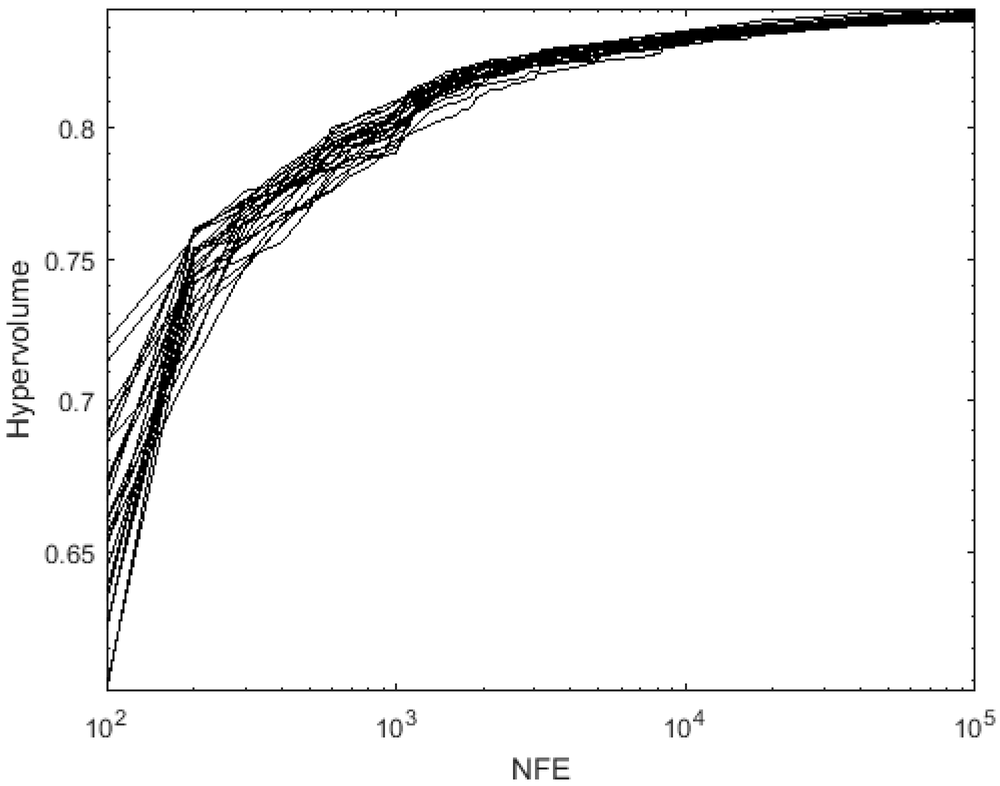


Figure A3.2: The hypervolume trajectory of 25 optimisation runs up to 10^5 Number of Function Evaluations (NFE), each with a different seed for the Borg MOEA and evaluated on forcing with different seeds. See [*Salazar et al.*, 2016] for a more detailed discussion.

Results

To create Figure A3.3, we take sets of Pareto solutions (where a solution is an operating policy) that have been optimized to a given framing (as indicated by the legend) and re-evaluate them in other framings (as indicated by the row number). We then plot a CDF of the differences in objective values (each column shows a different objective) between optimization and re-evaluation. This enables us to visualise how estimates of objective values change under different realizations of uncertainty (i.e. under framings that are different from those used in optimization). CDFs that lie to the right of the X=0 line have performed better in the framing used for re-evaluation than the framing used for optimization, and vice versa if they lie to the left of the X=0 line.

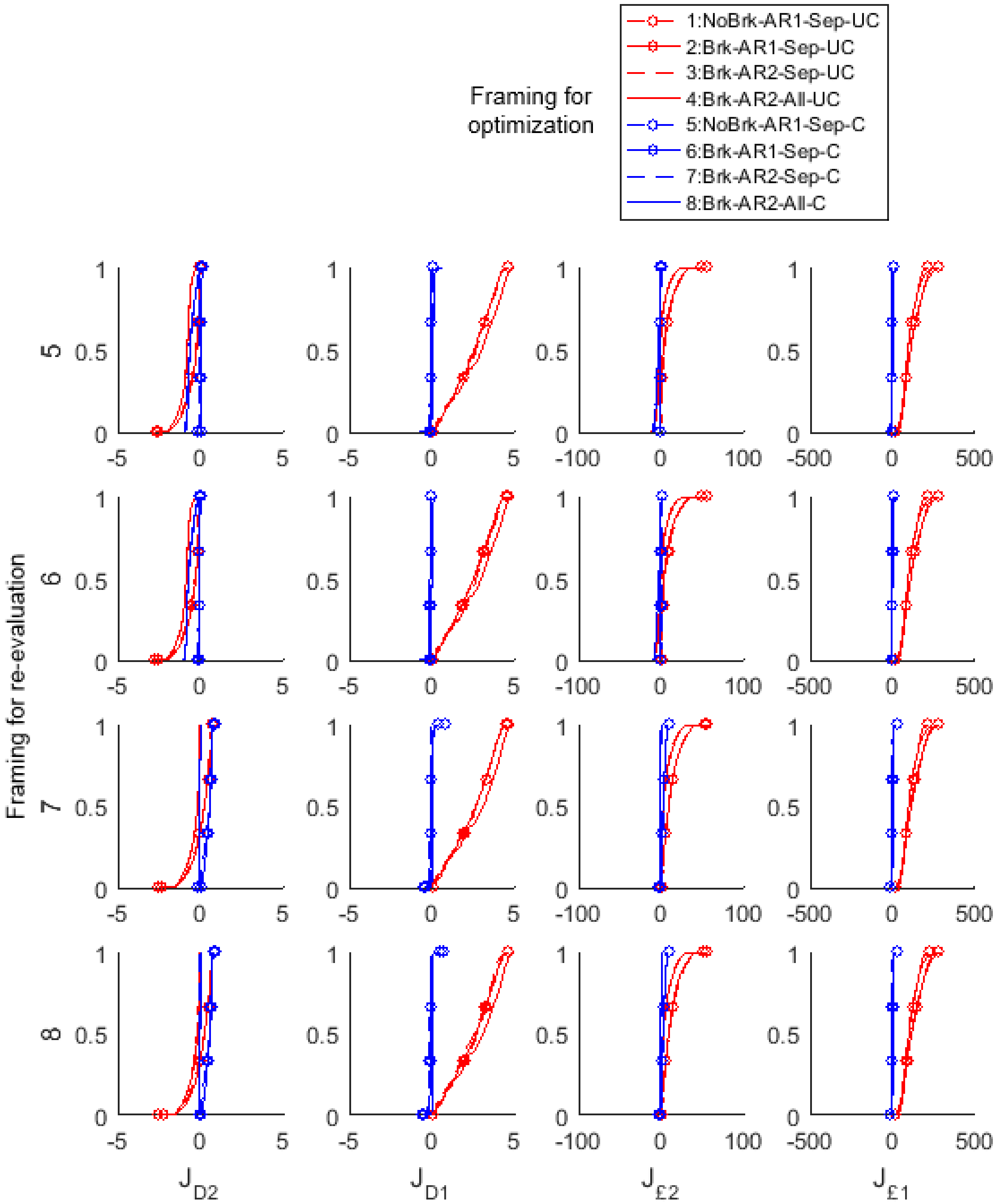


Figure A3.3: CDFs of the difference between re-evaluation and calibration of an objective (positive indicates improvement on re-evaluation and negative indicates deterioration). Each line indicates the framing number as given in the legend, cooperative framings are shown in blue and non-cooperative in red.

In Figure A3.4, for the red points we take sets of Pareto solutions that have been optimized in non-cooperative framings (each non-cooperative framing is a different column) and evaluate them in cooperative framings (each cooperative framing is a different row). Within the full set of Pareto solutions, we extract and plot the objective values in the framing used for re-evaluation that form a trade-off front between the objectives $J_{£1}$ (the X-axis in all plots) and J_{D2} (the Y-axis in all plots). For the black points we do the same but instead of using the non-cooperative framings for optimization, we use the cooperative framings.

These plots are included to show that, regardless of the specific framing, non-cooperative framings always degrade in a specific region of objective space ($J_{£1}$ vs J_{D2}) when evaluated in cooperative framings. This is visible in the plot because the red points are higher than the black points.

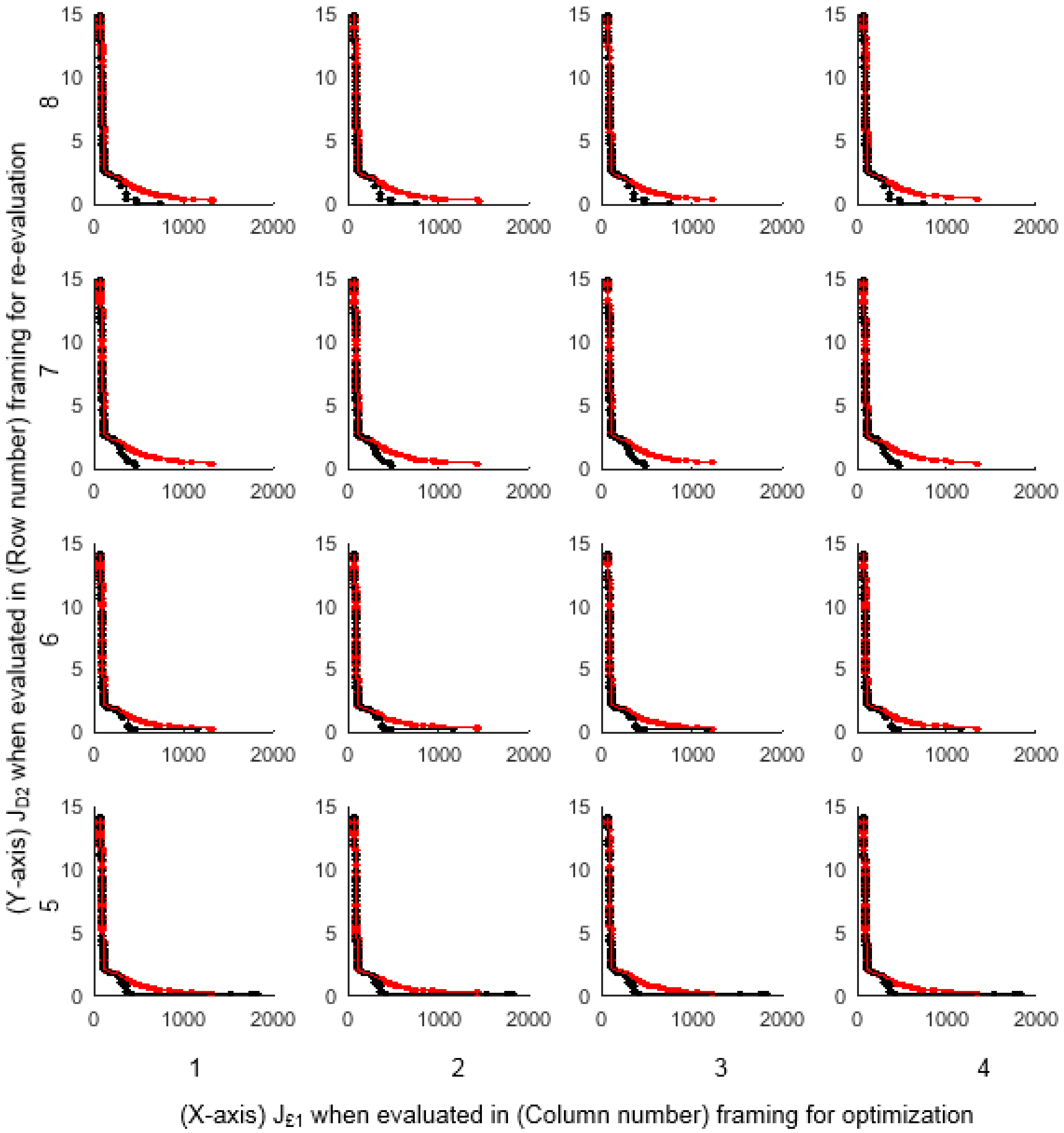


Figure A3.4: (Red points) estimated performances of the set of Pareto solutions optimised under framing indicated by the column numbers re-evaluated using the framing indicated by the row number. Only policies that lie on the trade-off between the pump costs for company 1 (X-axis) and deficit for company 2 (Y-axis) are shown for clarity. (Black points) the same but with the Pareto set of operating policies optimised in the framing indicated by the row number.

Figure A3.5 shows the same as Figure A3.3 but uses a smaller (30 year, 10^4 days) sample size. This figure should be contrasted with Figure A3.3 to show that there are no consistent patterns when a small sample size is used – the differences between framings appear to be random (as opposed to the consistency seen in Figure A3.3).

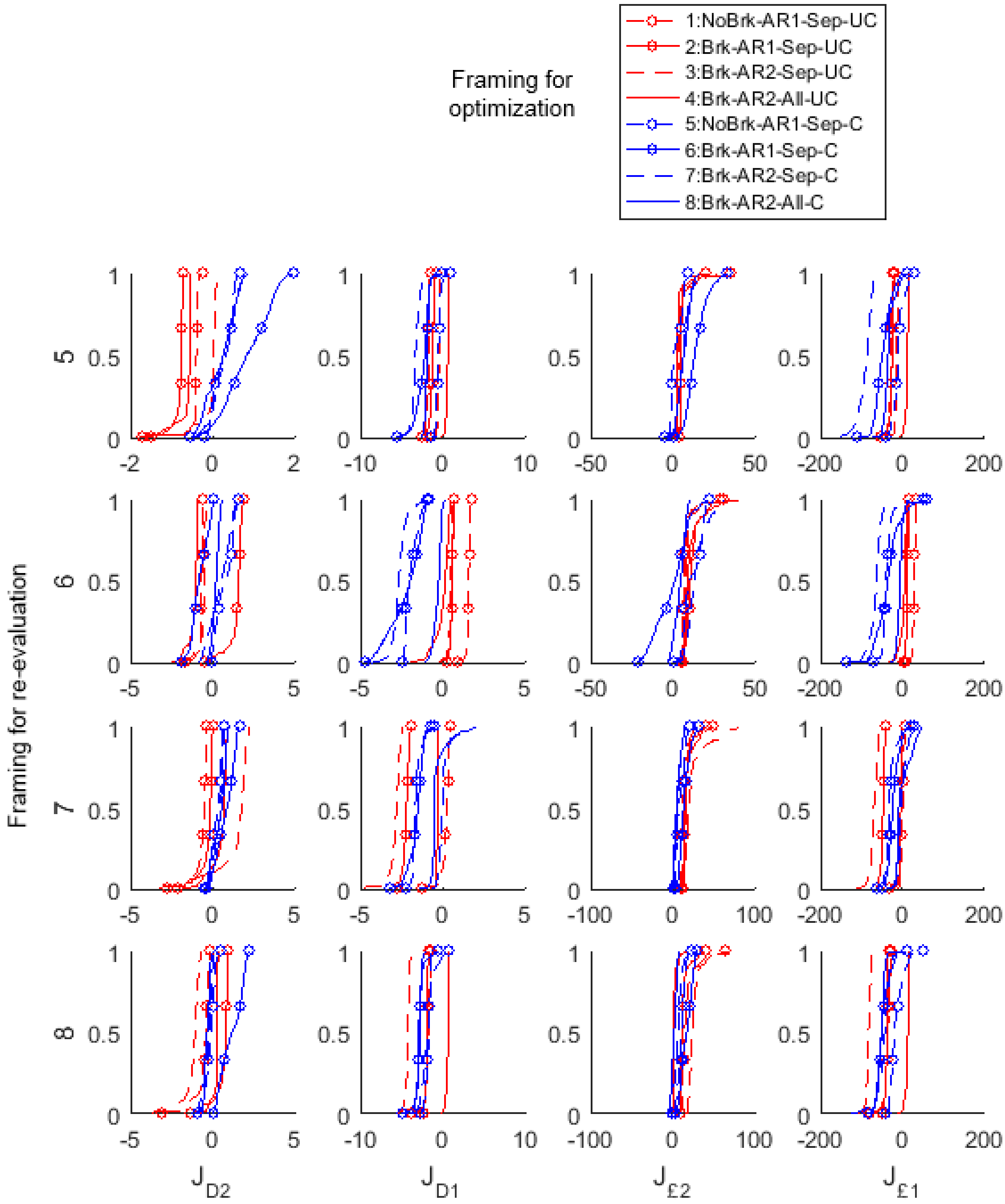


Figure A3.5: Same as in Figure A3.3 but using a much shorter (30 year) simulation length in both optimisation and re-evaluation.

Appendix B: Supporting results for Chapter 5

Figure B.1 shows the performance of individual operating policies when re-evaluated on the three different framings (each row is a different framing). The three colours indicate which framing was used for the purpose of optimization, while the black lines indicate that all three framings were used for optimization in a 'robust optimisation'. There are many lines of each colour because each line represents an individual operating policy's performance over the four objectives and a set of Pareto solutions consists of many (thousands) of operating policies.

This figure shows that the AR2-Reg and Robust sets of Pareto solutions seem to perform better than those optimized in AR0-Reg and AR2-Low for the objective J_{D2}, but worse on objectives J_{D1}, $J_{£2}$ and $J_{£1}$. This plot is included because parallel coordinates plots are a common method to present the objective values of sets of Pareto solutions.

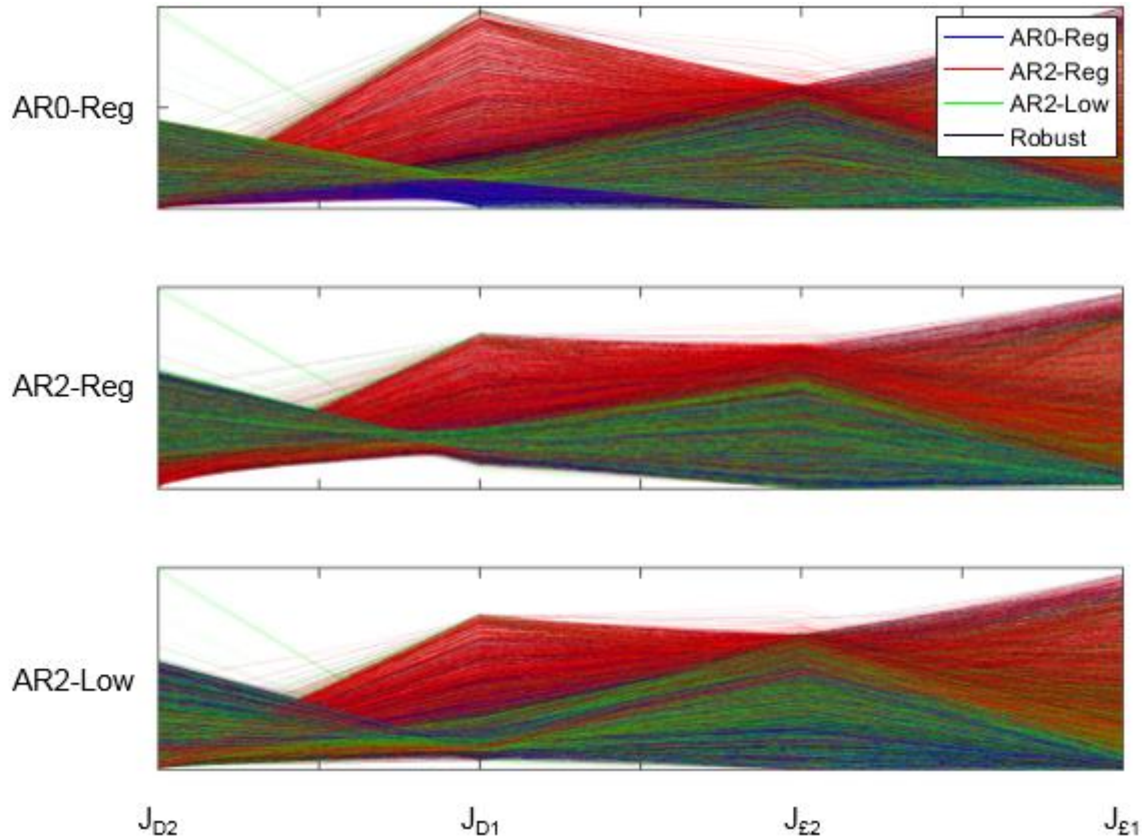


Figure B.1: Parallel coordinates plots of the different sets of solutions. The row indicates the framing used for re-evaluation. The Y-axis is the normalized objective value for each of the 4 objectives (x-axis). Each line denotes the 4 objective values of each specific solution. Colours differentiate the set of solutions. Solutions are plot in a random order to prevent one set of solutions being given a disproportionate amount of visual area.

To create Figure B.2 we take the objective values described for Figure B.1 and make pairwise comparisons between all objectives. We plot the trade-off front that forms when each pairwise comparison is made. These plots show that there are many trade-offs to be made between different objectives but do not do much to distinguish between the different sets of Pareto solutions.

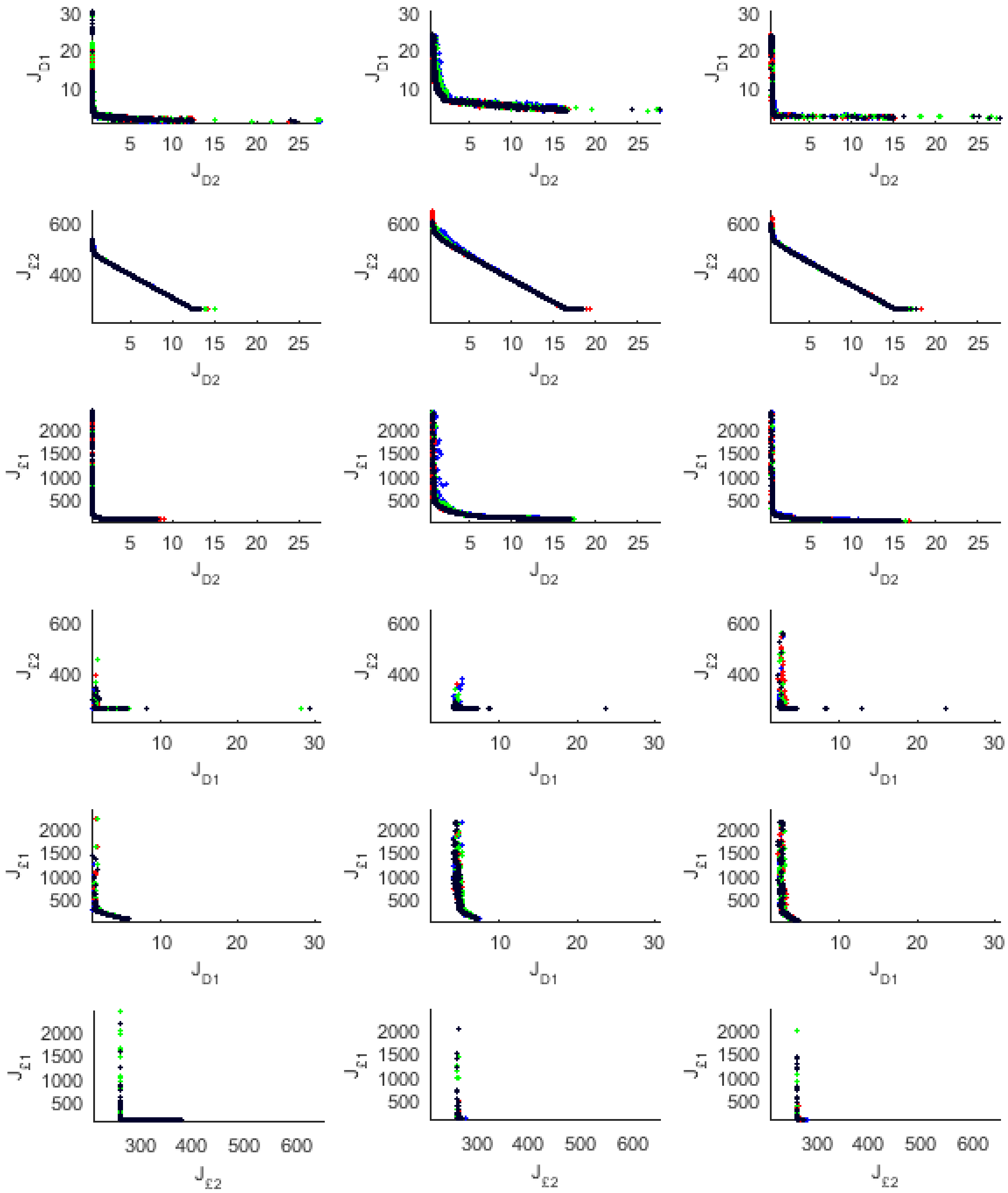


Figure B.2. Pairwise trade-offs between each of the 4 objectives (rows), for each of the 3 framings (columns). These can be considered 2-objective 'slices' of the 4-objective space. Note that these plots show points from 21 repetitions of the experiment (to account for variability in the optimisation process).