On considering robustness in the search phase of Robust Decision Making

A comparison of Many-Objective Robust Decision Making, multi-scenario Many-Objective Robust Decision Making, and Many Objective Robust Optimization

Bartholomew, Erin; Kwakkel, Jan H.

DOI
10.1016/j.envsoft.2020.104699

Publication date
2020

Document Version
Accepted author manuscript

Published in
Environmental Modelling and Software

On Considering Robustness in the Search Phase of Robust Decision Making: a comparison of Many-Objective Robust Decision Making, multi-scenario Many-Objective Robust Decision Making, and Many Objective Robust Optimization

Erin Bartholomew [1], Jan H. Kwakkel [2]

[1] *independent researcher*

[2] *Delft University of Technology, Faculty of Technology, Policy and Management, Section Policy Analysis, Jaffalaan 5 2628 BX Delft, The Netherlands*

Abstract

In recent years, a family of approaches has emerged for supporting decision-making on complex environmental problems characterised by deep uncertainties and competing priorities. Many-Objective Robust Decision Making (MORDM), Multi-scenario MORDM and. Many-Objective Robust Optimization (MORO) differ with respect to the degree to which robustness is considered during the search for promising candidate solutions. To assess the efficacy of these three methods, we compare them using three different policy formulations of the lake problem: inter-temporal, planned adaptive, and direct policy search. The more robustness is considered in the search phase, the more robust solutions are also after re-evaluation but also the lower the performance in individual reference scenarios. Adaptive policy formulations positively affect robustness, but do not reduce the price for robustness. Multi-scenario MORDM strikes a pragmatic balance between robustness considerations and optimality in individual scenarios, at reasonable computational costs.

Keywords: Deep Uncertainty, Many Objective Robust Decision Making, Many Objective Robust Optimization

software availability

All code used for this research can be found at `https://github.com/eebart/RobustDecisionSupportComparison`. The underlying data, because of its shear size, is available upon request from the corresponding author.

1. Introduction

Decision-making and planning of complex environmental systems typically involves various actors with competing preferences, different understandings of the system, and diverging beliefs about the future. To support the decision making on such wicked-problems under deep uncertainty (Rittel and Webber, 1973, Kwakkel et al., 2016c), a variety of decision support approaches, rooted in exploratory modeling (Bankes, 1993, Bankes et al., 2013), have emerged in recent years (Walker et al., 2013a, Kwakkel and Haasnoot, 2019). Given that analysis of deeply uncertain problems cannot reliably depend on a single description of the system under consideration (Quinn et al., 2017a), exploratory modeling uses a series of potential explanations, in the form of computational experiments, to analyze a wicked problem and support the decision making process (Bankes, 2002).

The various robust decision support methods seek a set of robust solutions that achieve satisfactory performance across multiple possible realization of the deep uncertainties (Herman et al., 2015, Bankes, 2002, Kwakkel et al., 2016b). Since many problems often have both deep uncertainties and well-characterized uncertainties, a given realization of the deep uncertainties can have a range of possible outcomes conditional on the exact stochastic realization. To address this, it is quite common to evaluate a given realization of the deep uncertainties for a number of different stochastic realizations of the well characterized uncertainties and take descriptive statistics over these different stochastic realizations as the performance in the given realization of the deep uncertainties. In this paper, a given realization of the various deep uncertainties is called a scenario. The word realization is reserved in the remainder for a stochastic realization of the well characterized uncertainties.

Given the existence of various robust decision support methods, the question is when which method is most appropriate. In an attempt to develop an answer to this, there is an emerging body of literature comparing them (Hall et al., 2012, Kwakkel et al., 2016b, Matrosov et al., 2013, Roach et al., 2015, 2016, Moallemi et al., 2019). This study adds to this literature by comparing three different variations of Robust Decision Making (RDM), a foundational robust decision making method (Lempert et al., 2006, Groves and Lempert, 2007). These variants are Many Objective Robust Decision Making (MORDM) (Kasprzyk et al., 2013), Multi-Scenario MORDM (Watson and Kasprzyk, 2017), and Many Objective Robust Optimization (MORO) (Hamarat et al., 2014, Kwakkel et al., 2015, Trindade et al., 2017).

RDM in essence is an iterative approach where candidate policy alternatives are stress tested over a large ensemble of plausible scenarios (Lempert, 2002). Next, using Scenario Discovery (Bryant and Lempert, 2010, Kwakkel and Jaxa-Rozen, 2016) the conditions under which the candidate policies are failing to meet prespecified performance thresholds are identified. In light of these vulnerabilities, the candidate policies can be refined (Lempert et al., 2006, Groves and Lempert, 2007). RDM provides a structure for comparing previously identified policy alternatives and for discovering how various deeply uncertain factors affect each alternative's performance. That information can then be used to refine the initially identified set of policy alternatives to yield a more robust set of alternatives. This structure is iterative and interactive, allowing analysts and decision makers to work together to stress-test and refine potential policies. The fact that RDM requires a list of promising policy alternatives from the start can prove a difficult challenge when considering problems with multiple conflicting objectives (Kasprzyk et al., 2013). MORDM addresses this by searching for promising policy alternatives using many-objective evolutionary algorithms (MOEA) in a single reference scenario (Kasprzyk et al., 2013).

Multi-Scenario MORDM addresses a recognized weaknesses in MORDM. Namely, that MORDM uses only a single reference scenario for the deeply uncertain factors when searching for promising policy alternatives (Watson and Kasprzyk, 2017). Doing so may yield policy alternatives that perform poorly in future states of the world that deviate from the baseline used during the search. Multi-scenario MORDM reduces this risk by repeating the search for several alternative future states of the world (Watson and Kasprzyk, 2017). These alternative future states are selected to represent conditions that are challenging to address by solutions found for the reference scenario (Eker and Kwakkel, 2018).

Around the same time that MORDM was put forward, an alternative approach was also being pursued. We suggest to label this approach Many Objective Robust Optimization (MORO). Where MORDM and multi-scenario

MORDM are optimizing for a single scenario, MORO considers a set of scenarios and optimizes the robustness of strategies over this set of scenarios. Hamarat et al. (2014) used MORO to find appropriate signposts and triggers for an adaptive energy transition policy. Kwakkel et al. (2015) used MORO to identify the Pareto approximate set of robust policy pathways for climate adaptation. Trindade et al. (2017) explicitly positioned MORO within the MORDM framework, while searching for robust policy pathways for water resources management in the Research Triangle. In essence, MORO generalizes the robustness approach suggested by Deb and Gupta (2006) who suggested optimizing the mean effective objective functions by averaging over a set of neighboring solutions. In MORO, this set of neighboring solutions is replaced with a set of scenarios, while it is recognized that robustness functions other than the mean can be used (McPhail et al., 2018, Kwakkel et al., 2016a).

MORDM, multi-scenario MORDM, and MORO representing increasing considerations of robustness within the search phase (Eker and Kwakkel, 2018). MORDM only considers robustness during the testing over a very large ensemble of scenarios. Multi-scenario MORDM increases the consideration of robustness in the search phase by performing search for multiple scenarios which are selected because they represent conditions under which solutions found for the first reference scenario performs poorly. MORO goes one step further by shifting from optimizing performance in a given scenario to optimizing robustness over a set of scenarios. The trade-off is that by increasing robustness considerations in the search phase, optimally in a reference scenario might decline.

We want to assess the efficacy of the three RDM methods in finding robust solutions as well as the consequences of this for the performance in baseline scenarios. For this, we use three policy formulations of the shallow lake problem (Carpenter et al., 2001), an established case for testing and bench-marking RDM methods. In the lake problem, the aim is to decide on the amount of pollution to put into a lake which maximizes utility, while minimizing the overall pollution in the lake and the chance that the lake is permanently shifted to a different state. The first policy formulation is an inter temporal formulation where one tries to find for each time step the appropriate amount of pollution to put into the lake. The second policy formulation is about finding a decision rule for making a decision for the coming ten years on how much pollution to put into the lake each year. The third policy formulation also searches for a decision rule, but one that is used each year. These three formulations span a continuum from static, via planned adaptive, to a fully adaptive policy architecture (Kwakkel and Haasnoot, 2019).

The remainder of this paper is structured accordingly. In Section 2, we provide a more detailed description of the three robust decision support methods. Section 3 provides more background on the lake problem test case and details the three policy formulations that will be used for comparing the three robust decision support approaches. Section 4 provides the methodological details for how the comparison will be performed. Section 5 contains the results. In Section 6, we present the main conclusions.

2. Model-based approaches for supporting decision making under deep uncertainty

The search for an optimal solution is recognized as an impossible task when faced with deeply uncertain problems. Policy makers have instead looked to an alternate mechanism to analyze the goodness of potential solutions: robustness (Maier et al., 2016). A robust solution is one that performs well across a variety of possible future states of a system, due to both internal and external changes (Herman et al., 2015, Walker et al., 2013a,b). By searching for robust policies one aims to find policies that are not overly sensitive to changes in uncertain parameter values. It is possible that the optimal policy belongs to the set of robust policies (which is known as a super-robust solution). However, it is much more common that robust policies are not optimal under any individual state of the world (Sniedovich, 2016). This is known as *the price of robustness* (Bertsimas and Sim, 2004).

Robustness of a policy can be analyzed from multiple perspectives: resistance to change, avoidance of change, recovery from change, and adaptability in response to change (Durach et al., 2015, de Goede et al., 2013). Because of these various perspectives, there exists many established robustness metrics, each prioritizing a different perspective. Calculating any of these metrics generally involves the same three elements: a set of decision alternatives, several outcomes of interest or performance metrics, and the scenarios or possible future states of the world that will be considered (McPhail et al., 2018). Robustness metrics may determine performance as an absolute calculation or relative to the performance of the other decision alternatives. Each metric also employs differing levels of risk aversion: include more extreme scenarios in calculations to have a higher level of risk aversion (Giuliani and

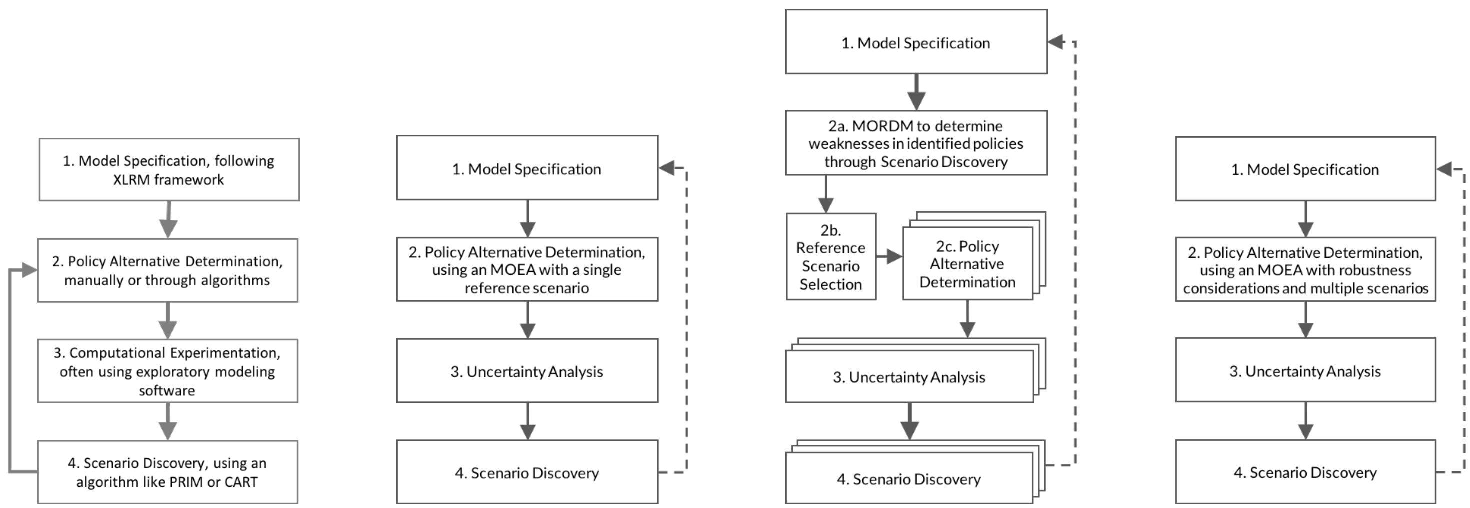
Castelletti, 2016). Finally, each metric has a different method of combining robustness calculations across scenarios for a specific policy option, including mean, standard deviation, skewness, or kurtosis (McPhail et al., 2018).

The search for robust solutions requires assessment of different potential solutions over a large ensemble of scenarios. This set of potential futures cannot be represented by a small number of possibilities (given the large amount of uncertainty that is frequently influenced by multiple input variables, it is generally impossible to codify a short list of possible futures for a problem), but has to instead be described using large ensembles of potential futures, with the number of scenarios stretching anywhere from a few hundred to several million. Lempert et al. (2006) proposed RDM as a method for supporting decision making under deep uncertainty. RDM is an iterative process of model and policy specification, computer aided experimentation that involves the generation and execution of a large ensemble of scenarios that span the defined uncertainty space, development of interactive visualizations, and decision maker input and refinement based on the results of computational experimentation and generated visualization (Lempert et al., 2006). Figure 1a shows this approach.

2.1. Many-Objective Robust Decision Making

Building on RDM, Kasprzyk et al. (2013) propose Multi-Objective Robust Decision Making (MORDM), which provides a structure for managing a wide spectrum of decision maker perspectives and conflicting objectives. Figure 1b indicates how MORDM modifies step 2 of the RDM process. MORDM introduces a formal process to determine a rich set of policy alternatives with different trade-offs on the competing objectives in step 2 through the application of a many-objective evolutionary algorithm (MOEA).

The MORDM method also codifies the process with which to help select a preferred solution from the set of solutions generated with the MOEA, through uncertainty analysis, scenario discovery, and interactive visualizations (Kasprzyk et al., 2013). After model specification and a MOEA search for policy alternatives, the performance of the list of alternatives is re-evaluated or stress tested over a set of possible future states of the world. This set should capture the relevant deeply uncertain factors. This involves building a set of alternative scenarios by sampling across the set of deeply uncertainty parameters. Kasprzyk et al. (2013) recommends using Latin Hypercube Sampling, which ensures that each member of the uncertainty set is represented evenly across the sampled set of scenarios (McKay et al., 1979).



(a) Robust Decision Making (RDM)

(b) Multi-Objective Robust Decision Making (MORDM)

(c) Multi-Scenario MORDM

(d) Many-Objective Robust Optimization (MORO)

Figure 1: four robust decision making approaches

Given the performance of each candidate solution in each scenario, the next step is to analyze this data. The focus is on assessing the robustness of the candidate solutions, and the determination of how the various deeply uncertain factors alone or in combination affect this robustness (Herman et al., 2015). The results of this analysis are communicated through interactive visualizations that decision makers can leverage to examine the robustness of policy alternatives and to better understand the trade-offs that exist between the various objectives. Like RDM, MORDM is intrinsically iterative. If the trade-offs of the various decision alternatives or their robustness is deemed unacceptable, a next iteration starts. However, as MORDM leverages an MOEA to determine alternatives, any refinements occur at the model specification level, where new insights can be used to adjust the uncertainty space, to change the set of decision levers, or modify the objectives.

2.2. Multi-Scenario Many-Objective Robust Decision Making

Multi-scenario MORDM (Watson and Kasprzyk, 2017) is a further refinement of MORDM. The main contribution of MORDM was the use of a MOEA for finding a set of promising candidate solutions which together capture the key trade-offs amongst competing objectives. However, this search uses a single reference scenario, and it is unlikely that solutions that are optimal in a given scenario are also optimally robust. Multi-scenario MORDM (fig. 1c) addresses this by performing a search for candidate strategies for several different reference scenarios. The additional scenarios for which search is performed are selected from regions in the deep uncertainty space where candidate solutions found for the first reference scenario are failing to meet their objectives. So, one performs the four MORDM steps, and based on the insights from scenario discovery, additional scenarios are selected for which search is also performed. The goal of this is to build a more diverse set of policy alternatives which are Pareto optimal under different scenarios.

The selection of scenarios after the first MORDM iteration is a critical step in multi-scenario MORDM (Eker and Kwakkel, 2018). Watson and Kasprzyk (2017) suggest picking scenarios based on the scenario discovery results. The number of scenarios to select is left to the analyst. Clearly, if the number of scenarios for which a search is conducted increases, the chance of finding solutions that are robust during the re-evaluation also increases. However, this comes at a substantial computational cost. To assist in balancing comprehensiveness and computational cost, while making scenario selection transparent and reproducible, Eker and Kwakkel (2018) present an approach for finding the most policy relevant and maximally diverse set of scenarios. Policy relevance is defined as scenarios that lead to poor outcomes and the diversity criterion is based on Carlsen et al. (2016).

2.3. Many-Objective Robust Optimization

While MORDM and multi-scenario MORDM were being developed, a variety of authors under different labels were investigating the role of Many Objective Robust Optimization for supporting planning and design under deep uncertainty (Hamarat et al., 2014, Kwakkel et al., 2015, 2016b, Trindade et al., 2017, Beh et al., 2017). We suggest to label this strand of literature as MORO and explicitly structure it using the RDM framework (see fig. 1d). The main idea uniting this literature is the observation that solutions found through optimization for a reference scenario can have very poor performance in other scenarios. In fact, given the price of robustness, it is unlikely that a solution optimal in any particular scenario is also very robust over a large number of scenarios. Since the overarching aim of supporting decision making under deep uncertainty is the identification of robust strategies that offer an acceptable performance across multiple competing objectives, why not include these robustness considerations already in the search phase for candidate solutions?

In the search phase of MORO, typically, one uses a sampling approach to generate a test set of scenarios over which the robustness of candidate solutions is calculated. One thus approximates the robustness metric over the entire domain by calculating them using an ensemble of scenarios sampled from this domain. So, a candidate solution is evaluated for each scenario. Next, for each outcome of interest, an aggregation function is applied over the performance in each scenario to arrive at a single robustness score for each outcome of interest (Beyer and Sendhoff, 2007, McPhail et al., 2018).

3. The Lake Problem

In order to compare MORDM, multi-scenario MORDM, and MORO, there must be a usable problem that is representative for the class of problems for which these methods have been suggested. Relevant characteristics include, a wicked problem subject to deep uncertainty, a threshold point of no return, where behavior of the system changes dramatically, and the consideration of multiple decision makers with multiple conflicting criteria. The shallow lake problem (Carpenter et al., 1999), a common reference problem in policy analysis research, incorporates all of these characteristics. Over the last decade, the shallow lake problem has repeatedly been used in developing and testing methods for supporting decision making under deep uncertainty (Lempert and Collins, 2007, Quinn et al., 2017b, Singh et al., 2015, Ward et al., 2015, Kwakkel, 2017)

The shallow lake problem is a stylized decision problem in which a town must decide the amount of pollution to release into a nearby shallow lake over time. This hypothetical problem involves two sources of pollution: anthropogenic pollution generated by the town through industrial and agricultural waste, and natural inflows that are uncontrollable and come from the environment. There is also a natural outflow process based on the capability of the lake to recycle resources that is capable of naturally reducing pollution over time in the lake (Hadka et al., 2015). Pollution levels are determined through eq. (1), where X represents the concentration of pollution in the lake, a is the anthropogenic pollution input for the time period, Y refers to the natural inflows of pollution which is described using a lognormal distribution, q refers to the rate at which pollution is recycled the lake's sediment, and b refers to the loss of pollution from the lake through natural outflows. The exact specifications for each of the parameters are based on the lake model developed by Quinn et al. (2017b).

$$X_{t+1} = X_t + a_t + Y_t + \frac{X_t^q}{1 + X_t^q} - bX_t \tag{1}$$

The behavior of the lake problem has a tipping point. If the critical threshold of pollution concentration is surpassed, the trend transitions toward eutrophic equilibrium, making it impossible to return to a healthier oligotrophic equilibrium without active human intervention reducing pollution in the lake (Quinn et al., 2017b).

3.1. Objectives

In the typical setup of the shallow lake problem, there are four conflicting objectives: minimize the maximum pollution level, while maximizing the utility of the release policy to the town, the reliability of the policy, and policy inertia. The multi-objective form of this problem was introduced by Singh et al. (2015) and further developed by Ward et al. (2015), with the goal of introducing objectives that exemplify the conflicts that occur with a diverse group of decision makers and a problem characterized by both stochastic uncertainty (*i.e.,* the stochastic natural inflow), and deep uncertainty. To address the stochastic uncertainty, the model is run for N stochastic realizations and descriptive statistics are taken over these replications.

Maximum Pollution (minimize): Some decision makers such as environmental regulators are seeking to ensure that the maximum pollution level reached in the lake is kept as low as possible (Singh et al., 2015).

$$f_{max\ pollution} = \underset{t \in \{1, \cdots, T\}}{max} \frac{1}{N} \sum_{n=1}^{N} X_{t,n} \tag{2}$$

where $X_{t,n}$ is the concentration of the pollution in year t for stochastic realisation n.

Reliability (maximize): Reliability captures the desire of decision makers to keep the lake below the critical pollution threshold. At the same time, in contrast with the maximum pollution objective, a policy that has high reliability is also accepting of a small amount of pollution, as long as it remains below the critical threshold (Singh et al., 2015).The reliability of a policy is the average reliability for each time step over all realisations N, shown in eq. (3) (Ward et al., 2015).

$$f_{reliability} = \frac{1}{N} \sum_{n=1}^{N} \left(\frac{1}{T} \sum_{t \in T} \theta_{t,n} \right), \text{where } \theta_{t,n} = \begin{cases} 1 & X_{t,n} < P_{crit} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Utility (maximize): To contrast with the objectives that relate the goals common among environmental regulators, utility represents the interests of the town's agriculture and industry, with the goal being to maximize the utility of a policy for those decision makers. Here, α is the utility generated by one unit of antropogenic pollution, while δ is the discount rate. This objective naturally conflicts with the objective of minimizing the pollution level in the lake, providing a valuable dynamic for robust decision support analysis (Ward et al., 2015).

$$f_{utility} = \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{t \in T} \alpha a_{t,n} \delta^t \right) \tag{4}$$

Inertia (maximize): This objective captures the undesirability of large year-over-year changes to the anthropogenic inflow. The aim is to maximize the average inertia of a policy. Like utility, inertia of a policy and for an experiment is first calculated for every time step involved. The mean of that vector of values is what is used to determine inertia-based robustness. Inertia for a single time step in an experiment is determined with eq. (5).

$$f_{intertia} = \frac{1}{N} \sum_{n=1}^{N} \left(\frac{1}{T} \sum_{t \in T} \phi_{t,n} \right), \text{where } \phi_{t,n} = \begin{cases} 1 & |a_{t,n} - a_{t-1,n}| < 0.01 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

3.2. Deep Uncertainties

There are five sources of uncertainty in the definition of the lake problem used for this study. Table 1 shows the uncertainty ranges and base values which have been selected based on the most commonly used settings in literature (Carpenter et al., 1999, Eker and Kwakkel, 2018, Hadka et al., 2015, Quinn et al., 2017b, Ward et al., 2015).

Table 1: Deeply uncertainty variables

Name	**Description**	**Range**	**Reference scenario**
b	Pollution rate of removal through natural outflows	[0.1, 0.45]	0.42
q	Pollution recycling rate through natural processes	[2.0, 4.5]	2.0
μ	Mean of natural pollution inflows	[0.01, 0.05]	0.02
σ	Standard deviation of natural inflows	[0.001, 0.005]	0.0017
δ	Utility discount factor	[0.93, 0.99]	0.98

3.3. Policy Formulations

To ensure a thorough assessment of the relative merits of the three methods, we consider three alternative formulations of the policy problem.

Inter-temporal: Also known as open-loop control, this variation of the lake problem has been used in research several times and involves a series of pre-determined static decisions made every time-step (Hadka et al., 2015, Quinn et al., 2017b, Singh et al., 2015, Ward et al., 2015). This option represents a strictly static approach to solving the lake problem.

Direct Policy Search (DPS): Representing the other extreme in policy structure, direct policy search (DPS) (Giuliani et al., 2016), or closed-loop control. The DPS structure involves optimizing a set of parameters that form a state-aware pollution release rule. This control rule is used to update the level of pollution released at every time-step, giving this policy structure the ability to quickly respond to changes in system conditions. The DPS structure has also been used as a part of the lake problem in research before (Quinn et al., 2017b).

Planned Adaptive DPS: Given that both the inter-temporal and DPS policy structures adapt the pollution release every time period, they do not necessarily represent real-world decision strategy, where it takes time to implement changes. Therefore, this research is proposing a third policy structure that follows the same fundamental structure of the DPS policy, but only makes a decision every τ time steps about the level of pollution that is to be released at each time step, where τ is a number set by the decision makers or policy analysts . For this paper we use $\tau = 10$ (DPS uses $\tau = 1$). Note that Singh et al. (2015) do something similar but with the inter-temporal policy formulation and $\tau = 5$.

4. Approach

4.1. Many-Objective Evolutionary Algorithms

Many-Objective Evolutionary Algorithms (MOEAs) aim at identifying the Pareto approximate set in a multi-objective space (Maier et al., 2019). For this paper we use a novel generational version of BORG (Hadka and Reed, 2013). In essence, we use the auto-adaptive operator selection, adaptive population sizing, and restarts from BORG, but embed them into the ϵ-NSGAII algorithm (Kollat and Reed, 2007, 2006). The motivation for this generational version of BORG is twofold. First, steady-state algorithms like BORG might converge more slowly than generational algorithms such as ϵ-NSGAII (Vavak and Fogarty, 1996). Second, parallelization is possible for BORG (Hadka and Reed, 2014), but it requires some careful design considerations to align the parallelization with the available computing hardware and the nature of the optimization problem. In contrast, a generational algorithm is embarrassingly parallel and thus very easy to parallelize. The main drawback of using a generational algorithm in parallel is the potential of wasted compute resources. Imagine having 100 candidate solutions, where evaluating each solution takes essentially the same run time. If you evaluate this on *e.g.* 24 cores, it requires 4 rounds of evaluations after which 96 candidate solutions have been evaluated. While the last 4 solutions are being evaluated, the remaining 20 cores are idle. Depending on the computational cost of a single function evaluation, this can mean a substantial waste of compute hours. Given the very low run time of the lake problem, this is not a concern for this paper.

To ensure a fair comparison across the different methods and for each policy formulation, we focused on controlling for convergence. Convergence is evaluated based on hypervolume and ϵ-progress (Reed et al., 2013, Ward et al., 2015). For both MORDM and multi-scenario MORDM, 500,000 function evaluations are used. For MORO, 300,000 function evaluations are used. Based on several trails, and the analysis across seeds (see below), this number of function evaluations was adequate to guarantee convergence. In future work, a more formal stopping condition such as the number of unsuccesful restarts might be used for more rigour.

Because there is an element of randomness to the MOEA's process, it is best practice to perform a seed analysis where the algorithm is run multiple times using a different seed for the random number generator. We assessed the variation of identified solutions across seeds, and used this to balance computational costs. For MORDM, we used 50 repetitions; for multi scenario MORDM 20; and for MORO 10. Results were merged across repetitions and filtered using a non-dominated sort.

MORDM is applied using the reference scenario specified in table 1. For multi-scenario MORDM, we followed Eker and Kwakkel (2018) in selecting four additional reference scenarios given the results from MORDM and a re-evaluation over an ensemble of 500 scenarios. Since the way in which the solutions found through MORDM can fail to meet the desired performance thresholds differs across policy formulations, we identify different scenarios for each policy formulation. The values as used in this paper are given in table 2. For MORO, we determine robustness per outcome of interest using the domain criterion (see below, and table 3). To calculate this, we use a set of 50 scenarios sampled from the deep uncertainty space using Latin Hypercube sampling. The set is sampled once, prior to the optimization and stays the same throughout the optimization process. We kept this test set the same across the three policy formulations.

Table 2: Additional reference scenarios used in multi-scenario MORDM

Policy formulation	scenario	Parameters				
		b	q	μ	σ	δ
inter-temporal	1	0.2760	3.0490	0.0039	0.0039	0.9310
	2	0.1350	2.0255	0.0407	0.0030	0.9613
	3	0.2704	2.4783	0.0169	0.0039	0.9631
	4	0.1009	3.6789	0.0187	0.0037	0.9317
planned adaptive	1	0.1690	3.9163	0.0280	0.0024	0.9570
	2	0.2669	2.5997	0.0237	0.0016	0.9607
	3	0.1182	2.1082	0.0474	0.0030	0.9356
	4	0.1334	2.1351	0.0192	0.0029	0.9373
DPS	1	0.2683	3.5029	0.0430.	0.0027	0.9429
	2	0.1009	3.6998	0.0453	0.0044	0.9481
	3	0.2187	2.0506	0.0428	0.0025	0.9604
	4	0.1620	3.8685	0.0388	0.0022	0.9328

4.2. Robustness after re-evaluation under deep uncertainty

McPhail et al. (2018) describe a range of options for determining robustness of policies under conditions of deep uncertainty. To facilitate the comparison of results across methods in this study, a single robustness metric will be used: the domain criterion (Starr, 1963). The domain criterion provides an effective and straightforward way to focus on policies that ensure minimum thresholds of performance are met when considering conflicting objectives. This metric is suitable wherever robustness is considered in any of the three robust decision making approaches. It is also implicitly used when applying Scenario Discovery. Domain criterion satisficing is defined as the fraction of all considered scenarios in which a threshold of performance is met. This results in a metric value between 0 and 1, where 0 indicates that no scenario produced an outcome that met the defined threshold given a specific candidate solution, and 1 indicates that the candidate solution meets the threshold in all scenarios. The threshold values and goal for each outcome can be found in table 3. In order to calculate the robustness metrics, we re-evaluated all candidate solutions resulting from the search phase of each approach across the three policy formulations for the same set of 10,000 scenarios, sampled using Latin Hypercube sampling given the ranges in table 1.

Table 3: Robustness threshold values

Outcome	**Goal**	**Threshold**
Pollution Level	Minimize	Critical Pollution Level
Utility	Maximize	0.75
Inertia	Maximize	0.99
Reliability	Maximize	0.8

The thresholds in table 3 are, were possible, based on previous research (Quinn et al., 2017b, Singh et al., 2015). However, no established threshold has been used for the pollution objective. We therefore choose to use the critical pollution level as defined by Quinn et al. (2017b) as threshold. This means that for each deeply uncertain scenario, we assess whether the average maximum pollution over the stochastic realizations stays below the critical pollution threshold. This is subtly different from the reliability objective as used in an individual scenario, because this objective tracks in each stochatic realization if the threshold is actually crossed.

5. Results

5.1. Robustness after re-evaluation

In discussing the results, we first focus on the results of the re-evaluation under deep uncertainty. We compare the solutions across methods and policy formulations in terms of their robustness on each of the four objectives calculated using the domain criterion and thresholds specified in table 3. Figure 2 shows the robustness on each objective for each method over the rank sorted solutions. Each row corresponds to a different policy formulation. If we look at the inter-temporal policy formulation, we see that by an large, the more robustness is being considered in the search phase, the better robustness remains during re-evaluation. That is, multi-scenario MORDM largely dominates MORDM, and similarly is being dominated by MORO. A similar picture emerges from the DPS formulation. The planned adaptive formulation, however is quite different. On the pollution and reliability objective, multi-scenario MORDM dominates MORO, while for the utility objective it is the inverse. The likely explanation is that the set of 50 scenarios used in the MORO setup biases the optimization towards being more aggressive in exploiting the lake (resulting in better utility) but at the expense of being more likely to destroy the lake as found during the re-evaluation. In contrast, since multi-scenario MORDM optimizes for individual scenarios, and these scenarios have been selected to represent primarily challenging conditions, the approach produces many more candidate solutions that are more cautious in exploiting the lake. The reason that this happens for the planned adaptive formulation is that since you can only update your release decision every 10 time steps, solutions are biased towards more conservative solutions.

Figure 2 show the performance on the individual objectives, at the expense of hiding information on trade-offs across the objectives. A parallel coordinate visualization of the results is shown in Figure 3 to provide insight into these robustness trade-offs. Again, the policy formulation is on the rows, with each column now being a method. If we look at the inter-temporal policy formulation, we see roughly the same pattern across the three methods. The three methods produce solutions that after re-evaluation similarly span the robustness space. However, we can also see that by increasing the robustness considerations during the search phase we are able to improve the

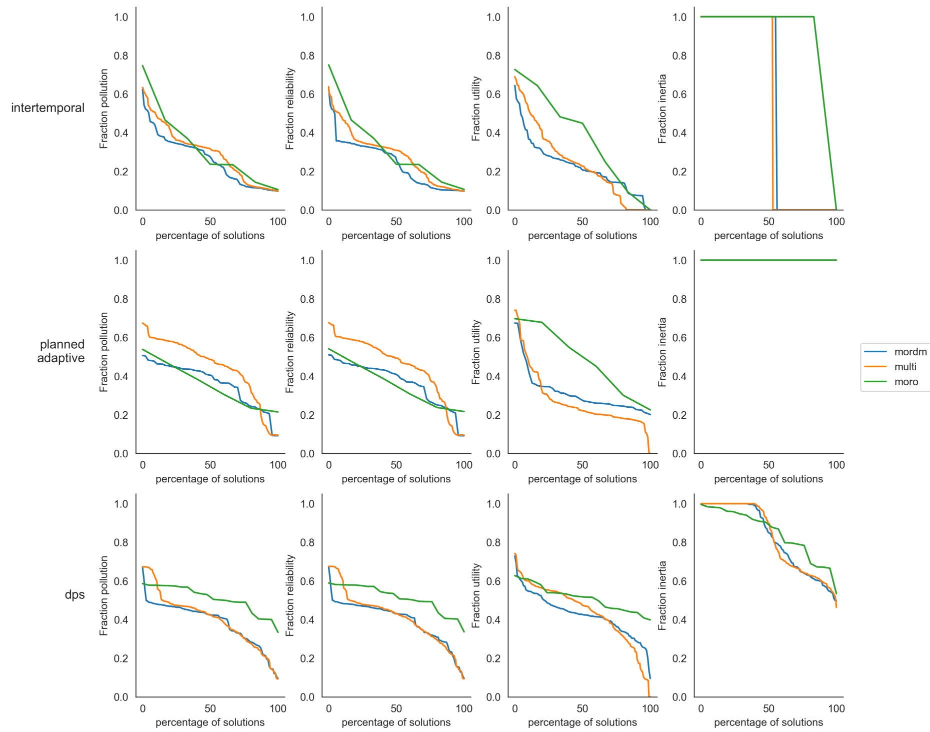


Figure 2: Rank sorted robustness scores using the domain criterion for the solutions found for each policy formulation, grouped by method

robustness trade-offs that we find. For example, multi-scenario MORDM finds solutions that can sustain a much higher robustness performance on pollution and reliability with a similar poor performance on utility as found with normal MORDM. Similarly, multi-scenario MORDM can combine the best robustness performance on utility with the best performance on inertia, something normal MORDM was unable to find. MORO in turn improves on this compared to multi-scenario MORDM, with higher robustness scores on pollution, reliability and utility. Note however, that the basic trade-offs do not change drastically across the three methods. A similar pattern of increasing robustness can be seen for planned adaptive and DPS. Although here, in particular on the utility objective, multi-scenario MORDM produces a much broader range of robustness scores. This suggests two things: multi-scenario MORDM helps finding promising solutions by performing the search phase for multiple different scenarios, but also that there seems to be a dependency between the scenario under which solutions are found and how robust they are when re-evaluated over a much larger set of scenarios.

Table 4 shows the hypervolume for each method across the three problem formulations. The hypervolume is based on the robustness scores for each of the four objectives after re-evaluation. This table reinforces the previous results. Also in terms of hypervolume, multi-scenario MORDM produces slightly better results than MORO. Interestingly, this is true across problem formulations. An important caveats here is that the number of solutions for multi-scenario MORDM is much larger than the the number of solutions for MORO, which can partly explain the difference.

Table 4: hypervolume in robustness space for each method across the three problem formulations

	inter-temporal	**planned adaptive**	**DPS**
MORDM	0.044	0.010	0.153
Multi-scenario MORDM	0.064	0.142	0.216
MORO	0.058	0.122	0.190

The results hitherto suggest that multi-scenario MORDM might be performing as good if not better than MORO. Is this really true? To assess this, we first merged all Pareto sets across methods for each policy formulation. Next, we performed a non dominated sort on this and counted the number of solutions from each method that are in the non dominated set. Table 5 shows these results. In between brackets, we also give the total number of solutions from each method. Note again that multi-scenario MORDM has a much higher number of solutions, because it is based on the results of performing separate optimizations for 5 scenarios. Interestingly, all the solutions identified through MORO are always present also in the combined Pareto set. MORO thus has much stronger guarantees of finding solutions in the Pareto optimal set in robustness space after re-evaluation, as compared to MORDM and multi-scenario MORDM.

Figure 4 visualizes the results of the combined Pareto set for each policy formulation, with colors denoting the different methods. If we focus on comparing multi-scenario MORDM and MORO, it appears that the solutions identified by MORO might offer a better way of balancing across objectives. For example, for the inter-temporal formulation (fig. 4a) typically MORO solutions appear to be quite similar in their robustness on the pollution and

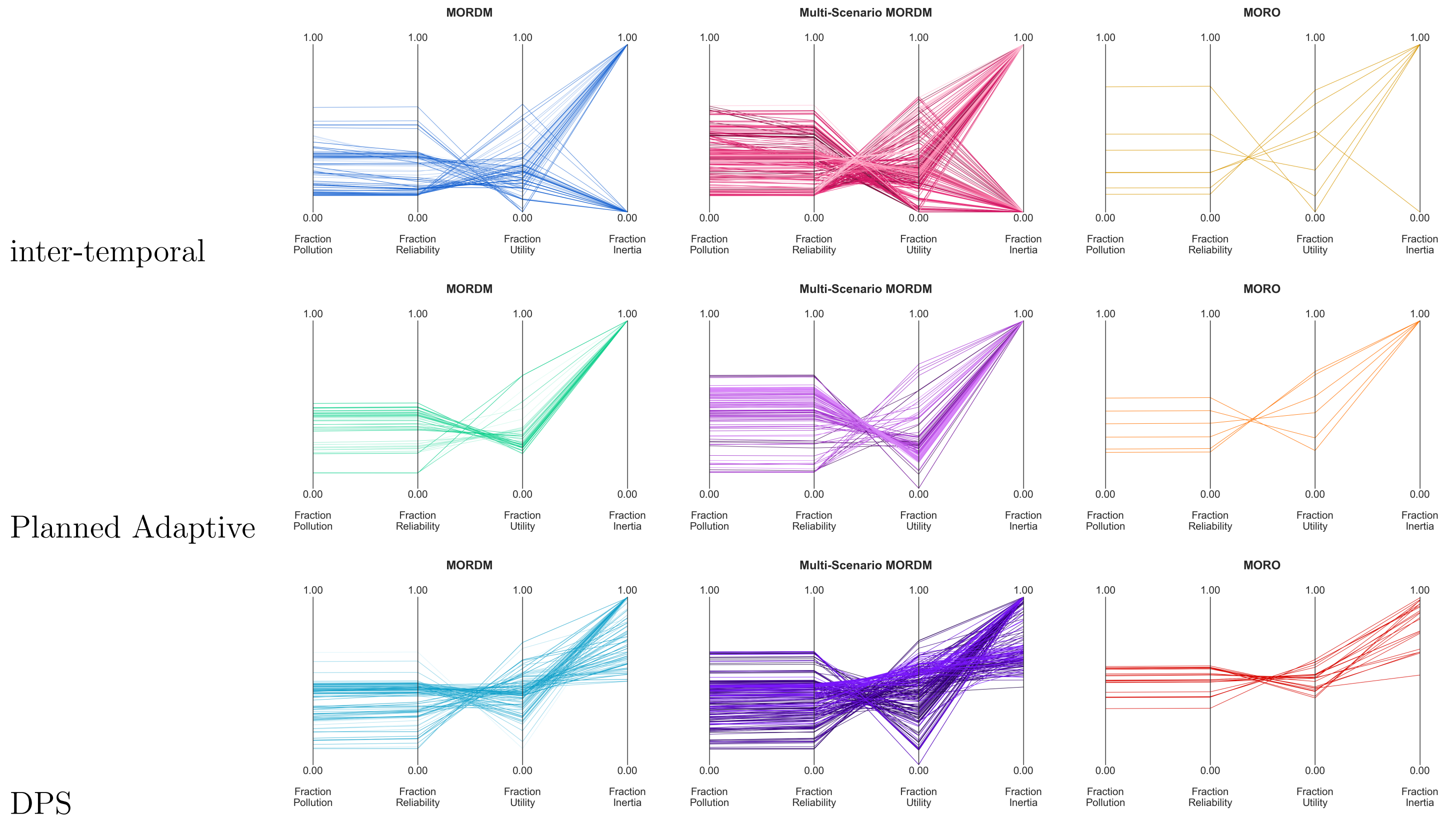


Figure 3: Parallel coordinate plots for robustness of the solutions found during the search phase

Table 5: Number of solutions in Pareto set when compared across methods per problem formulation, the number in brackets is the size of the original Pareto set

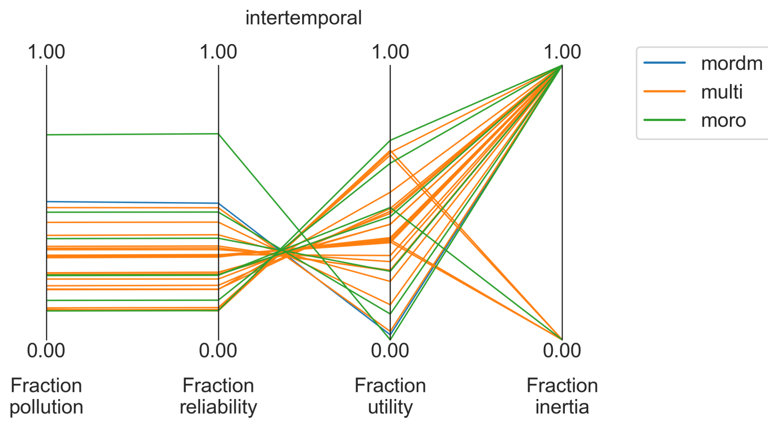
	inter-temporal	**planned adaptive**	**DPS**
MORDM	1 (90)	2 (48)	6 (110)
Multi-scenario MORDM	25 (291)	26 (113)	58 (209)
MORO	7 (7)	6 (6)	22 (22)

reliability objective as solutions found through multi-scenario MORDM, but offer clearly better robustness on utility. Or vice versa. This pattern persists across the other two policy formulations (fig. 4b and fig. 4c). Again, MORO is able to almost match robustness on either utility, or pollution and reliability, with a substantial increase in robustness on the other objective(s). This suggests that not only are all solutions found through MORO retained in the Pareto set if we combine the results across the three methods, it also seems that the solutions found through MORO might be more interesting compromise solutions in terms of robustness for the given case analyzed here.

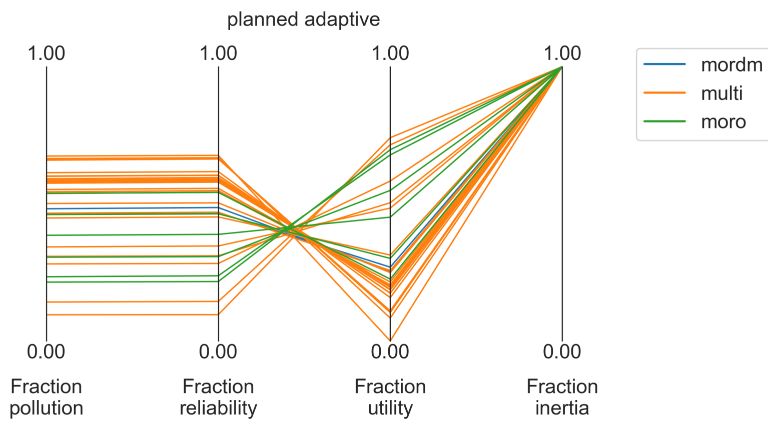
5.2. The price of robustness

In our analysis so far, we have focused on the robustness of solutions found through the three different methods across the different policy formulations. Robustness however often comes at the price of optimality in a given scenario. To assess this price of robustness, we compare the results found through the three methods for the reference scenario assumed by MORDM as shown in table 1 as well as the additional reference scenarios as used in multi-scenario MORDM as shown in table 2.

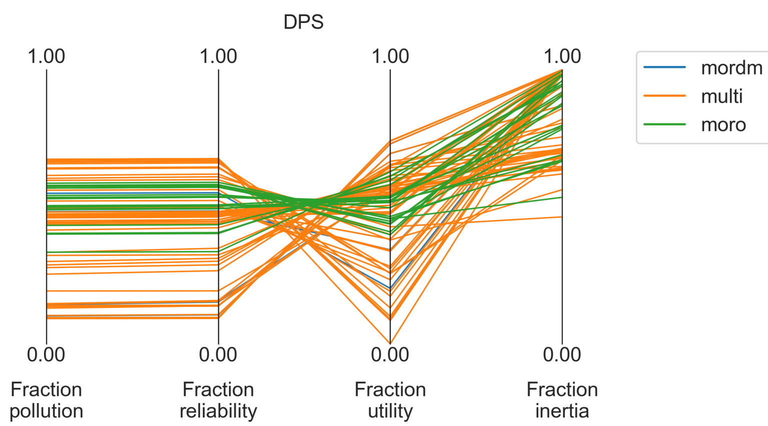
Table 6 shows the hypervolume of the solutions found by each method for each policy formulation when evaluation in each of the five reference scenarios. For this, each solution found by each method is re-evaluated for each of the five scenarios. Next, we identity the Pareto approximate set for each unique combination of method, policy formulation and scenario and calculate its hypervolume. To ensure comparisons, the hypervolume is normalized for each scenario per policy formulation. Scenario 0 is the baseline scenario, while the remainder are the additional scenarios as used in multi-scenario MORDM. For the reference scenario assumed by MORDM, MORDM al-



(a) inter-temporal policy formulation



(b) planned adaptive policy formulation



(c) direct policy search policy formulation

Figure 4: parallel coordinate plot of solutions after non-dominated sort on combined set of archives per policy formulation

Table 6: hypervolume per reference scenario for each policy formulation

		scenarios				
		0	1	2	3	4
static	MORDM	0.283	0.329	0.02	0.169	0.008
	multi-scenario MORDM	0.268	0.284	0.021	0.237	0.068
	MORO	0.053	0.454	0.016	0.061	0.053
		0	1	2	3	4
planned adaptive	MORDM	0.348	0.023	0.027	0.025	0.034
	multi-scenario MORDM	0.303	0.194	0.235	0.039	0.065
	MORO	0.25	0.009	0.007	0.01	0.01
		0	1	2	3	4
dps	MORDM	0.361	0.271	0.055	0.042	0.086
	multi-scenario MORDM	0.325	0.33	0.071	0.057	0.105
	MORO	0.237	0.35	0.01	0.003	0.025

ways finds the Pareto approximate set with the highest hypervolume, closely followed by multi-scenario MORDM. For the other four scenarios, typically multi-scenario MORDM has the highest hypervolume. There are however a few exceptions. For example, in case of the static formulation for scenario 1, both MORDM and MORO result in a higher hypervolume. Also, for the DPS formulation for scenario 1 and 2 MORO has a slightly higher hypervolume than multi-scenario MORDM. Remember that the reference scenarios are specific to the policy formulation. Outside these two exceptions, however, MORO results in a substantially lower hypervolume, suggesting there is a substantial loss in performance in individual scenarios if one tries to be maximally robust.

Table 7 shows the total number of solutions in the Pareto approximate set for each method for each policy formulation, as well as the number of solutions that remain in the Pareto set when evaluated only in one of the reference scenarios. Specifically, we merge the performance of the solutions on a scenario by scenario basis for each policy formulation. Next, we perform a non-dominated sort on this combined set. Finally we count the number of solutions found by each method that are in the resulting Pareto approximate set. Similar to the observations for hypervolume, in general the method which explicitly optimized for a given scenario has the highest number of solutions that remain in the Pareto approximate set for that scenario when compared with the solutions found by the other methods. In addition, for the static policy formulation, only a few solutions found by MORDM are also in the Pareto approximate set of the other four scenarios. For the planned adaptive and DPS formulation, this pattern persists but not to the extreme seen for the static formulation. For MORO, there seems to be always at least one scenario in which many of the solutions identified are also in the Pareto set.

So what do these results imply for the price of robustness. First, optimizing for robustness comes in general at the expense of attainable hypervolume in any given reference scenario. The nature of the policy formulation, ranging from static to adaptive does not seem to strongly affect this. For each policy formulation, examples of scenarios were the price is low (or even negative) exists, but there are also scenarios were the price of robustness is quite high. Similarly, the number of solutions found through MORO that are also in the Pareto approximate set for any given scenario is typically quite small, although for each policy formulation scenarios that are an exception to this exist as well.

5.3. Computational costs

Next to the trade-off between robustness over a set of scenarios and optimality in a given scenario, another major concern is the computational cost associated with finding these solutions. As indicated by table 8, a MORO analysis has a significantly higher computational cost then either MORDM or multi-scenario MORDM. For the inter-temporal problem, the difference between multi-scenario MORM and MORO is a factor 6, while for the other two policy formulations it is a factor 10. The increased computational cost had a significant impact on the time it took to complete the analysis even for a highly-stylized and relatively low computational cost problem like the lake problem used in this analysis and can have an even more significant impact

Table 7: Number of solutions that remain in the Pareto set for each reference scenario for each policy formulation

			scenarios				
			0	1	2	3	4
static	MORDM	90	85	9	4	41	13
	multi-scenario MORDM	200	44	46	50	90	103
	MORO	7	2	6	2	2	3
			0	1	2	3	4
planned adaptive	MORDM	48	46	34	31	34	35
	multi-scenario MORDM	77	52	61	58	43	51
	MORO	6	3	2	1	6	4
			0	1	2	3	4
dps	MORDM	110	109	73	83	30	75
	multi-scenario MORDM	94	66	83	43	30	70
	MORO	22	4	16	2	2	12

when considering policy problems that require significantly more complex models with more sources of uncertainty than are present in the lake problem.

6. Conclusions

In recent years various approaches have been put forward to aid multi-actor deliberation and decision-making on complex environmental problems characterized by deep uncertainty. One family of approaches relies on the iterative stress testing of candidate solutions. In this paper we considered three variants within this family which differ with respect to how they identify the candidate solutions to be stress tested. MORDM uses many-objective optimization for a reference scenario. Multi-scenario MORDM extents this by performing the optimization several times for different scenarios. MORO instead optimizes for robustness directly, where robustness is established based on the performance of solutions in a small ensemble of scenarios.

Table 8: Number of function evaluations for each the three methods for each policy formulation for a single run of the MOEA. The total computation costs expressed in function evaluations of the lake model is in the final row.

		MORDM	**Multi-Scenario MORDM**	**MORO**
NFE in MOEA	Inter-temporal	500,000	500,000	300,000
	Planned Adaptive	100,000	100,000	100,000
	DPS	100,000	100,000	100,000
Number of scenarios		1	1	50
Search repetitions		1	1+4	1
total NFE	Inter-temporal	500,000	2,500,000	15,000,000
	Planned Adaptive	100,000	500,000	5,000,000
	DPS	100,000	500,000	5,000,000

To assess the efficacy of MORDM, multi-scenario MORDM, and MORO, we applied them to three policy formulations of the shallow lake problem. These three formulations spanned the space from a static policy formulation, via a planned adaptive policy formulation, to a fully adaptive closed loop control policy formulation. We find that the more robustness is considered in the search phase of robust decision making, the higher the robustness attainment of the resulting solutions will be during re-evaluation. Vice versa, optimizing for robustness comes at the expense of optimality in any given scenario. There are however a few caveats.

First, the more adaptive the policy formulation, the more robust solutions are even if found through MORDM. Multi-scenario MORDM, by optimizing specifically for scenarios that represent conditions under which solutions found through normal MORDM perform poorly, is able to identify solutions which are substantially more robust also after re-evaluation. MORO has the strongest guarantees that its solutions are robust also after re-evaluation, irrespective of the policy formulation

Second, when analysing the price of robustness, we see that MORO pays a high price. Only few solutions are in the Pareto set for a specific scenarios, and the hypervolume of the MORO solutions in a given scenario is often quite low as well. Interestingly, the policy formulation seems to not have a clear influence here.

Third, a major challenge for both multi-scenario MORDM and MORO is the selection of the scenarios to use. Multi-scenario MORDM, by selecting scenarios from the region where the solutions found in the first search performed poorly, intrinsically biases subsequent results towards solutions that do well in this region. But there is no *a-priori* reason to assume that these resulting solutions might not be vulnerable in a different way. In the lake problem, the conditions under which any of the solutions, irrespective of the policy formulation and method, is vulnerable, is essentially the same. Yes, the volume of the space within which a given solution is vulnerable might be a bit larger or a bit smaller, but the dimensions which characterize this space stay the same. It is quite plausible that in many other infrastructure cases this does not hold: different adaptive strategies might be vulnerable to quite different conditions (see e.g., Hamarat et al., 2013).

MORO is in principle less vulnerable to the selection of scenarios, since it relies on sampling scenarios from the complete deep uncertainty space rather than a specific subspace. However, for such a sample to be representative of the entire space, often many more samples are required compared to multi-scenario MORDM. MORO thus in general will have substantially higher computational costs. Reducing this costs requires developing techniques to carefully select a small set of scenarios that enable an accurate estimation of the robustness found after re-evaluation. Giudici et al. (2020) offer a nice example of what such a scenario selection technique might entail. Since all solutions identified through MORO remained Pareto optimal after re-evaluation, using 50 scenarios during the robust optimization seems to be sufficient for lake problem as considered in this paper. There is no guarantee that this will hold in general. Research is needed into the selection scenarios which as a set contain the appropriately stressing conditions against which solutions have to be robust, while also capturing the scenarios under which one would like to have near optimal performance.

In light of these caveats, we suggest that in general multi-scenario MORDM is the preferred method. It offers a balance between optimality in various reference scenarios and robustness over a larger ensemble, while requiring only a relatively modest increase in computation costs as compared to MORDM. Only in case of a static policy formulation and a very clear emphasis on robustness, would MORO be the more appropriate method.

In this paper we used the ubiquitous shallow lake problem, but with an additional intermediate policy formulation. Interestingly, this intermediate policy formulation produces the more surprising results. Multi-scenario MORDM seems to work almost as well if not better than MORO for this case. This raises a more general concern. The inter-temporal and the DPS version of the lake problem are essentially control problems where at each step action can be taken. And although it can be useful to draw an analogy between optimal control and strategic planning (Herman et al., 2020), we suggest that real world decision making on infrastructure systems deviates from this in relevant ways highlighted in part by the planned adaptive policy formulation used in this paper. There can be multiple years between a decision and its implementation due to construction time. Budget considerations and financial risks can further limit the ability to implement actions if and when desired. The comparative literature on robust decision making approaches would benefit from having benchmark problems that better reflect the reality of infrastructure problems. The Waas case (Haasnoot et al., 2012, Kwakkel and Pruyt, 2015) and, with some adaptation, the Eldorado case (Smith et al., 2018) might potentially be used to further explore this.

References

Bankes, S. C. (1993). Exploratory Modeling for Policy Analysis. *Operations Research*, 4(3):435–449.

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(3):7263–7266.

Bankes, S. C., Walker, W. E., and Kwakkel, J. H. (2013). Exploratory Modeling and Analysis.

Beh, E. H., Zheng, F., Dandy, G. C., Maier, H. R., and Kapelan, Z. (2017). Robust optimization of water infrastructure planning under deep uncertainty using metamodels. *Environ. Model. Softw.*, 93(C):92–105.

Bertsimas, D. and Sim, M. (2004). The Price of Robustness. *Operations Research*, 52(1):35–53.

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

Bryant, B. P. and Lempert, R. J. (2010). Thinking Inside the Box: a participatory computer-assisted approach to scenario discovery. *Technological Forecasting and Social Change*, 77(1):34–49.

Carlsen, H., Eriksson, E. A., Dreborg, K. H., Johansson, B., and Bodin, O. (2016). Systematic exploration of scenario spaces. *Foresight*, 18(1):59–75.

Carpenter, S. R., Ludwig, D., and Brock, W. A. (1999). Management of Eutrophication for lakes subject to potentially irreversible change. *Ecological Applications*, 9(3):751–771.

Carpenter, S. R., Walker, B., Anderies, J. M., and Abel, N. (2001). From Metaphor to Measurement: Resilience of what to what. *Ecosystems*, 4:765–781.

de Goede, D., Gremmen, B., and Blom-Zandstra, M. (2013). Robust agriculture: Balancing between vulnerability and stability. *NJAS - Wageningen Journal of Life Sciences*, 64-65:1–7.

Deb, K. and Gupta, H. (2006). Introducing robustness in multi-objective optimization. *Evolutionary Computation*, 14(4):463–494.

Durach, C. F., Wieland, A., and Machuca, J. A. (2015). Antecedents and dimensions of supply chain robustness: a systematic literature review. *International Journal of Physical Distribution & Logistics Management*, 45(1/2):118–137.

Eker, S. and Kwakkel, J. H. (2018). Including robustness considerations in the search phase of Many-Objective Robust Decision Making. *Environmental Modelling & Software*, 105:201–2016.

Giudici, F., Castelletti, A., Giuliani, M., and Maier, H. R. (2020). An active learning approach for identifying the smallest subset of informative scenarios for robust planning under deep uncertainty. *Environmental Modelling & Software*, page 104681.

Giuliani, M. and Castelletti, A. (2016). Is robustness really robust? How different definitions of robustness impact decision-making under climate change. *Climatic Change*, 135(3-4):409–424.

Giuliani, M., Castelletti, A., Pianosi, F., Mason, E., and Reed, P. M. (2016). Curses, Tradeoffs, and Scalable Management: Advancing Evolutionary Multiobjective Direct Policy Search to Improve Water Reservoir Operations. *Journal of Water Resources Planning and Management*, 142(2).

Groves, D. G. and Lempert, R. J. (2007). A New Analytic Method for Finding Policy-Relevant Scenarios. *Global Environmental Change*, 17(1):73–85.

Haasnoot, M., Middelkoop, H., Offermans, A., van Beek, E., and van Deursen, W. P. A. (2012). Exploring pathways for sustainable water management in river deltas in a changing environment. *Climatic Change*, 115(3-4):795–819.

Hadka, D., Herman, J. D., Reed, P. M., and Keller, K. (2015). OpenMORDM: An open source framework for many objective robust decision making. *Environmental Modelling & Software*, 74:114–129.

Hadka, D. and Reed, P. M. (2013). Borg: An Auto-Adaptive Many-Objective Evolutionary Computing Framework. *Evolutionary Computation*, 21(2):231–259.

Hadka, D. and Reed, P. M. (2014). Large-scale parallelization of the Borg multiobjective evolutionary algorithm to enhance the management of complex environmental systems. *Environmental Modelling & Software*.

Hall, J. W., Lempert, R. J., Keller, A., Hackbarth, A., Mijere, C., and McInerney, D. (2012). Robust Climate Policies Under Uncertainty: A Comparison of Robust Decision Making and Info-Gap Methods. *Risk Analysis*, 32(10):1527–1672.

Hamarat, C., Kwakkel, J. H., and Pruyt, E. (2013). Adaptive Robust Design under Deep Uncertainty. *Technological Forecasting and Social Change*, 80(3):408–418.

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

Herman, J. D., Quinn, J. D., Steinschneider, S., Giuliani, M., and Fletcher, S. (2020). Climate adaptation as a control problem: Review and perspectives on dynamic water resources planning under uncertainty. *Water Resources Research*, 56(2):e24389. e24389 2019WR025502.

Herman, J. D., Reed, P. M., Zeff, H. B., and Characklis, G. W. (2015). How should robustness be defined for water systems planning under change. *Journal of Water Resources Planning and Management*, 141(10).

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

Kollat, J. B. and Reed, P. M. (2006). Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. *Advances in Water Resources*, 29(6):792–807.

Kollat, J. B. and Reed, P. M. (2007). A computational scaling analysis of multiobjective evolutionary algorithms in long-term groundwater monitoring applications. *Advances in Water Resources*, 30(3):408–419.

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

Kwakkel, J. H., Eker, S., and Pruyt, E. (2016a). How Robust is a Robust Policy? Comparing Alternative Robustness Metrics for Robust Decision-making. In Doumpos, M., Zopounidis, C., and Grigoroudis, E., editors, *Robustness Analysis in Decision Aiding, Optimization, and Analytics*. Springer.

Kwakkel, J. H. and Haasnoot, M. (2019). Supporting dmdu: A taxonomy of approaches and tools. In Marchau, V. A. W. J., Walker, W. E., Bloemen, P. J. T. M., and Popper, S. W., editors, *Decision Making under Deep Uncertainty: From Theory to Practice*, pages 355–374. Springer International Publishing, Cham.

Kwakkel, J. H., Haasnoot, M., and Walker, W. E. (2015). Developing Dynamic Adaptive Policy Pathways: A computer-assisted approach for developing adaptive strategies for a deeply uncertain world. *Climatic Change*, 132(3):373–386.

Kwakkel, J. H., Haasnoot, M., and Walker, W. E. (2016b). Comparing Robust Decision-Making and Dynamic Adaptive Policy Pathways for Model-Based Decision Support under Deep Uncertainty. *Environmental Modelling & Software*, 86:168–183.

Kwakkel, J. H. and Jaxa-Rozen, M. (2016). Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes. *Environmental Modelling & Software*, 79:311–321.

Kwakkel, J. H. and Pruyt, E. (2015). Using System Dynamics for Grand Challenges: The ESDMA Approach. *Systems Research and Behavioral Science*, 32(3):358–375.

Kwakkel, J. H., Walker, W. E., and Haasnoot, M. (2016c). Coping with the Wickedness of Public Policy Problems: Approaches for Decision Making under Deep Uncertainty. *Journal of Water Resources Planning and Management*, 142(3).

Lempert, R. J. (2002). A New Decision Sciences for Complex Systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3):7309–7313.

Lempert, R. J. and Collins, M. (2007). Managing the Risk of Uncertain Threshold Response: Comparison of Robust, Optimum, and Precautionary Approaches. *Risk Analysis*, 24(4):1009–1026.

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

Maier, H., Razavi, S., Kapelan, Z., Matott, L., Kasprzyk, J., and Tolson, B. (2019). Introductory overview: Optimization using evolutionary algorithms and other metaheuristics. *Environmental Modelling & Software*, 114:195–213.

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

Matrosov, E. S., Woords, A. M., and Harou, J. J. (2013). Robust Decision Making and Info-Gap Decision Theory for water resource system planning. *Journal of Hydrology*, 494(28 June 2013):43–58.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2):239–245.

McPhail, C., Maier, H. R., Kwakkel, J. H., Giuliani, E., Castelletti, A., and Westra, S. (2018). Robustness metrics: How are they calculated, when should they be used and why do they give different results? *Earth's Future*.

Moallemi, E., Elsawah, S., and Ryan, M. (2019). Robust decision making and epoch–era analysis: A comparison of two robustness frameworks for decision-making under uncertainty. *Technological Forecasting and Social Change*.

Quinn, J. D., Reed, P. M., Giuliani, M., and Castelletti, A. (2017a). Rival framings: A framework for discovering how problem formulation uncertainties shape risk management trade-offs in water resources systems. *Water Resources Research*.

Quinn, J. D., Reed, P. M., and Keller, K. (2017b). Direct policy search for robust multi-objective management of deeply uncertain socio-ecological tipping points. *Environmental Modelling & Software*, 92:125–141.

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

Rittel, H. W. J. and Webber, M. W. (1973). Dilemmas in a General Theory of Planning. *Policy Sciences*, 4(2):155–169.

Roach, T., Kapelan, Z., and Ledbetter, R. (2015). Comparison of info-gap and robust optimisation methods for integrated water resource management under severe uncertainty. *Procedia Engineering*, 119:874–883.

Roach, T., Kapelan, Z., Ledbetter, R., and Ledbetter, M. (2016). Comparison of Robust Optimization and Info-Gap Methods for Water Resource Management under Deep Uncertainty. *Journal of Water Resources Planning and Management*, 142(9).

Singh, R., Reed, P. M., and Keller, K. (2015). Many-objective robust decision making for managing an ecosystem with a deeply uncertain threshold response. *Ecology and Society*, 20(3).

Smith, R., Kasprzyk, J., and Basdekas, L. (2018). Experimenting with water supply planning objectives using the eldorado utility planning model multireservoir testbed. *Journal of Water Resources Planning and Management*, 144(8).

Sniedovich, M. (2016). From statistical decision theory to robust optimization: A maximin perspective on robust decision-making. In *Robustness Analysis in Decision Aiding, Optimization, and Analytics*, pages 59–87. Springer International Publishing.

Starr, M. K. (1963). *Product design and decision theory.* Prentice-Hall, Englewood Cliffs, NJ.

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

Vavak, F. and Fogarty, T. (1996). Comparison of steady state and generational genetic algorithms for use in nonstationary environments. In *Proceedings of IEEE International Conference on Evolutionary Computation*. IEEE.

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

Walker, W. E., Lempert, R., and Kwakkel, J. H. (2013b). Deep Uncertainty.

Ward, V. L., Singh, R., Reed, P. M., and Keller, K. (2015). Confronting tipping points: Can multi-objective evolutionary algorithms discover pollution control tradeoffs given environmental thresholds? *Environmental Modelling & Software*, 73(1):27–43.

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