Aggregation in bottom-up vulnerability assessments and equity implications: the case of Jordanian households' water supply

Thibaut Lachaut[1], **Jim Yoon**[2], **Christian Klassert**[3], **and Amaury Tilmant**[1]

[1]Département de Génie Civil et de Génie des Eaux, Université Laval, Québec, QC G1V 0A6, Canada;
[2]Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA 99352
[3]Department of Economics, Helmholtz Centre for Environmental Research–UFZ, 04318 Leipzig, Germany

Correspondence: Thibaut Lachaut (thibaut.lachaut@gmail.com)

Abstract.
Bottom-up methods for water resources modeling rely on acceptability thresholds to find, through a response surface, which deeply uncertain futures lead to system failure. They commonly treat water users as aggregate actors, which may preclude analysis of the equity impacts of interventions. This paper explores how aggregation choices for large groups of water users lead to different policy recommendations in response surface assessments. Two aggregation methods with varying parameters are considered: percentile satisfaction targets and generalized mean. A 2-dimensional stress-test assessment across groundwater availability and population is applied to household water supply in Jordan. The study compares six different policies covering supply enhancement and rebalancing, using a country-wide multi-agent model that characterizes households across socioeconomic strata. For different aggregation levels, policies are ordered by their associated robustness index. Results show that aggregation choices strongly determine policy preference. A focus on the most vulnerable households favors the equalization of access to water, in terms of regional allocation and weekly supply durations, as it substantially reduces robustness disparity. Combined policies with additional resources allow to withstand higher levels of stress under most aggregation choices. Preferences defined by aggregation intervals provide a finer understanding of trade-offs among water users and may improve deliberation over equity under deep uncertainty.

1 Introduction

Water resources modeling and management can be hampered by the difficulties to anticipate the future state of a given water system. The economic, demographic or geopolitical upheavals of human societies are drivers of water demand fraught with deep uncertainties (Maier et al., 2016), while climate change challenges the assumption of hydro-climatic stationarity under which water systems could be designed (Milly et al., 2008). At the same time, model-based planning under such uncertainties also needs to consider the fairness of any policy recommendation, based on distributional outcomes of uncertain futures (Hallegatte and Rozenberg, 2017; Jafino et al., 2021).

Complementary approaches exist to assess and plan water systems under uncertainty. The most common relies on building a discrete set of scenarios to explore internally coherent, representative sets of future trajectories for climate, economic growth, land use or demographics (Riahi et al., 2017). Such scenarios are built upon different categories of projections, often informed by models. Such approaches are often called "top-down" (Mastrandrea et al., 2010; Brown and Wilby, 2012), or forward-oriented (Maier et al., 2016).

Another group of approaches, often called “bottom-up”, flip the procedure instead focusing on the robustness of current decisions to deeply uncertain assumptions, reducing reliance on predictive approaches or probabilistic assumptions. Bottom-up approaches have been used across a diverse set of methodologies including inverse climate impact response functions (Füssel et al., 2003; Marcos-Garcia et al., 2020), robust decision making (Lempert et al., 2006; Lempert, 2019), info-gap (Ben-Haim, 2006), or decision-scaling (Brown and Wilby, 2012), often in combination with other methods. Instead of calculating the impacts of projected changes on different performance indicators, inverse approaches generally seek to identify the range of possible changes that can lead to adverse outcomes, and thus usually expand the range of uncertainty in comparison to classic scenarios (Maier et al., 2016). They do not seek to find the impacts of specific conditions, but the conditions that lead to specific impacts on a system's performance. This range of conditions typically supports the construction of a response function or surface (Prudhomme et al., 2010): possible conditions are sampled through a few stressor variables, which define an exposure space. System performance is simulated over this exposure space with a computer-based model. An acceptability threshold then divides the domain of performance values into acceptable and unacceptable sub-sets. This allows one to draw acceptable and unacceptable sub-spaces of the exposure space, which are further used to compare the robustness of different policies and interventions, based on their respective areas.

Importantly, water systems are inherently complex and entail a number of actors with diverse objectives that are often conflicting (Loucks and van Beek, 2017). In particular, assessments using some form of robust or inverse approach have been considering increasingly large numbers of stakeholders or objectives, such as in Poff et al. (2016); Culley et al. (2016); Trindade et al. (2017); Kim et al. (2019); Gold et al. (2019). But each considered objective aggregates the stakes of multiple water users belonging to the same category of water use; e.g. households supplied by the same utility, farmers from the same irrigation scheme. Whereas in reality, users can experience differential impacts based on physical, geographic, and socioeconomic characteristics. Thus response surfaces can be substantially different among water users of a same category. This was illustrated in Hadjimichael et al. (2020) with the disparities of vulnerability profiles among farmers in the Colorado River Basin, showing the need for case-by-case analysis. For systems with a large number of actors, model aggregation risks hiding potential inequalities and undermining the relevance of the vulnerability assessment and public support for selected policies.

Eventually though, if a group of water users is very large, case-by-case assessments become impractical, requiring some form of aggregation to evaluate system-wide performances. For example, intermittent water supply systems can involve large numbers of households with very unequal access to water. In such cases, aggregation remains necessary to quantify the unequal vulnerabilities of different segments of the population. A key issue at the heart of distributional assessments and fairness considerations is the adequacy of the aggregation method (Jafino et al., 2021).

Aggregation does not only shape the description of a problem, but also the preferred policies to solve it. Aggregation of potentially misaligned individual preferences is thus arguably central to political theory, and more explicitly at the heart of social choice theory (Arrow, 1951). A strictly egalitarian worldview such as J. Rawls' *maximin* principle could consider a policy choice as fair if it maximizes the outcome for the worse-off individual among a group (Rawls, 1970). A more utilitarian worldview, as often found with average-based performance indicators, would seek to maximize the sum of individual outcomes, accepting that better and worse outcomes even each other out.

In the present paper, the question of aggregation specifically applies to robustness of water availability, understood as the acceptable share of the exposure space. Our goal is thus to explore how aggregation among the same type of water users affects response surfaces and policy recommendation in an inverse or bottom-up framework. We analyze a range of aggregation choices, translating different attitudes towards inequalities, and how it can affect response surfaces and the policy recommendation of a bottom-up assessment. Furthermore, in a similar manner to the inverse paradigm of the response function itself, we identify the aggregation ranges that lead to preferring one policy over another, to support equity and trade-off analysis under uncertainty within a group of similar water users.

In section 2, the conceptual methodology of the paper presents how to parameterize the aggregation and to identify the aggregation ranges that lead to certain policy preferences. Section 3 presents the studied system – the Jordanian household water supply – using the Jordan Water Model (Yoon et al., 2021), and describes the experimental design to apply an inverse approach using the model. Results are detailed in section 4, followed by a discussion regarding their potential implications and shortcomings.

2 Methodology

We explore multiple approaches to vary the aggregation level of a response function, in order to assess (i) the distribution of acceptable outcomes among a large group of water users and (ii) the effect of such aggregation choices on the policy recommendation. Just as the inverse approach looks for the conditions that lead to certain outcomes, here the question is what levels or types of aggregation lead to certain options being favored over others.

The proposed methodology relies on a simple version of the response surface as a common tool among bottom-up methodologies. In its simplest expression, a response function maps the values of a performance indicator, r, to a discrete number of stressor variables, $(x_1, x_2, \dots x_n)$, which define the exposure space or "states of the world" (fig. 1). The performance indicator r such as average consumption, reliability, resilience or vulnerability (Loucks and van Beek, 2017) is measured over a single time series. An acceptability threshold θ separates performance values between acceptable and unacceptable, and thus allows one to trace a frontier between acceptable and unacceptable shares of the exposure space. The response surface can be calculated for different policies or interventions that modify the system. Comparing the positions of the frontiers associated with different policies allows for the selection of preferred policy options.

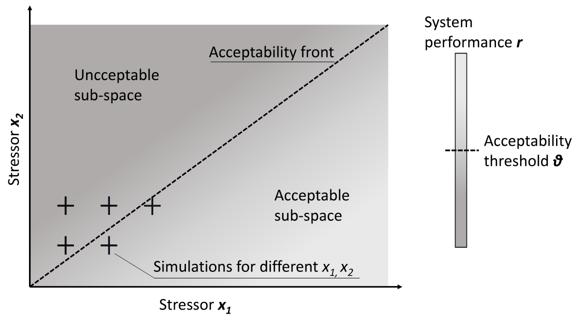


Figure 1. Conceptual response function, mapping system performance r as function of two variables x_1,x_2. An acceptability threshold θ qualifies performance r, and thus subsets of the exposure space, as acceptable or unacceptable. Alternative decisions can be compared based on their respective divisions of the exposure space

Here we consider a response surface specific to a water user, or an "agent". n agents can be grouped into a specific category of water users (e.g., households with unequal access to water supply), who share the same performance indicator, r, and acceptability threshold, θ. For a given agent, i, an individual response surface $r_i(x_1,x_2)$ is obtained by expressing the agent-specific performance of each simulation as a function of the two stressors, x_1,x_2 (e.g. changes in precipitation, temperature, demography, etc). The response surface is transformed into a binary acceptability surface. The acceptability a_i equals 1 if the measured performance r_i for agent i satisfies an acceptability threshold θ, 0 otherwise.

$$a_i(x_1,\ x_2) = \begin{cases} 1 & r_i(x_1,x_2) \geq \theta \\ 0 & r_i(x_1,x_2) < \theta \end{cases} \tag{1}$$

While the acceptability a_i is specific to a given agent, the objective is to produce aggregated response surfaces for the group of n agents, and understand the effect of different aggregation choices on the acceptability surface. Exploring a range of aggregation options allows the representation of different social priorities. For example, if all agent performances are aggregated through an arithmetic mean, extreme values will compensate for each other and the assessment will produce a policy recommendation that would ignore strong inequalities. In contrast, focusing on the 5 percent most vulnerable might lead to a different aggregation choice, that might lead to policy recommendations that are more equalitarian but might not benefit the majority of users.

Comparing a few isolated aggregation options would reveal the potential effect they can have on the assessment outcome. However, here we want to characterize (i) how unequally distributed the acceptability fronts can be depending on the aggregation and (ii) which exact aggregation choices produce different policy recommendations - similarly to how inverse approaches look for the range of conditions that lead to a specific impact. If a continuous parameter controls the aggregation, it is possible to answer to explore points (i) and (ii) by regularly sampling the aggregation parameter. It is thus a way to represent the effect of social preferences with a continuous approximation.

Two parameterized aggregations are thus selected for this study: a percentile-based approach and a generalized mean approach. They can be understood as generalizations (or parameterizations) of the particular cases that are the (arithmetic) mean and the median.

The percentile-based method provides a simple composite indicator to control the aggregation choice. A percentile operator considers a given position within a ranked sample as an adequate level of representation of the population. For example, the objective can be to satisfy a target of 90% of the population according to the threshold θ. In that case, the acceptability space is defined as the share of the exposure space where less than 10% of the population experiences an unacceptable performance r.

Defining $S(x_1,x_2)$ as the percentage of the population whose performance r does not meet the threshold θ, the parameterized acceptability function can then be defined for any percentile level L.

$$A_L(x_1,\ x_2) = \begin{cases} 1 & S(x_1,x_2) < L \\ 0 & S(x_1,x_2) \geq L \end{cases} \tag{2}$$

For example, if the acceptability front should be drawn as to satisfy *at least* 90% of the population, then L=10%, the sub-space of the response surface where more than 10% does not reach the threshold is deemed unacceptable. By sampling L at different levels, the distribution of acceptability ranges for different parts of the population are explored. A composite response can thus be displayed, tracing in the same exposure space the acceptability fronts corresponding to different percentiles of the population (fig. 2). Over the exposure space, this allows for the assessment of (i) the spread between levels, an indication of how unequal the water users can be in terms of vulnerability, (ii) the relative effect of percentile targets and policy choice on the front position, possibly indicating that the policy is relatively ineffective for parts of the population, and (iii) the possibility that preference between policies (the respective position of their fronts) switches for different percentile targets.

This aggregation method also allows for an explicit distributional assessment. The difference with highly disaggregated impact assessments such as in, e;g., Hallegatte and Rozenberg (2017) or Jaeger et al. (2017), is that in an inverse approach, the key metric is not so much the impacts under certain conditions, but the range of conditions before a specific impact is reached. Thus here the distribution measures the spread of robustness rather than the spread of impacts.

The second parameterization method, the generalized mean, is sometimes used in economic development research,

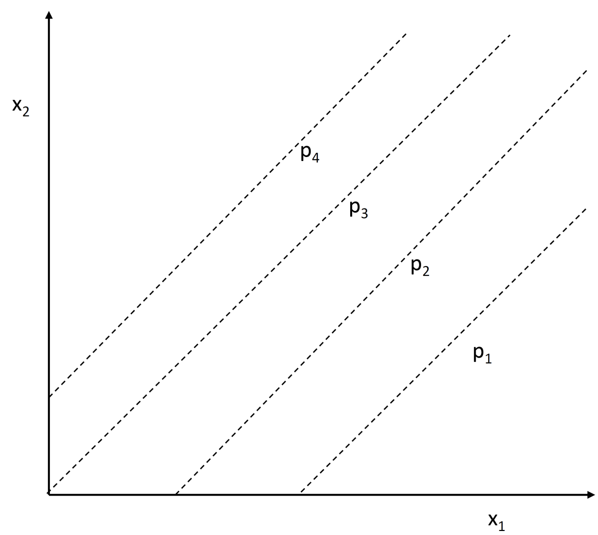


Figure 2. Aggregated response surface: acceptability fronts for different values taken by an aggregation parameter p.

for example to monitor Sustainable Development Goals – SDGs (Rickels et al., 2016) or design human development indices (Kawada et al., 2019). For example, the Human Poverty Index for developing countries computed by UNDP uses the generalized mean with parameter p = 3 (Mariani and Ciommi, 2022).

This method first aggregates the performance values before tracing the acceptability front. For any coordinate (x_1, x_2), and for a number of agents n, the generalized mean M_p with parameter p of the n performance values $r_i(x_1, x_2)$ is defined for positive values of r as:

$$M_p\left(r_1, ...r_n\right) = \left(\frac{1}{n}\sum_{i=1}^{n} r_i^p\right)^{\frac{1}{p}} \tag{3}$$

For $p = 0$, the generalized mean is defined as equal to its limit when p approaches zero:

$$M_0\left(r_1, ...r_n\right) = \sqrt[n]{\prod_{i=1}^{n} r_i} \tag{4}$$

A weighted version allows for further modulation of the generalized mean, either to introduce additional priorities, or in case each single agent represents a larger population.

$$M_p\left(r_1, ...r_n\right) = \left(\sum_{i=1}^{n} w_i r_i^p\right)^{\frac{1}{p}} \tag{5}$$

$$M_0\left(r_1, ...r_n\right) = \prod_{i=1}^{n} r_i^{w_i} \tag{6}$$

The parameter p controls how skewed the aggregation is towards lower or higher performance values r_i. For each value of p, the aggregate acceptability at any given coordinate of the exposure space is given by:

$$A_p(x_1,\ x_2) = \begin{cases} 1 & M_p(x1, x2) \geq \theta \\ 0 & M_p(x1, x2) < \theta \end{cases} \tag{7}$$

The aggregated acceptability function over the response surface is thus the satisfaction of the threshold by the generalized mean, at different values of p. The front between accepted and rejected sub-spaces is drawn for different values of p, translating a different weighting given to performance values at different positions within a ranked sample. A notable drawback is that the generalized mean is not defined for $r = 0$.

A few special cases illustrate the effect of the parameter p. When p tends towards negative infinity, the generalized mean is equal to the minimum value of the sample. When $p = 1$, it becomes the arithmetic mean. When p tends towards positive infinity, it gives the maximum performance value. The generalized mean thus allows to parameterize in an almost continuous manner different aggregation choices between minimum, mean, and maximum. Just like choosing an acceptability percentile target, choosing a value of p when applying a generalized mean operator also translates different collective choice paradigms (Tilmant et al., 2007), that can be linked to social choice theory (Arrow, 1951; Moulin, 1985). Considering only the minimum water use ($p \to +\infty$) of the entire sample when drawing the acceptability surface would correspond to a strictly egalitarian worldview: policies would be designed to improve the least robust water use, following a maximin rule, the Rawlsian definition of justice (Rawls, 1970). Respectively, considering only the maximum consumption would be considered as "dictatorial", as policies are selected to increase the robustness of the single agent with the highest performance. In between, different values of p express different degrees of utilitarianism. The arithmetic mean ($p = 1$) corresponds to a fully utilitarian worldview, seeking to improve the average performance among a population indifferent to the statistical distribution of such performance.

For both approaches, percentile-based and generalized mean, interventions are then compared based on the respective position of acceptability fronts on the aggregated response surfaces. This comparison is done for different levels of the controlling parameter, either the percentile of unacceptability L, or the generalized mean parameter p (fig. 2).

The final goal of this method is to express policy preference as a function of the aggregation parameters. To do so, a single metric should represent the acceptability space. As opposed to the performance indicator r, this metric is not calculated over a time series but should be suited to qualify the whole set of simulations constituting the response sur-

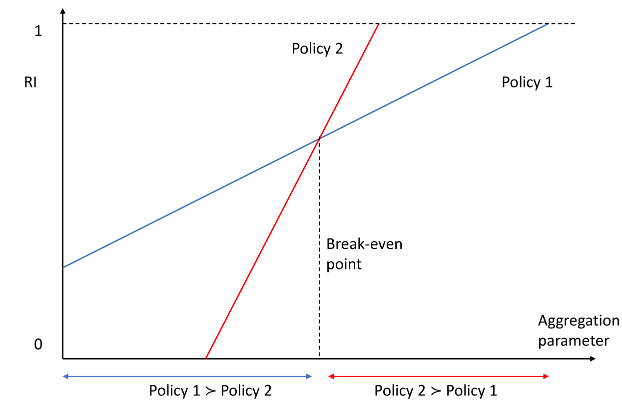


Figure 3. Comparing aggregation-robustness functions for alternative policies allows to identify the aggregation ranges that favor one policy over another ("$Policy1 \succ Policy2$" means policy 1 is preferred to policy 2)

face. For each acceptability surface, a robustness index (RI) is calculated (Moody and Brown, 2013) to represent the distribution of the acceptable surface area across the population. Similar indices exist with additional weights to measure robustness over scenario ensembles or if the stressor domains can be weighted with probabilities. In this case, we consider the range of stressor values as equiprobable and select the simplest variant of the RI. For any tested policy, and for a given aggregation parameter value,

$$RI = \frac{\iint A(x_1,\, x_2)dx_1dx_2}{\iint dx_1dx_2} \tag{8}$$

With acceptable ranges being simplified to a single metric, alternative policies and interventions can be quantitatively compared either through the empirical cumulative distribution function (ECDF) across agent percentiles, or as a function of the parameter p with the generalized mean aggregation. RI thus becomes dependent on a parameterized acceptability A_L or A_P, and can therefore be expressed as a function of L or p.

For any value taken by an aggregation parameter, policies can thus be ordered by preference, by comparing their robustness index. Break-even points can then be identified for aggregation values where robustness is the same for two policies, and thus policy ordering is indifferent. Such aggregation values form the boundaries of aggregation ranges. Each of these aggregation bins is thus defined by a specific ordering of policy preference based on their RI values ordering (fig. 3).

It is thus possible to define the aggregation ranges that would lead to favor a policy over another, when applied to a large number of agents. In the end, the objective is to acknowledge and quantify the winners and losers associated with each policy option, the trade-offs within a group of similar water users in the face of deep uncertainty, and promote informed dialogue among stakeholders.

3 Application

3.1 Case study: the Jordanian water system

As a prime example of a tense water situation and looming uncertainties, the country of Jordan (fig. 4.a) faces a widening gap between dwindling freshwater resources and rapidly increasing demand, with difficult trade-offs among water uses (Whitman, 2019; Yoon et al., 2021). With an overall dry climate ranging from Mediterranean to arid, Jordan relies on limited natural freshwater resources (Gunkel and Lange, 2012). Its almost exclusive source of surface water, the Jordan River Basin, is shared with the neighboring countries, with Israel and Syria using an important part of the upstream flow (Courcier et al.; Avisse, 2018; Avisse et al., 2020). Groundwater resources are heavily overexploited, leading to a rapid decline of water tables that can reach 3.5 meters per year (Goode et al., 2013; Ministry of Water and Irrigation - MWI). Ecosystems are strongly affected, with the disappearing of the Ramsar-classified Azraq oasis (Al-Kharabsheh, 2000; Mustafa and Tillotson, 2019), and the shrinking of the Dead Sea (Salameh et al., 2019). Meanwhile, water demand persistently increases. Agriculture remains a major water consumer despite efforts to curb groundwater abstraction for irrigation (Ministry of Water and Irrigation - MWI). Demographic changes have been sudden, with a population increase of 50% since 2010 in part due to migration from the Syrian civil war, reaching about 11 million today (Central Intelligence Agency (CIA), 2021), even while the population growth rate has declined to 1%. Urban water consumption also includes industries and services, with tourism playing an important role in the country's economy. As a result, Jordan has one of the lowest per capita water availabilities in the world.

Jordan has few options for developing new water resources. Wastewater is reused at 90% for agriculture (Ministry of Water and Irrigation - MWI). All fossil aquifers are now being exploited, including the deep Disi aquifer shared with Saudi Arabia (Müller et al., 2017). Desalination and conveyance from the Red Sea is expensive and depends on uncertain international financing. Increasing Jordan's share of transboundary surface water requires complex negotiations with upstream countries (Haddadin, 2009), though water imports from Israel are substantially increasing under recent agreements.

Important uncertainties are attached to many of the stressors, external or internal, that are relevant for Jordan's water system. Since 1947, demographic growth has not followed a steady and predictable rate but has been punctuated by sudden increases from populations displaced by neighboring conflicts in Israel, Lebanon, Iraq, Yemen, and importantly Syria since 2011 (Courcier et al.; Müller et al., 2016). Rainfall has decreased over the 20th century (Rahman et al., 2015), and climate change is expected to be particularly severe, with droughts becoming twice as frequent, long, and intense by the end of the 21st century (Rajsekhar and Gorelick, 2017). Meanwhile, the state of groundwater resources at any point in the future is hard to predict, as it depends on many factors and decisions made today.

Jordan's water system, characterized by such severe uncertainties, is a prime candidate for analysis using a deep uncertainty paradigm with a stress-test approach. For example, the time needed to reach current population levels was impossible to project before the Syrian war, thus hampering any form of predictive water planning in the view of high-ranking officials (Mustafa and Tillotson, 2019). The Disi aquifer and conveyance project, developed with the objective of satisfying a projected demand, strongly underestimated the demographic changes to come. Water availability at any given time in the future will also depend on climate change, transboundary renegotiations, and the previous trajectory of groundwater depletions. Decisions taken now, be they infrastructure projects or reallocation policies, can be hard to change later given their financial and political cost. Selecting a course of action based on its robustness to highly uncertain factors would thus make sense.

Once propagated through the system, these uncertainties affect a spatially and socially heterogeneous water supply. The case of Jordanian households is an example of a specific category of users (domestic water consumption) that can experience high disparity levels in terms of supply performance. Like many countries in the world, Jordan implements a rationing policy through intermittent water supply over most of the country (Rosenberg et al., 2008; Klassert et al., 2018a). Such intermittency varies strongly between neighborhoods, from less than one day per week in poorer districts to five days in wealthiest neighborhoods (Talozi, 2018), increasing reliance on private vendors.

The system comprised of Jordanian households and their sources of water supply thus provides an adequate case to explore the question of representativeness of bottom-up water vulnerability assessments for large numbers of water users.

3.2 The Jordan Water Model

This work builds on the Jordan Water Model - JWM (Yoon et al., 2021). Watershed and groundwater modules are process-based and spatially explicit. Watershed rainfall-runoff is computed with SWAT, providing inflows for the major reservoirs. A groundwater response function is precomputed with a detailed MODFLOW model at the sub-district level and dynamically reacts to pumping decisions with a drawdown response. The coupled multi-agent, hydro-economic model employs an object-oriented software architecture (Knox et al. 2018). Here we focus on the simulation of dynamic interactions between a hierarchy of diverse actors and the natural/engineered water system primarily involving the piped water supply system (fig. 4.b). Components and features that are particularly relevant for this study are summarized in this section.

Using monthly time-steps, the 1,923 human agents make autonomous decisions based on inputs from natural-engineered modules and other human agents, in a hierarchical manner. Government bodies define high-level constraints and decisions, such as transboundary water availability or groundwater extraction limits. Among them, the Water Authority of Jordan (WAJ) determines monthly allocation and transfers of bulk water volumes between the twelve governorates of Jordan, based on regional per-capita targets and physical/topological constraints from the conveyance network.

From there, local piped supply institutions distribute the available water among sub-districts and among different and competing categories of households and commercial establishments. The quantity of water made available to each sub-district in the JWM is based on the number of agents and the rationing schedule, following Klassert et al. (2015). Weekly supply durations can range between 7.5 hours and uninterrupted supply. Agents buy a certain amount from the public supply based on tariffs and their respective demand function estimates derived from 16,153 observations (Sigel et al., 2017; Klassert et al., 2018b). Urban consumers can supplement piped water with purchases from private vendors who largely obtain groundwater sourced by farmers (Selby et al., 2016).

Each of the 800 household agents represents a certain share of the Jordanian population for specific characteristics, such as location (sub-district), income, refugee status, etc. Households decide purchases of piped water based on econometric demand estimates. Water demand functions notably depend on each household conservation options. These can rely on storage capacities or water saving behaviors depending on the education level of the female household head (Klassert et al., 2018b))".

This multi-agent framework provides an opportunity to test a response surface approach for Jordanian households in a highly disaggregated manner. As economic decisions are dynamic at the agent level, one can interrogate how the frequency distribution function of water use changes under stress, in a coherent, calibrated manner.

The model also allows for the evaluation of results based on other household characteristics that shape the dynamics that are simulated. For example, income, which is particularly relevant for equity-oriented questions, plays a central role in the amount of purchased water, the district of residence, and the rationing pattern of household users. Socio-economic causes of vulnerability underlying the analyses here, such as disparities in income and price elasticities, are described in Klassert et al. (2018b).

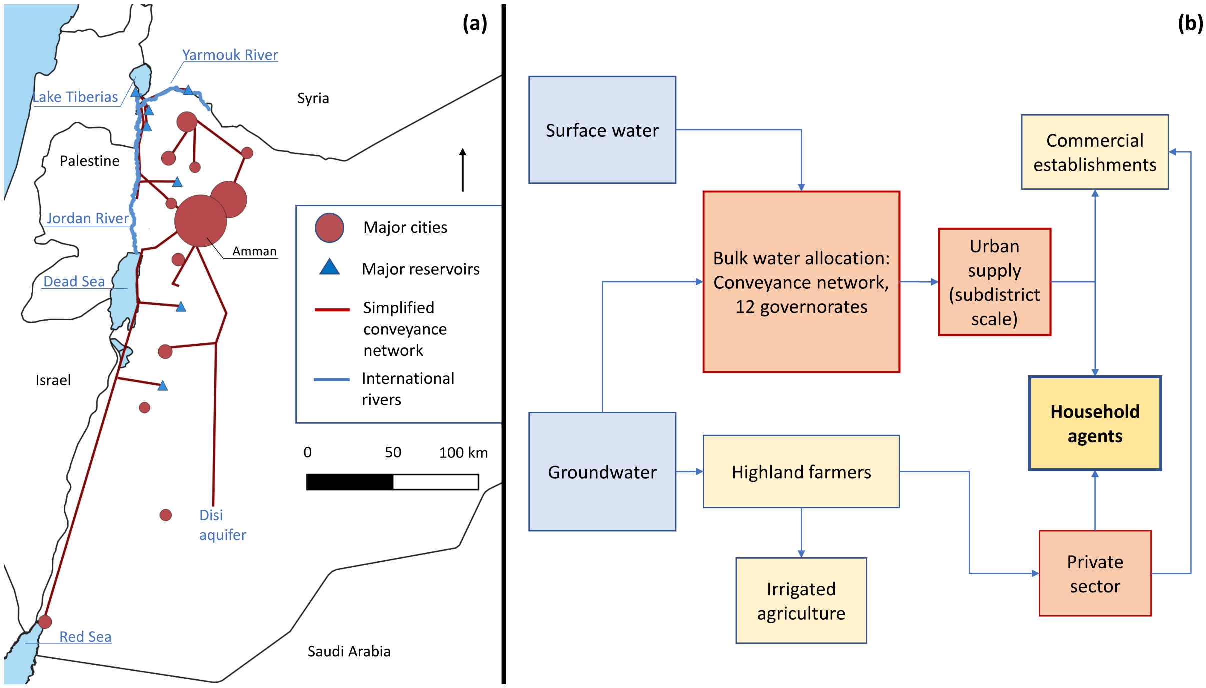


Figure 4. Map of Jordan and simplified model concept (a) The Jordanian population is concentrated in its north-west area, particularly in the capital Amman and its larger agglomeration. Water supply comes mostly from different aquifers throughout the country, the Yarmouk River, shared with Syria, and transfers from Israel through Lake Tiberias. A bulk water conveyance network connects most of the cities. Future desalination projects could convey water from the Red Sea to the northern cities. (b) A simplified view of the water flows through the hierarchy of agents on the JWM. The schematic focuses on the modules most relevant to the present study. Blue: water resources / natural modules. Red: intermediary institutions. Yellow: consumers. Households are the studied system

3.3 Experimental design

Using the multi-agent framework, with a focus on household water use, provides an opportunity to test a response surface approach for Jordanian households in a highly disaggregated manner. As economic decisions are dynamic at the agent level, one can interrogate the distribution of agent water use to discover the impact of changes in system stresses. The model allows for the evaluation of results based on other household characteristics that shape the dynamics that are simulated. For example, income, which is particularly relevant for equity-oriented questions, plays a central role in the amount of purchased water, the district of residence, and the rationing pattern of household users.

To illustrate the general approach, the bottom-up methodology is implemented as a linear change applied to two variables of the system without any associated probabilities (i.e. assuming a uniform probability distribution for all sampled states of the world). In practice, stress-testing is often only a first screening step along more complete decision frameworks with probabilistic approaches, directed exploration, adaptive planning, robust optimization, etc.

This exploratory work is complementary to the scenario approach deployed with the JWM in Yoon et al. (2021), where many more variables were considered in a consistent set of time-dependent narratives. For example, in the present stress-test approach, the time required to reach certain degrees of change on the selected variables is treated as a deep uncertainty. For consistency and comparability, the acceptable consumption threshold and the tested policies are adopted from Yoon et al. (2021) with some modifications that are described further below. The experimental design is further described in the following sub-sections.

3.3.1 Problem delineation

While the JWM simulations involve many more modules and other agents that have dynamic effects on household water use, such as highland farmers deciding to sell water to urban consumers, this study focuses on household agents. The two selected variables are groundwater availability and total population. While it is also subject to deeply uncertain factors like climate change or geopolitical upheavals, surface water is not selected as stressor in this study, as it has a limited impact on household water use specifically. Urban water supply relies in majority on groundwater. The surface water share comes from the Jordan Valley and takes precedence over other uses, and is thus secured to a large degree from climatic or geopolitical perturbations. Groundwater availability can still reflect unknown changes in precipitation and tem-

Table 1. Labels and description of tested policies

Policy label	Policy effects
B Baseline	No intervention besides all projects that were already planned in 2017. They are all set as active in the first year.
R Rebalancing	Supply and demand management - water availability per sub-district is modified as to represent equalized rationing patterns, now only depending on the number of agents. - tariffs are doubled on the higher tier blocks - administrative losses (theft or wrong billing) are halved - Per capita targets for bulk water allocation have a floor set at 50 $m^3/cap/year$ to tone down geographic disparities.
S / S+ Supply enhancement	- New projects are developed at either half (S) or full (S+) capacity (Red Sea desalination, increased transfers from Lake Tiberias ...)/ Available resource before losses is thus increased from 365 million cubic meters (MCM) each year to 500 MCM (+37%) or 624 MCM (+70%). - For both cases, physical losses are halved (from 25% to 12.5%) through pipe replacement and better management of system pressure to prolong the system lifespan.
RS / RS+ Combined policies	Rebalancing policy R combined with new supply S or S+ - Again, new supply at either half (RS) or full capacity (RS+)

perature that would reduce the natural recharge and increase irrigation needs.

These two variables are considered as stressors, in the sense that at any moment in time, water availability in the system is affected by both variables. Stressors represent sets of possible future conditions. The trajectory that led to any given condition, or its associated probabilities, are considered as unknown. Here, groundwater availability can be the result, at an unknown date, of past depletion rates, of political decisions, without having to make a statement about which of these factors lead to a specific level of availability. Similarly, the stress-test assumes that population reaches a certain level at an unknown moment in time, without the need to know if it comes from higher or lower growth rates, or sudden shifts due to war or peace. However, with a different system delineation and approach, those variables might not be considered as independent external stressors as they are heavily path-dependant. Such a difference with time-dependent simulations will be further addressed in the discussions section.

For groundwater, a single capacity reduction factor is applied to all groundwater nodes, reducing in such proportion the maximum allowed monthly extraction. The model still dynamically determines abstractions within this limit and the drawdown response. Similarly, for the population variable, the same increase factor is applied to all representative households, regardless of location, income, etc. (in practice, demographic changes have been, and will be, much more heterogeneously applied).

3.3.2 Simulations, policies, and post-processing

For each tested intervention or policy, 72 simulations of the Jordan Water Model are performed. They combine nine levels of groundwater extraction decline (from 0 to 40%) and eight levels of population growth (from 0 to 175%). Such changes are consistent with those considered in the previous work with the JWM for the 2100 horizon. Simulations are performed over two years and results are recorded for the second year only. This allows agents to adapt to the circumstances as applied to the first year (typically expected market prices). The baseline year is 2016, the last one for which the supporting data were available when developing the JWM. The specific hydrological intra-year variability has little impact in the present study, though it would have to be considered if agriculture were included.

Response surfaces seek to compare options based on the respective position of their acceptability fronts. Six different interventions or policies are stress-tested, consistent with those that were simulated in Yoon et al. (2021). As presented in table 1, the tested policies focus either on supply improvement (adding new resources to the system, in two stages), supply and demand management (reshaping the distribution without increasing the total available resources), or a combination of both.

For each simulation, we record monthly water use for each household agent. To build response surfaces, the common performance indicator is the average water consumption over a year, in liters per capita per day [L/c/d], calculated over the

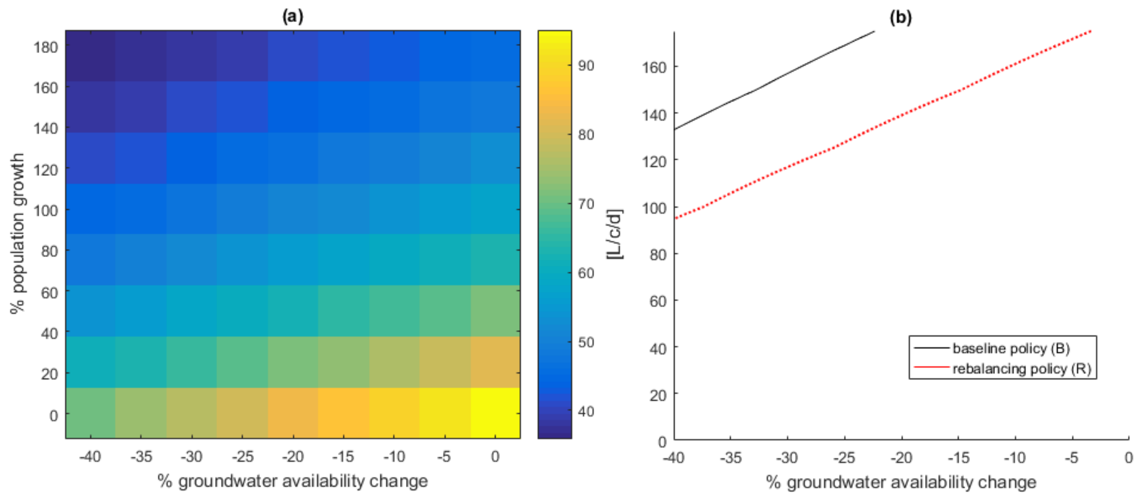


Figure 5. (a) average water use response surface, L/cap/d, for all 800 household agents, baseline policy B. Results from the initial 72 simulations (9 groundwater reduction levels x 8 population growth levels). (b) average after linear interpolation (113x129) is performed. Policies are compared based on the average acceptability front. Only the baseline policy front (solid black line) and the rebalancing policy front (dotted red line) appear on the response surface, the other policies show an acceptable average performance over the entire exposure space and, thus, their acceptability does not show here

second year. The acceptability threshold is set at 40 L/c/d following Yoon et al. (2021).

For a given household i, the response surface $r_i(x_1, x_2)$ is obtained by expressing the performance of each simulation (the average water consumption per capita par day) as a function of the two stressors x_1, x_2 (here groundwater and population changes). Composite response surfaces are then calculated by sampling aggregation parameters (satisfaction percentiles or generalized mean parameter). Finally, a robustness index RI is expressed as a function of the aggregation parameters, revealing the aggregation ranges that correspond to specific preference orderings of the 6 tested policies. A sensitivity analysis showing how the value of the acceptability threshold θ (testing 30, 50 and 60 L/c/d) affects policy preference is provided in supplementary information (tables S.I.2 and S.I.3). The S.I. also includes additional results assessing the spread of the robustness index for different income deciles and governorates, and how the Gini coefficient of water use changes over the response surface.

4 Results

Across the 72 simulations, sampling 9 levels of groundwater availability decline and 8 levels of population growth, average water use declines are as expected along with average water per capita (fig. 5a). The average consumption only gets below the acceptability threshold of 40 L/cap/d in the most extreme combinations of groundwater reduction and population growth. To trace the frontier between acceptable and unacceptable subspaces, linear interpolation is performed for each of the 800 individual response surfaces. The average is then recalculated, and the exposure space is divided between acceptable and unacceptable average use (fig. 5b). The acceptability gradient mostly follows a constant anisotropy in all tested sub-spaces, thus in all figures hereafter the acceptable sub-space is southeast of the front and the unacceptable sub-space is northwest of the front. The effect of different policies and interventions can be compared based on the position of their respective fronts, though this policy comparison reveals the limitations of evaluating acceptability with an aggregate measure of average water use. Fig. 5, for example, shows that the baseline policy (B) seems to be preferable to the rebalancing policy (R), since households have a larger average water use under policy B for any given combination of stressors, and the sub-space that would be evaluated as acceptable under policy B is correspondingly larger. The reason for this, however, is not a better supply situation under policy B. Rather, policy B reflects the highly unequal distribution of piped water supply durations that currently prevails in Jordan.

Under the unequal distribution resulting from policy B, some households are unable to meet their essential water demands with piped water and have to purchase expensive supplementary water from private vendors, while others receive much more water. When policy R distributes about the same overall quantity of piped water more equitably, more households are satisfied with the amount of piped water they receive and fewer have additional demand for expensive supplementary water purchases. This leads to a lower total water use quantity. As a result, the higher aggregate measure of average water use under policy B seems to indicate that policy B is strictly preferable, while in most cases, policy R is actually better at meeting households' demands, as the subsequent analyses show.

4.1 Percentile-based approaches

To further explore the disparity of policy preference among Jordanian households through the 800 representative agents, different aggregation levels are sampled for the two methods presented in section 2: percentile-based and generalized mean.

We first proceed by percentile slicing. The percentage of households with insufficient water use is calculated over the exposure space, i.e., every combination of population and groundwater change. Acceptability fronts are defined by drawing contour plots for specific percentiles of unsatisfied users. Percentiles are weighted by the number of households that each agent represents. Thus, the "5%" line delineates the border of the region in which 95% of the population is satisfied, the 10% line is the limit where 90% of households are satisfied, etc. In Fig. 6, the alternative policies and interventions are compared based on their acceptability frontiers, for different percentiles. With the baseline intervention (fig. 6a), the acceptability fronts for different percentiles are widely spaced. The 50% front is the median acceptability front, where half the households have an acceptable water consumption. The 5% front is not visible; thus, the corresponding share of the population is already in an unacceptable state under initial (2017) conditions.

Gradient slopes slightly change across percentiles, indicating that the more vulnerable percentage of households is also more vulnerable to demographic growth, while more robust percentiles are more sensitive to groundwater availability decline, as they rely more on private water sales and thus private wells. Another notable feature is that the spread between percentiles under baseline policy can be much wider than the difference between policies B and R based on the average water use, while such a difference would have been used to select one option above another in fig. 5. The local gradient change for the 25$^{\text{th}}$ percentile front (southwest corner) is due to the response surface of one particular agent, which locally becomes the 25$^{\text{th}}$ percentile and thus modifies the aggregated front. The location of that agent in Aqaba governorate means it is more sensitive to groundwater changes than most others and thus shows a different performance gradient

The following sub-plots compare this baseline policy response with alternative policies and interventions. Figure 6b shows the effect of the supply and demand rebalancing policy (detailed in table 1) on the distance between acceptability fronts for different percentiles. It is much more compact than under baseline policy, with obvious winners and losers. By providing more water to households with lower water use, the equalization of supply durations within rationing schedules, combined with the raise in minimum regional targets, massively expands the acceptable space for the 40% most vulnerable households, while it decreases for the median or above. This also shows that, for moderate levels of stress, changes in allocation rules are extremely effective at protecting the most vulnerable households, while for higher levels of stress, most households fall under unacceptable consumption if no additional resources are added. The change in rationing schedule also means supplementary purchases from private vendors are decreased, as described for fig. 5.b. The respective roles played by the rationing schedule and the bulk water allocation targets are further separated and discussed in S.I., additional results. This difference between percentiles is also further analyzed in subsection 4.3. Additional resource interventions, at either half (6c) or full (6e) capacity, shift the distribution away from the axis origin ("current" conditions), effectively increasing the acceptable space for all percentiles while slightly increasing the spread between them. Combined policies (6d) drastically increase the acceptable space for all household categories, as well as reducing the spread between them. In the case of the combined policy at full capacity (6f), no household reaches unacceptable water use in the sampled exposure space. This is also a case where the rebalancing policy, through supply duration equalization and increase in minimum bulk water regional supply, considerably improves the robustness equity, this time compared to the supply enhancement.

For most household percentiles, increasing supply with new projects at half capacity provides a larger acceptable space than the rebalancing policy. Both interventions provide about the same acceptable space for the 10-percentile. For the 5% most vulnerable share of households, the rebalancing policy increases the acceptable space further than the new supply policy at half capacity.

Combining policies has a massive effect in expanding the acceptable space for the most vulnerable percentiles, while remaining positive for most percentiles. A combined policy with new supply at half capacity (6d) provides more acceptable space than a full supply expansion policy without rebalancing (6e) for at least 25% of the population.

A more explicit distribution function of the varied responses in the household population can be obtained if the acceptable sub-space area is computed for each agent. This abstraction can be a loss of information, as a single area value can hide varied shapes of acceptability fronts, but in the present case the gradients of the acceptability range remain quite similar. To quantify the acceptability ranges we use the simplest version of the Robustness Index (RI, Moody and Brown, 2013), which is the ratio of the acceptable sub-space over the entire exposure sub-space. The robustness index RI (section 2, eq. 8) is calculated for each policy and for different values of the aggregation parameters. Figure 7 shows the quantile function of the RI distribution (weighted by the number of real households that each agent represents). Each distribution corresponds to a given policy or intervention. It allows us to quantify the difference between interventions in terms of acceptability space. Break-even points can be identified when comparing interventions to see the percentage of

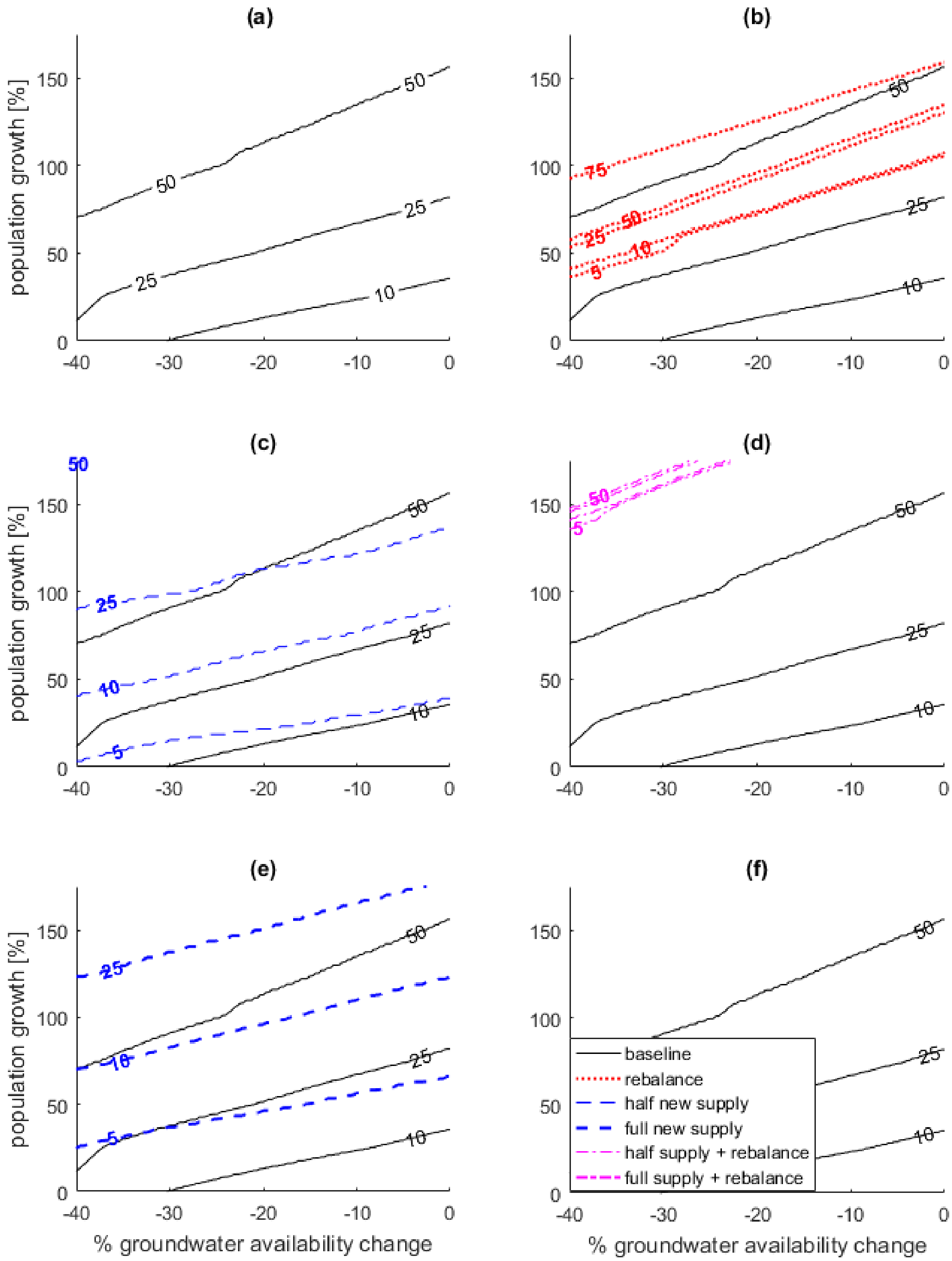


Figure 6. percentile-based acceptability fronts. Ex: a front with a value of 10 delineates the acceptability space for the 10% most vulnerable share in terms of water use. (a) baseline - no intervention. (b) rebalancing supply/demand (c) new supply at half capacity (d) half capacity + rebalancing (e) new supply at full capacity (f) new supply + rebalancing (fully acceptable response).

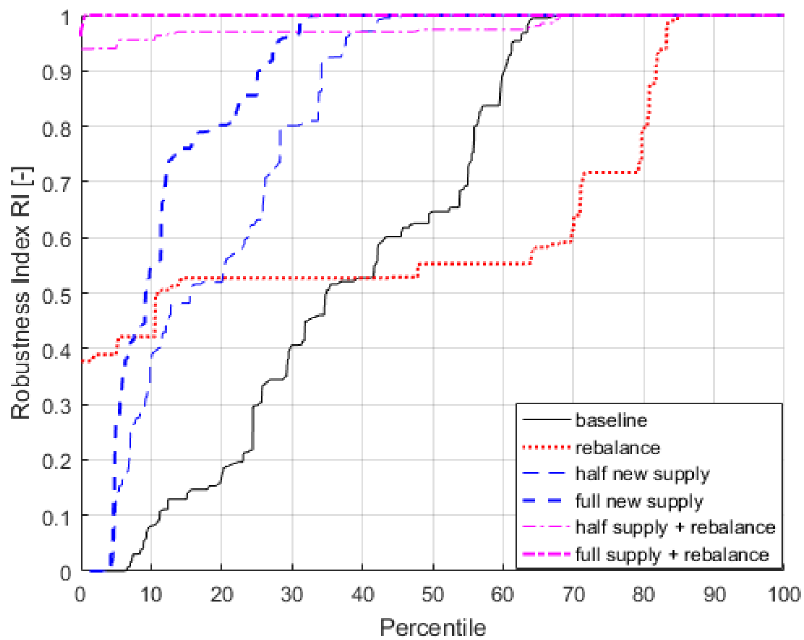


Figure 7. Quantile functions of the individual Robustness Index (i.e. share of the exposure space or sampled simulations that yield an acceptable outcome) for the different interventions.

the population that benefits or is penalized by switching policies. For example, the rebalancing policy strongly increases the RI for the 40% of the population with lower water use and decreases it for the remaining 60%. Figure 6 shows that a combined policy (rebalancing + new supply at half capacity) was more beneficial than new supply at full capacity for the low consumption households. In fig. 7, we see the break-even point is at 30% (thin dash-dotted magenta line vs thick dashed blue line). Break-even points can be considered as the boundaries of preference groups: shares of the population defined by how they prioritize policies based on the robustness index metric (table 2). Percentile intervals corresponding to specific preferences can also be found in pair-wise tables in the Supplementary Information appendix, additional results (table SI2), among a sensitivity analysis changing the value of the threshold θ.

For certain metrics, fig. 7 also shows the compounding effect of combining policies, as noted in Yoon et al. (2021). With the baseline scenario, 60% of the population have a RI below 0.9. The rebalancing policy (red line) is detrimental in that regard, increasing the share to 80%. The supply enhancement at half capacity (thin blue dashed line) lowers the share to 35%. Combining both (thin magenta dash-dotted line) leads to 0% of the population below 0.9, thus having far more than additive effects compared to the baseline and outperforming the full supply enhancement policy.

Finally, we further disaggregate results based on other household characteristics that shape the dynamics that are

Table 2. preference ordering for specific ranges of the household population

Percentile range	Policy preference ordering
$0\% < L < 8\%$	$RS+ \succ RS \succ R \succ others$
$8\% < L < 20\%$	$RS+ \succ RS \succ S+ \succ R \succ S \succ B$
$20\% < L < 31\%$	$RS+ \succ RS \succ S+ \succ S \succ R \succ B$
$31\% < L < 34\%$	$RS+ \succ S+ \succ RS \succ S \succ R \succ B$
$34 < L < 40\%$	$others \succ RS \succ S \succ R \succ B$
$40\% < L < 48\%$	$others \succ S \succ RS \succ B \succ R$
$48\% < L < 63\%$	$others \succ RS \succ B \succ R$
$63\% < L < 68\%$	$others \succ B \succ RS \succ R$
$68\% < L < 84\%$	$others \succ R$
$84\% < L < 100\%$	$indifferent$

With: B: baseline; R: rebalancing; S: new supply (half); S+: new supply (full); RS: combined (half); RS+ combined (full). L: robustness-ordered percentile level Full/combined policy excluded.

simulated, in particular household income. In fig. 8, the same percentile-based fronts are used for 3 interventions, but only for households at the top and bottom 10% of incomes. Under the baseline policy (8a) there is a large spread between households within each income category, thus strongly overlapping with the other income group over the exposure space despite a notable difference. For example, the 75th percentile of poor households (dotted red line, "75") is roughly as robust as the 50th percentile of rich housheolds (solid black line, "50"). In a similar way to the overall population figures, the rebalancing policy erases most of the differences (8b), while

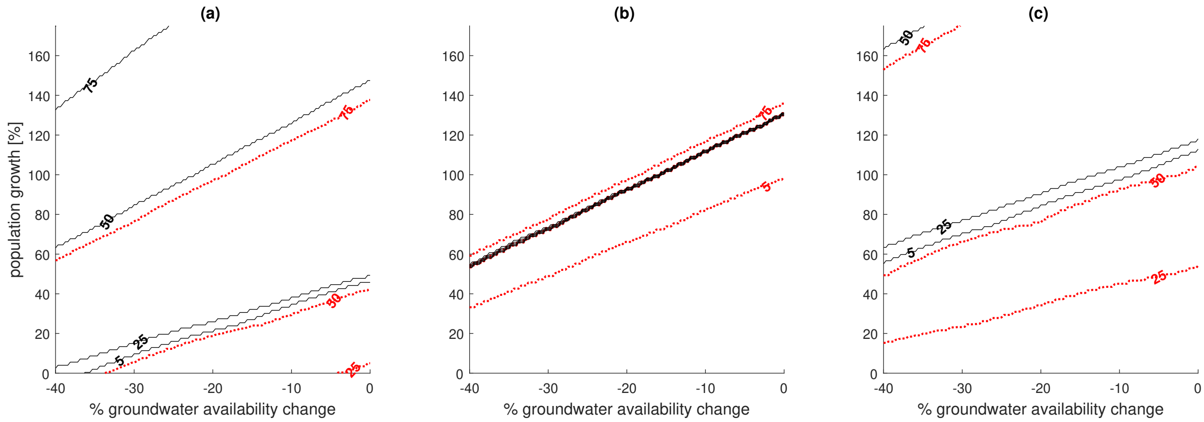


Figure 8. further disaggregation for top (solid black) and bottom (dotted red) income deciles. (a) baseline - no intervention. (b) rebalancing supply/demand (c) new supply at half capacity

a new supply policy (half capacity, 8c) shifts the fronts towards higher stress levels without changing the spread. Further results in supplementary information show the spread of the robustness index for different income deciles and for the different governorates.

4.2 Generalized mean

The alternative way to explore the disparity of the response surfaces is to first aggregate individual water use, in a similar way to the mean operator but controlled by a parameter that can take different values for skewness. Figure 9 shows the use of the generalized mean. Interventions are compared for an array of values of p that control the skewness of the generalized mean towards lower or higher values. As covered in section 2, the parameter p in eq.5 and eq. 6 can be seen as a continuous cursor between the household with the smallest water use ($p \to -\infty$), the arithmetic mean of water use ($p = 1$), and the household with highest water use ($p \to +\infty$), all particular cases of the generalized mean. The acceptability threshold is then simply applied afterwards to divide the exposure space for each value of p. Results show similar dynamics compared to a percentile-based approach. The parameter p plays a similar role to the percentile L, inferior values of p mean more weight is given to the households with lowest use, while high values of p give more importance to the households with higher water use. Rebalancing reduces the spread between levels of aggregation while new supply shifts the fronts and tends to slightly increase the spread between levels. Fig 9b shows the effect that a single household agent can have when getting closer to a MIN operator, as one outlier has different front slopes than the others.

The robustness index RI can similarly be computed for different values of p (fig. 10). Policies can be compared based on their RI, representing different social choices. A more egalitarian approach (lowest values of p) favors rebalancing over new supply, while a more utilitarian approach (p around 1) prefers even the baseline policy over the rebalancing one (same result as fig. 5). High values of p give more importance to households with high water use, which are more indifferent towards policy choice. This figure shows a discrepancy with the percentile-based method, as this time combined policies are preferred in any case. Again, break-even points can be used to identify decision-specific intervals of p, each defined by a given policy ordering (table 3). This classification is more abstract than percentages of the population, it rather integrates all household consumptions like the arithmetic mean, but with varying degrees of skewedness towards those with lowest use or those with highest use. Aggregation intervals corresponding to specific preferences can also be found in pair-wise tables in the Supplementary Information appendix, additional results (table SI3), among a sensitivity analysis changing the value of the threshold θ.

Table 3. Preference ordering for p intervals (generalized mean)

Parameter range	Policy preference ordering
$-\infty < p < -17$	$RS+ \succ RS \succ R \succ others$
$-17 < p < -5.3$	$RS+ \succ RS \succ R \succ S+ \succ S \succ B$
$-5.3 < p < -3.6$	$RS+ \succ RS \succ S+ \succ R \succ S \succ B$
$-3.6 < p < -1.8$	$RS+ \succ RS \succ S+ \succ S \succ R \succ B$
$-1.8 < p < -0.4$	$others \succ S \succ R \succ B$
$-0.4 < p < -0.2$	$others \succ R \succ B$
$-0.2 < p < 1.6$	$others \succ B \succ R$
$1.6 < p < 2.6$	$others \succ R$
$2.6 < p < +\infty$	$indifferent$

With: B: baseline; R: rebalancing; S: new supply (half); S+: new supply (full); RS: combined (half); RS+ combined (full). L: robustness-ordered percentile level Full/combined policy excluded.

4.3 Distribution of water use

To better understand the structure of the response surfaces for different segments of the population and the role played by the model dynamics, we look at the consumption distribution

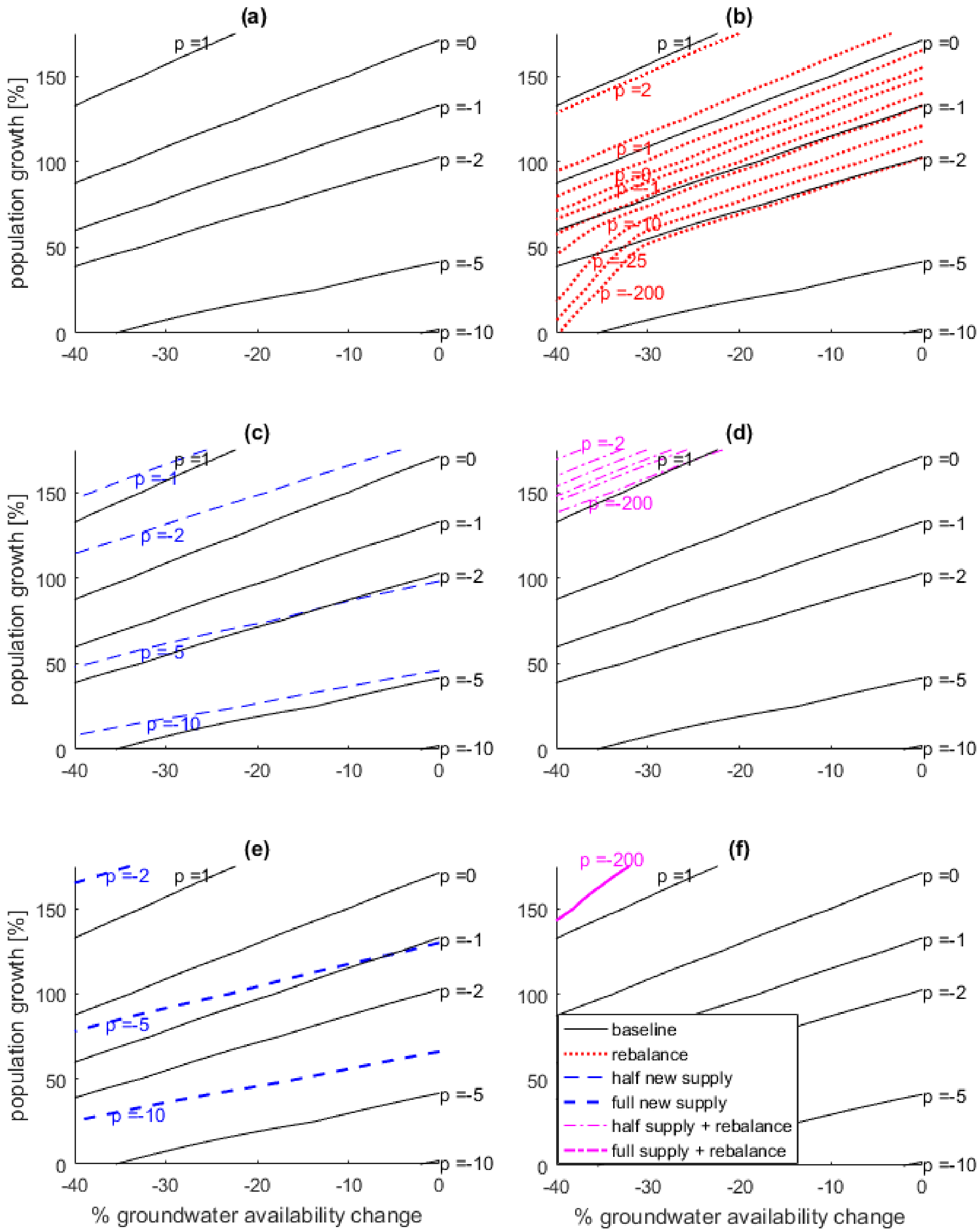


Figure 9. aggregation of response surfaces by generalized mean. The parameter p controls the importance given to higher or lower values. A very high p is equivalent to MAX, a very low p equivalent to MIN. p=1 equals the arithmetic mean. (a) baseline - no intervention. (b) rebalancing supply/demand (c) new supply at half capacity (d) half capacity + rebalancing (e) new supply at full capacity (f) new supply + rebalancing

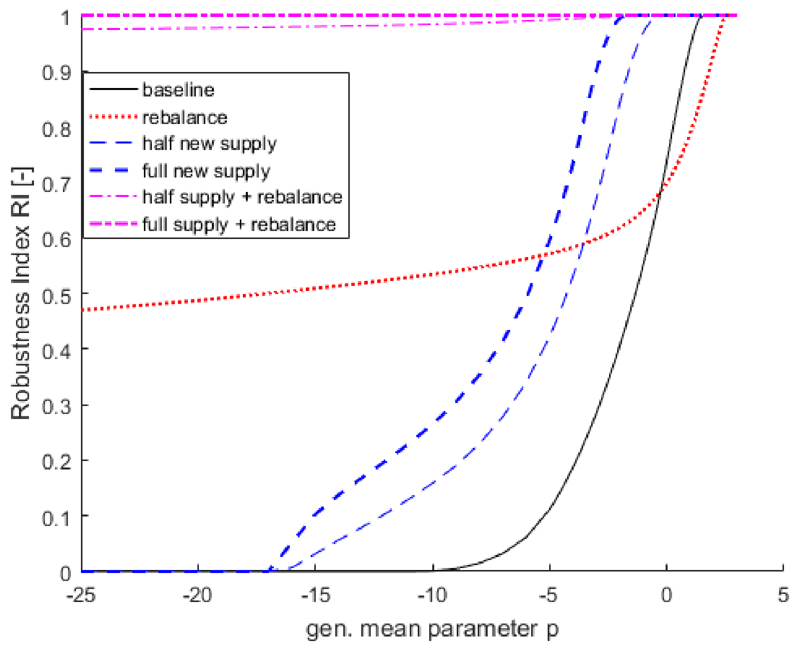


Figure 10. robustness index RI vs generalized mean parameter p

functions sampled within the exposure space at different levels of stress: groundwater availability decline of -15%, population growth of +75% (fig. 11a), respectively -30% / +150% (fig 11b).

One apparent dynamics is that the benefits of supply enhancement policies (blue, dashed lines) are shared unevenly across the population compared to the baseline policy. This explains the low preference ranking of these policies on the lower end of the aggregation spectrum, for both aggregation methods. As new supply follows existing rationing patterns, it tends to increase availability within neighborhoods with already good supply duration. Such policies thus benefit more the upper half of the population in terms of water consumption and is relatively inefficient at increasing the acceptability sub-space for the population with the lowest use.

Importantly, fig. 11 also exemplifies one of the reasons for the compounding effect of combining policies (redistributing and increasing the resource pool are more effective combined than alone). This is closely linked to the evaluation of policy performance based on the share of a population that is above a given threshold. The more vertical is the distribution, the more egalitarian is the water use (dotted red and dash-dotted magenta lines). This mechanically makes those policies much more susceptible to resource fluctuations as it leads to very large shares of the population suddenly crossing a threshold, in one way or the other, thus producing strong non-linearities with such metrics. Most households currently below the threshold generally benefit from more equal supply durations, while those above the threshold see their consumption reduced. Without additional supply there is a point where many households fall below the threshold, even though having the policy is still beneficial to the most vulnerable households. Combined policies benefit at the same time from the somewhat homogeneous shift upwards that added supply brings, and from the more egalitarian distribution that lifts low use households much more effectively.

5 Discussion

5.1 Significance

Even within the same category of water users, aggregation choices can lead to different preferences when comparing possible policies in a water system. This can be particularly relevant for bottom-up methods in water vulnerability assessments, as those commonly rely on limited numbers of acceptability thresholds in order to establish policy preferences under uncertainty.

This study shows how different aggregated response functions can be obtained using a multi-agent hydro-economic model for large groups of water users, by continuously shifting an aggregation parameter either through percentile ordering or generalized mean. Results not only illustrate that water users within the same category can have differing preferences among a set of possible policies and interventions, but also reveal how aggregation choices, and thus socio-political attitudes towards equity, lead to the selection of one

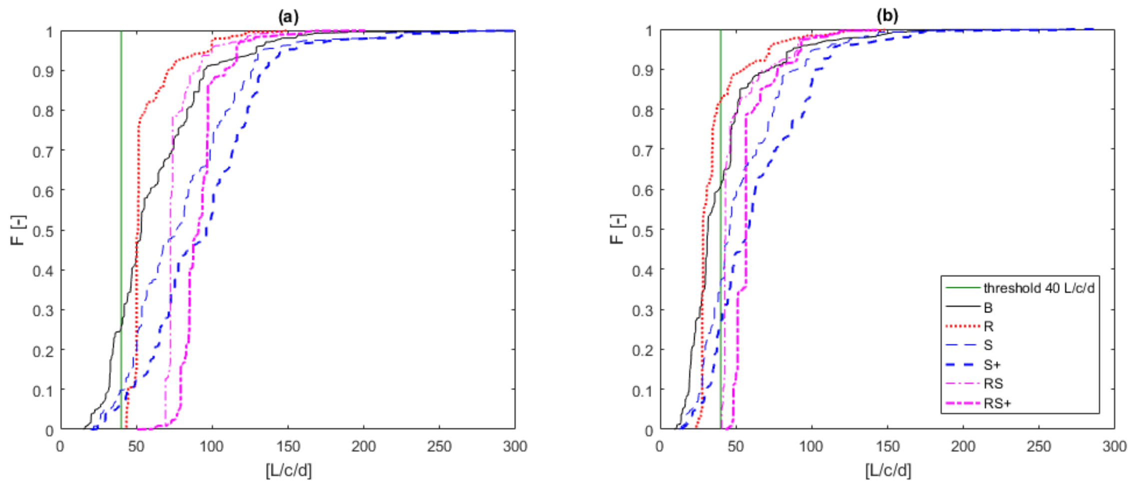


Figure 11. empirical CDFs of water use for different levels of stress (a) -15% groundwater availability, +75% population growth (b) -30% groundwater availability, +150% population growth.

course of action over another. One advantage the generalized mean has over the percentile-based approach is that the generalized mean is affected by the actual value of water use, not only its position compared to the threshold. With the percentile-based approach, being slightly below the threshold or having no water consumption at all makes no difference. However, the generalized mean is also a more abstract method, while the percentile-based approach provides an explicit distribution of robustness. Both approaches do produce similar results, in terms of how relative preferences change along with aggregation parameters. This illustrates how, in this given case study, the aggregation parameter can matter more than the aggregation method. It can also facilitate the interpretability of the generalized mean.

The Jordan Water Model offers a high level of complexity and detail that allows for the exploration of such robustness distributions across a large number of representative agents, and to design policies that modify the demand and the supply distribution apart from the development of new water resources. Statistical distributions are also affected by the internal dynamics of the model, such as the rationing structure or the private water sales, allowing them to evolve under stress and further justifying the use of a distributed stress-test.

The disaggregation of the JWM enables analysis on the compounding benefits from combined policies, compared to their standalone performance, in this case within a bottom-up framing. The present stress-test, with its inherent emphasis on a satisficing metric (meeting an availability threshold), underlines the advantages and drawbacks of changing the slope of the water use distribution through policy. It thus partially explains how it can lead to non-linear benefits if combined with supply enhancement that shifts the distribution away from the acceptability threshold. Developing new water resources or reallocating existing ones often represent conflicting narratives, with different national or international institutions favoring one or the other (Hussein, 2018). In addition to their standalone or combined benefits, the present work further asks "for whom?", exploring how aggregation determines preference.

With the tested aggregation functions, an aggregation parameter can modify the preference ordering for the set of different considered policies. Once such a divergence is acknowledged, society is faced with a trade-off between increasing the acceptable space for the most vulnerable households or increasing that for average or median households. This exploratory work can inform such discussions, and more generally the concept of equity in the face of uncertain change, by quantifying such trade-offs within groups of similar water users. By applying different aggregation metrics, managers can identify more effective solutions to reduce vulnerabilities more equally, notably by combining new investments with changes in allocation rules. In turn, informing vulnerabilities for different socioeconomic strata may also facilitate a broader negotiation process to identify acceptable policies in face of deeply uncertain stress.

When producing a response surface at the scale of a country, exploring aggregation ranges also allows one to circumvent loaded narratives about tracing a country's “safe space” with a single threshold, that could be seen as an excessively Malthusian perspective at best - particularly when water already feeds into internal tensions over migration and Jordanian identity (Mustafa and Tillotson, 2019) - while still considering the tangible benefits of increasing available resources for specific levels of demographic growth. Disaggregating the notion of “acceptable space“, of the “time left“ for Jordan as a whole before reaching some levels deemed as catastrophic, shows that such levels will be reached at very different times by different parts of the population, and that this is strongly influenced by allocation policies. This can help design the most effective solutions to ensure equitable robustness under uncertainty.

5.2 Caveats and future research

The readability of results is here favored by several circumstances and assumptions. Slopes and gradients can slightly change but the surfaces are still roughly oriented in the same direction for all household agents, while the performance indicator itself remains the same for all. In other cases, the diversity of indicators, of relevant stressors, of response shapes and gradients, as in Hadjimichael et al. (2020), can make aggregation much more challenging. While this study remains a proof of concept, the selected indicators, the problem framing and their underlying values (Jafino et al., 2021) should receive further scrutiny within an actual policy recommendation paper. In particular, the selection of the satisfaction threshold can shape the results considerably, all the more so for the percentile aggregation method.

For the scope of this paper, we have only considered a few levels of complexity that the JWM can handle, leaving many others for future work. For example, it is highly unlikely that population growth would happen homogeneously over the country or social strata, as was assumed in this simple stress test. Among the notable factors that were not considered but would affect the results, income is considered as constant, thus effectively assuming a stable, null economic growth outside of population changes. GDP per capita in Jordan could increase 8 times by 2100 according to the SSP2 scenario from the Shared Socio-economic Pathways (Riahi et al., 2017). Different growth or crisis trajectories would have an impact on many levels of the models such as the ability of households to purchase water from the private sector. Besides, integrating cost-benefit analyses and socio-political assessments of the tested policies within the present approach should be an avenue for future research. The implementation of a cost-benefit analysis would require estimated costs on the various intervention strategies (including both supply infrastructure and demand management), while benefits from the interventions can be estimated via modeling results. The benefits analysis could further be enhanced to account for the distributive effects among the population utilizing the aggregation approaches introduced in this work.

Another economic aspect that should be incorporated in a complete vulnerability analysis is the relationship between income and robustness of water use, which in this modelization are only partially correlated over the entire sample (geographic disparities being an important factor). For example, here the 10% most vulnerable share of the population in terms of water use does not correspond entirely to the bottom 10% of incomes. Households with acceptable water use might face other difficulties due to their low income, households with average income might have other ways to mitigate a low water availability. A related research continuation would be to assess the effects of household's conservation options (technical and behavioral) on their water vulnerability using the present framework.

Fluctuations in surface water were not considered either as they have limited impact on direct drinking water supply. However, there would certainly be an influence through changes in the agriculture sector and the effect on mobile providers. And while this study focuses on the household sub-system, a complete, multi-sectoral assessment should include agriculture, with climate change as an additional stressor, and rural-to-urban transfers as an additional policy.

These limitations highlight the trade-offs and complementarity between a narrative, scenario-focused forward approach as originally used with the JWM, and the present inverse stress-test approach. The bottom-up method can identify the exact levels of stress from a few variables that would lead to unacceptable performance, independently of time or without needing a mechanistic explanation to reach such levels. However, one of the challenges in applying a bottom-up sensitivity framework to a large group of water users is that it requires a binary outcome (acceptable/unacceptable water use) excluding information on the magnitude of the deficit. Besides, if more variables were to be considered (even after a preliminary selection of the most impactful ones), a bottom-up assessment would quickly run into a curse of dimensionality. Not only in terms of computational resources, where each added dimension increases the number of simulations by an order of magnitude, but also in terms of visualization for policy makers. Besides, the non-temporal stress-test also precludes analysis of path dependent dynamics, which are particularly important in the Jordanian case (e.g. groundwater depletion). It is important to note that there is a likely degree of dependence between the stressor variables. More population at any given time may prevent curbing groundwater abstractions and lead to a reduced availability later. Or a collapse in water availability could have dire economic impacts and lead to emigration. For the present experiment under a deep uncertainty assumption, we choose to apply a veil of ignorance on the relative likelihood of stressor combinations, but further weighting could be applied to the response surfaces based on trustworthiness of future scenarios. While the conceptual simplicity of the stress-test is convenient for use with a complex model, it should be be viewed as complementary to other decision frameworks such as adaptive planning (Haasnoot et al., 2019). Future avenues of research would also involve using this framework with the option to screen more intermediate degrees of intervention, in order explore trade-offs more methodically and strategically design new policy portfolios that target specific robustness and equity outcomes.

5.3 Conclusions

This study explores the effect of aggregation choices on water vulnerability assessments that rely on response surfaces, when applied to a large number of water users. To do so, it relies on a dynamic, multi-agent model of the Jordanian water system, and tests combinations of supply enhancement and distributional policies under groundwater decline and population growth. Response functions are aggregated with percentile targets or generalized mean. By relating the acceptable share of the exposure space to an aggregation parameter, this work illustrates how the safe range provided by different supply enhancement and rebalancing polices depends on aggregation assumptions, but also allows one to identify specific ranges of aggregation - and thus social choices - that lead to each different policy preference ordering. The proposed methodology can be used to quantify the benefits of more equitable policy design under a deep uncertainty framework. In the case of Jordan, different policy portfolios have different equity implications, and changes in allocation and rationing patterns can be particularly effective to equitably reduce water vulnerabilities. This exploratory work provides a proof of concept for more theoretical frameworks to define distributed freshwater security, and thus formulate equity and trade-offs within a given type of water user in the face of deeply uncertain changes.

Acknowledgements. We thank staff in the Jordanian Ministry of Water and Irrigation (MWI), Water Authority of Jordan (WAJ), Jordan Valley Authority, Ministry of Agriculture, and Department of Statistics for provision of data and reports for the analysis. We are particularly grateful for support provided by Dr. Hazim El-Naser, Ali Subah and Susan Kilani at MWI, and Refaat Bani Khalafat at WAJ. We also thank Steve Gorelick for bringing comments and advice to this paper; Samer Talozi, Yazzan Haddad, Marwan Shamekh and Nicolas Avisse for the help extended in Jordan, as well as all the contributors to the Jordan Water Model. We also thank Mohammad Bataineh and Amin Badr El-Din for their support over the course of the project. Additional data and information were provided by the United States Geological Survey and the United States Agency for International Development. The Economic Research Forum and the Jordanian Department of Statistics granted the researchers access to relevant data, after subjecting data to processing aiming to preserve the confidentiality of individual data. The researchers are solely responsible for the conclusions and inferences drawn upon available data. The authors acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC) through grant G8PJ-437384-2012. This work was also supported by the US National Science Foundation (NSF) under Grants GEO/OAD-1342869 and ICER/EAR-1829999 as part of the Belmont Forum – Sustainable Urbanisation Global Initiative (SUGI)/ Food-Water-Energy Nexus theme). The NERC Belmont Forum provided UK funding (NE/L009285/1). Funding to the Helmholtz Centre for Environmental Research (UFZ) was provided by the Deutsche Forschungsgemeinschaft (DFG) (KL 2764/1-1) and the German Federal Ministry of Education and Research (BMBF) (033WU002) as part of the Belmont Forum. The University of Manchester's Computational Shared Facility is acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are solely those of the authors and do not necessarily reflect the views of the NSERC, NSF and other funding or governmental agencies that provided funding or data.

Author contributions. TL and AT designed research, JY, CK and TL developed the relevant modules, TL performed simulations, analyzed data and wrote the first draft, JY, CK and AT reviewed the paper and provided substantial improvements, AT secured funding and provided supervision. All authors contributed to the paper revision

Competing interests. The authors declare that they have no conflict of interest.

References

Al-Kharabsheh, A.: Ground-water modelling and long-term management of the Azraq basin as an example of arid area conditions (Jordan), Journal of Arid Environments, 44, 143–153, https://doi.org/10.1006/jare.1999.0580, 2000.

Arrow, K. J.: Social choice and individual values, yale university press - 2nd ed., 1963 edn., 1951.

Avisse, N.: Quantitative analysis of non-cooperative transboundary river basins, Ph.D. thesis, Université Laval, Québec, Canada, https://corpus.ulaval.ca/jspui/handle/20.500.11794/30278, 2018.

Avisse, N., Tilmant, A., Rosenberg, D., and Talozi, S.: Quantitative assessment of contested water uses and management in the conflict-torn yarmouk river basin, Journal of Water Resources Planning and Management, 146, 05020 010, https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29WR.1943-5452.0001240, 2020.

Ben-Haim, Y.: Info-gap decision theory: Decisions under severe uncertainty (2nd ed.)., London: Academic Press, https://doi.org/10.1016/B978-0-12-373552-2.X5000-0, 2006.

Brown, C. and Wilby, R. L.: An alternate approach to assessing climate risks, Eos, Transactions American Geophysical Union, 93, 401–402, https://doi.org/10.1029/2012EO410001, 2012.

Central Intelligence Agency (CIA): World Factbook: Middle East -Jordan, https://www.cia.gov/the-world-factbook/countries/jordan/, 2021.

Courcier, R., Venot, J., Molle, F., Suleiman, R., and Jridi, A.: "Historical transformations of the Lower Jordan River Basin (in Jordan): Changes in water use and projections (1950-2025)" (Comprehensive Assessment of Water Management in Agriculture Research Report 9, International Water Management Institute, Colombo, Sri Lanka, 2005).

Culley, S., Noble, S., Yates, A., Timbs, M., Westra, S., Maier, H. R., Giuliani, M., and Castelletti, A.: A bottom-up approach to identifying the maximum operational adaptive capacity of water resource systems to a changing climate, Water Resources Research, 52, 6751–6768, https://doi.org/10.1002/2015WR018253, 2016.

Füssel, H.-M., Toth, F. L., van Minnen, J. G., and Kaspar, F.: Climate impact response functions as impact tools in the tolerable windows approach, Climatic Change, 56, 91–117, https://doi.org/10.1023/A:1021340513936, 2003.

Gold, D. F., Reed, P. M., Trindade, B. C., and Characklis, G. W.: Identifying actionable compromises: navigating multi-city robustness conflicts to discover cooperative safe operating spaces for regional water supply portfolios, Water Resources Research, 55, 9024–9050, https://doi.org/10.1029/2019WR025462, 2019.

Goode, D. J., Senior, L. A., Subah, A., and Jaber, A.: Groundwater-level trends and forecasts, and salinity trends, in the Azraq, Dead Sea, Hammad, Jordan Side Valleys, Yarmouk, and Zarqa groundwater basins, Jordan, USGS Numbered Series 2013-1061, U.S. Geological Survey, Reston, VA, http://pubs.er.usgs.gov/publication/ofr20131061, 2013.

Gunkel, A. and Lange, J.: New insights into the natural variability of water resources in the lower jordan river basin, Water Resources Management, 26, 963–980, https://doi.org/10.1007/s11269-011-9903-1, 2012.

Haasnoot, M., Warren, A., and Kwakkel, J. H.: Dynamic Adaptive Policy Pathways (DAPP), in: Decision Making under Deep Uncertainty: From Theory to Practice, edited by Marchau, V. A. W. J., Walker, W. E., Bloemen, P. J. T. M., and Popper, S. W., pp. 71–92, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-05252-2_4, 2019.

Haddadin, M.: Cooperation and lack thereof on management of the Yarmouk River, Water International, 34, 420–431, https://doi.org/10.1080/02508060903396148, 2009.

Hadjimichael, A., Quinn, J., Wilson, E., Reed, P., Basdekas, L., Yates, D., and Garrison, M.: Defining robustness, vulnerabilities, and consequential scenarios for diverse stakeholder interests in institutionally complex river basins, Earth's Future, 8, https://doi.org/10.1029/2020EF001503, 2020.

Hallegatte, S. and Rozenberg, J.: Climate change through a poverty lens, Nature Climate Change, 7, 250–256, https://doi.org/10.1038/nclimate3253, 2017.

Hussein, H.: An analysis of the discourse of water scarcity and hydropolitical dynamics in the case of Jordan, Ph.D. thesis, University of East Anglia, 2018.

Jaeger, W. K., Amos, A., Bigelow, D. P., Chang, H., Conklin, D. R., Haggerty, R., Langpap, C., Moore, K., Mote, P. W., Nolin, A. W., Plantinga, A. J., Schwartz, C. L., Tullos, D., and Turner, D. P.: Finding water scarcity amid abundance using human–natural system models, Proceedings of the National Academy of Sciences, 114, 11 884–11 889, https://doi.org/10.1073/pnas.1706847114, 2017.

Jafino, B. A., Kwakkel, J. H., and Taebi, B.: Enabling assessment of distributive justice through models for climate change planning: A review of recent advances and a research agenda, WIREs Climate Change, 12, e721, https://doi.org/10.1002/wcc.721, 2021.

Kawada, Y., Nakamura, Y., and Otani, S.: An Axiomatic Foundation of the Multiplicative Human Development Index, Review of Income and Wealth, 65, 771–784, https://ideas.repec.org/a/bla/revinw/v65y2019i4p771-784.html, publisher: International Association for Research in Income and Wealth, 2019.

Kim, D., Chun, J. A., and Choi, S. J.: Incorporating the logistic regression into a decision-centric assessment of climate change impacts on a complex river system, Hydrology and Earth System Sciences, 23, 1145–1162, https://doi.org/10.5194/hess-23-1145-2019, 2019.

Klassert, C., Sigel, K., Gawel, E., and Klauer, B.: Modeling Residential Water Consumption in Amman: The Role of Intermittency, Storage, and Pricing for Piped and Tanker Water, Water, 7, 3643–3670, https://doi.org/10.3390/w7073643, 2015.

Klassert, C., Gawel, E., Sigel, K., and Klauer, B.: Sustainable transformation of urban water infrastructure in Amman, Jordan – Meeting residential water demand in the face of deficient public supply and alternative private water markets, in: Urban Transformations: Sustainable Urban Development Through Resource Efficiency, Quality of Life and Resilience, edited by Kabisch, S., Koch, F., Gawel, E., Haase, A., Knapp, S., Krellenberg, K., Nivala, J., and Zehnsdorf, A., Future City, pp. 93–115, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-59324-1_6, 2018a.

Klassert, C., Sigel, K., Klauer, B., and Gawel, E.: Increasing Block Tariffs in an Arid Developing Country: A Discrete/Continuous Choice Model of Residential Water Demand in Jordan, Water, 10, 248, https://doi.org/10.3390/w10030248, 2018b.

Lempert, R. J.: Robust Decision Making (RDM), in: Decision Making under Deep Uncertainty: From Theory to Practice, edited by Marchau, V. A. W. J., Walker, W. E., Bloemen, P. J. T. M., and Popper, S. W., pp. 23–51, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-05252-2_2, 2019.

Lempert, R. J., Groves, D. G., Popper, S. W., and Bankes, S. C.: A general, analytic method for generating robust strategies and narrative scenarios, Management Science, 52, 514–528, https://doi.org/10.1287/mnsc.1050.0472, 2006.

Loucks, D. P. and van Beek, E.: Water Resource Systems Planning and Management: An Introduction to Methods, Models, and Applications, Springer, Cham, https://doi.org/10.1007/978-3-319-44234-1, 2017.

Maier, H. R., Guillaume, J. H. A., van Delden, H., Riddell, G. A., Haasnoot, M., and Kwakkel, J. H.: An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together?, Environmental Modelling & Software, 81, 154–164, https://doi.org/10.1016/j.envsoft.2016.03.014, 2016.

Marcos-Garcia, P., Brown, C., and Pulido-Velazquez, M.: Development of Climate Impact Response Functions for highly regulated water resource systems, Journal of Hydrology, 590, 125 251, https://doi.org/10.1016/j.jhydrol.2020.125251, 2020.

Mariani, F. and Ciommi, M.: Aggregating Composite Indicators through the Geometric Mean: A Penalization Approach, Computation, 10, 64, https://doi.org/10.3390/computation10040064, number: 4 Publisher: Multidisciplinary Digital Publishing Institute, 2022.

Mastrandrea, M. D., Heller, N. E., Root, T. L., and Schneider, S. H.: Bridging the gap: linking climate-impacts research with adaptation planning and management, Climatic Change, 100, 87–101, https://doi.org/10.1007/s10584-010-9827-4, 2010.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Stationarity Is Dead: Whither Water Management?, Science, 319, 573–574, https://doi.org/10.1126/science.1151915, 2008.

Ministry of Water and Irrigation - MWI: Water Yearbook 2016-17 (MWI, Amman, Jordan, 2019).

Moody, P. and Brown, C.: Robustness indicators for evaluation under climate change: Application to the upper Great Lakes, Water Resources Research, 49, 3576–3588, https://doi.org/10.1002/wrcr.20228, 2013.

Moulin, H.: Egalitarianism and Utilitarianism in Quasi-Linear Bargaining, Econometrica, 53, 49–67, https://doi.org/10.2307/1911723, 1985.

Mustafa, D. and Tillotson, M.: The topologies and topographies of hydro-social territorialisation in Jordan, Political Geography, 70, 74–82, https://doi.org/10.1016/j.polgeo.2019.01.011, 2019.

Müller, M. F., Yoon, J., Gorelick, S. M., Avisse, N., and Tilmant, A.: Impact of the Syrian refugee crisis on land use and transboundary freshwater resources, Proceedings of the National Academy of Sciences, 113, 14 932–14 937, 2016.

Müller, M. F., Müller-Itten, M. C., and Gorelick, S. M.: How Jordan and Saudi Arabia are avoiding a tragedy of the commons over shared groundwater, Water Resources Research, 53, 5451–5468, https://doi.org/10.1002/2016WR020261, 2017.

Poff, N. L., Brown, C. M., Grantham, T. E., Matthews, J. H., Palmer, M. A., Spence, C. M., Wilby, R. L., Haasnoot, M., Mendoza, G. F., Dominique, K. C., and Baeza-Castro, A.: Sustainable water management under future uncertainty with eco-engineering decision scaling, Nature Climate Change, 6, 25–34, https://doi.org/10.1038/nclimate2765, 2016.

Prudhomme, C., Wilby, R., Crooks, S., Kay, A., and Reynard, N.: Scenario-neutral approach to climate change impact studies: Application to flood risk, Journal of Hydrology, 390, 198–209, https://doi.org/10.1016/j.jhydrol.2010.06.043, 2010.

Rahman, K., Gorelick, S. M., Dennedy-Frank, P. J., Yoon, J., and Rajaratnam, B.: Declining rainfall and regional variability changes in Jordan, Water Resources Research, 51, 3828–3835, https://doi.org/10.1002/2015WR017153, 2015.

Rajsekhar, D. and Gorelick, S. M.: Increasing drought in Jordan: Climate change and cascading Syrian land-use impacts on reducing transboundary flow, Science Advances, 3, e1700 581, https://doi.org/10.1126/sciadv.1700581, 2017.

Rawls, J.: A theory of justice, Mass, USA,, cambridge university press edn., 1970.

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Da Silva, L. A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J. C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., and Tavoni, M.: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview, Global Environmental Change, 42, 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009, 2017.

Rickels, W., Dovern, J., Hoffmann, J., Quaas, M. F., Schmidt, J. O., and Visbeck, M.: Indicators for monitoring sustainable development goals: An application to oceanic development in the European Union, Earth's Future, 4, 252–267, https://doi.org/10.1002/2016EF000353, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016EF000353, 2016.

Rosenberg, D. E., Talozi, S., and Lund, J. R.: Intermittent water supplies: challenges and opportunities for residential water users in Jordan, Water International, 33, 488–504, https://doi.org/10.1080/02508060802474574, 2008.

Salameh, E., Alraggad, M., and Amaireh, M.: Degradation processes along the new northeastern shores of the Dead Sea, Environmental Earth Sciences, 78, 164, https://doi.org/10.1007/s12665-019-8155-x, 2019.

Selby, P., Medellín-Azuara, J., Klassert, C., Yoon, J., and Harou, J.: An agent-based model of Jordan highland farmer decision making, in: undefined, Toulouse, France, 2016.

Sigel, K., Klasser, C., Zozmann, H., Talozi, S., Klauer, B., and Gawel, E.: Socioeconomic surveys on private tanker water markets in Jordan: Objectives, design and methodology, Tech. Rep. 4/2017, Helmholtz Centre for Environmental Research (UFZ), Division of Social Sciences (ÖKUS), https://ideas.repec.org/p/zbw/ufzdps/42017.html, 2017.

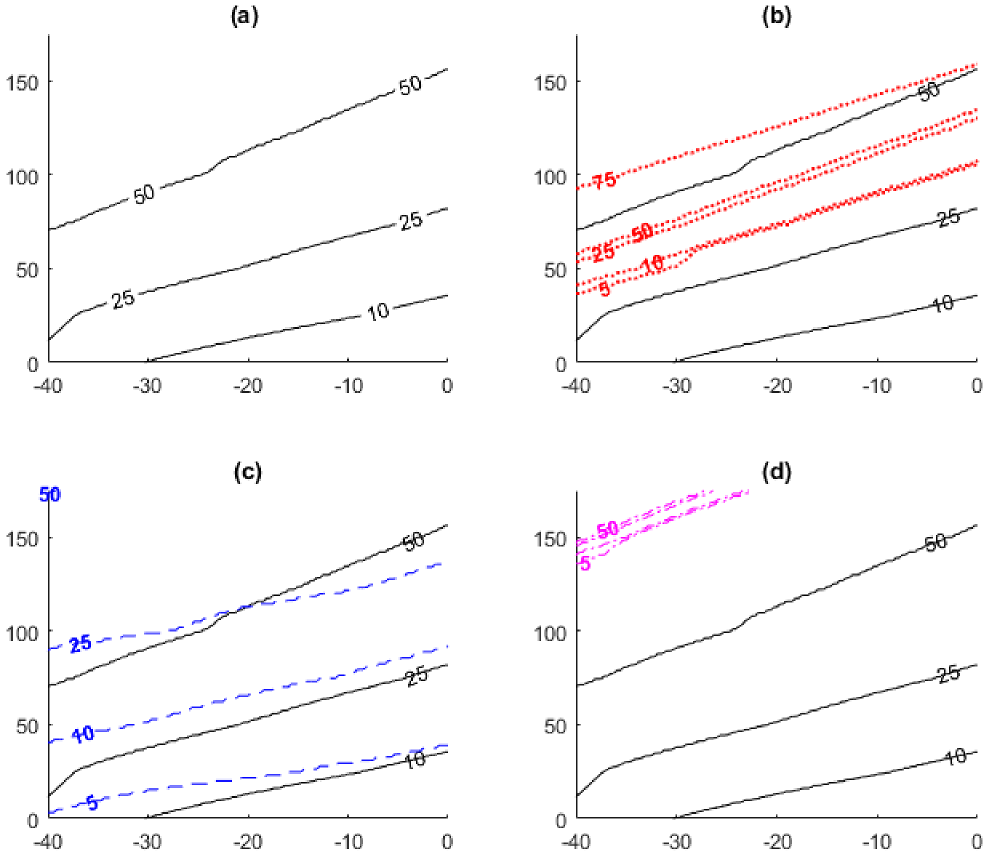
Talozi, D. M. S.: Tankers, Wells, Pipes and Pumps: Agents and Mediators of Water Geographies in Amman, Jordan, Water Alternatives, 11, 916–932, 2018.

Tilmant, A., van der Zaag, P., and Fortemps, P.: Modeling and analysis of collective management of water resources, Hydrology and Earth System Sciences, 11, 711–720, https://doi.org/10.5194/hess-11-711-2007, 2007.

Trindade, B. C., Reed, P. M., Herman, J. D., Zeff, H. B., and Characklis, G. W.: Reducing regional drought vulnerabilities and multi-city robustness conflicts using many-objective optimization under deep uncertainty, Advances in Water Resources, 104, 195–209, https://doi.org/10.1016/j.advwatres.2017.03.023, 2017.

Whitman, E.: A land without water: the scramble to stop Jordan from running dry, Nature, 573, 20–23, https://doi.org/10.1038/d41586-019-02600-w, 2019.

Yoon, J., Klassert, C., Selby, P., Lachaut, T., Knox, S., Avisse, N., Harou, J., Tilmant, A., Klauer, B., Mustafa, D., Sigel, K., Talozi, S., Gawel, E., Medellín-Azuara, J., Bataineh, B., Zhang, H., and Gorelick, S. M.: A coupled human–natural system analysis of freshwater security under climate and population change, Proceedings of the National Academy of Sciences, 118, https://doi.org/10.1073/pnas.2020431118, 2021.


(a)
(b)
(c)
(d)
150
100
50
0
-40
-30
-20
-10
0
50
25
10
75
5