Possibilistic response surfaces~~combining~~ : incorporating fuzzy ~~targets and hydro-climatic uncertainty in~~ thresholds into bottom-up flood vulnerability ~~assessment~~analysis

Thibaut Lachaut[1] and Amaury Tilmant[1]

[1]Laval University, Québec, QC G1V 0A6, Canada

Correspondence: Thibaut Lachaut (thibaut.lachaut.1@ulaval.ca)

Abstract. Several alternatives have been proposed to shift the paradigms of water management under uncertainty from predictive to decision-centric. An ~~often mentioned~~ often-mentioned tool is the stress-test response surface~~;~~ , mapping system performance to a large sample of future hydro-climatic conditions. Dividing this exposure space between ~~success and failure requires clear performance targets~~acceptable and unacceptable states requires a criterion of acceptable performance defined by a threshold. In practice, however, stakeholders and decision-makers may be confronted with ambiguous objectives for which ~~there are no clearly-defined~~ the the acceptability threshold is not clearly defined (crisp)~~performance thresholds. Furthermore, response surfaces can be non-deterministic, as they do not fully capture all possible sources of~~ . To accommodate such situations, this paper integrates fuzzy thresholds to the response surface tool. Such integration is not straightforward when response surfaces also have their own irreducible uncertainty, from the limited number of descriptors and the stochasticity of hydro-climatic ~~uncertainty. The challenge is thus to combine two different types of uncertainty~~conditions. Incorporating fuzzy thresholds therefore requires articulating uncertainties that are different in nature: the irreducible uncertainty of the response itself relative to the variables that describe change, and the ~~fuzziness of the performance target. We~~ ambiguity of the acceptability threshold. We thus propose possibilistic surfaces to assess flood vulnerability with fuzzy ~~performance thresholds. Three approaches are tested and compared on a un-gridded sample of the exposure space: (i) an aggregation of logistic regressions based on α-cuts combines the uncertainty of the response itself~~ acceptability thresholds. An adaptation of the logistic regression for fuzzy set theory combines the probability of acceptable outcome and the ambiguity of the ~~target~~ acceptability criterion within a single possibility measure~~; (ii) an alternative approximates the response with a fuzzy analytical surface; and (iii) a convex delineation expresses the largest range of failure specific to a given management rule without probabilistic assumptions. To illustrate the proposed approaches, we~~ . We use the flood-prone reservoir system of the Upper Saint-François River Basin in Canada as a case study ~~. This study shows that ambiguity can be effectively be considered~~ to illustrate the proposed approach. Results show how a fuzzy threshold can be quantitatively integrated when generating a response surface~~and suggests how further research could build a possibilistic framework~~ , and how ignoring it might lead to different decisions. This study suggests that further theoretical development should link the decision-making under deep uncertainty framework with the existing experience of fuzzy set theory, notably for hydro-climatic ~~uncertainty~~vulnerability analysis.

1 Introduction

Uncertainty is a ~~driving force of the transformations of human societies. Settlement, agriculture and animal husbandry, storage facilities, many innovations seek to make the future more predictable, and thus allow for investments with a better knowledge of associated risks. For some social scientists, uncertainty and risk become more central, even constitutive of society, following the industrial revolutions (U. Beck, 1985). As the scope of human activity expands exponentially and meets the boundaries of its functional environment (J. Rockström et al. 2009), the adverse externalities shape a new layer of human-induced risks. Nuclear and chemical catastrophes were the first to attract global attention, later joined by the perturbation of climatic and biophysical mechanisms at a global scale.~~

~~The relationship between society and water follows the same path. Five millennia of engineering have seen the development of reservoirs, irrigation, levees and aqueducts in order to counter the uncertainty inherent to precipitations and river flows. The water domain is consequently confronted to the same new layer of uncertainty and risks stemming from human externalities. Climate change is a looming threat on current investment or planning decisions in water resources (IPCC, 2014), while increased pressure at the basin level makes hazards, appropriations and conflicts all the more impactful (Srinivasan et al., 2012), as illustrated by the closed basin concept (Molle et al., 2010).~~

~~The way water~~ defining feature of water management - both science and practice~~- handles uncertainty is crucial~~. Not only does uncertainty justify ~~intervention~~ the need for infrastructures, interventions and planning, but the way decisions are taken is also based on different interpretations of uncertainty. The dominant paradigm has been to optimize investments or management plans according to the most probable future, ~~a knowledge based on the statistical analysis of historical time series and assuming their stationarity. This assumption of stationarity~~ assuming that the underlying processes are stationary. However, the stationary assumption has been contested ~~however~~ as anthropogenic activities do affect the very ~~climatic processes that led to past hydrologic behavior~~ processes that govern the water cycle (Milly et al., 2008).

Water ~~planning can also rely on the modeling of future scenarios and~~ resource planning approaches usually requires a prediction of those processes affecting both sides of the supply-demand relationship. In particular, the assessment of climate change impacts ~~over existing hydro-climatic conditions. This widely used method~~ commonly relies on General Circulation Models (GCM) that ~~simulates~~ simulate future global climates ~~depending on assumptions (Representative Concentration Pathways, RCP) about the CO2~~ based on assumptions about greenhouse gas concentrations in the atmosphere, or more generally, the radiative forcing (Brown and Wilby, 2012, Weaver et al, 2013). Assumptions about greenhouse gas emissions, land cover or population are grouped under representative concentration pathways (RCP). Results from global simulations are translated into local hydro-climatic projections through a downscaling process. Hydrological modelling then translates climatic variables into run-off time series. Such an approach has its own limitations, however. CO_2 emission pathways depend on worldwide future policy choices which are not yet known nor even predictable. Moreover, climate models carry their own structural uncertainties, and so ~~are~~ do the downscaling processes (Prudhomme et al., 2010, Mastrandrea et al., 2010, Kay, et al., 2013, Weaver et al., 2013, Kim et al., 2019). Besides, a discrete set of projections is not suited to find the hydro-climatic thresholds beyond which a system fails to reach its target (Culley et al., 2016). Such a risk assessment process is also increasingly unreliable with systems that operate with shorter time steps and extreme events, like flood control operations (Knighton et al., 2017).

In the last 15 years there has thus been a widespread effort to find new paradigms to make decisions under deep uncertainty, notably through a greater focus on the decision process rather than on improving predictions (Lempert et al., 2006, Maier et al., 2016, Lempert, 2019). Switching to a robust or decision-centric paradigm always seeks to increase the sampling of hydro-climatic conditions, and relies on a sensitivity analysis of a water system to ~~driving variables~~ stressors rather than evaluating the consequences of the most probable future and optimizing accordingly (Weaver et al., 2013). A consolidation of the field is proposed under the decision making under deep uncertainty (DMDU) denomination (Marchau et al., 2019).

One of the most common tools within the decision-centric framework is the response function or surface (Prudhomme et al., 2010, Brown et al., 2012, Culley et. al., 2016, Brown et al., 2019, Nazemi et al., 2020, Di Francesco et al., 2020). Through a stress-test, "bottom-up" approach, a water system is simulated for a large set of conditions representing possible evolutions of some uncertain hydro-climatic variables (or stressors), establishing a relationship between such stressors and the performance of the system. When specifically addressing climate change, it corresponds to the reverse CIRF (Climate Impact Response Function, Fussel et al., 2003, Marcos-Garcia et al., 2020). Such an approach is sometimes called scenario-neutral (Prudhomme et al., 2010, Broderick et al., 2019) as it ~~doesn't intrinsically rely on GCM outputs and RCP assumptions~~separates the system response from the likelihood of each scenario. Alternatives, like making new investments, changing management schemes, are compared through their respective performance outcome over a whole ~~space of possibilities~~ range of possibilities (states of the world), or exposure space (Culley et al., 2016). The response surface can be used to measure an uncertainty horizon between a first estimate of the state of the world and an acceptability frontier (Info-gap decision theory, Ben-Haim, 2006). In the Decision Scaling approach (Brown et al., 2012, Brown et al., 2019) GCM projections can then be introduced as weights on the response surface to inform probabilities associated to climate states. GCMs can thus remain useful without conditioning the decision process, and ~~once updated their outcomes on the system can be mapped on the response~~ their weights can be updated as uncertainty is resolved, resulting in a refined estimate of the expected system outcome over the response surface without the need for new simulations of the water system. The intention shared within the overall decision-centric framework is to adapt classic risk assessment to the “death of stationarity” (Milly et al., 2008) while producing information more useful and engaging than a fully descriptive scenario approach (Weaver 2013). Response surfaces have been illustrated by many case studies (e.g. Nazemi et al., 2013, Turner et al., 2014, Whateley et al., 2014, Herman et al., 2015, Steinschneider et al., 2015, Spence et al., 2016, Pirttioja et al., 2019, Ray et al., 2020), expanded to many-objectives or stakeholder systems (Poff et al., 2016; Culley et al., 2016, Kim et al., 2019) and sometimes officially adopted in management processes (Moody and Brown, 2013, Weaver et al., 2013, Brown et al., 2019).

Although the response surface is a powerful and efficient tool to circumvent the problems ~~and arbitrariness~~ brought by “top-down”, GCM-based assessments, the ~~applications~~ incorporation of such tools in actual decision-making processes to date remain relatively recent and scarce (Guo et al., 2018). Moreover, many assumptions associated with the stress test approach can introduce additional uncertainty.

One source can be the ambiguity of the user-defined ~~performance targets~~ acceptability thresholds (Maier et al., 2016). The stress-test approach needs performance target values ~~(thresholds)~~ in order to separate the exposure space between ~~accepted and rejected~~ acceptable and unacceptable domains. However such ~~targets~~ thresholds are often unclear or arbitrary ~~; and are heavily reliant on political, sociological and institutional processes~~ (El-Baroudy and Simonovic, 2004). Recently, Hadjimichel et al. (2020) performed a sensitivity analysis on the definition of binary acceptability thresholds for a large number of stakeholders in a deep uncertainty framework, demonstrating its impact on decision making. Fuzzy set theory (Zadeh, 1965) provides an analytical framework to characterize and manipulate stakeholders' ambiguity (Huynh et al., 2007). It has been extensively used in the water domain (Tilmant et al., 2002, El-Baroudy and Simonovic, 2004, Qiu et al., 2018) in particular to solve multi-objective decision-making problems (e.g. Jun et al., 2013). However, to the best of our knowledge, fuzzy set theory has not yet been used to handle imprecise thresholds between satisfactory and ~~failure~~ unsatisfactory regions of a response surface. The very notion of an arbitrary threshold defining success, like flood control reliability above 0.95, can be considered as a departure from a strictly probabilistic framework and could justify a complementary possibilistic approach based on fuzzy sets (Dubois et al., 2004). This paper therefore introduces the use of fuzzy acceptability thresholds when building a response surface for decision-centric vulnerability assessment.

~~Independently from performance targets, response functions also have their own noise or internal uncertainty , as their selected driving~~ However, the internal uncertainty of the response surface hinders the direct application of a fuzzy threshold. The selected stressor variables can only partially explain ~~hydrological and climatic uncertainties~~hydro-climatic uncertainties, and stochastic realizations introduce noise in the resulting system performance. As such, performance is an expected value rather than a deterministic one, ~~hence possibly underestimating~~ and that estimate might underestimate real risks. Irreducible uncertainty usually requires adaptive management (Brown et al., 2011), but there is interest ~~to integrate part of this information~~ in integrating estimates of uncertainty into the response surface tool.

Kay et al. (2014) proposed the use of uncertainty allowances that could vary depending on the response type and catchment. More specifically, flood control systems operate on shorter time scales and are ~~even~~ harder to assess over long term climate shifts (Knighton et al. 2017), ~~thus also more challenging to evaluate with~~ increasing uncertainty in flood response functions. Kim et al. (2018) ~~stress~~ show how the choice of a longer modelling time scale (daily vs hourly) can lead to risk underestimation. The choice of ~~different~~ weather generator used to generate synthetic weather series in a scenario-neutral ~~methods can experiment~~ can also lead to different results (Keller et al., 2019~~), notably the choice of the synthetic series generator (~~, Nazemi et al., 2020). Steinschneider et al. (2015) and Whateley and Brown (2016) compare different sources of uncertainty in the response, acknowledging the strong impacts of hydrological modelling and internal climate variability compared to long term climate uncertainty~~, as well as Whateley and Brown (2016)~~. Testing a limited number of stressors as ~~explaining~~ explanatory variables therefore leads to a response function that returns ~~uncertain performance .~~ imprecise performance estimates. Quin et al. (2018), Kim et al. (2019)~~propose to associate probabilities to uncertain response functions through logistic regression , while~~ , Lamontagne et al. (2019), Hadjimichael et al. (2020), and Marcos-Garcia et al. (2020) use a logistic regression to divide the exposure space based on probability of success (often as a first step in a scenario-discovery approach). Tanner et al. (2019)

do so with a Bayesian belief network model. ~~It is here noted that most response surfaces follow gridded sampling, which can also be a loss of information (e.g. Huang, 2000 for elevation models) and thus risk under-estimation.~~

~~The objective of the present study is to combine with a possibilistic approach~~ Such internal uncertainty to the response function challenges the introduction of fuzzy thresholds, as the separation of the exposure space in acceptable and unacceptable regions is not obvious even with a binary definition.

Through a possibilistic approach, the present study combines two different types of uncertainty: the fuzziness of ~~performance targets~~ acceptability thresholds and the irreducible uncertainty of the response surface. The rationale ~~behind it and three tested implementations are presented~~ of the paper is further developed in section 2~~: a numerical approximation of a fuzzy-random~~ , followed by the incorporation of a fuzzy acceptability threshold to the logistic regression, ~~a fuzzy analytical approximation of the response itself and a convex delineation of the largest range of failure~~which has already been used to handle the response uncertainty. A case study is presented in section 3, a flood-prone reservoir system in southern Québec, Canada. Results are presented in part 4, followed by a discussion on the ~~respective~~ merits and limitations of the proposed ~~methods~~method.

2 Methods

2.1 Rationale

2.1.1 Uncertain response function

We first consider how a limited set of variables leads to an inherently uncertain response function, and how it relates to the partition of the exposure space and the decision process.

A stress test consists in assessing the performance of a system for a large enough number of situations, in order to identify ~~*which of these situations* leads to an unsatisfying performance, or overall failure~~which of these situations lead to an unacceptable performance.

~~Concept of the response surface as a stress-test with describing variables (x_1, x_2). Success and failure regions are defined by a threshold θ over performance p.~~

~~Often inspired from Hashimoto et al. (1982), a performance indicator is a statistical measure of local failureduration or amplitude, aggregated over a certain time period. Local failure is~~ We first define how success, failure, performance and acceptability are used in this study. Success or failure are a state of the system at a given time step~~: a state of flooding~~ . For example failure can be defined by a streamflow exceeding a threshold at any moment, characterizing a state of flooding. Inspired from Hashimoto et al. (1982), common performance indicators of a water system are statistical measures of the frequency, amplitude or duration of failures, aggregated over a certain time period. ~~The overall performance of the system over a period quantifies its ability to mitigate the number or amplitude of local failures.~~ For example, the reliability of a flood control system can be measured as the proportion of a given period (frequency) where no flooding happens.

While success/failure define the state of the system for a single time step, acceptable/unacceptable define its behavior over a time period. When performing a stress-test of a system, ~~overall success or failure~~ the criterion for acceptable/unacceptable outcomes is usually defined by ~~a performance target~~ performance satisfying a threshold θ, for example reliability above 0.95 over a given period can define ~~overall success~~an acceptable outcome.

A stress-test maps the performance ~~p~~ R on a response surface, to a limited number of descriptive variables or stressors x_i. ~~It aims at delineating~~ Each coordinate, or state of the world, is a combination of specific values taken by such stressors. The stress-test aims to delineate the subsets A and D of ~~overall success and failure~~ acceptable and unacceptable outcomes (Fig. 1). ~~Such variables~~ Alternative options (management rules, infrastructure design) can be ranked based on the respective size of subsets A and D. The more states of the world lead to acceptable outcomes, the more robust an option is, thus the more preferable in this approach.

The descriptive variables or stressors, like the mean flow, the peak flow, or temporal autocorrelations, are aggregations of the time series that are the inputs of a water system ~~simulation~~model. Because a limited number of descriptors do not capture all possible fluctuations of a time series, a term of irreducible uncertainty remains. The response surface, R, is then given by:

$$\cancel{p}R = g(x_1, x_2 ...) + \cancel{R}\epsilon \tag{1}$$

In a risk-averse approach, the objective is to find the range of ~~failure (more than success)~~unacceptable outcomes, the space over which a system fails to satisfy ~~a performance target~~ an acceptability threshold θ. With 2 variables, this space is the set of solutions $D = (x_1^*, x_2^*)$ to the ~~inequation~~ inequality $p < \theta$, so

$$g(x_1, x_2) + \cancel{R}\epsilon < \theta \tag{2}$$

Simplifying the response surface by ~~averaging it over *p* (*vertically*)~~ e.g. its average estimate can thus under-estimate the ~~failure~~ unacceptability domain. Irreducible uncertainty can be addressed through adaptive management (Brown et al., 2011), uncertainty allowances (Kay et al., 2014), and extensive Monte-Carlo sampling (Whateley and Brown, 2016). If possible though, it can be convenient to directly integrate information about the remaining uncertainty within the response surface itself. It can be represented through a transition zone between success and failure domains ~~, as performed by Kim et al. (2019)~~ with a logistic regression. ~~Besides, most studies use gridded sampling of the exposure space, which is a *horizontal* aggregation that also results in information loss like in the case of digital elevation models (Huang, 2000), and which in this case can also under-estimate risks. A simple un-gridded alternative is proposed in section 3.2.~~

2.1.2 Fuzzy ~~performance targetsThe performance target~~ acceptability thresholds

The acceptability criterion based on a threshold θ defines the set of ~~successful~~ acceptable outcomes. It is a subjective or arbitrary opinion from stakeholders or decision makers to attribute a normative value to a certain performance level. The vast majority of the studies reported in the literature assume that the threshold between satisfactory and unsatisfactory ~~states~~ outcomes is crisp (Brown et al., 2012, Culley et al., 2016, Kim et al., 2019). As such a threshold shapes directly the partition of the response function, with a crisp value the exposure space can be subdivided in only two sub-spaces: ~~failure versus success~~acceptable versus unacceptable.

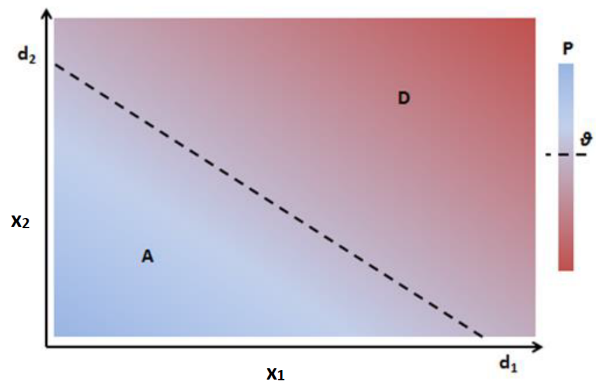


Figure 1. Concept of the response surface as a stress-test with descriptive variables (x_1, x_2). Acceptable and unacceptable regions are defined by a threshold θ over performance R.

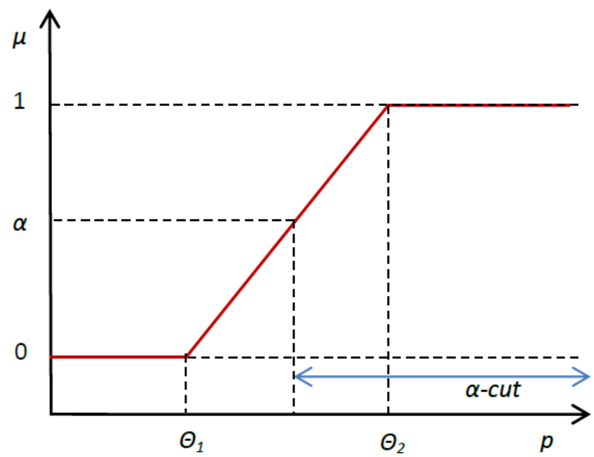


Figure 2. Concept for a fuzzy set of acceptable outcomes A_μ over performance R.

The very existence of a ~~target~~ threshold is the basis of satisficing behaviors (Simon, 1955) that differ from utility maximizing behaviors as coined by Von Neumann and Morgenstern (1944). In practice however, ~~while clearly following a satisficing model,~~ there might be situations whereby the water manager is unable (or unwilling) to provide a crisp, well-defined ~~target~~threshold, or when such threshold is disagreed upon by stakeholders. For example, when controlling water levels in a reservoir to prevent ~~inundations~~floods, the operator can handle certain tolerances above the maximum desired level. Of course, the greater the deviation from the desired level, the less acceptable it becomes.

Mathematically, fuzzy sets theory ~~(FST)~~ handles imprecisely-defined or ambiguous quantities. Introduced by Zadeh in 1965, fuzzy sets theory has become a common tool in decision making analysis or computational sciences when non-probabilistic uncertainty stemming from ambiguity or vagueness must be considered (Yu et al., 2002). In our case, ~~FST~~ fuzzy sets theory allows us to introduce vagueness in target-based decision making, without forsaking a target-based model in favor of an unbounded maximizing behavior (although a fuzzy target can also be seen as a generalization of both maximizing and satisficing behaviors – see Castagnoli and LiCalzi, 1996, and Huynh et al., 2007).

~~Concept for a fuzzy set of success A_μ over performance p.~~

We consider here the case where such a ~~target~~ threshold θ may not be precisely defined by stakeholders but can take many subjective qualifications from acceptable to unbearable, hence relaxing (without fully removing) the arbitrary condition of satisfying a crisp value. A fuzzy set A_μ of acceptable states therefore qualifies the performance ~~p~~ R with a membership value comprised between 0 and 1. The membership function ~~μ~~ μ associated to the fuzzy set A describes the degree to which any value of ~~p~~ R more or less belongs to A (Figure 2, eq 3).

$$\begin{cases} \mu(R) = 0 & R < \theta_1 \\ 0 < \mu(R) < 1 & \theta_1 \geq R < \theta_2 \\ \mu(R) = 1 & R \geq \theta_2 \end{cases} \tag{3}$$

When a threshold corresponds to a fuzzy set, it means that there is a transition zone between ~~success and failure states~~ acceptable and unacceptable outcomes where intermediate levels of membership exist. Conversely, another interpretation is that the membership function is the distribution of the *possibilities* (Zadeh 1978, Dubois and Prade, 1988) that any given performance ~~p represents a success~~ R represents an acceptable outcome.

An α-cut A_α is the crisp set over A_μ for which the membership degree to A_μ is equal or above α. The largest α-cut is called the *support* of the fuzzy set A_μ (~~$p \geq \theta_1$~~ $R \geq \theta_1$). The smallest α-cut is the *core* of the fuzzy set (~~$p \geq \theta_2$).~~

$R \geq \theta_2$).

$$A_\alpha = \{pR \in A_\mu \mid \mu(pR) \geq \alpha\} \tag{4}$$

A fuzzy definition of acceptability is not only a way to accommodate ambiguity as a stakeholder-based constraint, it can also alter the outcome of the analysis. Theoretically, it can happen when the slope of a response as function of stressors is different for the compared alternatives (infrastructure investments, management rules…), as illustrated in Fig. 3 for a single stressor variable. Rule 2 has a larger region of success with a crisp threshold, but the result is mixed with a fuzzy definition of acceptability. In that case, a trade-off appears between minimizing a loss of any sort (i.e. any type of flooding), and minimizing the maximum loss (min-max).

For example, in Quinn et al. (2017a), an attempt at reducing flooding most of the time leads to worse results under extreme events. Hadjimichael et al. (2020) perform a sensitivity analysis on binary acceptability thresholds, show the impact such definition has on the outcome of a vulnerability assessment. Criterion ambiguity could lead to similar effects: the preferred option might not be the same depending on the value of the threshold, and in the present study depending on the degree of acceptability.

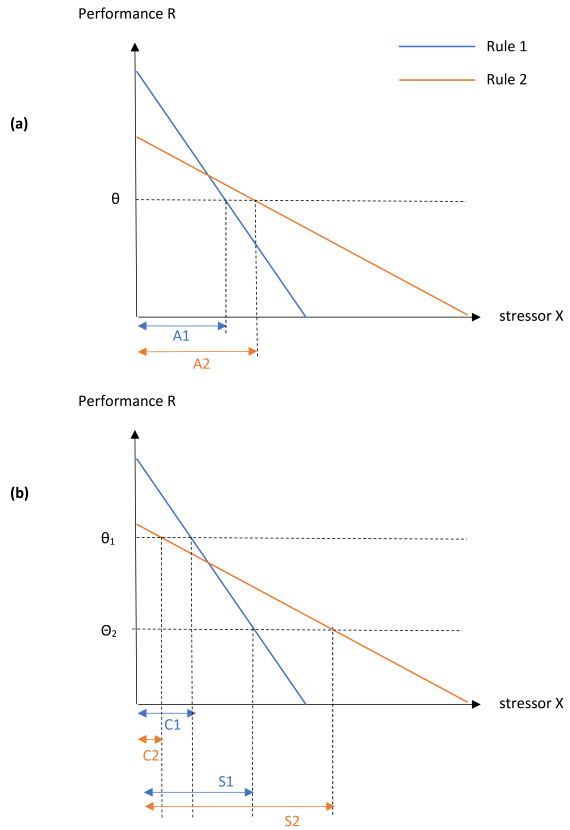


Figure 3. The case of different performance slopes, as function of a single stressor X. (a) With a crisp threshold θ, rule 2 has a larger acceptable region A2. (b) With a fuzzy threshold (θ_1, θ_2), the fuzzy set of acceptable outcomes over performance R has a core ($R \geq \theta_1$, where acceptability $\mu = 1$) and a support ($R \geq \theta_2$, where $\mu \geq 0$), to which respective regions C and S are associated. Rule 2 has a larger "at least partial" acceptability domain S2, but a smaller "full" acceptability domain C2, than Rule 1.

2.2 Combination of fuzzy ~~targets~~ thresholds and uncertain response function

~~The~~

When incorporating a fuzzy threshold, the challenge is to combine two different sources of uncertainty described in section 2.1: the uncertainty ~~or low quality~~ of the response itself relative to the variables that describe change, and the ~~fuzziness of the performance target. In order to integrate both, three methods are suggested: an~~ ambiguity of the acceptability threshold. An approximated fuzzy-random logistic regression ~~, an analytical approximation of the response surface and a convex delineation of the space of failure~~is proposed in order to integrate both.

2.2.1 Approximation of a fuzzy-random logistic regression

As the goal of the response surface is to divide the exposure space between ~~success and failure~~acceptable and unacceptable outcomes, the value associated to any combination of variables can be either 0 or 1 if a specific ~~performance target~~ acceptability threshold θ is reached or not. As seen in section 2.1, an intrinsic uncertainty remains in response surfaces. Quin et al. (2018), Kim et al. (2019)~~introduce the~~ , Lamontagne et al. (2019), Hadjimichael et al. (2020), and Marcos-Garcia et al. (2020) use a logistic regression to ~~incorporate probabilistic information into the response surface~~divide the exposure space based on probability of success. The logistic regression is used to explain a binary outcome from independent variables (x_1, x_2), and returns a probability of success π :

$$\pi_\theta = \frac{1}{1 + \exp\left(-\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\right)\right)} \tag{5}$$

$$\pi_\theta\left(x_1, x_2\right) = P\left(\cancel{p}R \geq \theta\right) \tag{6}$$

where x_i are the defining variables of the exposure space and β_i the regression coefficients. The logistic response surface therefore provides the probability π of meeting the ~~target~~ threshold θ over the (x_1, x_2) exposure space. The logistic regression also has its own uncertainty but it is not considered here. While the response surface considers a range of states of the world without knowing their probability of occurrence, the logistic regression still provides a conditional probability of acceptable outcome once a given state of the world is reached. Partitions of the space between ~~success and failure~~ acceptable and unacceptable sub-spaces, that can be defined as $\pi - cuts$, are now relative to a specific probability of success $\pi*$ taken by π_θ:

$$S_{\pi^*} = \left\{x_1, x_2 \mid \pi\ \left(x_1, x_2\right) \geq \pi^*\right\} \tag{7}$$

By considering the domain of successful outcomes as a fuzzy set, we introduce a layer of uncertainty that is different in nature from the irreducible hydro-climatic uncertainty. While the logistic regression returns a *probability* of surpassing any given ~~performance target~~ acceptability threshold for a combination of variables (eq. 5 and 6), the fuzzy set of success returns the *possibility* of any such performance ~~target~~ value being actually considered as a success (eq. 7).

Fuzzy regression models, including fuzzy logistic regression (e.g. Pourahmad et al., 2011, Namdari et al., 2014) replace probabilities by fuzzy numbers; they usually do not combine them. Fuzzy probabilities (Zadeh, 1984) are considered within the so-called fuzzy random regression field, however no fuzzy random logistic regression seems to have been developed to date (see Chukhrova and Johannssen, 2019, for a review of the fuzzy regression field).

Here we use a discretised approximation of a fuzzy random logistic regression based on α-cuts. ~~A single target θ defining a crisp set of success~~ As illustrated in Fig. 2 and Fig. 4, a fuzzy set A_μ can be decomposed in alpha-cuts. Each α-cut is a crisp set, and the values belonging to an alpha-cut also belong to the fuzzy set A_μ with a membership degree equal or above α.

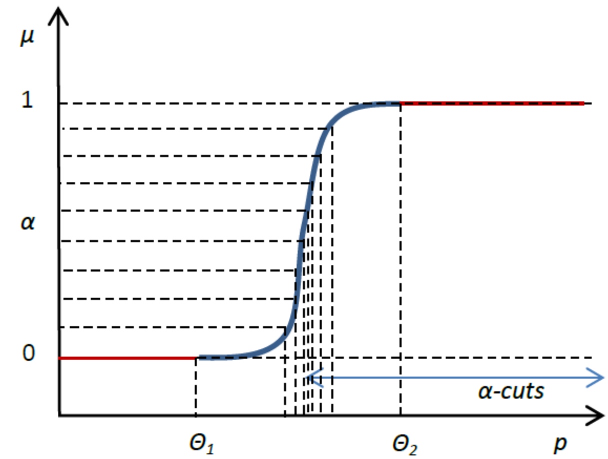


Figure 4. Concept for α-cut sampling, sigmoid function

Therefore, any crisp set of acceptable outcomes A, defined by a single threshold θ, is also an α-cut of the fuzzy set of success A_μ. Then a single logistic regression for any success threshold θ is also the probability of belonging to the α-cut of the fuzzy set of success defined by θ:

$$\pi_\theta(x_1,x_2) = P\left(pR \in A_\alpha\right) = P\left(pR \in A_\mu || \mu\left(pR\right) \geq \alpha\right) \tag{8}$$

with $\alpha = \mu(\theta)$.

Following the interpretation of Huynh et al. (2007), the overall possibility Π of the random variable ~~p~~ R belonging to the fuzzy set A_μ can be given by the integral over α of the probabilities of success defined at every α-cut.

$$\Pi(x_1,x_2) = P\left(pR \in A_\mu\right) = \int_0^1 P\left(p \in A_\mu \mid \mu(p) \geq \alpha\right) d\alpha_0 P\left(R \in A_\mu \mid \mu(R) \geq \alpha\right) d\alpha \tag{9}$$

And thus

$$\Pi(x_1,x_2) = \int_0^1 \pi_{\mu^{-1}(\alpha)}(x_1,x_2)\, d\alpha \tag{10}$$

The approximated logistic regression for a fuzzy set of success is therefore the average of the logistic regressions for all the associated α-cuts. With a uniform discretization of 10 alpha levels, the spacing of every α-cut, defined with $\theta = \mu^{-1}(\alpha)$, relies on the shape of the membership function. A linear shape of ~~$\mu(p)$~~ $\mu(R)$ leads to a uniform sampling of the α-cuts, while a sigmoid ~~shape~~ error function leads to a Gaussian sampling of α-cuts centered on ~~θ~~ $\theta^* = \mu^{-1}(0.5)$ (Fig. ~~3~~4).

~~Concept for α-cut sampling, sigmoid function~~

2.2.2 ~~Analytical approximation of the response function and fuzzy set intersection~~

~~(a) Fuzzy set A_μ. (b) Fitting error set Sp. (c) Intersection between A_μ and S_p (shaded area). (d) resulting membership function μ_R.~~

~~Instead of fitting a logistic function to the binary outcomes of success and failure, performance itself can be directly approximated as an analytical response surface (eq. 3). The final outcome is then a direct mapping from the performance approximation to a 0 1degree of success with the membership function of the fuzzy set of success A_μ (eq. 5). For every (x_1, x_2), a single approximated performance is given by $p^* = g^*(x_1, x_2)$, so the possibilistic response surface is defined by $\mu_R(p^*)$. The membership function μ is modified as follows to account for the fitting error R.~~

~~To any value of p^* is associated the membership degree μ between 0 and 1 depending on θ_1 and θ_2, 1 defining complete success. The fitting error, or uncertainty around performance p^* can be expressed as another set S_p, centered on any value p with a 2R-sized support (Figure 4a). This set can also follow any shape depending on the user's risk aversion. With a risk-averse attitude, a crisp set defines here R assuming a uniform possibility distribution. But an actual distribution of R around the approximation could also be used.~~

~~The modified membership degree $\mu_R(p^*)$ should account for the 2R-large interval that represents the possibility domain around p^*. So at any given p^*, the possible acceptability values are represented by the intersection between the sets S_p and A_μ, given by the MIN operator. The resulting value $\mu_R(p^*)$ is the average over this intersection.~~

~~The new membership function μ_R over the entire domain of performance is then the moving average of μ with 2R window size (Figure 4b). A single possibility surface is thus obtained for any (x_1, x_2) coordinate (eq. 11).~~

$$\mu_R(x_1, x_2) = \mu_R(p^*) = \frac{1}{2R} \int_{p^*-R}^{p^*+R} \mu \; (p^*) \, dp^*$$

2.2.2 ~~Convex hulls as range of success and failure~~

~~The climate stress test seeks to identify accepted and rejected sub-spaces A and D within the exposure space. As seen in section 2.1.1, gridded sampling can result in risk under-estimation. With an uncertain, noisy response function and a non-gridded sampling of the exposure space, the sets A and D of accepted and rejected points do not form two cohesive, identifiable and mutually exclusive sub-spaces. The methods described in sections 2.2.1 and 2.2.2 are regression or surface fits that incorporate the remaining errors but are still approximations and might not represent all possibilities of success or failure. In a risk-averse approach, decision-relevant outliers could also be considered, in order to prepare for the most unlikely, but possible failures. The question is which performance values should be attributed at any location between sampled points.~~

~~One simple way to conservatively identify a sub-space from a set of points is their convex hull, the smallest possible convex space that contains the set. Convex hulls are extensively used in point process analysis and notably decision theory and risk analysis (Harris 1971). They can be used to identify failure regions when a response surface is inadequate, e.g. in mechanical engineering (Missoum et al., 2007).~~

~~Concept for largest convex ranges of success and failure~~

~~The underlying assumption is that, for any triangle of points contained in a set, any point within the triangle also belongs to the set. Following further a possibility-centric approach, what is sought here is the largest convex range of failure (LCRF). While less impactful for a risk-averse decision process, the largest range of success can also be expressed to further differentiate the regions of the response function. With a deterministic response function a single threshold will discriminate a space between two complementary sub-spaces, accepted and rejected. With a noisy response and a crisp target, both sub-spaces will overlap, creating a transition zone (Fig. 5).~~

~~Considering a fuzzy performance target we modify the definition of both accepted and rejected sub-spaces. The loosely accepted sub-space is the convex hull of all performance values *superior to the lower threshold*. The loosely rejected sub-space is the convex hull of all performance values *inferior to the higher threshold*.~~

~~Said otherwise, because only the upper bound of possibility is sought for success and failure, the largest range of failure (resp. success) is simply defined by *the smallest convex set of points where p belongs to the smallest (resp. largest) α-cut of A_μ also called core (resp. support)*.~~

~~This method gives more weight to outliers, as they define the convex hull. It is a simple measure of possibility, and does not discriminate points within the transition zone. Different management rules are compared according to the relative position and overlap areas of their respective transition zones.~~

~~As decision-centric methods rely on large number of simulations, computing power parsimony is an applicability concern (e.g. Whateley et al., 2016, Zatarain Salazar et al., 2017). The LRCF is only defined by its vertices, and only the border closer to the success region matters for decision purposes. Large parts of the response surface could potentially be ignored, saving computation time. An iterative sampling of the response function can thus complement the LCRF method.~~

3 Application

A ~~reservoir~~ two-reservoir system in eastern Canada is used as a case study to illustrate the applicability of the possibilistic response surfaces.

3.1 Upper Saint-François River Basin features

The Upper Saint-François River Basin (USFRB) is located in the province of ~~Quebec~~Québec, Canada. The selected gauging point, near the agglomeration of Weedon, drains an area of 2940 km^2 with an average annual flow of 2.1 billion cubic meters. The system (Fig. ~~6~~5) involves the Saint François River, controlled by two reservoirs Lake Saint-François and Lake Aylmer with a combined storage capacity of 941 million cubic meters, and the uncontrolled affluent Saumon River.

Both reservoirs are managed by the Québec Water Agency which is part of the Ministry of Environment ~~and Fight against Climate Change through the CEHQ water agency~~(Ministère de l'Environnement et de la Lutte contre les changements climatiques - MELCC). The main operational objectives are: (i) to protect the municipality of Weedon and several residential areas around the lakes from floods, (ii) to ensure minimum river discharges and water levels in the lakes to preserve aquatic ecosystems, (iii)

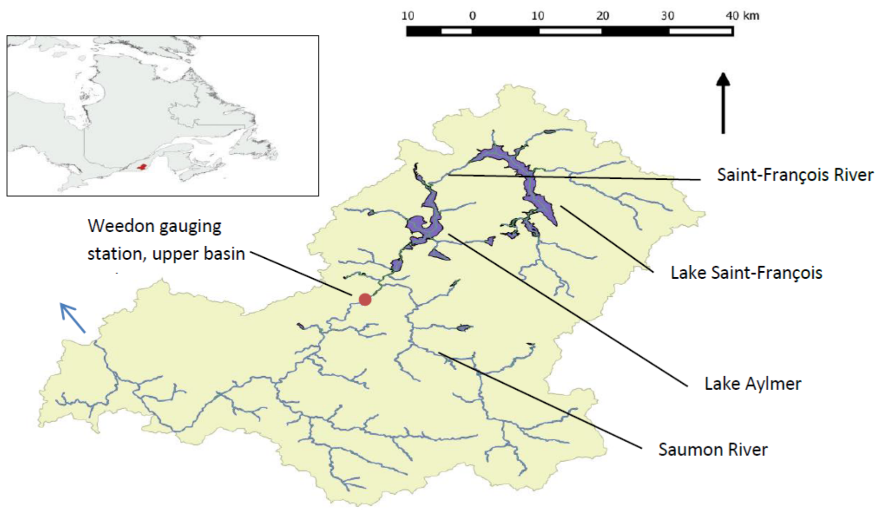


Figure 5. Layout of the Upper Saint-François River Basin, Québec, Canada.

to ~~provide the downstream run off river power station with a reliable water discharge~~regulate the floods for downstream power plants; and (iv) to maintain desired water levels in the lakes for recreational uses during the summer.

This multipurpose reservoir system thus follows a refill-drawdown cycle accordingly. With a ~~snowmelt dominated~~ snowmelt-dominated flow regime, the reservoirs are emptied in winter, filled ~~in spring and aim at a stable level in summer. Difficult snowpack volume estimation and variable precipitations can lead to under or over-estimation of the spring flood, leading to either insufficient summerlevels or frequent flooding of the downstream agglomeration of Weedon as experienced in 2018 and 2019.~~ during the spring and kept at constant pool elevation during the summer.

3.2 Inflow time series

~~Following~~ At the request of system operators we use hydrologic stressors instead of climatic ones. Other authors have also used hydrologic stressors, see e.g. Nazemi et al.~~(~~, 2013~~)~~, Borgomeo et al.~~(~~, 2015~~)~~, Herman et al.~~(~~, 2016~~)~~, Zeng et al.~~(~~, 2017~~)~~, or Nazemi et al.~~(2020), we use streamflow stressors instead of climatic ones, as the present study does not aim at differentiating between several sources of uncertainty – climate change, climate variability, run-off modelling – but proposes a method that accommodates any of them.~~

~~In Québec, the CEHQ water agency regularly produces projections of river discharges throughout the province as part of the Hydroclimatic Atlas based on climate models (see Hydro-climatic Atlas, 2015~~, ~~2017). In this study, available time series (MELCC, 2018) can be used as basis for synthetic streamflow generation,~~ 2020. Readily available inflow time series from GCM weather projections are used to generate additional synthetic streamflow series as in Vormoor et al., 2017. Results are then directly plotted on ~~a response function~~ the exposure space according to their own (x_1,x_2) coordinates~~, filling the response function without following a gridded sampling. Besides avoiding the spatial simplification entailed by gridded sampling (section 2.1.1.),~~

~~such a method makes~~ . Such a method seeks to make a greater use of hydro-climatic future scenarios when many are already available, ~~providing an already high diversity of~~ to obtain a higher diversity of synthetic times series (based on different GCM simulations, RCP scenarios and downscaling techniques)~~, which can then be expanded with additional synthetic generation as in Vormoor et al. , 2017.~~ . We first describe the initially available time series, then how they are perturbed and reused to create synthetic time series.

Historical daily measurements are available for the 2000-2014 period (MELCC, 2018). They include lakes inflows, levels and reservoir releases, and river discharges from the tributary and at the basin outlet.

Streamflow scenarios are ~~made available by the CEHQ~~ provided by MELCC through the Quebec Water Atlas 2015 (CEHQ, 2015, MELCC, 2018). ~~These~~ Those hydrologic projections are based on climatic projections from the Natural Resources Canada data base of GCM simulations (CMIP5, Hydro-climatic Atlas, 2015) that were downscaled by the ~~CEHQ.~~

Québec Water Agency. A set of 501 time series was made available, spanning 30 years of daily inflows. The set contains 135 scenarios for a 1971-2000 reference period; and 366 scenarios for the 2041-2070 period. The 366 scenarios are based on 122 GCM projections, from which 3 different downscaling techniques were applied: without bias correction, with quantile mapping or with delta quantile mapping (based on Mpelasoka and Chiew, 2009). In order to obtain the largest degree of variability, and find as many failure configurations as possible, all 501 time series are used indistinctively, first perturbed to increase the sample, then used as input for the synthetic time series generation. ~~The synthetic generator is the Kirsh-Nowak method (Nowak et al., 2010, Kirsh et al., 2013), made available online as Matlab® code by Quinn et al. (2017b), employed e.g. in Quinn et al., 2017a.~~

In order to expand the sample of the exposure space and explore less favorable conditions, the perturbation of ~~synthetic~~ available inflows is performed by either modifying ~~only the flow average without affecting the dispersion , or by affecting~~ the average annual flow, the dispersion of daily flows, or both. To increase the range of tested inflow volumes, a single change factor is applied in the first case, arbitrarily increasing all flow values at every time step by 50%. ~~In the second case (perturbed dispersion)~~ To perturb the dispersion, a varying factor multiplies flow values depending on their rank in the series distribution (factor 1 for the lowest, factor 1.5 for the highest flow). There are then 4 categories of perturbation: volume only, dispersion, volume and dispersion, and none. ~~Moreover each synthetic~~

This expanded set of time series is then used as input of the synthetic generator. The generator is the Kirsh-Nowak method (Nowak et al., 2010, Kirsh et al., 2013), made available online as Matlab code by Quinn et al. (2017b), employed e.g. in Quinn et al., 2017a. Each synthetic generation is performed twice for each available time series. We then get ~~$501 \times 4 \times 2 = 4008$ synthetic~~ 501 (initial tie series) $\times 4$ (different perturbations) $\times 2$ (synthetic realizations) = 4008 synthetic time series, each containing 30 years of daily river discharges.

$$G = \frac{1}{N}\left(N + 1 - 2\frac{\sum_{i=1}^{N}(N+1-i)\,q_i}{\sum_{i=1}^{N} q_i}\right) \tag{11}$$

Similarly to other stress-test studies that generate inflow instead of climate time series (Feng et al., 2017), the selected driving variables (axes x and y of the response function) are the total annual inflow volume and a measure of the intra-annual variability of streamflow. The intra-annual variability is here measured with the dispersion coefficient G, a measure also known as Gini coefficient in economics but employed too in hydrology (Masaki et al., 2014). It is similar to the ~~variation coefficient used in other studies~~ coefficient of variation used in e.g. Nazemi et al. (2020) but bound between 0 and 1, which offers convenient interpretation: at G=0 all daily discharges in a year are equal, if G=1 the entire yearly run-off happens in a single day. Like the variation coefficient it allows for a second variable statistically independent of the total annual run-off volume. Here q_i are the ordered daily discharges of a given year, N=365 days.

$$G = \frac{1}{N}\left(N + 1 - 2\frac{\sum_{i=1}^{N}(N+1-i)\,q_i}{\sum_{i=1}^{N} q_i}\right)$$

3.3 Simulation and response surface

The model is built with HEC-ResSim, the Reservoir System Simulation software developed by the US Army Corps of Engineers (Klipsch and Hurst, 2007). It relies on a network of elements representing the physical system (reservoirs, junctions, routing reaches), as well as the sets of operating rules. HEC-ResSim replicates the decision-making process applied to many actual reservoirs through a rule-based modeling of operational constraints and targets.

Hydrologic inputs consist of 30 years long, daily river discharges for each sub-basin. The main outputs are daily water levels in lakes, reservoir releases, as well as the ~~discharge~~ discharges at the outlet. A complementary Jython routine is developed in order to run HEC-ResSim in a loop to systematically load a large set of different hydro-climatic scenarios. Dam characteristics and operational rules were provided by the ~~Quebec~~ Québec Water Agency (MELCC, 2018).

The model is developed with a first set of operating rules (rule 1) expected to mimic the current operation of the system. It reproduces measured daily releases over the 2000-2014 period. 4008 simulations are then run, each taking an input of synthetic daily flow series spanning 30 years. In order to increase the density of the un-gridded exposure space sampling, results are divided in 5 years periods. Such decomposition is deemed acceptable based on the reservoir system, which storage capacity is designed for seasonal regulation, not multi-year, mitigating the effects of boundary conditions. It leads to a sample of 24’048 points, each one representing a five-year simulation. ~~Observation is independence not considered here, as the prime objective is to maximize the diversity and noise of the sample.~~

Although the operating rules were designed taking into account all operating objectives, the present study focuses on the flood control performance $\cancel{p}R$. More specifically, it is the reliability (Hashimoto et al., 1982) of the system keeping the river discharge at Weedon below 300 $\mathrm{m^3s^{-1}}$. Mathematically, if $F(t)$ is the state of flooding at time step t, then $\cancel{p}\ R$ is given by:

$$F(t) = \begin{cases} 0 \;\; if \;\; Q(t) \leq 300 \\ 1 \;\; if \;\; Q(t) > 300 \end{cases} \tag{12}$$

$$pR = 1 - \frac{1}{T}\sum_{t=1}^{T} F(t) \tag{13}$$

The response function is built by representing ~~p~~ performance R as a function of the selected inflow characteristics (yearly volume and dispersion). ~~As developed in section 2, the~~ We consider the case where the threshold between acceptable and unacceptable performance is not clearly defined, but is bounded between $\theta_1 = 0.93$ and $\theta_2 = 0.97$.

The separation of the exposure ~~subspace is first performed through a performance target θ set at $p \geq 0.95$. Alternatively the acceptability condition is defined by the fuzzy set within the bounds~~ space between acceptable and unacceptable regions is calculated following section 2, combining a logistic regression with a fuzzy acceptability domain, its support being [0.93, ~~0.97~~1] and its core [~~, considering a 0.02 tolerance~~0.97,1]. Consequently, any given performance value ~~p~~ R has a membership degree of 0 for ~~p~~ $R < 0.93$, and equal to 1 for ~~p~~ $R \geq 0.97$. ~~Two different rules sets are tested: rule 1 that replicates the current management; and an instrumental set rule 2 which slightly alters the anticipation and emergency release algorithm of the reservoirs. The current and alternative sets of operational rules are compared based on the aggregated logistic regression,~~ A counterfactual exercise is also run with a crisp threshold θ^*=0.95, where the ambiguity is ignored and only the median between bounds is selected. Performance is also calculated for a sub-set of GCM-based projections deemed more trustworthy by the ~~analytical approximation and the error-accounting membership function, or the relative position of the largest convex ranges of failure.~~ CEHQ (with quantile mapping downscaling for the 2041-2070 period), each one divided by 5-years periods.

~~Concept for an iterative sampling of the LCRF. Simulations are first run for available hydr-climatic scenarios (a), then for successive random samples of synthetic time series (b).~~

~~The LRCF method is also tried through an iterative sampling to evaluate its potential for computational parsimony. The convex range of failure is thus first calculated on downscaled time series, including raw ones without bias correction (Fig. 7a). Then each iteration expands the range by sampling pre-generated synthetic time series around its boundaries, and simulating the water system with only those. Failure points constitute the new hull for the next iteration (Fig. 7b). Not finding failure points is the exit condition. Results~~ These rules are compared to ~~the LRCF calculation on the full sample~~an instrumental set rule 2 which slightly alters the anticipation and emergency release algorithm of the reservoirs.

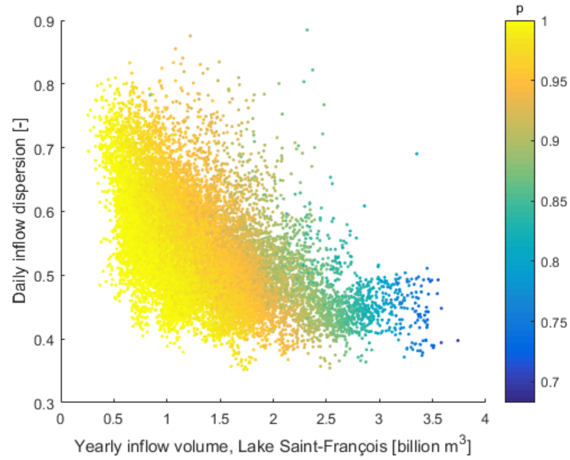
4 ~~Results~~

3.1 ~~Simulations~~

The simulation is first run with 122 of the original time series made available by the ~~CEHQ~~Québec Water Agency. These are the bias-corrected rainfall / run-off simulations considered as the most reliable for scenario-driven adaptation plans, corresponding to different radiative forcing scenarios. Taken by 5 year periods (thus 610 time series), all lead to flood control reliabilities superior to 0.97, above any considered ~~performance target~~acceptability threshold. So both rule sets are considered successful in all these time series.

~~Response surface (rule 1). Performance p : flood control reliability~~

Simulations are then run for the much larger ~~, and diverse , un-gridded~~ and diverse sample of 4008 synthetic time series. The performance R, measured as the reliability of flood control, is evaluated for each 5-years period contained in the 4008 simulations of 30 years (24'048 evaluations)~~, for each of the two operation rules~~. The color scale represents the performance



~~Logistic regression: probability of success π for crisp target 0.95~~

Figure 6. Response surface (rule 1). Performance R : flood control reliability

(reliability R against floods) in Fig. ~~8~~ 6 for each 5 year time series. ~~Describing variables are~~ The axes are the stressors x_1, x_2: the average annual inflow volume at Lake Saint-François ~~as x-axis~~ and the dispersion ~~of daily inflows~~ (or Gini coefficient) ~~as y-axis. The large sample size and lack of gridded sampling reveal the noise of the response~~ of the daily inflows. The response shows considerable noise, although a north-east / south-west anisotropy or gradient can be visually noticed.

3.1 ~~Logistic surface aggregated over fuzzy target~~

~~The logistic regression is first performed with the response surface converted into crisp binary outcomes. Success is defined by p $\geq$ 0.95, failure by p $<$ 0.95. The logistic surface provides the probability of success π at any coordinates (Fig. 9). Depending on the risk attitude of stakeholders or decision-maker, the surface can be divided in success and failure regions for specific probabilities of success (π-cuts, eq. 9), like 0.95 or 0.99 as shown on the figure. The approximation was done with the Matlabs® function *mnrfit*. The deviance of the fit – an expression of likelihood of the model, 0 being a perfect model – is 0.0085.~~

~~As in Kim et al.(2019) probability cuts can be contrasted with the projection of downscaled time series from GCMs on the response surface (Fig. 10). A way to compare the two management rules, besides the relative position of their π-cuts, is the number of projections falling out the π-cuts. For rule 1, no GCM projection – taken as 5 year portions - falls out of the 0.99 π-cut, i. e. the space in which π(p $\geq$ 0.95) $\geq$ 0.99. For rule 2, 17 projections fall out of the 0.99 π-cut, one of them falling too out of the 0.95 π-cut.~~

4 Results

~~Compared logistic regressions, rules 1 and 2, vs GCM hydroclimatic scenarios. Crisp target : p $\geq$ 0.95~~

~~While all these downscaled time series lead to successful performances, showing reliability values above 0.97 for any rule, their coordinates, thus their corresponding driving variables, can still fall outside of a π-cut. With rule 2, a scenario sharing the same properties x_1, x_2 - yearly inflow volume and daily inflow dispersion – with a successful GCM projection could still lead to failure.~~

~~Compared logistic surfaces, rule 1 (a) 0.95 crisp target (b) 0.93 0.97linear fuzzy target (c) 0.93 0.97sigmoid fuzzy target~~

~~The fuzzy performance target is then integrated to the calculation. Success is not anymore a set defined with the satisfaction of a crisp target of 0.95 for flood reliability, but with a fuzzy target between 0.93~~ An acceptability value μ is then associated to each dot (time series) in the sample depending on the value of performance (reliability) R (Fig. 7.a). The acceptability value μ is the membership degree of R to the fuzzy set of acceptable outcomes, with [0.93, 1] as support and [0.97, 1] as core (as in Fig. 2). The sample of simulations thus leads to acceptability values between 0 and ~~0.97. Reliability values above 0.97 are considered full success, and below 0.93 full failure. In between, the membership function can be either linear or sigmoid. For both cases, and for both rule sets, the~~ 1 in Fig. 7a.

To solve the problem of combining two uncertainties that are different in nature (probability of to meet a threshold vs possibility that this threshold is acceptable), the aggregated logistic regression presented in section 2.2.1. is performed for the fuzzy outcomes, thus proposing a continuous mapping for a case where the outcomes are not available as binary categories. The logistic regression is performed 10 times for 10 α-cuts corresponding to a uniform sampling of α-levels~~(see section 2.2.1.).~~. The aggregated logistic regression at every coordinate is the average of the 10 logistic regressions, each one considering a single α-cut as the crisp set over ~~p that defines successful outcomes. Figure 11 compares the crisp logistic regression and the aggregations over linear and sigmoid fuzzy targets. The averaging effect can be noted in the wider transition zone, which becomes steeper at its center when applying a sigmoid fuzzy target.~~ R that defines acceptable outcomes.

~~Both rule sets can be compared through their aggregated possibility (~~It provides at each coordinate of the exposure space (or state of the world) a possibility value Π ~~) cuts, e.g. at 0.95 and 0.99 (fig 12) . With rule 1 applied, zero GCM projections fall outside of the 0.95 possibility cut, and 8 fall out of the 0.99 cut. With rule 2, 6 projections fall out of the 0.95 cut, and 46 out of the 0.99 cut.~~

~~compared possibility cuts (Π-cuts, rules 1 and 2) vs GCM hydro-climatic projections, sigmoid fuzzy target~~

4.1 ~~Bivariate surface approximation~~

~~The second method consists in computing analytical functions, one for each rule, to fit the available sample, in this case with a bivariate quadratic approximation~~ of the outcome (reliability against floods) being deemed as acceptable given the realization of the state of the world. This – conditional – possibility measure expresses both the ambiguity of the acceptability criterion, and the probability of an acceptable outcome at any location on the response surface. The surface can be divided in acceptable and unacceptable regions (Fig. ~~13) . The resulting error R = 0.03 (selected here as the 95% quantile in the error distribution)is used to modify the membership function μ of the success fuzzy set A_μ, with a moving average with 2R-sized window (section 2.2.2.)~~7) based on any desired level of possibility (Π-cut).

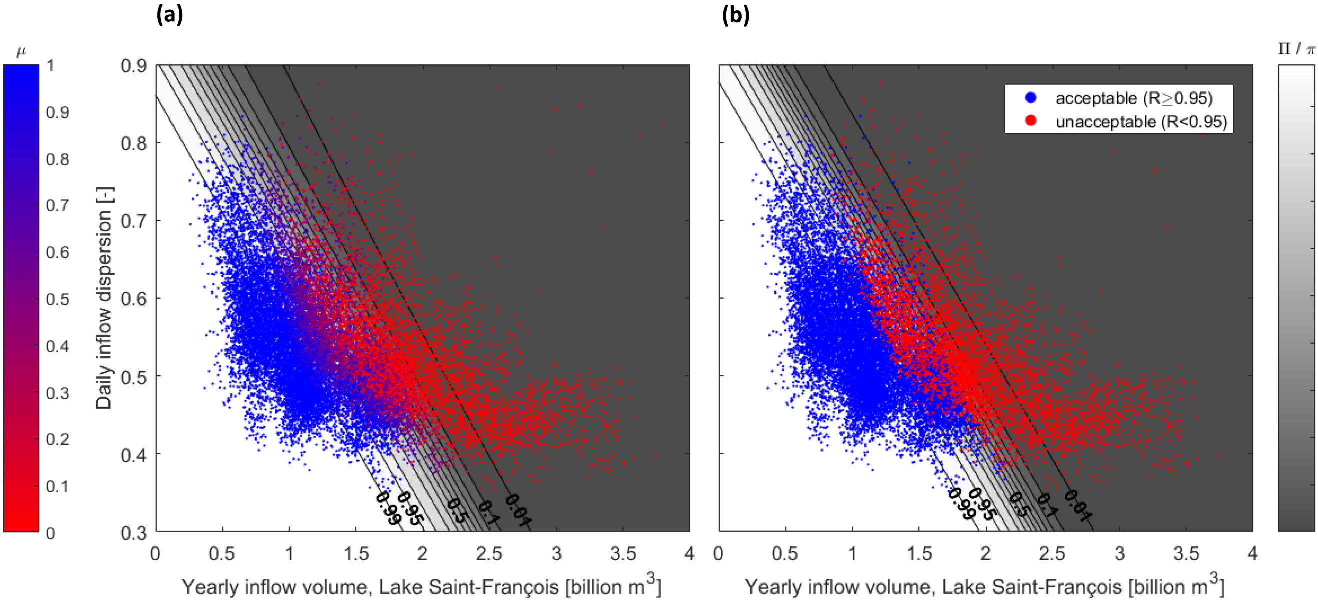


Figure 7. ~~quadratic approximation~~ Acceptability of ~~the response surface.~~sampled outcomes and logistic regressions (a) Fuzzy outcomes and possibility of acceptable performance Π (b) Binary outcomes and probability of acceptable performance π

As a counterfactual, we also compute a an alternative where the ambiguity of the threshold is ignored with the response surface converted into binary outcomes – acceptable or unacceptable frequency of flooding – based on a median crisp threshold of 0.95 (Fig. 7b). A simple logistic regression is performed for the counterfactual binary outcomes, leading to a probability π of acceptable outcome.

~~With an explicit, deterministic function and a modified membership function μ_R, the membership degree to the fuzzy set $A_{\mu R}$ allows to map success (white, membership degree 1) and failure (black, membership degree 0) sub-spaces with a continuous transition (Fig. 14a).~~ The approximation was done with the Matlabs® function *mnrfit*. The McFadden pseudo R^2 of the median threshold logistic regression is 0.7531. The relation between explanatory variables is kept linear, as introducing an interaction term only increased the pseudo R^2 to 0.7562.

~~GCM time series , again, while by themselves showing fully accepted performances (membership degree equal to 1), are partially located in the 9^{th} decile of the response surface .~~

~~Comparison between operation rules seems however less conclusive in this case~~ Π-cuts producing frontiers between acceptability regions can be contrasted with the mapping of time series from GCM projections on the response surface (Fig. ~~14b), at least around the GCM-based projections. No projection falls out of the 0.95 Π-cut~~8). ~~4 projections fail to meet the 0.99 Π-cut for rule 1, 6 for rule 2 (rule 1 thus being slightly better). For rule 1, 44 projections do not reach an aggregated membership of 1, 40 for rule 2, swapping positions in this case. It can be noted that, while both rules are similar in the vicinity of GCM-based projections, rule 1 performs better in the low inflow, high dispersion zone, and rule 2 in the high inflow, low dispersion. The~~

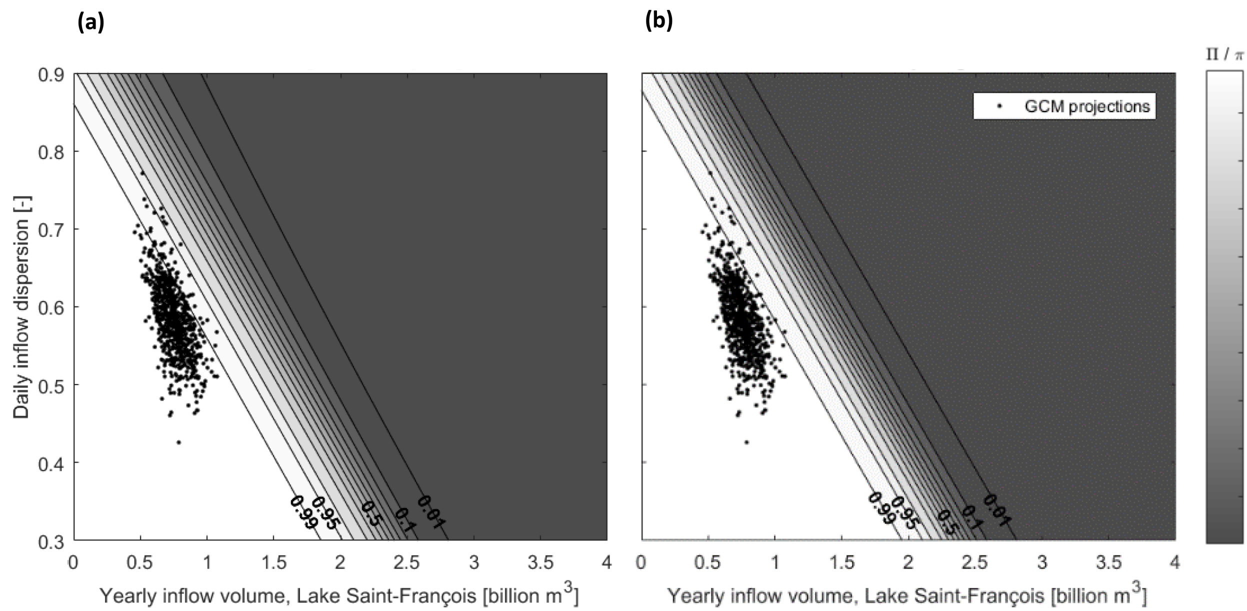


Figure 8. ~~Possibilistic surface, quadratic approximation~~ Logistic regressions and GCM projections (a) ~~Π-cuts, rule 1~~ for fuzzy outcomes (b) ~~compared Π-cuts vs GCM hydro-climatic projections~~for binary outcomes

~~method allows however for a non-linear relation between~~ While all these downscaled time series lead to fully acceptable performances, showing reliability values above 0.97, their coordinates, thus their corresponding state of the world, can still fall below a Π-cut.

This is because for any of these projections, the evaluated sequence is one realization of those conditions x_1, x_2 ~~variables and performance, as opposed to~~. Assuming the logistic regression ~~. Dispersion has a varying effect on performance depending on total inflow, with an increasing slope as yearly inflow grows~~model is accurate, with possibility 1-Π, alternative realizations of those conditions may not be seen as satisfactory. A scenario sharing the same properties x_1, x_2 - yearly inflow volume and daily inflow dispersion – with a satisfactory GCM projection could still lead to unacceptable frequency of flooding if its possibility of acceptable outcome Π is below the acceptable level.

4.1 ~~Largest convex range of failure (LCRF)~~

~~The third approach to identify accepted and rejected sub-spaces for each management rule is the comparison of respective convex hulls, each hull representing a possibility sub-space of success/failure of the system respective to flood reliability. It heavily relies on outliers and thus represents an upper possibility bound, here called the maximum convex range of success or failure (LCRF). It answers the question "where can the system possibly fail or succeed" given a sample of points. Figure 15a shows the maximum range of success (dashed line) for both rule 1 (blue)and rule 2 (red) , that includes all values with flood reliability superior or equal to~~ Respectively, the binary counterfactual model (Fig. 8b), provides a degree of probability π of an unacceptable outcome for the same state of the ~~acceptability threshold, 0.95, here considered as a crisp value. The solid line hulls represent the maximum range of failure, containing all values with flood reliability inferior to 0.95. The overlap between maximum ranges is a transition zone. With~~ world, as in the previous studies using the logistic regression. Fig 8. illustrates the straightforward difference between adapting to the ambiguity of the acceptability criterion (Fig. 8a) and ignoring it (Fig. 8b). For any given state of the world (coordinate) x_1, x_2, the aggregated logistic regression not only considers the probability of a realization leading to a certain performance, but also the possibility that such performance, the reliability against floods, would be considered an acceptable outcome. Accepting a fuzzy acceptability criterion thus mechanically widens the range of the continuous transition resulting from the logistic regressions. A state of the world with a ~~similar range of failure and a larger range of success, the rule set 2 (red) would this time be considered as superior to the rule set 1.~~ 100% probability of meeting a 0.95 reliability threshold might still have a possibility of this threshold not being *actually* accepted.

~~Compared largest convex ranges of success (LCRS) and failure (LCRF) vs GCM hydro-climatic projections (a) for crisp target 0.95 (b) for fuzzy target 0.93-0.97.~~

~~In Fig. 15b, sub-spaces are defined with a fuzzy performance target, with a +/- 0.02 tolerance and bounds defined as 0.93, 0.97in section 2.1.2. The largest range of success now accepts candidate values that are partially accepted (partial successes, $p \geq 0.93$) , so it considers success as the largest α-cut of A_{μ}. The shape of~~ Such differences are more noticeable in this case study when using GCM projections as ex-post weights. With a fuzzy target, 46 projections (6.3%) fall out of the ~~membership function has no influence here. Respectively the largest convex range of failure considers as success the smallest α-cut of A_{μ}, it now accepts candidate values considered as partial failures of the system ($p<0.97$)~~ 0.99 Π-cut, i.e. the sub-space where the possibility of acceptable outcome is at least 0.99. Said otherwise, there is a possibility of at least 0.01 that a realization leading to the same state of the world (coordinates) would produce an flooding pattern considered as unacceptable, according to the aggregated logistic regression. With a crisp simplification, thus with less information, 17 projections (2.3%) fall out of the 0.99 π-cut. There is a probability of at least 0.01 that a realization leading to the same state of the world would produce an unacceptable outcome.

~~Figure 15 also highlights (black dots)~~These frontiers are specific to a reservoir operation rule. When using a stress-test with a response surface, alternative rules are compared based on the relative position of the ~~position of the downscaled, bias-corrected hydro-climatic scenarios based on GCM projections. Again in this case, even if none of such time series is considered as even partial failure (flood reliability $p>0.97$ for all of them), with a fuzzy definition of target performance some of these time series can fall within a maximum range of failure~~frontiers between acceptable and unacceptable regions. Two of them are here compared: an approximation of current reservoir operations (rule 1) and an alternative, instrumental set of rules (rule 2). Fig. 9 compares the two rules based on selected Π-cuts (Fig. 9a). The counterfactual calculation with binary outcomes and π-cuts is also provided (Fig. 9b).

~~Figure 15b shows a varying relative performance of both rules depending~~ Fig. 9 shows a situation partially similar to the theoretical situation of Fig. 2, where the relative advantage of each rule depends on the location ~~in~~ on the exposure space ~~. Rule 1 has a larger maximum range of success. Maximum ranges of failure for both rule sets sometimes switch their relative position: rule~~ and the preferred level of possibility. With a fuzzy acceptability criterion (Fig. 9a), rule 2 ~~usually performs worse,~~

~~but not systematically. Here again the GCM projections allow for an additional weighting, pointing at the priority region to analyze. In this case rule 1 shows a better performance in the vicinity of climate projections. 60 segments of climate projections out of 610 (9.8%)fall within the maximum space of failure with rule~~ is preferred to rule 1 for high possibility of acceptable outcome ($\Pi \geq 0.95$), because the region defined by this frontier for rule 2 ~~while it is only the case for 12 of them with rule 1 (~~is larger than for rule 1. It means that rule 2 ~~%) .~~ leads to acceptable outcomes in a larger range of states of the world than rule 1. However, for low possibilities of acceptable outcome (below 0.05), the comparison depends on the stressors x_1, x_2. Rule 2 is preferred for very high daily inflow dispersion (or Gini coefficient, y axis) but moderate yearly inflow (x axis), while rule 1 is preferred for low dispersion and very high yearly inflows (again assuming the logistic regression model is accurate).

~~Given that a limited number of outliers define the maximum ranges, an ex-post sampling algorithm is tested to see if the number of simulations can be reduced and reach similar results. Only the range of failure is tried here. Starting from the original 3006 time series from 501 available run-off simulations (including those based on raw downscaled rainfall models without bias correction), the convex range of failure is calculated then "stretched" by iterative sampling. At each iteration, a new random sample of 1000 synthetic time series is selected within a normalized radius of 0.5 from both the previous hull and the bias-corrected scenarios. New failure points are added to the previous set of failures and constitute the new range of failure, other points are discarded. The number of bias-corrected scenarios included in the final hull is an evaluation of the quality of the algorithm; the closer to the result obtained with the full sample of synthetic time series the better.~~ Using a counterfactual with binary outcomes (Fig 9b), and thus frontiers defined only by probabilities, modifies the above results. While rule 2 remains preferable for high probabilities of acceptable outcomes, it becomes worse than rule 1 for low probability-cuts, this time independently of the location on the exposure space.

~~Iterative sampling of the LCRF, rule 1.~~

~~Figure 16 shows an example of how the range of failure grows at each iteration with rule 1. Three iterations are enough in this caseto reach 11 GCM scenarios within the range of failure. It is the same number as obtained with the full sample (Fig. 15b) so it is the best possible result for the algorithm. The number of required 5-year simulations of the water system is 6006, 3006 for the first guess plus 3000 from the iterations, compared to~~ If the decision-makers choose to use GCM projections as ex-post weights, the preference order for low possibility levels becomes less important. It narrows the relevance of the exposure space to the vicinity of the ~~24048 from the full synthetic sample. Figure 17 shows a sensitivity analysis with 100 runs for both rules. While a majority of runs reach a number of GCM scenarios within failure range that is close to the full sample (10-11 for rule 1~~projections, thus to high possibilities of acceptable outcomes. In this case, ~~59-60~~ 8 scenarios for rule 2 (1.1%) fall below the 0.99 Π-cut (meaning other realizations have a 0.01 possibility of unacceptable outcome), against 46 scenarios for rule 1 (6.3%). Again, all GCM-based scenario lead to fully acceptable outcomes ($R \geq 0.97$). Rule 2 ~~), a considerable number of runs falls short, although still far from switching performance between the two rules. Numbers on top of columns show the range of required iterations, between 3 and 9 overall. The iterative process can thus divide the overall simulation time by 2 to 4.~~

~~Sensitivity test of the iterative LCRF: normalized search radius 0.05, 1000 simulations by iteration. Numbers on columns give the range of required iterations. (a) rule 1 (b) rule 2.~~

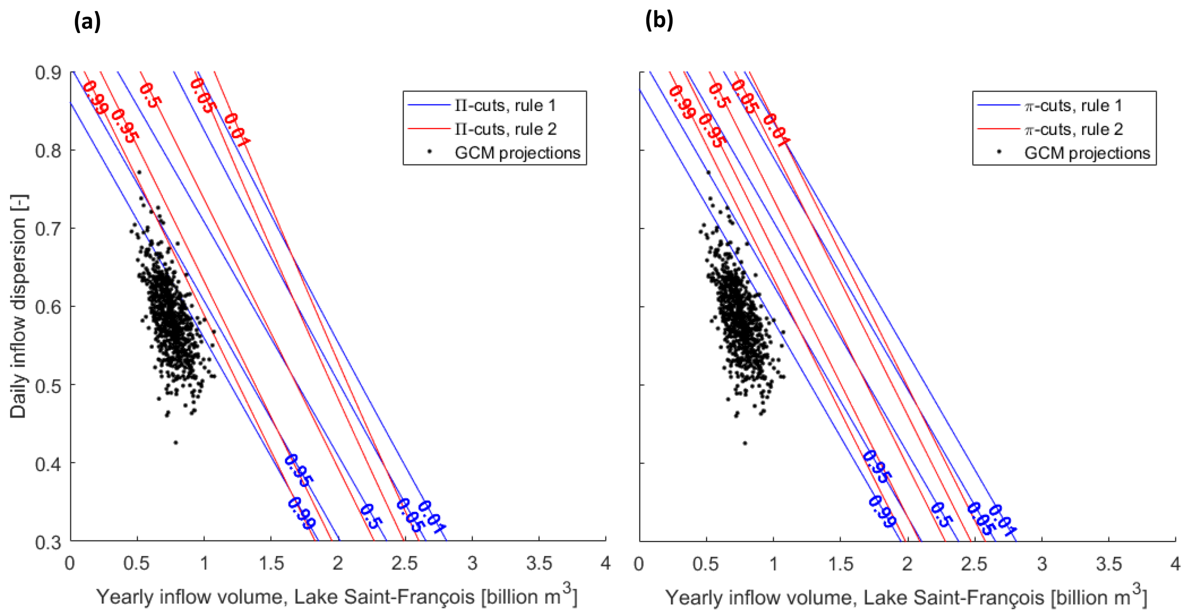


Figure 9. ~~Logistic possibility cuts~~Compared logistic regressions, ~~sigmoid fuzzy target vs LCRF~~rules 1 and 2, and GCM projections. (a) ~~rule 1~~ for fuzzy outcomes (b) ~~rule 2.~~for binary outcomes

~~Finally, Fig. 18 combines the largest convex ranges of failure with the fuzzy logistic surfaces. As the logistic regression is itself a sigmoid approximation that cannot take values of 0 or 1,~~ would then be preferred to rule 1, but there would still be a possibility of unacceptable outcome superior to 0.01 with this rule, for the same states of the ~~LCRF can be a complement to remind the actual position of the point of failure that is~~ world (coordinates x_1, x_2) sampled by the ~~farthest on the gradient line. An empirical distribution projected on the gradient line of the logistic regression would reach a value of 1 at that point~~GCM-based scenarios.

Using only binary outcomes, thus only probability-based frontiers, produces a slightly different result. Rule 2 is not only preferred in the vicinity of the GCM projections, but also no such projection falls below the 0.99 π-cut. The probability of unacceptable outcome is thus less than 0.01 at the vicinity of any GCM projection. Based on such projections, rule 2 would be adopted with less reservations with a binary model than with a fuzzy model.

5 Discussion

By itself, ~~the~~ a stress-test approach ~~is~~ can be seen as a departure from a probabilistic framework towards a possibilistic one. ~~It asks what situations lead to a system failure~~The stress-test of a water system, through a response surface, asks which states of the world could possibly lead to an unacceptable outcome, instead of evaluating the system ~~for the most probable future.~~

~~Since response surfaces are not deterministic, further information of irreducible uncertainty must be incorporated through e.g. the use of logistic regression (Kim et al., 2019).~~ performance with a given probability.

In this paper, we ~~further~~ consider that the threshold employed to define ~~success might be itself~~ acceptable outcomes might be ambiguous or contentious. The fuzzy or possibilistic framework (Zadeh 1965, 1978; Dubois and Prade, 1988), often used in decision-making analysis, provides the analytical tools to incorporate an uncertainty that is not probabilistic~~in nature~~, the ambiguity of a decision ~~target~~threshold, within the increasingly popular stress-test tool~~that itself seeks to depart from probabilistic approaches~~.

Applying a fuzzy ~~target~~ threshold would be straightforward for a deterministic response surface, each performance value on the exposure space being mapped to a degree of ~~success~~ acceptability between 0 and 1. This study explores how to ~~combine~~ solve the problem of combining a fuzzy definition of ~~success or failure~~ acceptability with the remaining hydro-climatic uncertainty of the response surface ~~, and compares different methods and interpretations.~~

~~As a first option, the~~ itself. The two sources of uncertainty are different in nature: one applies to the performance of the system, the other to the qualification of this performance as acceptable or not. To integrate them in a same response surface, the methodology relies on the concept of alpha-cuts to produce an aggregated logistic regression ~~measures within a single possibility value the probabilistic information of the regressions and the fuzzy definition of the performance target. The shape of the membership function also affects how the α-cuts are sampled, thus allowing for different interpretations of ambiguity and decision theories. A linear membership function translates a form of neutrality towards marginal gains or losses within the fuzzy boundary. A sigmoid shape gives more weight to the median α-cut , corresponding to a degree of success of 0.5, and diminishing marginal improvement or loss the further the α-cut is from the median, which can be thought as based on prospect theory (Kahneman and Tversky, 1979). The relative performance of the compared rules, however, is not here altered by the inclusion of the fuzzy target, so the resulting decision is not affected. It still might be~~ from a membership function.

The case study of the Upper Saint-François reservoir system illustrates the implementation of the aggregated logistic regression and the ~~case when the response surfaces have different slopes and gradient directions depending on the tested alternative~~conceptual framework behind it. Just like for previous uses of logistic regression, in the present method the response surface does not show a single frontier that divides the exposure space between acceptable and unacceptable flooding outcomes, but rather a parametric frontier depending, in this case, on a desired level of possibility. While possibility and probability levels cannot be directly compared (the first comprises the latter), their difference is illustrated by the wider spread of the transition zone in the response function. This wider spread is to be expected as more sources of ambiguity are considered in the possibilistic approach, and a consequence can be that GCM projections may fall below an acceptability frontier when they do not for a probabilistic logistic regression.

~~The analytic approximation by a quadratic function similarly proposes a continuous measure of possibility over the exposure space, but has the advantage of identifying non-linearity between describing variables, which the logistic regression cannot. Possible drawback are that , while the logistic regression considers equally all results that fall out of~~ Although the main goal of this study is to propose a practical adaptation to a stakeholder-driven constraint (the absence of a clear threshold), results also explore the effect that threshold ambiguity can have on final decisions. Compared response surfaces show that ignoring the ambiguity of a criterion can alter the comparison between options, either based on the size of acceptable domains or based on the position of GCM projections in the response surface. While applying specifically for a fuzzy approach with varying degrees of acceptability, this type of result is comparable to more general sensitivity studies over binary thresholds as in Hadjimichael et al. (2020). Not accounting for the ~~fuzzy target, extreme performance values here shape the fitted surface and might have an influence that they do not have in the decision process. And of course, the sample might be unadapted to any fitting attempt. As for comparing options, the method remains less conclusive, in the vicinity of GCM projections both rules cannot be easily sorted out. Importantly, it confirms however the possibility of diverging slopes and directions, as the preference between the two rules can switch~~ criterion ambiguity may thus lead, in some cases and for some actors, to worse perceived floods with the selected option than with the discarded option.

Results show that the preference between options can change depending on the ~~position in the exposure space, and therefore that considering fuzzy targets could very well alter the preference between rules. Varying preference depending on the exposure space is also a case for adaptive management.~~

~~The largest convex ranges of success or failure provide an upper possibility bound and thus easily integrate the fuzzy target by either maximizing or minimizing the α-cut of the fuzzy set of success. The shape of the membership function has no impact here, only its bounds. An advantage of the method is its consistency: if the whole philosophy of a stress test is asking where the system can *possibly* fail, then it is good to look and to prepare for the least probable cases of failure, those in a region where success is *almost* guaranteed~~possibility level (a feature that that may also be found the probabilistic-only logistic regressions). When it happens, selecting the appropriate level is highly consequential and depends on the involved actors. Previous studies have linked this choice to degrees of risk aversion (Quinn et al., 2018). In the present case, the choice of the right possibility level also depends on the qualification of what degree of performance is acceptable. The ambiguity can expresses that some might consider 0.93 as acceptable, while others would only settle for 0.97. The trade-off is thus between selecting the most robust rule for different degrees of performance. Analogous to risk aversion, this trade-off can instead be linked to loss aversion as developed by Kahneman and Tversky (1979) in prospect theory. While this study focuses on the practical integration of ambiguity as a real-world constraint, further theoretical research should focus on linking risk and loss attitudes to hydro-climatic response functions.

~~However,~~ The loss aversion function would be a useful concept to shape the membership function, which affects how the α-cuts are sampled, and thus the position of boundaries between regions in the response surface. When maximizing the range of states of the ~~convex hulls are obviously highly reliant on the generation of synthetic scenarios, which has its intrinsic randomness. The largest convex range of failure focuses on a very limited number of vertices as hydro-climatic situations of interest, the least probable, but still possible, configurations of failure; but this limited set also entails a strong sensitivity to a specific realization, with a specific generator. More generally, the impact of the choice of synthetic generator is a growing concern (Nazemi~~ world that meet a performance criterion, selecting either a higher threshold or a lower threshold would correspond to different shapes of loss aversion functions, i.e. the weights attributed to a smaller or larger loss. A centered sigmoid shape gives more weight to the median α-cut, corresponding to a degree of success of 0.5, and diminishing marginal improvement or loss the further the α-cut is from the median, which is characteristic of neutral loss aversion functions, but loss-averse actors might express a more asymetrical membership function. Other studies have linked membership functions for fuzzy sets and prospect theory (e.g. Liu et al., 2014, Andrade et al., 2014, Gu et al., 2020)~~and should be further studied, possibly integrating the potential differences within a possibilistic approach~~.

~~The fact that they are defined by a very small number of points could allow for much shorter simulation times with the right sampling method. However a first trial with a simple search algorithm shows for now an important spread in results for limited computational gains. Convex hulls also remain a straightforward tool in spatial or point process analysis and , like the logistic regression, are not always suited for non-linear relations between describing variables. Refinements would be needed, possibly at the cost of more degrees of freedom or assumptions if more parameters are needed, like in alpha-shapes or more advanced clustering tools. The time series that constitute the vertices of the hulls can be further perturbed, to see if the range of success or failure can be "stretched" and look for a *physical* possibility (where is it physically possible for the system to succeed or fail).~~ Defining the membership function does introduce an additional layer of complexity in the modelling process. It is ultimately up to the modeler and stakeholders to decide if it is a necessary translation of the social reality, keeping in mind how it can affect the results. The elaboration of membership functions from linguistic information is well studied in many fields (Zimmermann, 2001; Garibaldi and John, 2003; Sadollah, 2018), including in water resources management (Khazaeni et al, 2012). Future works should further explore how to elaborate adapted membership functions specific to the linguistic inputs that characterize satisfaction thresholds in the case of flood control systems, notably linking the membership function to risk and loss attitudes.

An important caveat is that the response surface relies on a specific set of realizations from a synthetic generator and a starting data set that is perturbed and re-shuffled. The choices and assumptions that lead to a realization deserve further scrutiny in future works. Un-gridded~~and un-aggregated~~ , on-the-fly sampling here allows exploring more ~~comprehensively~~ freely the variability of the response, ~~which can be more consistent with the whole stress test approach for certain systems. It also makes use of existing streamflow scenarios, but it has drawbacks. One advantage of usual stress-tests is their scenario-neutral property (Prudhomme et al., 2010): the extensive computation of the response surface only needs to be done once, further information on future conditions can be projected directly on it. Here stream-flow scenarios, based on GCM projections, were used to generate synthetic time series in order to obtain the granular, non-gridded response. This creates a link between downscaled scenarios and the response that is usually avoided. The underlying assumption is that between the number of scenarios,~~ as the focus of the study is the ~~different types of bias correction (or lack of thereof) and the imposed perturbations, the diversity of the synthetic time seriesleads to a relative independence from the initial bias of GCM simulations. Besides, the generation of synthetic time series always relies on available data on one way or another; a "scenario-neutral" generation could rely on historical observations and be skewed towards a conservative bias in face of for example, brutal climate shifts. The present generation method prioritizes sample diversity, but its assumptions could~~ diversity of outcomes for a given coordinate. The sampling should be improved to cover more evenly the exposure space, but without constraining too much the diversity of time series. The impact of the choice of a given synthetic generator, of the sample size and the perturbation process should be further examined~~, such as the lack of independence between observations and thus the applicability of the logistic regression~~. Likewise, the choice of describing variables was not the focus of the study but should be subject to an initial comparison ~~of~~

~~predicting values within~~ among a larger number of ~~predictors.~~ candidate predictors, and the quality of the logistic model should be further analysed and integrated as uncertainty.

~~Another trade-off from such synthetic generation is the uneven sampling, denser at the vicinity of available streamflow scenarios from downscaled rainfall series. The tool to generate un-gridded sampling should ideally ensure a balanced sample density over the response surface . This study used like others the streamflow scenarios from GCM projections as a prioritization tool and focuses on their vicinity, thus paying less attention to the sampling density in other areas.~~ A possibilistic framework could integrate within a response surface many more of such uncertainties when probabilities are not relevant, as done in other water management studies (El-Baroudy and Simonovic, 2004, Afshar et al., 2011, Jun et al., 2013, Qiu et al., 2018, Wang et al., 2020). One particularly suitable use of fuzzy logic should be to consider as fuzzy values the ex-post expert judgement on the possibility or likeliness of the obtained synthetic time series in a given river basin. The synthetic generator explores time series configurations but those may not always correspond to the range of outcomes expected in a watershed.

The integration of uncertainty and ambiguity quantification within the response surface tool could allow for ~~further~~ aggregation options in a multi-objective problem (like in Poff et al., 2016, Kim et al., 2019), while easily controlling its two separate components, response uncertainty and ~~target ambiguity. Other sources of uncertaintycould also be added and combined, like ambiguity about the streamflow threshold that defines a state of flooding, the goodness of fit for the approximations, or the expert judgement or trust on data quality~~threshold ambiguity.

Importantly, the response surface is here considered as a generic tool for decision-making under deep uncertainty, but it is used within more complex frameworks. Further research should also analyse how fuzzy thresholds can be inserted within a more complete set of methods, along with e.g. scenario discovery or adaptive approaches.

6 Conclusions

We explore in this study how to integrate ~~fuzzy performance targets~~ ambiguous acceptability thresholds within uncertain response surfaces in decision-centric vulnerability assessments. ~~Three methods are proposed~~ We propose a method to produce a possibilistic surface when the fuzzy threshold is applied to an uncertain surface. Aggregating logistic regressions over α-cuts ~~combines probability of success and target ambiguity~~ incorporates in a single measure ~~. Using a quadratic approximation of the response surface itself allows for non-linear relations. The largest convex ranges seek upper bounds for the possibility of success or failure. Two possible management rules are compared for the~~ the ambiguity of the acceptability threshold and the probability to meet such threshold, for a given state of the world. The method is illustrated with the Upper Saint-François reservoir system in Canada. ~~Aggregated logistic regression and largest range of failure show complementary ways to integrate fuzzy targets and differentiate failure domains, with respective advantage and limitations. For continuous approximations, fit quality could be integrated in the final uncertainty measure. The largest convex range method can be refined by further perturbation of the streamflow series on the vertices, in order to find a physical boundary to success and failure~~We show how a fuzzy threshold shapes the response surface, and how the way this ambiguity is treated can affect the vulnerability assessment.

Challenging old probabilistic assumptions, notably in a climate crisis context, brings new tools that also imply new choices and degrees of arbitrariness. How to transparently elaborate fuzzy ~~targets~~ thresholds jointly with stakeholders, or the choice of a synthetic scenario generator, are necessary research continuations. The presented approach enables further work on stakeholder attitudes, multi-objective problems and aggregation choices. The framework here introduced to solve a practical challenge can be consolidated from a more theoretical perspective, from both possibility theory and decision making under deep uncertainty.

Code and data availability. The data can be provided upon authorization from the MELCC, Québec, Canada (Ministère de l'Environnement et de la Lutte contre les Changements Climatiques). The codes required to reproduce the results are available upon request (thibaut.lachaut1@ulaval.ca).

Author contributions. TL and AT conceptualized the study. TL developed the methods, models and simulations, and drafted the manuscript. AT acquired the funding and provided extensive supervision.

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

Acknowledgements. The work was supported by a project from Ministère de l'Environnement et de la Lutte contre les Changements Climatiques (MELCC, Québec, Canada) entitled "Étude visant l'adaptation de la gestion des barrages du système hydrique du Haut-Saint-François à l'impact des changements climatiques dans le cadre du Plan d'action 2013-2020 sur les changements climatiques (PACC 2020)". This study does not represent the views of MELCC. The authors would like to thank Louis-Guillaume Fortin, Richard Turcotte and Julie Lafleur from the CEHQ for the fruitful discussions and the knowledge of the reservoir system, Alexandre Mercille and Xavier Faucher who contributed to earlier versions of the HEC-ResSim model, and Jean-Philippe Marceau who developed the iterative Jython routine.

References

Afshar, A., Mariño, M. A., Saadatpour, M., and Afshar, A.: Fuzzy TOPSIS Multi-Criteria Decision Analysis Applied to Karun Reservoirs System, Water Resources Management, 25, 545–563, https://doi.org/10.1007/s11269-010-9713-x, 2011.

Andrade, R. A. E., González, E., Fernández, E., and Gutiérrez, S. M.: A Fuzzy Approach to Prospect Theory, in: Soft Computing for Business Intelligence, edited by Espin, R., Pérez, R. B., Cobo, A., Marx, J., and Valdés, A. R., Studies in Computational Intelligence, pp. 45–66, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-53737-0_3, 2014.

Ben-Haim, Y.: Info-Gap Decision Theory: Decisions Under Severe Uncertainty, Elsevier, 2006.

Borgomeo, E., Farmer, C. L., and Hall, J. W.: Numerical rivers: A synthetic streamflow generator for water resources vulnerability assessments, Water Resources Research, 51, 5382–5405, https://doi.org/10.1002/2014WR016827, 2015.

Bouchon-Meunier, B., Dotoli, M., Maione, B., and Bari, D.: On The Choice Of Membership Functions In A Mamdani-Type Fuzzy Controller, 1996.

Broderick, C., Murphy, C., Wilby, R. L., Matthews, T., Prudhomme, C., and Adamson, M.: Using a scenario-neutral framework to avoid potential maladaptation to future flood risk, Water Resources Research, 55, 1079–1104, https://doi.org/10.1029/2018WR023623, 2019.

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.

Brown, C., Ghile, Y., Laverty, M., and Li, K.: Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector, Water Resources Research, 48, https://doi.org/10.1029/2011WR011212, 2012.

Brown, C., Steinschneider, S., Ray, P., Wi, S., Basdekas, L., and Yates, D.: Decision Scaling (DS): decision support for climate change, 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. 255–287, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-05252-2_12, 2019.

Castagnoli, E. and Li Calzi, M. L.: Expected utility without utility, Theory and Decision, 41, 281–301, https://doi.org/10.1007/BF00136129, 1996.

CEHQ (Centre d'Expertise Hydrique du Québec)): Atlas hydroclimatique du Québec méridional – Impact des changements climatiques sur les régimes de crue, d'étiage et d'hydraulicité à l'horizon 2050., 2015.

Chukhrova, N. and Johannssen, A.: Fuzzy regression analysis: Systematic review and bibliography, Applied Soft Computing, 84, 105 708, https://doi.org/10.1016/j.asoc.2019.105708, 2019.

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: identifying the maximum operational adaptive capacity, Water Resources Research, 52, 6751–6768, https://doi.org/10.1002/2015WR018253, 2016.

DiFrancesco, K., Gitelman, A., and Purkey, D.: Bottom-Up Assessment of Climate Risk and the Robustness of Proposed Flood Management Strategies in the American River, CA, Water, 12, 907, https://doi.org/10.3390/w12030907, 2020.

Dubois, D. and Prade, H.: Measures of possibility and fuzzy sets, in: Possibility Theory: An Approach to Computerized Processing of Uncertainty, edited by Dubois, D. and Prade, H., pp. 1–31, Springer US, Boston, MA, https://doi.org/10.1007/978-1-4684-5287-7_1, 1988.

Dubois, D., Foulloy, L., Mauris, G., and Prade, H.: Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities, Reliable Computing, 10, 273–297, https://doi.org/10.1023/B:REOM.0000032115.22510.b5, 2004.

El-Baroudy, I. and Simonovic, S. P.: Fuzzy criteria for the evaluation of water resource systems performance, Water Resources Research, 40, https://doi.org/10.1029/2003WR002828, 2004.

Feng, M., Liu, P., Guo, S., Gui, Z., Zhang, X., Zhang, W., and Xiong, L.: Identifying changing patterns of reservoir operating rules under various inflow alteration scenarios, Advances in Water Resources, 104, 23–36, https://doi.org/10.1016/j.advwatres.2017.03.003, 2017.

Garibaldi, J. and John, R.: Choosing membership functions of linguistic terms, in: The 12th IEEE International Conference on Fuzzy Systems, 2003. FUZZ '03., vol. 1, pp. 578–583 vol.1, https://doi.org/10.1109/FUZZ.2003.1209428, 2003.

Gu, J., Zheng, Y., Tian, X., and Xu, Z.: A decision-making framework based on prospect theory with probabilistic linguistic term sets, Journal of the Operational Research Society, 0, 1–10, https://doi.org/10.1080/01605682.2019.1701957, https://doi.org/10.1080/01605682.2019.1701957, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01605682.2019.1701957, 2020.

Guo, D., Westra, S., and Maier, H. R.: An inverse approach to perturb historical rainfall data for scenario-neutral climate impact studies, Journal of Hydrology, 556, 877–890, https://doi.org/10.1016/j.jhydrol.2016.03.025, 2018.

Haber Guerra, R. E., Haber, R., Alique, A., and Ros, S.: Application Of Knowledge Based Systems For Supervision And Control Of Machining Processes, 2, https://doi.org/10.1142/9789812389701_0029, 2001.

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.

Harris, B.: Mathematical models for statistical decision theory, in: Optimizing Methods in Statistics, edited by Rustagi, J. S., pp. 369–389, Academic Press, https://doi.org/10.1016/B978-0-12-604550-5.50021-3, 1971.

Hashimoto, T., Stedinger, J. R., and Loucks, D. P.: Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation, Water Resources Research, 18, 14–20, https://doi.org/10.1029/WR018i001p00014, 1982.

Herman, J. D., Zeff, H. B., Reed, P. M., and Characklis, G. W.: Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty, Water Resources Research, 50, 7692–7713, https://doi.org/10.1002/2014WR015338, 2014.

Herman, J. D., Quinn, J. D., Steinschneider, S., Giuliani, M., and Fletcher, S.: Climate adaptation as a control problem: review and perspectives on dynamic water resources planning under uncertainty, Water Resources Research, 56, e24 389, https://doi.org/10.1029/2019WR025502, 2020.

Huynh, V.-N., Nakamori, Y., Ryoke, M., and Ho, T.-B.: Decision making under uncertainty with fuzzy targets, Fuzzy Optimization and Decision Making, 6, 255–278, https://doi.org/10.1007/s10700-007-9011-0, 2007.

IPCC: AR5 Climate Change 2014: Impacts, adaptation, and vulnerability, https://www.ipcc.ch/report/ar5/wg2/, 2014.

Jun, K.-S., Chung, E.-S., Kim, Y.-G., and Kim, Y.: A fuzzy multi-criteria approach to flood risk vulnerability in South Korea by considering climate change impacts, Expert Systems with Applications, 40, 1003–1013, https://doi.org/10.1016/j.eswa.2012.08.013, 2013.

Kahneman, D. and Tversky, A.: Prospect theory: an analysis of decision under risk, Econometrica, 47, 263–291, https://doi.org/10.2307/1914185, 1979.

Kay, A. L., Crooks, S. M., and Reynard, N. S.: Using response surfaces to estimate impacts of climate change on flood peaks: assessment of uncertainty, Hydrological Processes, 28, 5273–5287, https://doi.org/10.1002/hyp.10000, 2014.

Keller, L., Rössler, O., Martius, O., and Weingartner, R.: Comparison of scenario-neutral approaches for estimation of climate change impacts on flood characteristics, Hydrological Processes, 33, 535–550, https://doi.org/10.1002/hyp.13341, 2019.

Khazaeni, G., Khanzadi, M., and Afshar, A.: Fuzzy adaptive decision making model for selection balanced risk allocation, International Journal of Project Management, 30, 511–522, https://doi.org/10.1016/j.ijproman.2011.10.003, 2012.

Kim, D., Chun, J. A., and Aikins, C. M.: An hourly-scale scenario-neutral flood risk assessment in a mesoscale catchment under climate change, Hydrological Processes, 32, 3416–3430, https://doi.org/10.1002/hyp.13273, 2018.

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.

Kirsch, B. R., Characklis, G. W., and Zeff, H. B.: Evaluating the impact of alternative hydro-climate scenarios on transfer agreements: practical improvement for generating synthetic streamflows, Journal of Water Resources Planning and Management, 139, 396–406, https://doi.org/10.1061/(ASCE)WR.1943-5452.0000287, 2013.

Klipsch, J. and Hurst, M.: HEC-ResSim: reservoir system simulation, User's manual version 3.0 CPD-82, USACE, Hydrologic Engineering Center, Davis, CA, 2007.

Knighton, J., Steinschneider, S., and Walter, M. T.: A vulnerability-based, bottom-up assessment of future riverine flood risk using a modified peaks-over-threshold approach and a physically based hydrologic model, Water Resources Research, 53, 10 043–10 064, https://doi.org/10.1002/2017WR021036, 2017.

Knighton, J. O., Tsuda, O., Elliott, R., and Walter, M. T.: Challenges to implementing bottom-up flood risk decision analysis frameworks: how strong are social networks of flooding professionals?, Hydrology and Earth System Sciences, 22, 5657–5673, https://doi.org/10.5194/hess-22-5657-2018, 2018.

Lamontagne, J. R., Reed, P. M., Marangoni, G., Keller, K., and Garner, G. G.: Robust abatement pathways to tolerable climate futures require immediate global action, Nature Climate Change, 9, 290–294, https://doi.org/10.1038/s41558-019-0426-8, 2019.

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.

Liu, Y., Fan, Z.-P., and Zhang, Y.: Risk decision analysis in emergency response: A method based on cumulative prospect theory, Computers & Operations Research, 42, 75–82, https://doi.org/10.1016/j.cor.2012.08.008, http://www.sciencedirect.com/science/article/pii/S0305054812001736, 2014.

Maier, H., Guillaume, J., van Delden, H., Riddell, G., Haasnoot, M., and Kwakkel, J.: 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.

Marchau, V. A. W. J., Walker, W. E., Bloemen, P. J. T. M., and Popper, S. W., eds.: Decision Making under Deep Uncertainty: from theory to practice, Springer International Publishing, https://doi.org/10.1007/978-3-030-05252-2, 2019.

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.

Masaki, Y., Hanasaki, N., Takahashi, K., and Hijioka, Y.: Global-scale analysis on future changes in flow regimes using Gini and Lorenz asymmetry coefficients, Water Resources Research, 50, 4054–4078, https://doi.org/10.1002/2013WR014266, 2014.

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.

MELCC (Ministère de l'Environnement et de la Lutte contre les Changements Climatiques): Données du Programme de surveillance du climat, Direction générale du suivi de l'état de l'environnement, Québec., 2018.

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.

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.

Mpelasoka, F. S. and Chiew, F. H. S.: Influence of rainfall scenario construction methods on runoff projections, Journal of Hydrometeorology, 10, 1168–1183, https://doi.org/10.1175/2009JHM1045.1, 2009.

Namdari, M., Taheri, S. M., Abadi, A., Rezaei, M., and Kalantari, N.: Possibilistic logistic regression for fuzzy categorical response data, in: 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6, https://doi.org/10.1109/FUZZ-IEEE.2013.6622457, 2013.

Nazemi, A., Wheater, H. S., Chun, K. P., and Elshorbagy, A.: A stochastic reconstruction framework for analysis of water resource system vulnerability to climate-induced changes in river flow regime, Water Resources Research, 49, 291–305, https://doi.org/10.1029/2012WR012755, 2013.

Nazemi, A., Masoud, Z., and Elmira, H.: Uncertainty in bottom-up vulnerability assessments of water supply systems due to regional streamflow generation under changing conditions, Journal of Water Resources Planning and Management, 146, https://doi.org/10.1061/(ASCE)WR.1943-5452.0001149, 2020.

Nowak, K., Prairie, J., Rajagopalan, B., and Lall, U.: A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow, Water Resources Research, 46, https://doi.org/10.1029/2009WR008530, 2010.

Pirttioja, N., Palosuo, T., Fronzek, S., Räisänen, J., Rötter, R. P., and Carter, T. R.: Using impact response surfaces to analyse the likelihood of impacts on crop yield under probabilistic climate change, Agricultural and Forest Meteorology, 264, 213–224, https://doi.org/10.1016/j.agrformet.2018.10.006, 2019.

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.

Pourahmad, S., Ayatollahi, S. M. T., Taheri, S. M., and Agahi, Z. H.: Fuzzy logistic regression based on the least squares approach with application in clinical studies, Computers & Mathematics with Applications, 62, 3353–3365, https://doi.org/10.1016/j.camwa.2011.08.050, 2011.

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.

Qiu, Q., Liu, J., Li, C., Yu, X., and Wang, Y.: The use of an integrated variable fuzzy sets in water resources management, Proceedings of the International Association of Hydrological Sciences, 379, 249–253, https://doi.org/10.5194/piahs-379-249-2018, 2018.

Quinn, J.: julianneq/Kirsch-Nowak_Streamflow_Generator, https://github.com/julianneq/Kirsch-Nowak_Streamflow_Generator, 2017.

Quinn, J. D., Reed, P. M., Giuliani, M., and Castelletti, A.: Rival framings: A framework for discovering how problem formulation uncertainties shape risk management trade-offs in water resources systems, Water Resources Research, 53, 7208–7233, https://doi.org/10.1002/2017WR020524, 2017.

Quinn, J. D., Reed, P. M., Giuliani, M., Castelletti, A., Oyler, J. W., and Nicholas, R. E.: Exploring How Changing Monsoonal Dynamics and Human Pressures Challenge Multireservoir Management for Flood Protection, Hydropower Production, and Agricultural Water Supply, Water Resources Research, 54, 4638–4662, https://doi.org/10.1029/2018WR022743, 2018.

Ray, P., Wi, S., Schwarz, A., Correa, M., He, M., and Brown, C.: Vulnerability and risk: climate change and water supply from California's Central Valley water system, Climatic Change, https://doi.org/10.1007/s10584-020-02655-z, 2020.

Sadollah, A.: Introductory Chapter: Which Membership Function is Appropriate in Fuzzy System?, Fuzzy Logic Based in Optimization Methods and Control Systems and Its Applications, https://doi.org/10.5772/intechopen.79552, 2018.

Simon, H. A.: A behavioral model of rational choice, The Quarterly Journal of Economics, 69, 99–118, https://doi.org/10.2307/1884852, 1955.

Steinschneider, S., Wi, S., and Brown, C.: The integrated effects of climate and hydrologic uncertainty on future flood risk assessments, Hydrological Processes, 29, 2823–2839, https://doi.org/10.1002/hyp.10409, 2015.

Taner, M. , Ray, P., and Brown, C.: Incorporating multidimensional probabilistic information into robustness-based water systems planning, Water Resources Research, https://doi.org/10.1029/2018WR022909, 2019.

Tilmant, A., Vanclooster, M., Duckstein, L., and Persoons, E.: Comparison of Fuzzy and Nonfuzzy Optimal Reservoir Operating Policies, Journal of Water Resources Planning and Management, 128, 390–398, https://doi.org/10.1061/(ASCE)0733-9496(2002)128:6(390), 2002.

Turner, S. W. D., Marlow, D., Ekström, M., Rhodes, B. G., Kularathna, U., and Jeffrey, P. J.: Linking climate projections to performance: A yield-based decision scaling assessment of a large urban water resources system, Water Resources Research, 50, 3553–3567, https://doi.org/10.1002/2013WR015156, 2014.

Von Neumann, J. and Morgenstern, O.: Theory of games and economic behavior, Theory of games and economic behavior, Princeton University Press, Princeton, NJ, US, 1944.

Vormoor, K., Rössler, O., Bürger, G., Bronstert, A., and Weingartner, R.: When timing matters-considering changing temporal structures in runoff response surfaces, Climatic Change, 142, 213–226, https://doi.org/10.1007/s10584-017-1940-1, 2017.

Wang, W., Zhou, H., and Guo, L.: Emergency water supply decision-making of transboundary river basin considering government– public perceived satisfaction, Journal of Intelligent & Fuzzy Systems, Preprint, 1–21, https://doi.org/10.3233/JIFS-191828, publisher: IOS Press, 2020.

Weaver, C. P., Lempert, R. J., Brown, C., Hall, J. A., Revell, D., and Sarewitz, D.: Improving the contribution of climate model information to decision making: the value and demands of robust decision frameworks, Wiley Interdisciplinary Reviews: Climate Change, 4, 39–60, https://doi.org/10.1002/wcc.202, 2013.

Whateley, S. and Brown, C.: Assessing the relative effects of emissions, climate means, and variability on large water supply systems, Geophysical Research Letters, 43, 11,329–11,338, https://doi.org/10.1002/2016GL070241, 2016.

Whateley, S., Steinschneider, S., and Brown, C.: A climate change range-based method for estimating robustness for water resources supply, Water Resources Research, 50, 8944–8961, https://doi.org/10.1002/2014WR015956, 2014.

Whateley, S., Steinschneider, S., and Brown, C.: Selecting Stochastic Climate Realizations to Efficiently Explore a Wide Range of Climate Risk to Water Resource Systems, Journal of Water Resources Planning and Management, 142, https://doi.org/10.1061/(ASCE)WR.1943-5452.0000631, 2016.

Wu, D.: Twelve considerations in choosing between Gaussian and trapezoidal membership functions in interval type-2 fuzzy logic controllers, in: 2012 IEEE International Conference on Fuzzy Systems, pp. 1–8, https://doi.org/10.1109/FUZZ-IEEE.2012.6251210, iSSN: 1098-7584, 2012.

Yu, C.-S.: A GP-AHP method for solving group decision-making fuzzy AHP problems, Computers & Operations Research, 29, 1969–2001, https://doi.org/10.1016/S0305-0548(01)00068-5, 2002.

Zadeh, L. A.: Fuzzy Sets, Information and Control, pp. 8, 338–353, 1965.

Zadeh, L. A.: Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems, 1, 3–28, https://doi.org/10.1016/0165-0114(78)90029-5, 1978.

Zadeh, L. A.: Fuzzy probabilities, Information Processing & Management, 20, 363–372, https://doi.org/10.1016/0306-4573(84)90067-0, 1984.

Zimmermann, H.-J.: Fuzzy Logic and Approximate Reasoning, in: Fuzzy Set Theory—and Its Applications, edited by Zimmermann, H.-J., pp. 141–183, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-010-0646-0_9, 2001.