Environmental Modelling and Software 159 (2023) 105560

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft

Belief-Informed Robust Decision Making (BIRDM): Assessing changes in decision robustness due to changing distributions of deep uncertainties

A. Ciullo [a,*], A. Domeneghetti [b], J.H. Kwakkel [c], K.M. De Bruijn [d], F. Klijn [c,d], A. Castellarin [b]

[a] Institute for Environmental Decisions, ETH Zurich, Switzerland
[b] University of Bologna, Department of Civil, Chemical, Environmental and Materials Engineering, Italy
[c] Delft University of Technology, Department of Technology Policy and Management, the Netherlands
[d] Deltares, Department of Flood Risk Management, the Netherlands

ARTICLE INFO

Keywords:
Robust decision making
Uncertain probabilistic information
Flood risk management planning

ABSTRACT

Robust Decision Making (RDM) is an established framework for decision making under deep uncertainty. RDM relies on the idea of *scenario neutrality*, namely that decision robustness is not affected by how scenarios are generated if these are uniformly distributed and span a sufficiently large range of future states of the world. Several authors have shown that *scenario neutrality* may not hold, but they did so by adopting either new or computationally expensive modeling. We introduce the Belief-Informed Robust Decision Making (BIRDM) framework to assess how robustness might change under an arbitrary large number of non-uniform distributions at virtually no additional costs with respect to RDM. We apply BIRDM to a flood management problem and find that alternative distributions change the robustness and ranking of measures. BIRDM allows identifying what distributions lead to these changes and under what set of distributions a measure has a specific robustness and rank.

1. Introduction

Long-term infrastructure planning requires making decisions amid climatic and socio-economic uncertainties. Many of these uncertainties are epistemic in nature as they are due to a lack of knowledge, and they might become better characterized as one acquires new information over time. However, decisions about future infrastructures are urgent and need to be made today such that chances that future infrastructures perform adequately are maximized, regardless of how exactly the future unfolds. Lempert et al. (2003) defined this challenge as a problem of decision making under deep uncertainty. That is, a situation where experts do not know or cannot agree upon the probability distributions of the uncertain factors that influence decision outcomes. In a context of deep uncertainty, decisions must be made based on their robustness, namely on their capability to perform satisfactorily under a plausible range of assumptions regarding deep uncertainties.

Various authors have proposed decision-support frameworks for decision making under deep uncertainty (DMDU) such as Robust Decision Making (Lempert et al., 2006), Many-Objective Robust Decision Making (Kasprzyk et al., 2013), Info-Gap decision theory (Ben-Haim 2006), Decision Scaling (Brown et al., 2012), and Dynamic Adaptive Policy Pathways (Haasnoot et al., 2013). These frameworks differ substantially, but they all rely on the idea of *scenario neutrality* (Quinn et al., 2020) as they generate plausible scenarios by sampling independently and uniformly over the widest range of physically plausible climatic and socio-economic states of the world. Scenarios are latter explored via factor mapping techniques to discover which ones are critical and lead to poor system performances. Robust decisions are then taken to allow maintaining satisfactory system performances if these critical scenarios manifest. It follows from *scenario neutrality* that decision robustness should be insensitive to how scenarios are sampled. However, Quinn et al. (2020) and McPhail et al. (2020) challenged this idea and showed that the selected experimental design, namely the way scenarios are constructed, does affect decision robustness.

Quinn et al. (2020) conducted a robustness study of the Upper Basin of the Colorado River by reconstructing discharge time series using four alternative information sources, namely historic data, paleo-data, future projections, and all three together. They found that the choice of which information is used for setting up the experimental design dramatically changed the range used to generate scenarios, the model parameters deemed critical for the system, and the resulting robustness values. McPhail et al. (2020) addressed a fictitious lake management problem

* Corresponding author.
E-mail address: alessio.ciullo@usys.ethz.ch (A. Ciullo).

https://doi.org/10.1016/j.envsoft.2022.105560
Received 30 November 2021; Received in revised form 31 May 2022; Accepted 17 October 2022
Available online 20 October 2022

using different scenario generation techniques. These techniques varied in the way the space of input parameters is covered during sampling and the number of considered scenarios. This resulted in different robustness values for the various decision criteria, although the final ranking of decision alternatives remained the same, at least for their case study.

Other authors (e.g. Shortridge and Zaitchik 2018; Taner et al. 2019; and Reis and Shortridge 2020) relaxed the main assumption underpinning the idea of *scenario neutrality,* namely that all scenarios are equally likely. They did so by exploring alternative approaches for when, where and how to integrate probabilistic information into DMDU frameworks. Shortridge and Zaitchik (2018) combined Robust Decision Making with a Bayesian statistical model. They first conducted a standard Robust Decision Making analysis to identify which subspaces in the model input space were critical. Next, they developed a Bayesian statistical model relying on projections from Global Climate Models (GCMs) to estimate the posterior probability of these subspaces. Taner et al. (2019) introduced a framework that integrates Decision Scaling with Bayesian Belief Networks, namely Bayesian Networks Decision Scaling. They proposed to build a Bayesian Belief Network to estimate the posterior joint probability distribution of the critical future scenarios identified through Decision Scaling. Reis and Shortridge (2020) carried out four Robust Decision Making analyses using four different probability densities for the input variables. Next, they showed that the parameters identified as critical, and their ranges, change according to the chosen distribution.

Shortridge and Zaitchik (2018), Taner et al. (2019), and Reis and Shortridge (2020) successfully integrated probabilistic information into scenario-neutral approaches. However, they do so either by resorting to additional modeling efforts (as done by Shortridge and Zaitchik (2018) or Taner et al. (2019)) or by re-running the same analysis multiple times (as done by Reis and Shortridge (2020)). This implies that either additional modeling efforts and expertise or high computational costs are required. Either undermines the practical feasibility and wider adoption of these approaches. To address this issue, we introduce the Belief-Informed Robust Decision Making (BIRDM) framework which allows incorporating probabilistic information into Robust Decision Making (RDM) without requiring any further modeling nor demanding additional computational costs. This is achieved by following the weighing method proposed by Beckman and McKay (1987), which allows exploring the influence of various choices about the distribution of inputs on the output statistics at virtually no cost compared to re-running the computer models for all desired distributions. We showcase BIRDM on a flood management problem along the Lower Po River in Northern Italy with the goal of exploring the effects of relaxing the assumption of *scenario neutrality* on decision robustness and the final ranking of measures.

The paper is structured as follows: section 2 describes the method, section 3 describes the case study and the simulation model, section 4 describes the analysis, section 5 reports results and discusses them, section 6 provides conclusions.

2. Belief-Informed Robust Decision Making

The Belief-Informed Robust Decision Making (BIRDM) framework builds on the Robust Decision Making (RDM) framework proposed by Lempert et al. (2006). Robust Decision Making aims at finding *intrinsic* system vulnerabilities which do not depend on the likelihood of the input factors revealing them. Yet, an *a posteriori* exploration of how likely these input factors might be is of high relevance for decision-making, even more so when different assumptions about these likelihoods might change decision robustness and the final ranking of measures. BIRDM expands RDM to account for this aspect.

BIRDM is presented in Fig. 1. In the figure, the standard RDM steps are contained in the dotted box, while BIRDM is represented by the full scheme. Basically, BIRDM adds an additional *Belief Analysis* step to RDM. Each step is described in detail in the following subsections.

2.1. Problem formulation

This step requires formulating the policy problem, specifying the scope, decision objectives, and possible measures. A suitable scheme to formulate a policy problem is the XLRM scheme proposed by Lempert et al. (2003). The XLRM scheme allows the analyst to clearly specify the uncertain factors (X), the policy levers (L), the system relationships, namely the system model, (R), and the performance metrics (M). In the present paper, we refer to levers (i.e., L) as measures, and to performance metrics (i.e., M) as outcomes.

2.2. Uncertainty analysis

This step involves generating scenarios and evaluating model outcomes for these scenarios. Typically, this step is conducted using a sampling strategy that maximizes the coverage of the input space, such as Latin Hypercube or low-discrepancy sequences like the Sobol

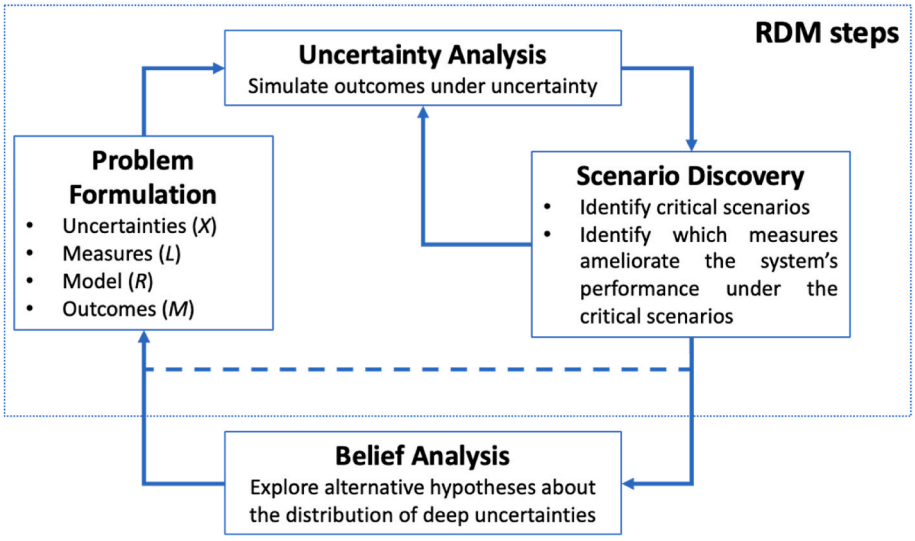

Fig. 1. Step-by-step representation of the Belief-Informed Robust Decision Making (BIRDM) framework. BIRDM builds on Robust Decision Making (RDM) (Lempert et al., 2006), whose original scheme is reported within the dotted box.

sequence (Sobol 1967). Compared to pseudo-random sequences, these sampling techniques more evenly distribute the sampled points over the domain of the input factors. Following the idea of *scenario neutrality*, the sampling of deeply uncertain factors is carried out assuming independence and a uniform distribution over the full range of physically plausible values. The outcome of this analysis is a database of model outputs which is investigated further through scenario discovery.

2.3. Scenario discovery

In this step, factor mapping techniques are used to identify regions in the input space corresponding to certain output values. Scenario discovery is typically used to identify regions in the input space that have a high concentration of failure scenarios, namely those under which system performances, y, are below a performance threshold, y^*, deemed critical for the functioning of the system.

Finding failure scenarios is instrumental to the identification of robust measures. In Robust Decision Making, robust measures are defined as those that improve system performances, y, under the failure scenarios and which therefore increase chances of meeting the critical performance threshold, y^*. It follows that in Robust Decision Making, robustness relies on the concept of *satisficing* performances (Lempert et al., 2006), namely on meeting certain performance requirements under the widest range of possible scenarios.

2.4. Belief analysis

This step explores how alternative assumptions regarding the likelihood of the failure scenarios affect the robustness of measures. We here refer to these alternative assumptions as *beliefs* that analysts, experts and decision makers may have about the likelihood of the failure scenarios. We do so by employing the weighting method proposed by Beckman and McKay (1987). The weighting method consists of replacing the original sampling distribution by an alternative one and then assessing – through reweighting – the output statistics as if the original probability distribution were used.

More formally, assuming a simulation model with input variables $\boldsymbol{x}$ and output variable y, and assuming that $\boldsymbol{x}$ has a multivariate distribution function F but it is sampled using an alternative distribution H, then the probability of y exceeding a threshold value y^* can be defined as (Diermanse et al., 2015):

$$P[y > y^*] = \frac{1}{n} \sum_{i=1}^{n} 1_{[y_i > y^*]} c_i \tag{1}$$

where n is the number of simulations, $1_{[\ldots]}$ is an indicator function which is equal to 1 if $y_i > y^*$ and 0 otherwise and c_i is a correction factor (weight) defined as follows:

$$c_i = \frac{f(\boldsymbol{x}_i)}{h(\boldsymbol{x}_i)} \tag{2}$$

with f and h being the density functions of F and H, which are called the *original* and *surrogate* distributions, respectively. In the method, the original distribution f is *unknown*, and one samples from a *known* surrogate distribution h to quantify output statistics had f been used. It is worth highlighting that the method is equivalent to the well-known Importance Sampling variance reduction technique (Tokdar and Kass 2010) but its intent is different as it reverses the use of the original and surrogate distributions. In Importance Sampling, the sampling distribution f is *known*, and one aims to find an *unknown* surrogate sampling distribution h such that more samples are drawn from important regions in the input space.

The weighting method of Beckman and McKay (1987) has been employed by different authors and for various purposes. Sparkman et al. (2016) used it to calculate global sensitivity indices from an existing sample of simulations and introduced the Importance Sampling-based Kernel Estimator to calculate the moments of the conditional distributions in the calculation of the sensitivity indices. Zhang and Shields (2018) used the method to deal with the epistemic uncertainties stemming from the inability to reliably specify a unique probability density function due to data scarcity. In such cases, they suggest to first identify an arbitrary large set of candidate probability densities through Bayesian inference. From the derived large set of densities, a single optimal density is selected which is representative of all candidate probability densities. This optimal density is the one used in the Monte Carlo analysis, and then the uncertainty in the selection of the distribution is explored by applying the weighting scheme as if any of the candidates were used. Zhang et al. (2021) expanded this framework within the context of global sensitivity analysis and derived imprecise Sobol indices (Sobol', 2001).

Following these applications, BIRDM applies the weighting method assuming a uniform surrogate distribution, h – in line with RDM and other *scenario neutral* frameworks – and evaluate robustness of all identified measures under alternative assumptions of the original distribution of deep uncertainties, f. The resulting outcome consists of as many robustness estimates for each measure as the number of alternative distributions being explored. The Belief Analysis step requires virtually no additional efforts irrespectively of the number of alternative distributions being explored, as no new Monte Carlo experiments are run.

3. The case study and the simulation model

With the aim of testing BIRDM, we apply it to support flood management planning along the lower reach of the Po River in Northern Italy (see Fig. 2). The goal is to find robust flood damage reduction measures and to assess the effect of alternative distributions regarding deep uncertainties on the ranking of measures. The considered river stretch ranges from the stream gauges of Borgoforte (upstream) to Pontelagoscuro (downstream), spanning across a total length of about 115 km and including the confluences of two main tributaries: the Secchia and Panaro rivers. The case study includes 17 levee-protected floodplains (purple areas in Fig. 2, also referred to as *compartments*) which are protected from the 0.5% annual probability flood event. Subsection 3.1 provides some background information regarding the choice of the case study and subsection 3.2 introduces the simulation model.

3.1. Rationale

The choice of the case study is driven by recent calls from the flood risk modeling and management community urging for decision support studies on systemic, large-scale, flood risk management planning (Vorogushyn et al., 2017). This systemic perspective was shown to lead to more accurate flood frequency analyses (Apel et al. 2009) and risk estimates (Courage et al., 2013; de Bruijn et al., 2016), and to ultimately widen the spectrum of flood risk management measures, potentially increasing optimality and fairness in the design of the system (Ciullo et al., 2019a; Ciullo et al., 2019b).

The BIRDM framework is used to explore beliefs about key epistemic uncertainties in dealing with embanked large-scale flood risk systems: the probability of levee failure as a function of hydraulic loads (Beven et al., 2018). This uncertainty is typically characterized by the so-called fragility curves (Bachmann et al., 2013; Curran et al., 2019). The generation of such curves, however, requires extensive knowledge of the geotechnical properties of the flood defenses which, in case of large-scale systems, may not be available or accurate at all locations of interest. Furthermore, even when assuming that fragility curves can be reliably derived, unexpected breaching can still occur and this is one of the main causes of disastrous river floods (Merz et al., 2021). In January 2014, for example, a levee failure occurred during a minor flood event

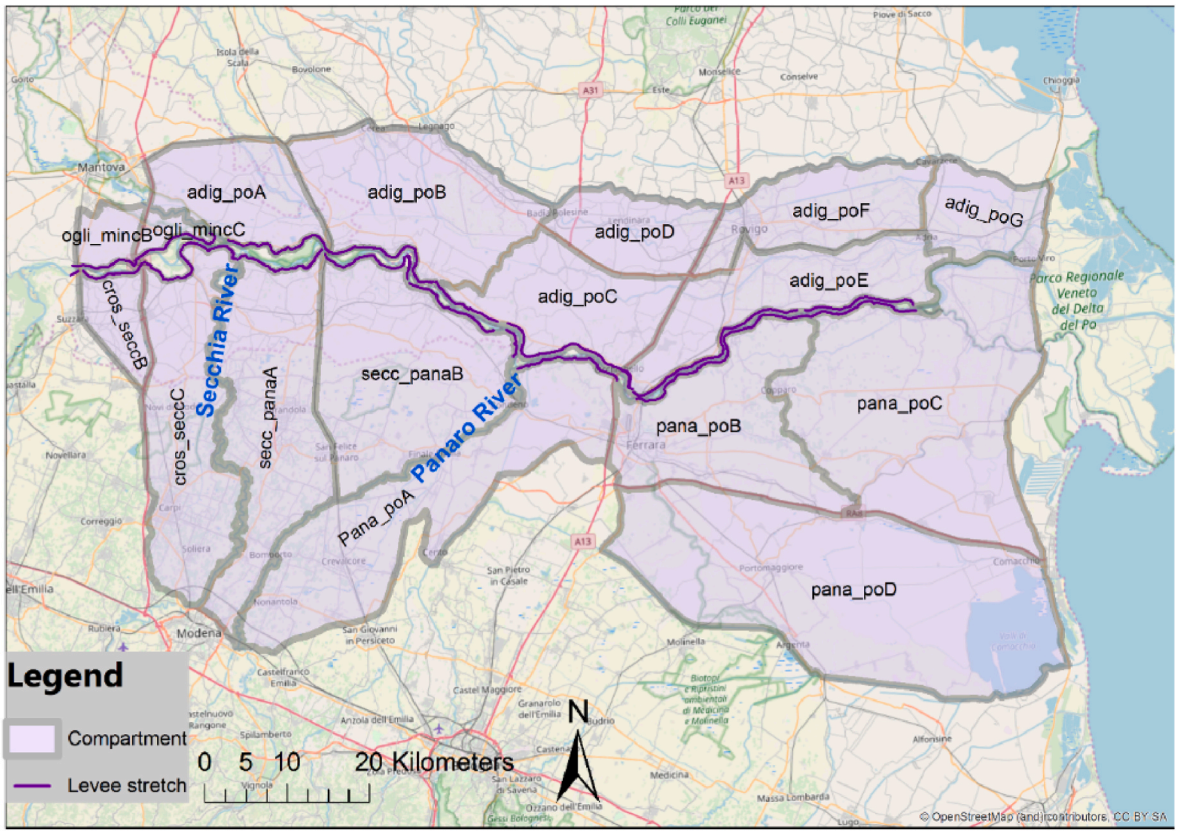

Fig. 2. The case study area. The thick grey lines delimit the flood-protected areas (i.e., compartments) and the purple lines depict the levee stretches where structural measures can be implemented.

along the Secchia River, Italy. The breach was unforeseen and it was caused by animal burrows and not by one of the failure mechanisms commonly modeled when deriving fragility curves, i.e., overtopping, piping or macro-instability (Orlandini et al. 2015). For these reasons, we treat breaching water levels, namely water levels triggering levee failures, as deeply uncertain.

3.2. Simulation model

The simulation model follows three steps: (1) the generation of hydrological events along the Po River and its main tributaries Secchia and Panaro, (2) the propagation of the generated events using a hydrodynamic model, (3) the assessment of economic damage in the flooded compartments. The following subsections briefly introduce the modeling steps. More detailed information is provided in the supplementary material.

3.2.1. Generation of hydrological events

We focus on the generation of alternative 0.2% annual probability flood events at the upstream Po River's cross-section of Borgoforte. These events are expected to cause flooding in the compartments, as the levee system is designed to withstand the 0.5% annual probability flood events.

The 0.2% annual probability events along the Po River at Borgoforte are generated based on the Flood-Duration-Frequency (FDF) curve derived by Maione et al. (2003) and the historical flood hydrographs of the Po at Borgoforte reported in Tanda et al. (2001). The aim is to generate 0.2% annual probability events of decreasing peak flow for increasing duration such that the FDF curve is always met.

After the 0.2% annual probability events are generated for the Po River, a Gaussian copula is used to capture the dependence between the generated events in the Po River and in its main tributaries Secchia and Panaro, similarly to what is done by Curran et al. (2020). The correlation between events in the Po River and those in the tributaries is modeled using data from the Hypeweb platform of the Swedish Meteorological and Hydrological Institute (SMHI). These data consist of 29 years of simulated daily mean discharges for the three rivers and are used to generate a Gaussian copula between volumes and discharges of the Po, Secchia and Panaro rivers. After generating the 0.2% annual probability flood event for the Po River as described above, events of the tributaries are found by conditional sampling of the Gaussian copula.

3.2.2. Hydrodynamic model

Flood hazard is simulated with a quasi-2D model implemented through the HEC-RAS software (Domeneghetti et al., 2015). In the model, the main river is represented through cross-sections retrieved from a LiDAR digital elevation model with a 2 m spatial resolution. The levee-protected floodplain is subdivided into 17 compartments, which are modeled as storage areas connected to each other and/or to the main channel by means of lateral structures or connections that reproduce existing levees. The hydrostatic behavior of each storage area is represented through volume-level curves. Thus, in case of flooding, water levels can be estimated from the water volumes exchanged with the main channel and/or adjacent storage areas. These curves are built using a 10 m resolution DEM available for all Italy (TINITALY, see Tarquini et al., 2012).

3.2.3. Economic impact assessment

Floods are often devastating events and the impact they cause includes loss of life, damage to buildings, infrastructures, and the deterioration of ecosystems. Moreover, flood impacts are larger for communities which are more socially and economically vulnerable. However, since the focus of this study is primarily methodological, for practical reasons we model flood impact only as economic damage to

residential buildings. To do so, we use the Hypsometric Vulnerability Curves approach proposed by Domeneghetti et al. (2015) and the damage function developed by Carisi et al. (2018).

Typically, the hypsometric curve of an area provides the percentage of the total area below a certain elevation. The Hypsometric Vulnerability Curve combines this information with the land-use of the area, thus providing the percentage of a land use class below a given elevation. Therefore, if combined with damage functions, the hypsometric vulnerability curve is a useful simplified graphical tool to quantify aggregated flood damage of large areas.

We calculate Hypsometric Vulnerability Curves relative to urban areas in all compartments. To do so, we use data of residential buildings available from the geodata web-platforms of the three Italian regions involved in the case study, namely Emilia-Romagna, Lombardia and Veneto. Data about asset values of buildings are retrieved from the Italian Revenue Agency. Asset value data are provided in terms of euros per square meter (€/m^2) for different types of buildings and all Italian municipalities every six months. We define an asset value per compartment as the weighted average of all asset values where weights are given by the extent of urban area of the municipalities in that compartment.

For deriving the depth-damage curve, we used the square root regression model developed by Carisi et al. (2018) using loss data collected after the Secchia River 2014 flood. We account for the uncertainty in this relationship by using alternative regression models generated from the distribution of the fitting parameter.

4. Description of the analysis

The analysis aims at applying the BIRDM framework introduced in section 2. The goal is to explore, for each compartment, how alternative assumptions regarding the distribution of deep uncertainties affect the robustness of damage reduction measures and their ranking. The description follows the steps introduced in section 2.

4.1. Problem framing

The adopted XLRM framework is shown in Table 1. Exogeneous uncertainties (X) relate to various hydrological features of the Po River and its tributaries, the damage function and the breaching water levels at each of the 52 levee stretches identified in the study area. The considered structural measures (L) correspond to either doing nothing (hereafter *status quo*), levees raising (hereafter *raising*) or levees strengthening (hereafter *strengthening*). The simulation model (R) is the one introduced in subsection 3.2. The outcomes of interest (M) are the flood damages from the 0.2% annual probability events in each compartment.

4.2. Uncertainty analysis

Uncertainty analysis is conducted performing 4500 runs of the simulation model using a Sobol sampling sequence (Sobol 1967). This is also known as quasi-Monte Carlo analysis. Table 2 reports the selected sampling distribution for each uncertainty.

We stress that, for the reasons outlined in subsection 3.1, uncertainties related to breaching water levels are considered as deeply uncertain. It follows from the *scenario neutral* assumption that these are uniformly distributed, with the range of physically plausible breaching water levels going from the levee crest height (upper limit) to the lowest level at which a breach can physically be triggered (lower limit). This latter is given by the highest point between the floodplain's height and the height of the levee-protected area such that the necessary hydraulic gradient required for a breach to plausibly develop is guaranteed. In the case study, this is estimated to be approximately 2 m below the initial crest level for all levee stretches. This yields a fragility curve as expressed by the thick blue line in Fig. 3.

As indicated in Fig. 1, this step and scenario discovery are iterated. Therefore, the uncertainty analysis is first carried out for the *status quo*. Next, system vulnerabilities are identified through scenario discovery and measures that limit such vulnerabilities are defined. Finally, uncertainty analysis is again performed assuming measures are implemented. Convergence plots are shown in the supplementary material as mean damage against number of simulations.

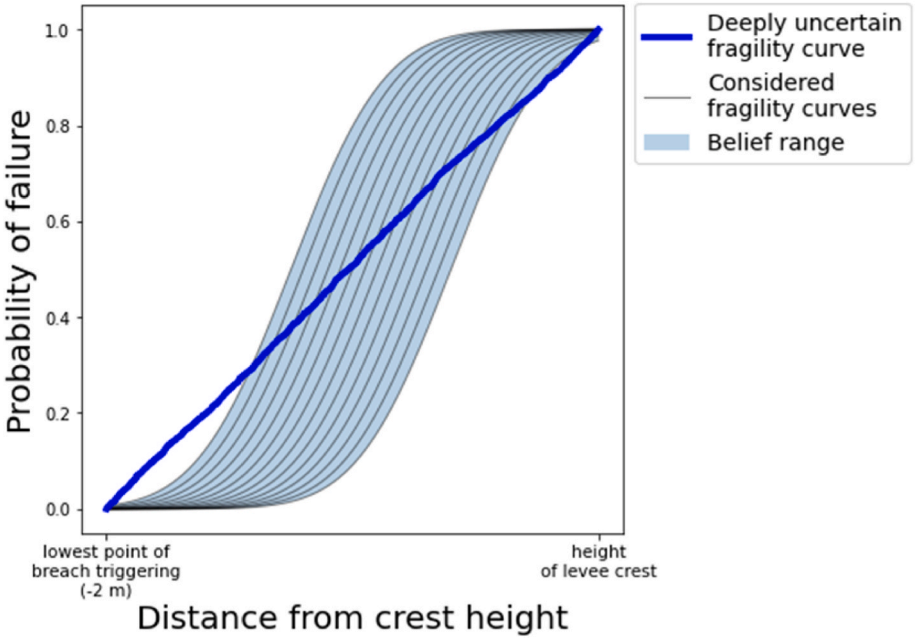

Fig. 3. The deeply uncertain fragility curve (thick blue line) and the 15 considered alternative fragility curves (grey lines) spanning the considered range (blue shaded area). All 15 alternative fragility curves follow a normal distribution with standard deviation of 0.3 but mean values spanning from 0.6 m to 1.3 m below the height of the levee crest, with a step of 5 cm.

4.3. Scenario discovery

Scenario discovery is performed using the Patient Rule Induction Method (PRIM) algorithm (Friedman and Fisher 1999). Scenario discovery is used to explore the uncertainty space generated by sampling uncertainties as reported in Table 2, except for the first three listed uncertainties, namely duration, shape class and shape of the Po River's upstream hydrograph which are used to generate flood waves. Instead, for the sake of interpretability, scenario discovery is run on the peak discharges and volumes of the generated flood waves. This implies the exploration of 59 variables: the peak discharge and hydrographs volumes of the 3 rivers, the breaching water levels at the 52 levees stretches and the damage model parameter. Scenario discovery is used to identify cases where flood damage in each compartment is larger than the 3rd damage quartile under *status quo*, which is the assumed critical performance threshold.

Scenario discovery provides an indication, for each compartment, of what levee stretches, when failing, are likely to lead to damages above the critical threshold. Based on scenario discovery results and a broader consideration of the geometry of the levee system, we identify critical levee stretches and define structural measures. After these measures are identified, their performance is re-assessed through uncertainty analysis (step 4.2).

4.4. Belief analysis

This step allows exploring how the ranking of structural measures changes under various alternative assumptions about the shape of the fragility curves. Structural measures are evaluated based on their robustness in providing satisficing performances. This is defined as the probability of not exceeding the critical performance threshold, which we define for each compartment as the flood damage equal to the 3rd damage quartile under *status quo* (see Section 4.3). This probability can be assessed, for each measure, via the method described in section 2.4. Obviously, the most robust measure is the one with the highest non-exceedance probability of the performance threshold.

Different hypotheses regarding the fragility curve may lead to different robustness values, and thus result into a different ranking of measures. In practice, the choice of the alternative fragility curves should be derived through expert elicitations and reflect the uncertainty around levee breaching for the levee stretch of interest. For simplicity, we assume all fragility curves follow a normal distribution with the same standard deviation of 0.3 but different mean values. These range from 0.6 m (upper curve) to 1.3 m (lower curve) below the height of the levee crest, with steps of 5 cm each, for a total of 15 fragility curves. The range (number of steps) could be made larger (finer) according to the application, as the computational cost of exploring a large number of alternative beliefs is very low. All considered fragility curves are shown in Fig. 3.

Table 1
The adopted XLRM framework.

Uncertainties (X)	Measures (L)	Model (R)	Outcome (M)
• Duration of the Po River's upstream flood hydrograph • Shape of the Po River's upstream flood hydrograph • Peak and volume of the Secchia and Panaro rivers' flood hydrograph • Breaching water level at each of the 52 levee stretches (deep uncertainty) • Damage model parameter	• Do nothing • Levee heightening • Levee strengthening	The flood impact modeling chain described in subsection 3.2.	The 0.2% annual probability damage at each compartment

Table 2
Description of the considered uncertainties, distributions, and parametrizations.

Uncertainties (X)	Distribution type	Parametrization
Hydrograph's duration for the Po River	Continuous Uniform	*Lower bound* = 24 h *Upper bound* = 192 h
Hydrograph's shape class for the Po River	Discrete Probability	See Table 1 in supplementary material
Hydrograph's shape for the Po River	Discrete Uniform	*Lower bound* = 1 *Upper bound* = number of hydrographs in the selected shape class
Hydrograph's peak discharge and volume for the Secchia and Panaro rivers	Conditioned Gaussian Copula	See Tables 3 and 4 in supplementary material
Breaching water level at the levee stretches (WL stretch) – deep uncertainty	Continuous Uniform	*Lower bound* = 2 m below the levee's height *Upper bound* = levee's height
Damage model parameter (DMRP)	Truncated Normal	*Mean* = 0.113 *Standard deviation* = 0.131 *Min* = 0 *Max* = $+\infty$

5. Results and discussion

This section reports the results of the analysis described in section 4. This section reports results from the Uncertainty Analysis and Scenario Discovery steps in subsection 5.1 and those from the Belief Analysis step in subsection 5.2.

5.1. Uncertainty analysis and scenario discovery

The output of the uncertainty analysis of the current system (*status quo,* blue boxplot in Fig. 4) reveals that five compartments are flooded by the 0.2% annual probability flood events, and they are all located upstream. This happens because breaching at these compartments results in flood peak attenuation and hence in an unloading effect for the downstream compartments, which are consequently not flooded.

For all the flooded compartments, failure scenarios are identified as those where flood damages under *status quo* are higher than the critical performance threshold. Table 3 shows the results of scenario discovery for these failure scenarios.

High values of the damage model parameter are responsible for large flood damage in all compartments. Flood waves with large volumes along the Po River are responsible for large flood damages only to the *Cross_seccB* compartment, one of the most upstream. For each compartment, one can identify the critical locations, namely those locations where, should the levees fail, large flood damage would occur. Identifying these critical locations provides a crucial indication on where structural damage reduction measures should be applied, and scenario discovery supports their identification.

In the case of *strengthening*, this information suffices as we simply assume that these stretches are strengthened in such a way that levee overtopping produces flooding without breaching the levee. In the case of *raising*, an additional choice needs to be made based on the height of the nearby levees. To make the two structural measures comparable, both *raising* and *strengthening* are applied to the same set of stretches.

In compartment *cross_seccB*, two stretches are identified as relevant for investing in flood protection, namely stretches number 3 and 5. The levee at stretch 5 has an average crest level of 25 m a.s.l., while the levee at stretch 3 is higher with an average crest level of 25.65 m a.s.l., Stretch 4, located between stretches 3 and 5, with an average levee crest level of 25.5 m a.s.l., is also considered part of the intervention. As the lowest average levee height along the compartment on the other side of the Po River is about 25.8 m a.s.l., levees of stretches 3, 4 and 5 are raised up to this level. Following similar reasoning, levee raising of 1.0, 0.7, 0.5 and

Table 3
Results from scenario discovery for each compartment. Results show which parameter and what range of values lead to flood damages higher than the 3rd quartile under the *status quo*. DMRP indicates the damage model regression parameter [.], WL stretch # [m a.s.l.] indicates the water level at a given stretch and Vol. Po is the flood volume of the Po River [m^3].

cross_seccB		cross_seccC		ogli_mincC		secc_panaA		ogli_mincB	
Params	*Range*	*Params*	*Range*	*Params*	*Range*	*Params*	*Range*	*Params*	*Range*
DMRP	(0.15, 0.58)	*DMRP*	(0.12, 0.58)	*DMRP*	(0.06, 0.58)	*DMRP*	(0.06, 0.58)	*DMRP*	(0.12, 0.58)
WL stretch #3	(23.7, 25.4)	*WL stretch #15*	(21.8, 22.9)	*WL stretch #6*	(23.3, 23.9)	*WL stretch #25*	(20.5, 21.2)	*WL stretch #1*	(23.8, 25.6)
WL stretch #5	(23.0, 24.6)								
Vol. Po	(7.4e9, 9.2e9)								

Table 4
Ranking of measures under the assumption of a deeply uncertain fragility curve. The number in parenthesis indicate the probability of non-exceeding the performance threshold. Status quo, raising and strengthening are indicated as SQ, R and S respectively. The number in brackets indicates the probability of non-exceeding the performance threshold.

Compartment/ Ranking	First-ranked measures	Second-ranked measures	Third-ranked measures
cros_seccB	*S (1.0)*	*R (0.80)*	*SQ (0.74)*
ogli_mincB	*S (0.81)*	*R/SQ (0.75)*	*R/SQ (0.75)*
cros_seccC	*R (0.79)*	*SQ (0.74)*	*S (0.69)*
ogli_mincC	*R (0.76)*	*SQ (0.74)*	*S (0.62)*
secc_panaA	*SQ (0.74)*	*R (0.72)*	*S (0.66)*

0.3 m is applied to stretches in compartments *cross_seccC, ogli_minC, ogli_minB, secc_panaA* respectively. In total, the procedure leads to the implementation of structural interventions at 7 stretches, as shown in Fig. 5.

The performance of both *raising* and *strengthening* is assessed by performing two uncertainty analyses and results are shown in Fig. 4. Looking at the total damage, *raising* and *strengthening* both improve the current situation, as they result in lower damage than *status quo*. In the case of *strengthening* this benefit is much more significant. Similar results are found for *cros_seccB*, the compartment where most damage occurs and the one where the most upstream stretches subject to intervention are located. Although with a less remarkable difference, *strengthening* is the best option also for compartment *ogli_minB*. For all other compartments it is found that either *raising* or *strengthening* may deteriorate performance with respect to *status quo,* as the number of cases above the performance threshold increases. This is due to right-left and upstream-downstream hydraulic interactions due to which structural measures reducing damage on a given river stretch may increase damage at the other side of the river or downstream (Ciullo et al., 2020). For example, for *ogli_mincB*, which is located opposite to *cros_seccB*, *raising* does not seem to bring much damage reduction as it performs similarly to *status quo*. For the more downstream compartments, i.e., *cros_seccC, ogli_mincC, secc_panaA* and *adig_poE, strengthening* may significantly deteriorate performance with respect to *status quo,* to the point that it may cause damage at compartment *adig_poE* where there was none initially. Interventions were designed via scenario discovery with the goal of reducing damage above the performance threshold at each compartment. To limit the effects of hydraulic interactions, the Uncertainty Analysis and Scenario Discovery steps would need to be run iteratively

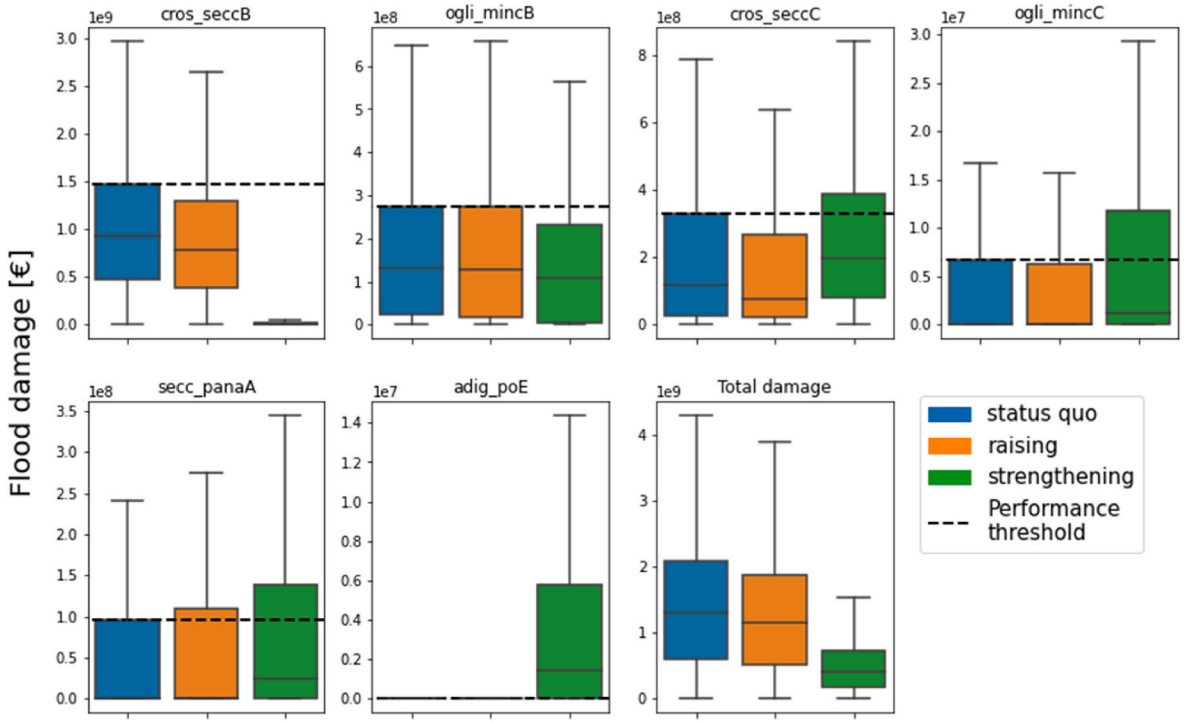

Fig. 4. Uncertainty analysis results for each compartment and the overall system. Each box shows flood damage from the 0.2% annual probability events under *status quo* (blue) and after raising (orange) and strengthening (green). The critical performance threshold (dotted line) corresponds to the 3rd quartile of the damages under *status quo*.

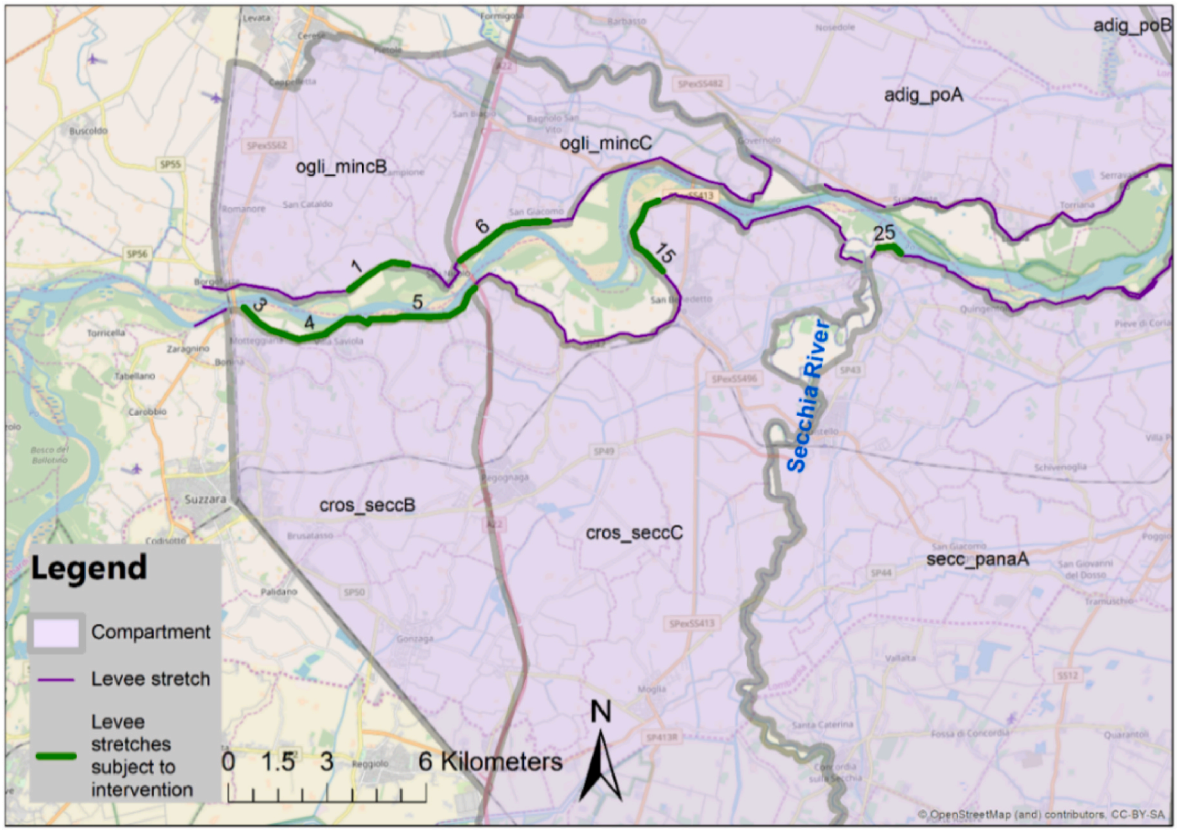

Fig. 5. Location of the stretches identified for the implementation of structural measures.

until a satisfactory set of measures is achieved for the entire system. A comprehensive design of the entire system requiring such an extensive iterative procedure of refinement of measures is beyond the scope of this paper.

5.2. Belief analysis

Following the description in Section 4.4, this section reports how robustness values and the ranking of measures change for each compartment utilizing each of the 15 alternative fragility curves reported in Fig. 3. The ranking of the three measures is assessed based on their robustness in achieving satisfactory performances, namely the probability of non-exceeding the performance threshold (see Section 4.4). The higher this probability, the higher a measure is ranked.

Results of the belief analysis are reported in Fig. 6. In each panel and for each of the 15 alternative fragility curves, colors indicate which measure ranks at the position indicated by the respective column. For comparison, Table 4 reports results under the assumption of a deeply uncertain fragility curve.

Compartment *cross_seccB* is the only one for which alternative beliefs do not affect the ranking of measures at all, as *strengthening* always ranks first, followed by *raising* and *status quo*. This is coherent with the overall performance under deep uncertainty.

For compartment *ogli_mincB* and *cross_seccC, raising* is instead the first-ranking measure regardless beliefs. For both compartments, the second- and third-ranked measures differ across beliefs and can be either *strengthening* or *status quo,* with the latter being preferred when levees are believed to be weak and prone to failing at low water levels. This is logical, as structural measures are most effective under circumstances of weak levees. For *ogli_mincB*, results are surprising since *strengthening,* which is the best measure under deep uncertainty (see Table 4), never ranks as the first measure under changing beliefs. This highlights how a high performance of measures under the *scenario neutrality* assumption may deteriorate when beliefs are introduced. For *cross_seccC, raising* is the first-ranked measure also under deep uncertainty, followed by *status quo* and *strengthening.*

For compartments *ogli_mincC* and *secc_panaA,* the first-ranked measure, likely to be the most relevant one from a practical viewpoint, is sensitive to changes in belief, as a shift from *status quo* to *raising* occurs depending on the belief about the strength of the levee. For both compartments, *raising* is preferable when the levee is believed to be weak, and *status quo* takes over when the levee is believed to be stronger. The opposite shift takes place as second ranked-measure, whereas the third ranked measure is always *strengthening* as this measure increases damages at these compartments due to upstream-downstream interactions (see previous section).

6. Conclusions

In the present paper we introduced the Belief-Informed Robust Decision Making (BIRDM) framework. BIRDM builds upon Robust Decision Making (RDM) and it enables the assessment of the effects of alternative distributions of deep uncertainties on the robustness values and ranking of measures. In principle, the same results provided by BIRDM could be achieved by running a standard RDM analysis multiple times with alternative assumptions regarding the distribution of deep uncertainties. This, however, would demand large computational power and may become unfeasible as soon as many alternative assumptions need to be explored. The proposed approach, instead, requires almost no additional computational costs nor any new modeling efforts compared to RDM, and it therefore allows the exploration of an arbitrarily large number of alternative distributions. As such, BIRDM extends and improves RDM as it allows assessing whether ambiguities, lack of knowledge or disagreement regarding the distribution of deep uncertainties matter in the first place.

We demonstrated BIRDM by applying it to a flood management

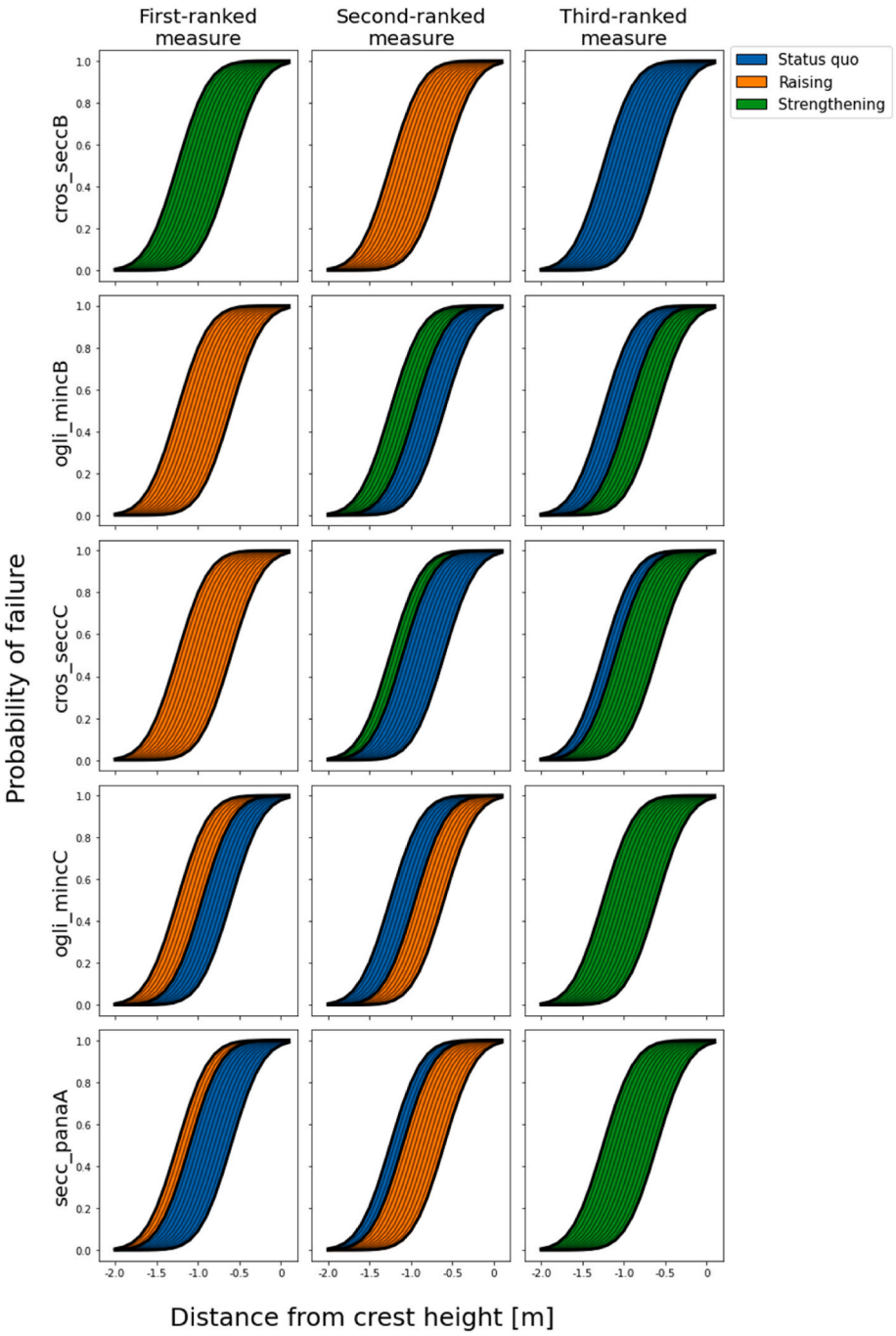

Fig. 6. Results from the Belief Analysis step. Each column indicates a ranking position and each row indicates a compartment. For each compartment and ranking position, colors indicate which measure rank at the given position across all 15 fragility curves introduced in Fig. 3. Fragility curves at which a rank shift occurs are reported in bold.

planning problem along the lower Po River, Italy. The goal was to identify the most robust structural measures and to assess whether, and to what extent, alternative assumptions about the levees' fragility curves affect robustness values and the ranking of measures. We find that BIRDM's results reveal changes in the ranking of measures because of changing assumptions about the levees' fragility curves. Therefore, the ranking emerging from the standard RDM is sensitive to alternative assumptions regarding such uncertainties. This can be decision-relevant, especially when these changes affect which measure ranks first. BIRDM also supports the identification of which belief triggers such a change in the ranking, thus enabling the assessment of what beliefs lead a given measure to have a specific rank. This may trigger further investigations regarding, for example, the reliability of such rank-changing beliefs. Finally, BIRDM can be used in a stress-test fashion to assess under what assumptions a given a measure no longer performs as desired.

Data and code availability

Results of the full analysis and the code need to perform the Scenario Discovery and Belief Analysis steps can be found at https://github.com/aleeciu/BIRDM_paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676027.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2022.105560.

References

Apel, Heiko, Merz, Bruno, Thieken, Annegret, 2009. Influence of dike breaches on flood frequency estimation. Comput. Geosci. https://doi.org/10.1016/j.cageo.2007.11.003.

Bachmann, D., Huber, N.P., Johann, G., Schüttrumpf, H., 2013. Fragility curves in operational dike reliability assessment. Georisk 7 (1), 49–60. https://doi.org/10.1080/17499518.2013.767664.

Beckman, R.J., McKay, M.D., 1987. Monte Carlo estimation under different distributions using the same simulation. Technometrics 29 (2), 153–160. https://doi.org/10.1080/00401706.1987.10488206.

Ben-Haim, Y., 2006. Info-Gap Decision Theory - 2nd Edition, 2006. https://www.elsevier.com/books/info-gap-decision-theory/ben-haim/978-0-12-373552-2.

Beven, K.J., Almeida, S., Aspinall, W.P., Bates, P.D., Blazkova, S., Borgomeo, E., Freer, J., et al., 2018. Epistemic uncertainties and natural hazard risk assessment – Part 1: a review of different natural hazard areas. Nat. Hazards Earth Syst. Sci. 18 (10), 2741–2768. https://doi.org/10.5194/nhess-18-2741-2018.

Brown, C., Ghile, Y., Laverty, M., Li, K., 2012. Decision scaling: linking bottom-up vulnerability analysis with climate projections in the water sector. Water Resour. Res. 48 (9) https://doi.org/10.1029/2011WR011212.

De Bruijn, K.M., Diermanse, F.L.M., van der Doef, M., Klijn, F., 2016. Hydrodynamic system behaviour : its analysis and implications for flood risk management. In: 3rd European Conference on Flood Risk Management, vol. 8. FLOODrisk 2016. https://doi.org/10.1051/e3sconf/20160711001.

Carisi, F., Schröter, K., Domeneghetti, A., Kreibich, H., Castellarin, A., 2018. Development and assessment of uni- and multivariable flood loss models for emilia-romagna (Italy). Nat. Hazards Earth Syst. Sci. 18 (7), 2057–2079. https://doi.org/10.5194/nhess-18-2057-2018.

Ciullo, A., de Bruijn, K.M., Kwakkel, J.H., Klijn, F., 2019a. Accounting for the uncertain effects of hydraulic interactions in optimising embankments heights: proof of principle for the IJssel river. J.Flood Risk.Manag. (December 2018) https://doi.org/10.1111/jfr3.12532.

Ciullo, A., De Bruijn, K.M., Kwakkel, J.H., Klijn, F., 2019b. Systemic flood risk management: the challenge of accounting for hydraulic interactions. Water 11 (12), 2530. https://doi.org/10.3390/w11122530.

Ciullo, A., Kwakkel, J.H., De Bruijn, K.M., Doorn, N., Klijn, F., 2020. Efficient or fair? Operationalizing ethical principles in flood risk management: a case study on the Dutch-German rhine. Risk Anal. https://doi.org/10.1111/risa.13527.

Courage, Wim, Vrouwenvelder, Ton, van Mierlo, Thieu, Schweckendiek, Timo, 2013. System behaviour in flood risk calculations. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards. https://doi.org/10.1080/17499518.2013.790732.

Curran, A., de Bruijn, K.M., Klerk, W.J., Kok, M., 2019. Large scale flood hazard analysis by including defence failures on the Dutch river system. Water 11 (8), 1732. https://doi.org/10.3390/w11081732.

Diermanse, F.L.M., De Bruijn, K.M., Beckers, J.V.L., Kramer, N.L., 2015. Importance sampling for efficient modelling of hydraulic loads in the rhine–Meuse Delta. Stoch. Environ. Res. Risk Assess. 29 (3), 637–652. https://doi.org/10.1007/s00477-014-0921-4.

Domeneghetti, A., Carisi, F., Castellarin, A., Brath, A., 2015. Evolution of flood risk over large areas: quantitative assessment for the Po river. J. Hydrol. 527, 809–823. https://doi.org/10.1016/j.jhydrol.2015.05.043.

Friedman, J.H., Fisher, N.I., 1999. Bump hunting in high-dimensional data. Stat. Comput. 9 (2), 123–143. https://doi.org/10.1023/A:1008894516817.

Haasnoot, M., Kwakkel, J.H., Walker, W.E., ter Maat, J., 2013. Dynamic adaptive policy Pathways: a method for crafting robust decisions for a deeply uncertain world. Global Environ. Change 23 (2), 485–498. https://doi.org/10.1016/j.gloenvcha.2012.12.006.

Kasprzyk, J.R., Nataraj, S., Reed, P.M., Lempert, R.J., 2013. Many objective robust decision making for complex environmental systems undergoing change. Environ. Model. Software 42 (April), 55–71. https://doi.org/10.1016/j.envsoft.2012.12.007.

Lempert, R.J., Groves, D.G., Popper, S.W., Bankes, S.C., 2006. A general, analytic method for generating robust strategies and narrative scenarios. Manag. Sci. 52 (4), 514–528. https://doi.org/10.1287/mnsc.1050.0472.

Lempert, R.J., Popper, S.W., Bankes, S.C., 2003. Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis. RAND, Santa Monica, CA.

Maione, Ugo, Mignosa, Paolo, Tomirotti, Massimo, 2003. Regional estimation of synthetic design hydrographs. Int. J. River Basin Manag. https://doi.org/10.1080/15715124.2003.9635202.

McPhail, C., Maier, H.R., Westra, S., Kwakkel, J.H., van der Linden, L., 2020. Impact of scenario selection on robustness. Water Resour. Res. 56 (9), e2019WR026515 https://doi.org/10.1029/2019WR026515.

Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J.C.J.H., Bates, P., Bertola, M., et al., 2021. Causes, impacts and patterns of disastrous river floods. Nat. Rev. Earth Environ. 1–18. https://doi.org/10.1038/s43017-021-00195-3. August.

Orlandini, S., Moretti, G., Albertson, J.D., 2015. Evidence of an Emerging Levee Failure Mechanism Causing Disastrous Floods in Italy. Water Resources Research, pp. 7995–8011. https://doi.org/10.1002/2015WR017426.Received.

Quinn, J.D., Hadjimichael, A., Reed, P.M., Steinschneider, S., 2020. Can exploratory modeling of water scarcity vulnerabilities and robustness Be scenario neutral? Earth's Future 8 (11), e2020EF001650. https://doi.org/10.1029/2020EF001650.

Reis, J., Shortridge, J., 2020. Impact of uncertainty parameter distribution on robust decision making outcomes for climate change adaptation under deep uncertainty. Risk Anal. 40 (3), 494–511. https://doi.org/10.1111/risa.13405.

Shortridge, J.E., Zaitchik, B.F., 2018. Characterizing climate change risks by linking robust decision frameworks and uncertain probabilistic projections. Climatic Change 151 (3), 525–539. https://doi.org/10.1007/s10584-018-2324-x.

Sobol', I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, The Second IMACS Seminar on Monte Carlo Methods 55 (1), 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6.

Sobol', I.M., 1967. Distribution of points in a cube and approximate evaluation of integrals. U.S.S.R Comput. Maths. Math. Phys. 7, 86–112.

Sparkman, Daniel, Milwater, Harry, Garza, Jose, Smarslok, Benjamin, 2016. Importance sampling-based post-processing method for global sensitivity analysis. In: 18th AIAA Non-deterministic Approaches Conference. https://doi.org/10.2514/6.2016-1440.

Taner, M.Ü., Ray, P., Brown, C., 2019. Incorporating multidimensional probabilistic information into robustness-based water systems planning. Water Resour. Res. 55 (5), 3659–3679. https://doi.org/10.1029/2018WR022909.

Tarquini, Simone, Vinci, Stefano, Favalli, Massimiliano, Doumaz, Fawzi, Fornaciai, Alessandro, Nannipieri, Luca, 2012. Release of a 10-m-resolution DEM for the Italian territory: comparison with global-coverage DEMs and anaglyph-mode exploration via the web. Comput. Geosci. https://doi.org/10.1016/j.cageo.2011.04.018.

Tokdar, S.T., Kass, R.E., 2010. Importance sampling: a review. WIREs Computational Statistics 2 (1), 54–60. https://doi.org/10.1002/wics.56.

Vorogushyn, S., Bates, P., De Bruijn, K., Castellarin, A., Kreibich, H., Priest, S., Schröter, K., et al., 2017. Evolutionary leap in large-scale flood risk assessment needed. WIREs Water 5 (2), 1–7. https://doi.org/10.1002/wat2.1266.

Zhang, Jiaxin, Shields, Michael, 2018. On the quantification and efficient propagation of imprecise probabilities resulting from small datasets. Mech. Syst. Signal Process. https://doi.org/10.1016/j.ymssp.2017.04.042.

Zhang, Jiaxin, TerMaath, Stephanie, Shields, Michael, 2021. Imprecise global sensitivity analysis using bayesian multimodel inference and importance sampling. Mech. Syst. Signal Process. https://doi.org/10.1016/j.ymssp.2020.107162.