Tools for risk assessment under uncertainty

National Science Challenges

SUSTAINABLE SEAS

Ko ngā moana whakauka

NIWA
Taihoro Nukurangi

Manaaki Whenua
Landcare Research

Managed Seas

Project 5.1.3: Risk and uncertainty

Deliver innovative and effective decision support tools that integrate the knowledge generated by the Challenge to allow ecosystem based management, ensuring sustainable utilisation of our marine resources.

For any information regarding this report please contact:

Graeme Inglis
Principal Scientist
Marine Biodiversity & Biosecurity
National Institute of Water & Atmospheric Research Ltd
PO Box 8602
Christchurch 8011
Phone +64-3-343 8036

Tarek Soliman
Economist
Landscape Policy and Governance
Manaaki Whenua – Landcare Research
Private Bag 92170
Auckland 1142

Utkur Djanibekov
Economist
Landscape Policy and Governance
Manaaki Whenua – Landcare Research
Private Bag 92170
Auckland 1142

NIWA CLIENT REPORT No: 2020259CH
Report date: August 2018
NIWA Project: SUSS18502

Quality Assurance Statement		
	Reviewed by:	Jeremy McKenzie, Principal Scientist Fisheries and Modelling
	Formatting checked by:	Rachel Wright
	Approved for release by:	Barbara Hayden, Chief Scientist Coasts & Oceans

Cover image – *Kapiti sunset. [Dave Allen, NIWA]*

NIWA Taihoro Nukurangi

Manaaki Whenua Landcare Research

Contents

Tables

Figures

Executive summary

Risk assessment is a process that is used to support decision-making about uncertain future events and their consequences for society. Fundamentally, risk occurs when the something of value is at stake and the outcome is uncertain. Common among most definitions of risk are at least three elements:

- an event or occurrence that may take place,
- the potential consequences (or outcomes) of that event and
- the likelihood that the event and its consequences will eventuate.

Risk management problems can be characterized according to their complexity and the level of knowledge that we have about the potentially hazardous event(s) and its consequences. They range from relatively well-known, recurrent hazards, such as car accidents and food contamination, in which the prospect of future events and their consequences can be predicted by analysis of past occurrences, to highly complex or novel risk problems where it is uncertain what threat the event poses to a range of values and interests and it is difficult to predict the likelihood that they will eventuate.

The varied contextual settings and complexity of risk problems means that are many different approaches to the assessment of risks, each involving different ways of structuring the collection and analysis of available information that are tailored to the problem under consideration and the kinds of data and knowledge that can be applied to it. Recent literature distinguishes between approaches where the purpose of the analysis is to provide the best available predictions of the occurrence of uncertain events and their associated consequences ('probabilistic' risk analyses or the 'predict-then-act' paradigm) and those where the focus is on identifying and reducing vulnerability to threats by ensuring that decisions are adaptable, robust to uncertainty and can withstand surprises (the 'monitor and adapt' paradigm). In this report, we review a range of analytical tools and processes that can be used to support each approach. The methods included here are not intended to be comprehensive, but rather to represent a range of 'best-practice' tools and approaches that can be applied across a spectrum of risk problems of differing complexity and uncertainty.

We begin by exploring the concepts of risk and uncertainty (Section 2) and expand on a typology of uncertainty to categorize risk problems according to their complexity, the extent of existing knowledge about the event and its consequences, and the level of consensus among decision-makers, analysts and stakeholders on the values that are at stake (Table 1).

Conventional probabilistic risk analyses express uncertainty about the occurrence of an event or consequence as probabilities. Two types of probability can be used, depending on the availability of data on previous occurrences of the event. Use of 'frequentist' probabilities, where the likelihood of an event can be estimated objectively from observational data, is generally limited to relatively simple risk problems where there are adequate data on past occurrences to characterize the risk (Level 1 uncertainty, Table 1). For most complex risk problems (Level 2 and 3 uncertainty, Table 1) there are few precedents and limited data available to describe the cause-effect relationships between a hazard and its consequences. In these circumstances, 'subjective' (or 'Bayesian') probabilities are used to bridge the gaps in observations and data. Subjective probabilities are based on the background knowledge and judgement of experts but can be beset by a range of heuristics and biases that diminish their validity. In Section 4 we provide best-practice guidance on protocols and methods for eliciting subjective probabilities from experts to reduce these known sources of bias. They include methods for reconciling vague linguistic descriptors of uncertainty as quantitative subjective probabilities (Section 4.3.3), and approaches for eliciting and aggregating judgements about highly uncertain events among expert groups (Sections 4.3 and 4.4).

In Section 5 we describe frameworks and methods for assessing cumulative and indirect risks from multiple activities. Most cumulative risk assessments fall into the realm of the Level 3 and Level 4 uncertainties described in Table 1. As the number of activities and stressors under consideration in the analysis increases and the pathways through which they may have an effect proliferate, so too does the range of values and assets that are potentially at risk and the uncertainties associated with each outcome. We review the use of tiered or 'semi-quantitative' (*sensu* Aven (2008)) approaches to structure probabilistic analysis of cumulative risks. These approaches combine qualitative description and analysis of the complex relationships among system components with quantitative analysis of prioritised subcomponents where uncertainty or risk is deemed to be greatest.

Conventional probabilistic risk assessments separate the analytical assessment of the likelihood of an event and its consequences from evaluation of the benefits and costs of different possible outcomes (a social process) and the design of management actions to mitigate risk or achieve the best outcomes (a policy process). In Section 6, we describe the use of scenario-based assessment for complex risk problems where these three components are combined in the analysis. Two quite distinct approaches are described. Decision-analytical methods (Section 6.1) can be used when uncertainties are able to be expressed probabilistically (Levels 2 and 3 in Table 1) and performance measures can be defined for the desired state of the system under consideration. In these approaches, optimization methods are used to iterate between examining modelled future scenarios (based on the probability of events and their consequences) and assessing their performance against the stated objectives for the system under consideration.

In deeply uncertain risk problems (Level 4 uncertainty, Table 1), when technical experts and others involved in the assessment do not know or cannot agree on appropriate ways to describe key relationships among the components of the system that affect future outcomes, when uncertainty cannot be represented reliably by probabilities, and when there is a range of perspectives of the value of predicted outcomes, there is not a clear path for choosing which scenarios to model or the appropriate model structure to analyse risks. Methods for Decision Making Under Deep Uncertainty (DMDU, Section 6.2) instead seek to examine vulnerabilities in policy settings, particularly the consequences that would occur if assumptions underlying the policy setting are violated. The purpose of DMDU methods is to identify policy options that best reduce the vulnerabilities to potential future risks and to hedge against the possibility of failure (i.e., to make the system robust to uncertainty). To deal with potential surprises and to avoid undesirable long-term outcomes, DMDU methods select policies on the basis of their performance across a wide range of potential future scenarios, not simply on what is considered the most likely future. In Section 6.2 we review a range of DMDU methods that have been developed to support risk management in deeply uncertain problems.

A central theme in our review is the need for more inclusive discourse in the framing, analysis and evaluation of highly complex and uncertain risk problems (Level 3 and 4 uncertainty in Table 1). Building trust in the analysis and management of risks requires greater transparency of the assumptions and judgements that underly them and better methods for facilitating participation and scrutiny by stakeholders, decision-makers and other experts in the design of analyses. As the complexity of the risk problem increases, so too does the range of interests that could be negatively affected and the diversity of perspectives on the risks and their management, necessitating increased roles for non-specialists in the governance of decisions about risk (Table 1). This will require more frequent involvement of skilled facilitators and structured participatory processes in risk assessments to enable framing of the risk problem, inputs to the choice and application of analytical tools and to facilitate evaluation of management options. Our review describes a range of participatory processes for informing probabilistic risk analyses (Section 3.1), eliciting and aggregating judgements about risk from diverse groups (Section 4.2), assessing cumulative risks (Section 5), and identifying 'robust' policies that will perform well over a wide range of plausible, uncertain future states of the world (Section 6.2).

Table 1: The spectrum of uncertainty and its relationship to risk analysis and stakeholder participation in risk governance. Adapted from (Aven & Renn 2009b; Kwakkel et al., 2010a; Walker et al., 2013b).

		Levels of uncertainty					
		Level 1	Level 2	Level 3	Level 4a	Level 4b	
Description		A single future state	Able to identify multiple alternatives and estimate probabilities (subjective or objective)	Able to identify multiple alternatives and rank order them in terms of perceived likelihood.	Able to identify multiple alternative future states without being able to rank order them in terms of how likely or plausible they are judged to be.	Unable to identify multiple alternatives while admitting the possibility of being surprised	
Nature of the risk	**Complete certainty**	Relatively well understood from precedents.	Knowledge and data adequate to describe the system probabilistically. Some scientific disagreement about how the cause-effect relationships should be described.	Complex cause-effect relationships between events and consequences. A range of potentially affected values that must be considered.	Unresolved complexity and uncertainty in cause-effect relationships. A range of different viewpoints about the relevance of technical information and predictions about risks and their management.	Surprising extreme events that cannot be predicted from our present knowledge, understanding or beliefs.	**Total ignorance**
System model		A single (deterministic) system model	A single (stochastic) system model	A few alternative system models	Many alternative system models	Unknown system model	

	Levels of uncertainty				
	Level 1	**Level 2**	**Level 3**	**Level 4a**	**Level 4b**
Types of analysis	Statistical models Deterministic risk models Sensitivity analysis of model parameters. Cost-benefit analyses	Quantitative risk analysis Bayesian (subjective) probabilities Some stochastic model parameters Tools for seeking consensus among experts (e.g., Delphi, IDEA protocols) Scenario-based optimisation	Quantitative and semi-quantitative risk analysis Scenario-based optimisation	DMDU methods Conflict resolution methods for reaching consensus on risk evaluation and management options	DMDU methods with explicit consideration of surprises
Level of stakeholder participation	Little dispute about values to be protected. Instrumental discourse aimed at finding the most cost-effective measures to make the risk acceptable or at least tolerable.	Epistemological uncertainty addressed through discourse aimed at obtaining the best estimates of the occurrence of events and their consequences	Reflective discourse. Stakeholders are involved in framing the problem and in evaluating the appropriate level of protection that they would be willing to accept to avoid potentially catastrophic consequences. Deliberation reflects on the relative consequences of over- and under-protection.	Participative discourse openly discusses competing arguments, beliefs and values about the risk problem and evaluates how well different management options are able to achieve equitable outcomes that are robust to uncertainty.	Participative discourse openly discusses strategies for reducing vulnerability to unforeseen hazards

1 Introduction

Decisions about the future utilisation of natural resources require predictions about how ecosystems will respond to prospective changes in use. Predicting the future state of complex natural systems where there are multiple stressors and interests involved is necessarily uncertain and contentious (Office of the Prime Minister's Chief Science Advisor 2016a).

Risk assessment is a process that is used to support decision-making about uncertain future events and their consequences for society. There are many different approaches to risk assessment, each involving different ways of structuring the collection and analysis of information that are tailored to the problem under consideration and the kinds of data and knowledge that can be applied to it (Burgman 2005). These range from simple, qualitative assessments of risk in which evaluations of uncertain outcomes are expressed as ratings or categories such as 'high', medium and 'low' (EFSA Scientific Committee et al., 2018a; OIE 2019), to quantitative assessments that use data or 'expert' judgements to parameterise a model to describe relationships between an activity and future uncertain outcomes, and approaches that explore and evaluate a broad range of possible future scenarios (Aven & Zio 2011).

Although risk assessment and the explicit treatment of uncertainty are recognised as important scientific inputs into ecosystem-based management of natural resources (Levin et al., 2009) there can be disagreement about how (and sometimes if) information about risks should be used in decision-making. To some extent, this reflects a measure of distrust in the outputs derived from technical assessments of complex risks (Slovic 1999). Risk problems can be characterized according to the level of knowledge that we have about a potentially hazardous event and its consequences (Aven & Renn 2009b). Where the causal relationships between the event and its possible negative impacts are well-established, as a result of past experience and observations, there is typically little dispute about what values are at stake, who will be affected and how risks should be assessed and managed (Oil & Gas UK 2014). In everyday life, well characterized risks include regularly occurring events such as car accidents, food and health hazards or landslides and floods. In each of these cases, the prospects for future events and their effects can be estimated reliably from data on past events using statistical analysis or well-validated models of the cause and effect relationships. However, when the hazard is novel or rare, when its effects are less certain or are indirect, and when existing data and observations are inadequate to predict future outcomes, the choice of approach to assess risk and the range of values and groups that could be affected is less clear.

For risk problems that are highly complex and uncertain or where there are few precedents, risk assessments have relied heavily on technical experts to determine the most appropriate methods for identifying and analysing risk and to bridge the gaps in observations and data (Aven & Renn 2009b). In these circumstances, validation of the technical assessment becomes problematic (Renn 2008). Scientists, stakeholders and decision-makers can have very different views of the legitimacy of the analysis and its implications for decision-making. Building trust in the analysis and management of risks requires greater transparency of these judgements and the assumptions that underly them and better methods for facilitating participation and scrutiny by stakeholders, decision-makers and other experts in the design, implementation and evaluation of the analysis (Burgman 2005; NASA 2010; Office of the Prime Minister's Chief Science Advisor 2016a).

In this report we review a range of methods that have been developed to support risk assessments of novel and complex hazards where there is a paucity of existing data and experience. We begin by exploring the concepts of risk and uncertainty as they are applied to risk assessment (Section 2) and discuss the increasing role of participatory processes to deal with decision-making about uncertain hazards (Section 3). Recent treatments of risk management distinguish between approaches where the purpose of the risk analysis is to provide the best available predictions of the occurrence of uncertain events and their associated consequences (‘probabilistic’ risk analyses) and those where the focus is on identifying and reducing vulnerability to threats by ensuring that decisions are adaptable, robust to uncertainty and can withstand surprises (Renn 2008; Marchau et al., 2019a). Our review covers both approaches. In Section 4, we describe methods that can be used to reduce bias in expert judgements about the probability of uncertain events. Section 5 describes the use of semi-quantitative approaches (*sensu* Aven (2008)) to structure probabilistic assessments of the cumulative and indirect effects caused by multiple stressors. By combining qualitative and quantitative analysis methods, semi-quantitative risk analyses allow a detailed conceptual model of the problem to be developed that incorporates complex relationships among system components and enables the rigour of quantitative assessments to be focussed on sub-components where there is the greatest uncertainty. In Section 6 we describe analyses that evaluate management scenarios in the context of risk and uncertainty. Section 6.1 describes methods designed to analyse and rank a set of decision alternatives against stated objectives for the resource. Methods described in Section 6.2 are designed to identify the vulnerabilities in management decisions – things that could go wrong if the assumptions are incorrect - and to evaluate options that best reduce susceptibility to uncertain events and hedge against the possibility of policy failure (i.e., to make it more robust to uncertainty). In the final section, we provide summary guidance on the types of risk problems that are most amenable to each approach and their advantages and constraints.

2 Understanding the concepts of risk and uncertainty

2.1 What is risk?

The concept of risk has had application across many facets of life, including safety, security, engineering, finance, human health, environmental management, and politics. Its varied use has meant that there are many ways that 'risk' has been defined within a specific context and a universally agreed upon definition has proved elusive (Haimes 2009; Aven 2010).

Fundamentally, risk occurs when the something of value is at stake and the outcome is uncertain (Rosa 2003). Common among most definitions of risk are at least three elements:

- an event or occurrence that may take place,
- the potential consequences (or outcomes) of that event and
- the likelihood that the event and its consequences will eventuate.

These three components are embodied in a simple triplet of questions proposed by Kaplan & Garrick (1981) to define risk:

- What can go wrong?
- If it does happen, what are the consequences?
- How likely is it that those consequences will eventuate?

Most situations that involve risk are, however, characterized by more than just a single possible future outcome. For example, the consequences of a hazardous event like an earthquake are influenced by (at least) the type of fault, its magnitude, geological location (i.e., depth, orientation of the fault, bedrock type, etc), proximity to human populations, and by the vulnerability of those populations to the effects of ground shaking or tsunami (e.g., presence of building strengthening, level of preparedness and emergency response, etc). In most complex, real-life situations, risk problems are inherently multi-dimensional.

For any given event or decision context, there are usually many plausible scenarios for 'what could go wrong', each with a different likelihood of occurrence and with a variety of social, economic, environmental and cultural things of value that could be affected (Aven & Zio 2011). In this view, the analysis of risk can be conceived collectively as identifying a set of *possible* scenarios, each of which has some likelihood that it will occur and consequences that would eventuate (NASA 2010; Aven & Zio 2011).

In formulating their "triplets" definition of risk, Kaplan & Garrick (1981) recognised this complexity by defining risk as the 'complete set of triplet' scenarios that could be envisioned, including their probabilities and consequences. The complete set of possible risk scenarios is, of course, infinite so that in most practical risk assessments the scenarios that are considered are a select subset of the full range of possibilities (Kaplan et al., 2001; NASA 2010; Aven & Zio 2011).

Conventional risk assessment typically includes activities that seek to:

1. define a set of adverse consequences (risk end points) that could eventuate,
2. identify a sequence of actions or "events" for each end point that if unmanaged could lead to it being realised ("risk scenarios"),
3. use data from similar events, model simulations, past experience or expert judgment to evaluate the probabilities of these scenarios,
4. rank the scenarios according to their expected likelihood of occurrence and the severity of their consequences (Apostolakis 2004).

The International Organization for Standardization (ISO) recently revised its definition of risk to shift the emphasis from a focus on the possibility of an *event* to the possibility of an *effect* on organizational objectives ((Purdy 2010),Table 2), where the objectives may be financial, health and safety, or environmental goals that apply at different levels (ISO 2009; ISO 2018). Supporting notes expand the definition by observing that risk "is often characterized by reference to potential events and consequences or a combination of these" and is "often expressed in terms of a combination of the consequences of an event and the associated likelihood of occurrence".

Table 2: Risk terminology used in the ISO 31000 standard on risk management[1] and ISO guide 73 on risk terminology[2].

Term	Definition
Consequence	Outcome of an event affecting objectives
Event	Occurrence or change of a particular set of circumstances
Likelihood	The chance of something happening
Objectives	Objectives can have different aspects (such as financial, health and safety, and environmental goals) and can apply at different levels (such as strategic, organization-wide, project, product and process).
Risk	The effect of uncertainty on objectives. Risk is often characterized by reference to potential events and consequences or a combination of these. Risk is often expressed in terms of a combination of the consequences of an event (including changes in circumstances) and the associated likelihood of occurrence.
Risk analysis	The process to comprehend the nature of risk and to determine the level of risk
Risk assessment	The overall process of risk identification, risk analysis and risk evaluation
Risk evaluation	The process of comparing the results of risk analysis with risk criteria to determine whether the risk and/or its magnitude is acceptable or tolerable
Risk identification	The process of finding, recognizing and describing risks
Risk management	Coordinated activities to direct and control an organization with regard to risk.

Term	Definition
Uncertainty	The state, even partial, of deficiency of information related to, understanding or knowledge of, an event, its consequence, or likelihood

[1]ISO (2018) ISO 31000:2018 Risk Management - Principles and Guidelines. International Organization for Standardization, Switzerland.
[2] ISO (2009) Guide 73. Risk management — Vocabulary. International Organization for Standardization, Switzerland: 15.

This definition of risk, which was adapted by ISO from the Australia/New Zealand Standard for Risk Management (Standards Australia & Standards New Zealand 2009), has also not received universal acceptance. It has been criticised by some practitioners as inconsistent with common usage, and imprecise (Leitch 2010; Aven 2011b). Greatest criticism has been directed at its emphasis on objectives rather than the societal consequences of events or policy decisions, with its implicit suggestion that risk management is an optimization process to support the achievement of organizational objectives by modifying the magnitude and likelihood of potential consequences, both positive and negative, to achieve a net benefit.

2.2 A typology of uncertainty

'Risk' and 'uncertainty' are intimately related concepts. Most definitions of risk include uncertainty as a component of the definition (Aven 2010). Like 'risk', the term 'uncertainty' also has many meanings and interpretations that depend on the context in which it is applied (Kwakkel et al., 2010a). In general usage, it describes the limits of our knowledge about past, future or current events (Walker et al., 2013b). The ISO standard for risk management defines 'uncertainty' as "the state, even partial, of deficiency of information related to, understanding or knowledge of, an event, its consequence, or likelihood".

Knight (1921) distinguished between those aspects of uncertainty that could be measured or estimated (which he defined as 'risk' but is more appropriately considered quantifiable or probabilistic uncertainty) and those which cannot (now referred to as 'Knightian uncertainty'; (Knight 1921)). According to Knight's definition, 'risk' (quantifiable uncertainty) can be estimated using data or existing knowledge. Uncertainties that cannot be resolved probabilistically include uncertainty about the structure of the model used to describe the decision problem and situations in which experts cannot agree upon the probabilities and their distributions (Kwakkel et al., 2010a). Other typologies have expanded on the types of uncertainty that can contribute to environmental decision-making (Kahneman & Tversky 1982; Regan et al., 2002; Ascough II et al., 2008; Aven 2011a; Hayes 2011; Walker et al., 2013b). Most distinguish four main sources:

1. *Linguistic uncertainty* - uncertainty that arises through the imprecise use of language to frame a problem,

2. *Epistemic uncertainty* - the uncertainty created by our limited understanding of complex natural systems,

3. *Variability* - the uncertainty created by the natural variability within these systems (sometimes referred to as 'aleatory' uncertainty), and

4. *Decision uncertainty* - the uncertainty associated with our value systems and management decisions (Regan et al., 2002; Ascough II et al., 2008; Hayes 2011).

Linguistic uncertainty refers to the vagueness, ambiguity, context dependence, and imprecision of language used in risk analysis and decision making (Regan et al., 2002). Ambiguity can arise when a word or phrase can be interpreted in different ways. Qualitative descriptors of the likelihood of events or the severity of consequences from an event may be interpreted in different ways by different people (see Section 4.3.3). Ambiguity can also occur because some words have more than one meaning and it is not clear which one is the true meaning (Hayes 2011). People's understanding of the descriptive terms used in risk assessment is also affected by the context of the description (Patt & Dessai 2005).

Epistemic uncertainty is the uncertainty associated with our current state of knowledge of complex events or systems. It can be due to insufficient information on, or our understanding (i.e., state of knowledge) of the system process ('process uncertainty' or 'functional uncertainty; (Kinzig & Starrett 2003)) or due to an inability to model or measure the system accurately ('model uncertainty'). Although epistemic uncertainty is sometimes characterized as the component of uncertainty that can theoretically be reduced by acquiring new information (Ascough II et al., 2008), this is not necessarily the case. New knowledge may reveal more about complex systems and processes, including new relationships and uncertainties that were not previously known or considered (Walker et al., 2013b).

Processes within natural ecosystems are complex and involve interactions between biophysical and human systems. Insufficient knowledge of the scale and scope of interactions within these systems is often an impediment to our understanding. For example, ignorance of important functional relationships between perturbations to the system and its response allows for the possibility of sudden transitions (often undesirable) from one system state to another. We are often uncertain about how much disturbance could trigger these transitions and how these thresholds ('tipping points') may be influenced by multiple interacting drivers and interactions (Kinzig & Starrett 2003).

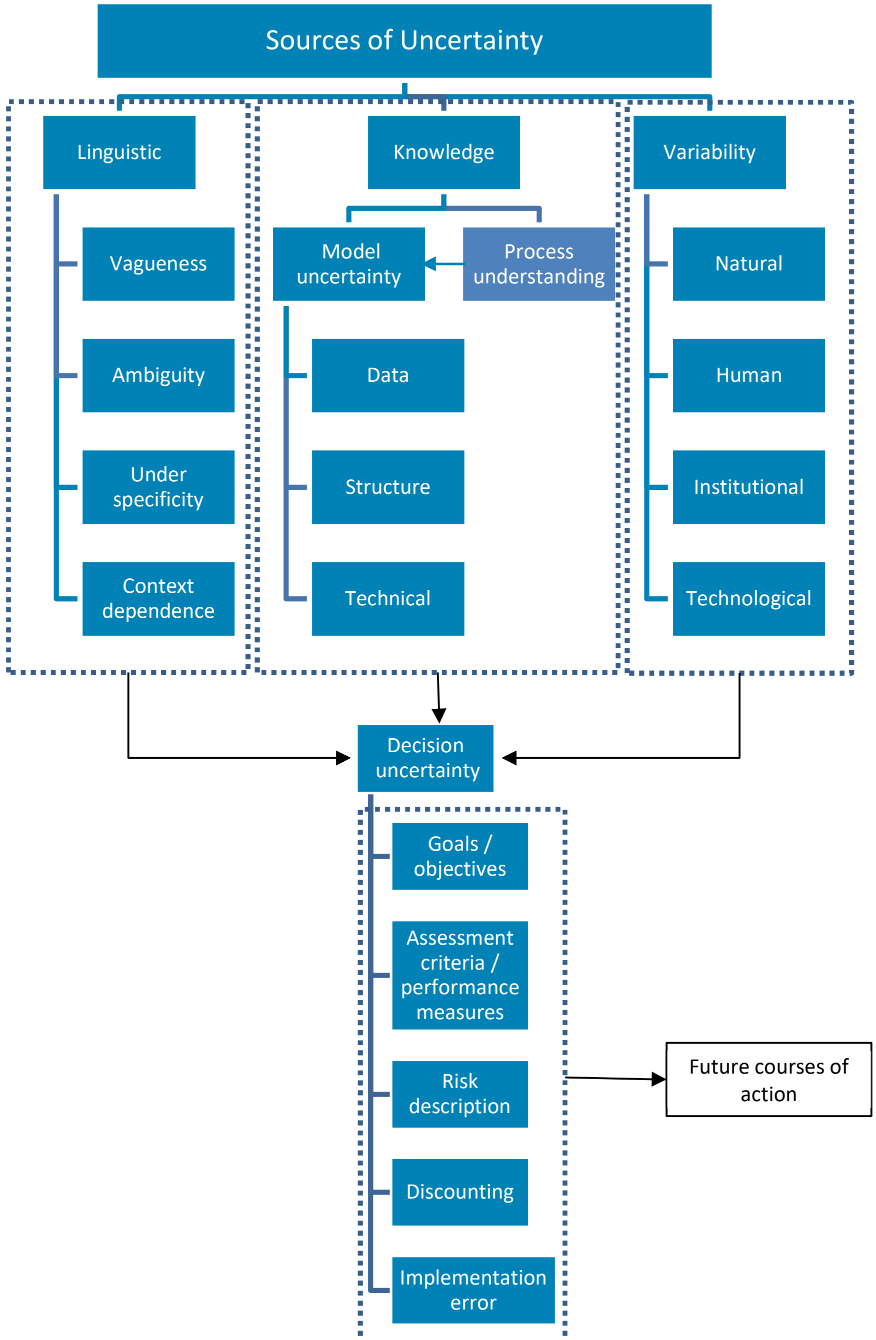

Figure 1: A taxonomy of uncertainty (based on Hayes (2011) and Ascough II et al., (2008))

System models are important tools to support decisions about risk. They help us understand complex processes and can allow experiments and predictions of how the system may behave in response to interventions that are not possible in real life. System models can take a variety of forms. They can be a conceptual formulation of the structure of the problem being considered, a statistical model that describes patterns in data or a mathematical model (algorithm) that characterizes cause-effect relationships within the system to simulate its behaviour under different circumstances. As models are abstractions of reality, they necessarily simplify complex systems. *Model uncertainty* can, therefore, arise from: (1) the structure of the model that is used for analysis, (2) the data that used as inputs, and (3) technicality of the model, which is uncertainty arising from the implementation of the model, usually by computer (Walker et al., 2003). *Structural uncertainties* may occur when components of the system are overlooked or when the relationships among components are inadequately characterized. In addition, alternative models might be available for the same processes and systems that incorporate a different structure, functional relationships and assumptions. This implies that one of the available models might be a more correct representation of the real-world process (or none of them may be).

Data plays a key role in environmental decision making. Data can be obtained through monitoring a process (quantitative) or through our understanding and experience about a process (qualitative). If data are collected through observations (quantitatively), uncertainty can arise from the measurement process (e.g., sampling errors or type of instrument used) and the type of data recorded, such as the unit, recorded time duration, and spatial scale, of the collected data. Subjective data, based on the knowledge or judgements of experts, is often used when observations are not available (see Section 4). Elicitation converts personal beliefs into data or inputs that are used in risk models. Elicitation can be conducted in a direct or predictive form. The direct form will specify the system of interest and then ask the expert to estimate the value(s) for the process under investigation, while the predictive form will specify a future scenario that describes a hypothetical behaviour of the process and then ask the expert to provide values of the system outcome. In the elicitation process, epistemic uncertainty can arise from judgements provided by the experts, as well as from fitting and aggregating the judgements. Structured elicitation methods such as Sheffield method, Cooke's method, and Delphi method, are often used to minimize uncertainty from the elicitation process (see Section 4).

Technical uncertainty can arise from errors from the software or hardware used. Software errors arise from defects in software, design errors in algorithms, and typing errors in model source code. Hardware errors could arise from flaws in the technical equipment (Ascough II et al., 2008).

Variability ('aleatory' uncertainty) is a distinctive type of uncertainty as it refers to natural heterogeneity and fluctuation in the real world (EFSA Scientific Committee et al., 2018a). Variability could be categorized as natural, human, institutional, and technological variability. For instance, the random variation in populations across time and space such as temporal variation in birth, death, and growth rates of organisms and the variation of chemical concentrations in a type of food are examples of natural variability. Human variability could arise due to different beliefs and experiences while institutional variability could be due to the political system. From a decision-making point of view, it is important to know two key points. First, unlike epistemic or linguistic uncertainty which can be reduced by additional research or more accurate wording, natural variability cannot be reduced. Second, levels of uncertainty in natural variability may themselves vary. For instance, human life span is more certain than chemical concentrations in food (EFSA Scientific Committee et al., 2018a).

Decision uncertainty. In the context of policy decisions, uncertainty describes the gap between the best knowledge available to the decision makers and the knowledge needed (in an ideal world) to make the optimal policy choice (Marchau et al., 2019b). It arises due to uncertainty about how social objectives should be valued and weighted (i.e., the relative importance of different and often conflicting economic, environmental, social and cultural values). In addition, the way in which outcomes from the risk analysis are communicated and presented to decision makers can contribute to this type of uncertainty.

Significant uncertainty in decision making is also related to how decision makers define an acceptable level of risk. People's perceptions of risk are affected by their own awareness and knowledge of hazards, as well as the individual and normative values and beliefs about the desirability of different outcomes (Office of the Prime Minister's Chief Science Advisor 2016a). A risk averse person will prefer to forgo some expected return for a reduction in risk that represents a trade-off (Kahneman 2011). Decision makers can have different attitudes towards risks. Risk averse behaviours may mean that decision makers' opt for conservative measures that are not optimal or the most preferred choice (Hardaker et al., 2004; Finnoff et al., 2007). Others may be more inclined to accept greater risk. The degree of risk aversion can also be different under different circumstances. For example, decision makers can be extremely risk averse and prefer to have lower benefits with lower risks, while others might be less risk averse and their decisions are only slightly influenced by risky activities.

Deep uncertainty is a term that has been used to describe situations where these multiple sources of uncertainty coincide in complex policy decisions. Lempert et al., (2003) defines 'deep uncertainty' as:

"*the condition in which analysts do not know or the parties to a decision cannot agree upon:*

- *the appropriate models to describe interactions among a system's variables,*
- *the probability distributions to represent uncertainty about key parameters in the models and/or*
- *how to value the desirability of alternative outcomes."*

These conditions arise when past decisions and data have limited value in predicting future events, when experts and stakeholders disagree about the likely consequences of different policies and how to weight them, and when parties to a decision cannot agree about the best course of actions to reduce risk and increase benefits to society (Cox 2012). Analysts, stakeholders and decision-makers can have very different views of the nature of the decision problem and of the system being considered, the environmental values to be protected and the desired future outcomes. This can occur where there are conflicting opinions, beliefs, and values about the system being considered. Deep uncertainty can also arise from complex, dynamic systems that change over time in unpredictable ways. In these circumstances, it can be difficult to specify appropriate system models or to parameterize them and the legitimacy of outputs from conventional probabilistic risk analysis becomes the source of debate.

In the marine environment, predicting the future effects of climate change, the effects of multiple stressors on natural ecosystems or the risks associated with novel technologies in underexplored environments such as the deep ocean, can be characterized as deeply uncertain policy settings. In these situations, the cause-effect relationships between the stressor(s) and change in the system are not unknown or are not well characterized and, consequently, there is potential for a range of possible negative effects on valued ecosystem components. The variety of social, economic and cultural interests in natural marine resources means that this complexity is compounded by disparate views on the level of protection that should be afforded to different ecosystem components and the costs that this might entail to society. Section 6.2 describes a range of methods that have been developed to identify policy settings that are robust to deep uncertainty (Marchau et al., 2019a).

2.2.1 Levels of uncertainty and their analysis

Walker et al., (2003) attempted to synthesise these various typologies and concepts of uncertainty into a single framework that comprised three dimensions:

- the location of uncertainty (i.e., where uncertainty occurs in the analysis framework),
- the level of uncertainty - an expression of the degree of uncertainty, and
- the nature of the uncertainty - whether the uncertainty is due to lack of knowledge (i.e., is 'epistemic') or to inherent variability in the phenomena ('aleatory' or 'stochastic' uncertainty).

Subsequent treatments have expanded on this framework to define five levels of uncertainty between the extremes of total certainty and complete ignorance (summarized in Table 3). The levels reflect increasing complexity in the problem under consideration that can entail different states of knowledge (and ignorance) of the system and its future behaviour, the model(s) used to assess risk, the outcomes that are sought by different stakeholders and the importance that they place on them (Kwakkel et al., 2010a; Walker et al., 2013b).

Level 1 uncertainty ('shallow uncertainty') occurs when the system is relatively well-known and predictable and can be described with a trusted model of its behaviour (which might simply be past behaviour). Uncertainty at this level can be explored through a simple sensitivity analysis of model parameters, where the impacts of small perturbations of model input parameters on the outcomes of a model are examined (Marchau et al., 2019b).

Level 2 uncertainty describes the situation in most conventional quantitative risk analyses where it is expected that the tools of probability and statistics can be used to resolve uncertainty in predicting future states and outcomes. In this situation, knowledge or data are considered adequate to describe the system model and its inputs probabilistically, or there are only a few likely future outcomes that can be predicted well enough (and to which probabilities can be assigned). In this case, the system model is trusted to estimate the probability distributions for the outcomes of interest accurately. Model outputs allow a policy to be chosen based on the expected outcomes and their associated likelihoods.

Level 3 uncertainty describes situations in which a limited set of plausible future states, system models, outcomes and weights can be identified. Available knowledge allows for several different possible parameterizations of the system models, alternative sets of outcomes and several alternative sets of weights that can be attributed to them. In these, more complex situations, alternative future states can be rank ordered in terms of their relative likelihood, but it is usually not possible (or appropriate) to quantify their likelihood probabilistically (Kwakkel et al., 2010a). Instead, scenario analysis is typically used to examine what could plausibly happen with implementation of each of several policy options. The optimal policy is the one that produces the most favourable outcomes across the scenarios (Marchau et al., 2019b).

Level 4 uncertainty describes situations of 'deep uncertainty' where there are many possible future states that the system could assume, and it is not possible to rank their likelihood or desirability reliably. This may be due to a lack of knowledge about the functional relationships being studies or because parties to the decision cannot agree on the importance to be given to each outcome. Consequently, it is difficult to specify appropriate models to describe functional relationships among system variables or to select probability distributions to represent key uncertainties.

Marchau et al., (2019b) also distinguish events that are outside the realm of our present experience and beyond our expectations. Surprising, extreme events have been labelled 'black swan' events (Taleb 2007). They differ from high consequence events with low probability in that 'black swan' events cannot be predicted from our present knowledge, understanding or beliefs (Aven 2013). For example, although the emergence of the global COVID-19 pandemic in 2020 caused severe financial and social shock across the world and highlighted a lack of preparedness to deal with risks from novel infectious diseases, it was not unanticipated and could not, therefore, be considered a 'black swan' event. This category of deep uncertainty is reserved for extreme events that are surprising and only explainable retrospectively. This is because nothing in our current knowledge suggested their possibility (Aven 2013).

Table 3: **Progressive levels of uncertainty.** After.(Kwakkel et al., 2010a; Walker et al., 2013b)

	Complete certainty	Level 1	Level 2	Level 3	Level 4a	Level 4b	Total ignorance
Description		A single future state	Able to identify multiple alternatives and estimate probabilities (subjective or objective)	Able to identify multiple alternatives and rank order them in terms of perceived likelihood.	Able to identify multiple alternatives without being able to rank order them in terms of how likely or plausible they are judged to be.	Unable to identify multiple alternatives while admitting the possibility of being surprised	
Context		A clear enough future	Alternate futures (with probabilities)	A few plausible futures	Many plausible futures	Unknown future	
System model		A single (deterministic) system model	A single (stochastic) system model	A few alternative system models	Many alternative system models	Unknown system model	
System outcomes		A point estimate for each outcome	A confidence interval for each outcome	A limited range of outcomes	A wide range of outcomes	Unknown outcomes	
Weights		A single set of weights	Several sets of weights, with a probability attached to each set	A limited range of weights	A wide range of weights	Unknown weights	

3 The role of risk assessment in decision-making

In general, the assessment of risk encompasses three types of activities (Aven & Renn 2009a; ISO 2009; Office of the Prime Minister's Chief Science Advisor 2016a):

- Identification of potential hazards and the sequence of events ('scenarios') that could lead to harm ('risk identification'),
- Characterizing potential consequences of the event(s) (both harmful and beneficial) and estimating the likelihood that they could occur ('risk analysis'), and
- Evaluating the desirability (or acceptability) of the consequences if they were to occur ('risk evaluation').

Risk assessment is intended to inform the management of risk. Risk management can include actions taken to avoid the occurrence of hazards or to reduce the harm that they could cause (Aven & Renn 2009b). Most early frameworks for using information from risk assessments in decision-making conceived a staged approach to the assessment in which risk identification and analysis were undertaken by technical specialists (scientists and analysts) and evaluation of the benefits and costs of different possible outcomes (a social process) and design of risk mitigating actions to achieve the best outcomes (a policy process) were the domain of risk managers (National Research Council 1982). Risk assessment and risk management were, therefore, conceived as distinct processes in which there was directional flow of information from the scientific process of analysis to the value-laden process of management (Jasanoff 1993; Amendola 2002).

An example is provided by the US Environmental Protection Agency (EPA) framework for Ecological Risk Assessments (Figure 2, U.S. Environmental Protection Agency (1992)). In it, risk assessment and risk management are conceptually separate processes with the former undertaken predominantly by scientists or technical experts, with some input from the risk manager to set the policy and regulatory context and objectives for the assessment. The results of the risk assessment are used as input to the risk management process, along with other inputs such as social and economic concerns to evaluate risk management options (U.S. Environmental Protection Agency 1992).

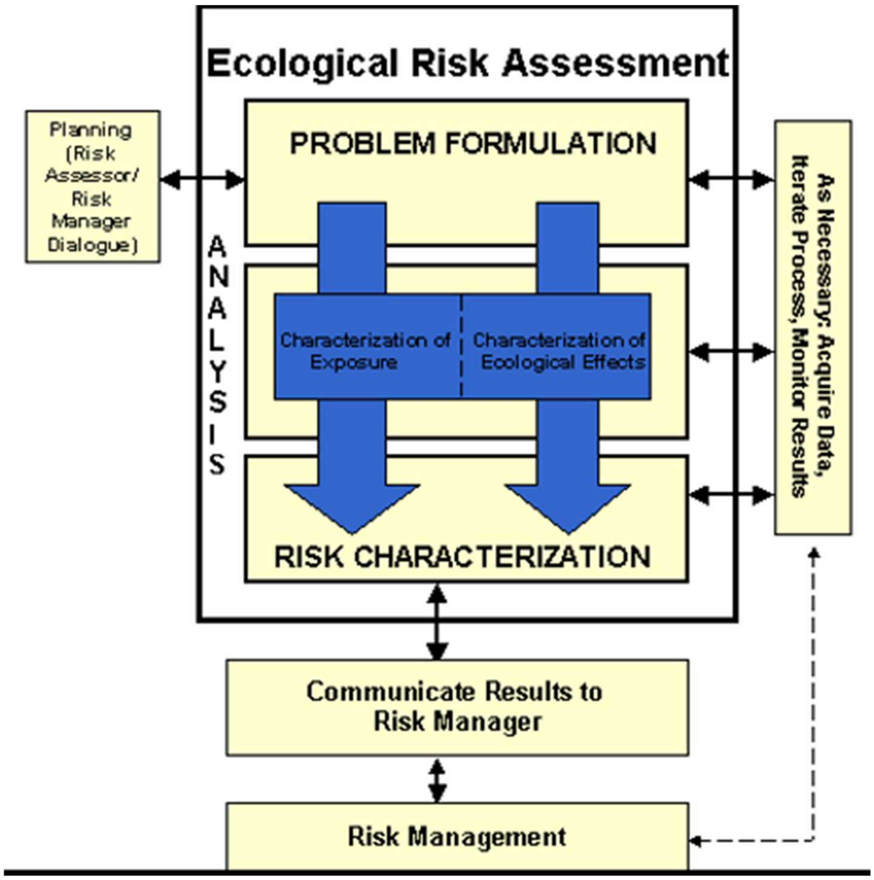

Figure 2: US EPA Framework for Ecological Risk Assessment. Source: (U.S. Environmental Protection Agency 1992).

This concept of risk assessment, as a largely technical process, makes little allowance for incorporating other perspectives and knowledge in framing the risk problem or in its analysis and evaluation. As the guidance document for the EPA framework noted, assessment of complex risk problems is almost invariably based on incomplete data and information. Because of this uncertainty, professional judgement plays a large part in the analysis (Box 1; U.S. Environmental Protection Agency (1992)). These judgements can be about the processes that are important for driving change in the system under consideration, choice of appropriate models to describe the system and potential hazards, the usefulness of different sources of data and the significance of predicted outcomes.

The original EPA framework and other, similar approaches emphasised the need for specialized scientific expertise to make these judgements (Box 1). Others have taken a different view. For example, nearly 40 years ago the US National Research Council Committee on Risk and Decision Making (NRC) acknowledged that decisions about risk almost invariably have a social context that is characterized by parties with diverse interests and stakes in the outcome (National Research Council 1982). Scientists, stakeholders and decision-makers can have very contrasting views of the nature of risk problems. Each brings different values, knowledge and experiences that shape their perspective on risk, the weighting that is given to protecting different values and, therefore, the trust that they have in outputs from the technical analyses (Amendola 2002). If these different perspectives are not considered in the analysis, the outcomes are likely to be challenged by one or more parties. Moreover, the judgements of scientists and other technical experts are not always accurate or appropriate, particularly when it comes to resolving the impacts of policy choices on conflicting values. As the NRC observed, the appropriate policy choice may not be obvious even after uncertainties in natural variation have been quantified (National Research Council 1982).

Even within the technical risk analysis experts frequently disagree about the best ways to approach the bulleted items in Box 1. Their judgements about risk, like those of non-specialists, are not immune to contextual biases and heuristics that are not always explicit in the analysis (Tversky & Kahneman 1975; Burgman 2015).

Box 1: The role of professional judgement in ecological risk assessments. (Source: U.S. Environmental Protection Agency 1992)

"Ecological risk assessments, like human health risk assessments, are based on scientific data that are frequently difficult and complex, conflicting or ambiguous, or incomplete. Analyses of such data for risk assessment purposes depends on professional judgment based on scientific expertise. Professional judgment is necessary to:

- *design and conceptualize the risk assessment;*
- *evaluate and select methods and models;*
- *determine the relevance of available data to the risk assessment;*
- *develop assumptions based on logic and scientific principles to fill data gaps; and*
- *interpret the ecological significance of predicted or observed effects.*

Because professional judgement is so important, specialized knowledge and experience in the various phases of ecological risk assessment is required. Thus, an interactive multidisciplinary team that includes biologists and ecologists is a prerequisite for a successful ecological risk assessment."

For risk decisions where there is high uncertainty and potentially high consequences there have been calls for more participatory processes that include stakeholders and decision-makers in constructing and reviewing the analyses (Burgman 2005; NASA 2010; Office of the Prime Minister's Chief Science Advisor 2016a). Funtowicz & Ravetz (1994) call this an extension of the 'peer community'. Rigorous quantitative analysis is still an important part of the assessment, but there is a need for better integration of the perspectives of stakeholders so that the dimensions of risk considered in the assessment and the underlying assumptions are transparent and accepted (Amendola 2002; Office of the Prime Minister's Chief Science Advisor 2016a). A considered and flexible process for reviewing the technical analysis of risk can improve its quality and acceptance (National Research Council 1982). Interactions among these different groups allows the scientific knowledge to be contextualized and for other forms of knowledge and values to be integrated into the assessment better to build consensus around the risk models, their outputs and subsequent policy choices (Horlick-Jones 1998).

3.1 Risk-informed Decision Making (RIDM)

Risk-Informed Decision Making describes a general approach that incorporates deliberation among stakeholders, risk analysts, subject matter experts, and decision-makers throughout the process to ensure that a range of objectives, values, knowledge and alternative strategies are considered in the technical risk analysis (Amendola 2002).

Organizations that deal with complex risk decisions like the US National Aeronautics and Space Administration (NASA), the US Nuclear Regulatory Commission and others have developed guidelines for incorporating RIDM into their operations (Drouin et al., 2009; NASA 2010; Gregory et al., 2012a; Zio & Pedroni 2012a). The NASA approach integrates two complementary processes: RIDM and Continuous Risk Management (Figure 3). The RIDM process is concerned with selecting from decision alternatives to meet agreed management objectives and the CRM process implements the selected alternative(s) with associated performance monitoring to ensure that the objectives are being met. Collectively, this integrated approach shares many features with Structured Decision Making (SDM), a participatory process that has been advocated for deliberations about natural resource management decisions (Gregory et al., 2012b). Participating decision-makers, analysts and stakeholders collaboratively define the decision context ('problem framing'), establish objectives and performance measures for the outcomes, design and evaluate management alternatives, and implement short term actions with associated monitoring of performance so that the management strategy can be modified over time. A similar, participatory approach to risk analysis and management has been proposed by the International Risk Governance Council (Renn 2008).

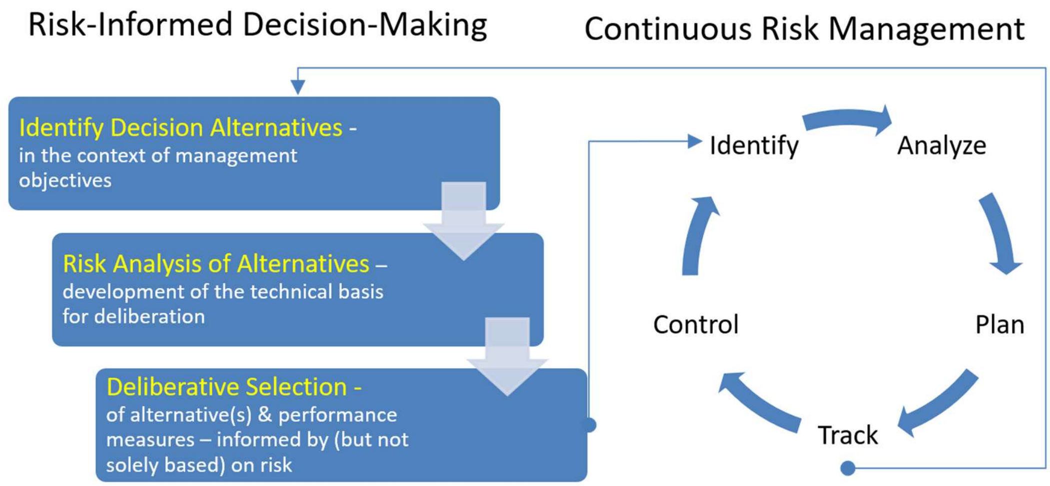

Figure 3: NASA's risk management framework integrates Risk-Informed Decision Making and Continuous Risk Management. Source: NASA (2010)

Participatory processes can be time-consuming and expensive to implement. As in other decision processes, the depth of analysis is scaled according to the stakes and complexity of the decision situations being addressed (Aven & Renn 2009b). In the NASA framework, the RIDM process is invoked when decisions have one or more of the following characteristics (NASA 2010):

- Potentially high consequences such as significant costs or serious outcomes for human health and the environment
- High complexity in the system or event being evaluated
- Significant uncertainty in the outcome of the decision alternatives and the risks that may need to be managed
- Many attributes that may require formal analysis, and/or
- A diversity of stakeholders with values, preferences, and perspectives.

RIDM at NASA consists of three parts (Figure 3):

5. Identification of decision alternatives ('decision framing')

The purpose of this stage is to understand the decision context and to a derive a set of objectives and measures for the performance of the system and feasible decision alternatives. It begins with a process of understanding the expectations of stakeholders for the outcomes which are then developed into a set of performance objectives.

The second step in this stage is to canvass a comprehensive list of feasible decision alternatives through a discussion of reasonable options. These are tested against the performance objectives and distilled into a series of options that can potentially achieve the objectives and which warrant further investigation. It is important at this stage to avoid constraining the range of proposed alternatives prematurely. Stakeholders should feel able to contribute their own ideas about what constitutes an optimal solution.

In the NASA process, decision trees are suggested as a tool for visualizing and evaluating alternative decision pathways (NASA 2010).

6. Risk analysis of decision alternatives

The second step is the technical analysis of the decision alternatives (the 'risk analysis' stage). In the NASA process, this involves selecting risk analysis methods (usually supported by probabilistic modelling) that are appropriate for each domain represented in the objectives. These may be physical, ecological, social, or economic objectives.

Because RIDM requires the inclusion of all parties at an early stage to frame the decision problem and contribute to collective decisions about the analysis, the technical methods must be:

- able to be understood (at least in general terms) by non-specialists
- amenable to participatory conceptualization, parameterization and evaluation
- scientifically robust when applied to complex problems where there are knowledge gaps
- able to incorporate different types of qualitative and quantitative information, and
- relatively accessible.

The analysis stage also requires design of an integrated analysis so that the performance of each decision alternative can be assessed against each of the objectives identified in phase 1. The analyses should address key uncertainties in each domain (e.g., budget uncertainties, environmental variability, etc). As is standard practice in risk analysis, uncertainties in the values of each alternative's performance parameters should be identified and propagated through the analysis to produce ranges of uncertainty in the performance outcomes.

The main output of the risk analysis stage is a document that lists the set of candidate alternatives, summarizes the methods that were used to quantify the performance measures, and presents the results. NASA refer to this as the Technical Basis for Deliberation (TBfD) as it is the input that informs subsequent deliberations to support decision-making. The TBfD describes the overall risk analysis framework, the methods used for each discipline-specific analysis and the outcomes of the analyses, including:

- the scenarios that were used to evaluate each decision alternative
- the probability density functions of the marginal performance measure
- estimated risks relative to constraints imposed on the analysis
- a discussion of significant drivers that contribute to the risk, and
- a summary of the uncertainty analyses and sensitivity analyses used for each alternative.

Analysts are also required to provide an assessment of the credibility of their analysis. In the NASA process, this covers:

- Development of the model - including the origin and quality of any data used, whether the model(s) was implemented according to any standards or specifications, and whether the results compare favourably to any reference sets (validation)
- Operation of the model and simulations - including the origin and quality of data inputs used to parameterize the model, how uncertainty was characterized and propagated, and what sensitivity analyses were done to identify components with high leverage on the results
- Supporting evidence – how similar the model and application are to previous applications and
- Management – how the modelling process was managed.

Guidance for defining levels of credibility in each of these components and for visualizing the overall credibility assessment is provided by NASA (2016).

Because the TBfD is a key tool for communicating risk to stakeholders and decision-makers, it must provide information that is relevant to the decision process and easily understood. Use of different forms of visualization and statistical summaries are recommended to aid understanding of the technical outputs (NASA 2010).

7. Risk-informed alternative selection

In this step, stakeholders, risk analysts, and decision-makers debate the relative benefits and disadvantages of each decision alternative, given the information in the TBfD and, where necessary, other sources of information. This step is iterative and may require additional risk analyses or other information gathering as the participants assess the alternatives and identify those that they consider to be worthy of consideration by the decision-maker. It also involves development of 'performance commitments' that represent consistent levels of risk tolerance across the decision alternatives. These are measures of the participants' risk tolerance relative to the performance expectations of each decision alternative and are expressed as a percentile of the probability density function for each performance measure for each decision alternative (NASA 2010; Zio & Pedroni 2012b).

Once the parties to the decision have decided on a course of action (a 'decision alternative') they finalise the performance characteristics and measures that are expected from the chosen actions and document the decision rationale (in the NASA process, this is referred to as a 'Risk-Informed Selection Report'). This provides the input into the CRM stage. The purpose of the CRM is to monitor performance of the outcomes and, where necessary, to implement mitigations to reduce risk further or improve performance. Information that arises from the CRM on emergent risks or additional measures of performance can be fed back into the RIDM process (i.e., it is an adaptive process).

4 Methods for eliciting, representing and incorporating uncertainty in risk assessments

Although they have quite distinct meanings in statistics (Schweder & Hjort 2016), the terms 'likelihood' and 'probability' are used interchangeably within the risk literature. Aven (2011b) has suggested that 'likelihood' should be used in a broader sense to mean a measure for representing uncertainty about the occurrence of an event or consequence. Probabilities are the most common measure used for this purpose in risk analysis, but they are not the only way that uncertainty can be expressed.

In risk analysis, probabilities are interpreted in two ways: (i) in the classic sense of the proportion of times that an event would occur if the situation was repeated an infinite number of times (also called 'frequentist' probabilities) and (ii) as 'subjective' (or 'Bayesian') probabilities that represent a measure of epistemic uncertainty about future events or consequences as expressed by an assessor or "expert" (Aven 2010). 'Subjective' probabilities represent the beliefs of the expert(s) conditioned on prior knowledge and background information about the likelihood of the event. Because such probabilities are subject to range of heuristics and biases (Tversky & Kahneman 1975), they are open to challenge (Burgman 2005; Aven 2013a; Shortridge et al., 2017). The use of frequentist probabilities (i.e., where the probability of an event can be estimated objectively from observational data) is the exception rather than the norm in most risk analyses to support natural resource decisions (Burgman 2005; Aven & Renn 2009a; Cox 2012). This is because long-run, historical data that can be used to estimate the relative frequency of events are often lacking. Even where they do exist, their relevance to estimating the likelihood of future outcomes in complex, dynamic systems is sometimes questionable, particularly for very novel or low frequency events. Subjective probabilities are, therefore, used extensively in risk analysis for ecological systems.

In this section, we will describe best practice guidance on:

- reconciling qualitative descriptors of uncertainty as quantitative subjective probabilities (Morgan et al., 2009; Ho et al., 2015)
- protocols for eliciting subjective probabilities from experts to reduce known sources of bias (Burgman et al., 2006; Kuhnert & Barry 2010) and
- methods for eliciting and propagating highly uncertain probabilities in risk analysis.

The last of these components includes specific treatment of probability bounds analysis (PBA) (Ferson & Ginzburg 1996) and elicitation and use of imprecise probabilities (Walley 1990; Kriegler et al., 2009). PBA is used to separate and represent uncertainty associated with natural variability and incomplete knowledge of a parameter simultaneously. Imprecise probability theory (Walley 1990) provides an approach for capturing ambiguous beliefs when knowledge is poor.

4.1 Bias in judgements about probabilities

Research has shown that people are poor at predicting the likelihood of uncertain events (Kahneman 2011). The subjective assessment of probability is a cognitively challenging task and we often use heuristics (mental short-cuts or 'rules of thumb') to help us draw upon our knowledge and experiences to express a belief about a probability. Some of these heuristics can also result in significant systematic bias (Table 4). Although scientists are more used to dealing with numeric probabilities than other members of the public, they are not immune to these biases (Slovic 1999).

‘Experts’ tend to be overconfident in their judgements, sensitive to how the question is presented (its ‘framing’) and to the order of presentation of evidence, and subject to cognitive and motivational biases (Burgman et al., 2006). Nevertheless, subject experts tend to provide better estimates within their domain of expertise than non-experts (Burgman et al., 2011a).

Table 4: Common heuristics (cognitive short-cuts or 'rules of thumb') and biases that occur in judgements about probabilities and other numeric quantities.

Heuristic or bias	Definition	Reference
Affect heuristic	where people make judgements and decisions that express their feelings about an issue without realising they are doing so	Finucane et al., (2000)
Anchoring and adjustment heuristic	where judgements about the value of an uncertain quantity are influenced by values that have been suggested or considered prior to making the judgement (‘anchors’). Anchor values serve as a reference for people to adjust the boundary of the range of plausible values for the uncertain value	Tversky & Kahneman (1975)
Availability heuristic	where people make judgements or decisions based on reference situations that come to mind easily (i.e., are easily recalled or imagined).	Tversky & Kahneman (1973)
Base rate neglect	where a judgement disregards or does not take account of the frequency at which similar cases normally occur (i.e., its ‘base rate’ or ‘prior probability’)	Tversky & Kahneman (1975) Gigerenzer & Hoffrage (1995)
Denominator neglect	when people make judgements or decisions based on the number of positive (or negative cases) and disregard the size of sample that it was drawn from	Slovic et al., (2000)
Framing effect	when a change in the way that a question is presented influences choice behaviour. Individuals may draw different conclusions from the same information, depending on how that information is presented	Kahneman & Tversky (1979) Kahneman & Tversky (1984)
Overconfidence	where a person’s confidence in their judgment or estimate does not correspond to the accuracy of that judgment or estimate.	Burgman et al., (2006)
Possibility effect	an example of ‘denominator neglect’ where very low-probability events are much more heavily weighted than might be expected because of the chance (however small) that they could occur.	Kahneman (2011)
Recognition heuristic	where people make judgements based on characteristics that they can recognise from memory and give greater weight to recognizable features.	Goldstein & Gigerenzer (1999)

Heuristic or bias	Definition	Reference
Representativeness heuristic	where people make judgements or decisions based on how much a case resembles a type or class of other situations that they have some experience with	Tversky & Kahneman (1982)

A large body of research has investigated the biases inherent in judgements of risk and methods that can be used to reduce them. Judicious selection of experts, careful structuring of the elicitation process and training in deliberation about probability assessments can substantially improve the quality of expert judgements of subjective probabilities (Kuhnert et al., 2010; Mellers et al., 2015b; Hemming et al., 2018a). The findings from this research have been summarised in a number of recent, excellent reviews to which readers are referred for more detail (Cooke 1991; Meyer & Booker 2001; Burgman et al., 2006; O'Hagan et al., 2006; Martin et al., 2012; Burgman 2015). Our purpose in this section is to highlight a few general approaches of good practice as guidance for eliciting subjective probabilities.

4.2 General steps in eliciting expert judgements

A number of different approaches have been proposed for organising elicitation processes. In general, they involve the following components (Renooij 2001; Burgman et al., 2006; Drescher et al., 2013; Sutherland & Burgman 2015; Hemming et al., 2018a):

- Clearly defining the judgement problem and the quantities that are to be elicited
- Recruitment of a diverse group of experts with knowledge of the system under consideration
- Motivation and training of the experts
- Appropriate framing of the questions and choice of an elicitation format
- Structured processes for implementing the elicitation
- Methods for weighting and aggregating the elicited values
- Feedback and verification of the results

Structured protocols for elicitation provide clarity about the decision problem and the quantities that are sought from the experts, a way of ordering evidence and information to support the judgements and a process for reducing individual and group bias in estimating the values. They can, however, be expensive and time-consuming to implement and should, therefore, be scaled according to the size and complexity of the decision problem. For example, Hemming et al., (2018a) estimate that the preparation phase for their ‘IDEA’ protocol (described later in Section 4.5.2) can take anywhere from 2 weeks to 4 months depending on the purpose of the elicitation.

4.2.1 Clear problem definition

Use of ambiguous phraseology and unconsidered framing of the question can cause unwanted bias. Questions should be precise and provide as much information as is needed to make a considered prediction (e.g., time, place, units of measurement, the level of precision required, etc.) while reducing any possible uncertainty about their meaning. The experts should be clear about what is expected of them and the format that the response is to be provided in. The format used should align closely with the domain knowledge and experiences of the experts. Often it is useful to draft the questions and pilot them with a small group of experts prior to the main elicitation to ensure that they are unambiguous.

4.2.2 Selecting experts

Identification and selection of appropriate expertise will depend on the project scope and objectives. For many studies, a range of expertise may be needed to ensure that different perspectives on the event are considered. Although experts are usually regarded as individuals who have specialist knowledge of a subject that they have gained through experience, education or training (Garthwaite et al., 2005), detailed subject knowledge is usually insufficient to make good judgements about uncertain events (Burgman et al., 2011b). Research has shown that the judgements that result from well-organised and diverse teams of experts routinely outperform assessments made by even the most knowledgeable individuals (Schoemaker & Tetlock 2016; Tetlock & Gardner 2016). Experts who perform best in these settings tend to:

- have good knowledge of the subject,
- be actively open-minded in their thinking,
- be willing to spend time deliberating about the problem and to receive training in the elicitation,
- work collaboratively,
- be willing to update their beliefs based on information they receive from others and
- communicate their knowledge and the limits of their knowledge effectively in numerical terms (Burgman et al., 2011a; Mellers et al., 2015a; Hemming et al., 2018b).

The best judgements are often made by experts who are good at inductive reasoning and who are willing to reconsider their initial estimates in the context of new information or knowledge provided as part of the assessment (Mellers et al., 2015a).

Hemming et al., (2018a) recommend a panel size of between 10-20 experts, but the number recruited will depend on the problem under consideration and the resources available for the study.

4.2.3 Training and facilitation

To get the best out of experts in group settings it is also important that they are instructed effectively about the elicitation process and the knowledge required and that they are provided with feedback on the process and results (Drescher et al., 2013). While it might seem counterintuitive to train people who have been selected for their specialist expertise, there is good evidence that providing training in probabilistic reasoning and cognitive debiasing can improve judgements of events

(O'Hagan et al., 2006; Mellers et al., 2014; Mellers et al., 2015a). The training may include discussion on common jargon and theoretical concepts, the use of case studies and other similar assessments, scenarios, and simulations to illustrate relevant processes, modules on ways of representing uncertainty and probabilities (Burgman et al., 2011a).

For example, Mellers et al., (2014) developed a short training module on probabilistic reasoning that provided panellists with information on the use of reference classes and base rates, the benefits of expert panels over individual judgements, averaging multiple estimates from existing models, extrapolating from continuous variables and how to avoid judgmental biases such as overconfidence, confirmation bias and base-rate neglect. Panellists who received the training consistently produced more accurate predictions of events than those who did not.

Because the purpose of an expert panel is to share information and for panellists to improve their own estimates by critically appraising the reasoning and judgements of others, interactions within the panel must be moderated to avoid dominant individuals steering the group in particular directions of thinking, to encourage quieter, reticent members of the panel to participate actively in discussions and to avoid ascribing greater credibility to individuals than they deserve based on their manner or professional experience. Structured, facilitated interactions counter factors such as these, which distort estimates (Burgman et al., 2011a; Sutherland & Burgman 2015).

4.2.4 Face-to-face vs remote

Structured elicitations can be implemented as face-to-face interviews and workshop-based methods, or remotely using online questionnaires and/or email (McBride et al., 2012). The latter are usually cheaper to implement and less demanding of time, but are more difficult to maintain the engagement of participants (Drescher et al., 2013).

To avoid fatigue and allow time for considered discussion and evaluation, Speirs-Bridge et al., (2010) recommend use of no more than 15-20 questions in a day of face-to-face elicitation.

4.3 Methods for eliciting judgements about subjective probabilities

Beliefs about subjective probabilities can be elicited directly or indirectly (Kuhnert et al., 2010). Direct elicitation requires experts to express their knowledge in quantitative terms. These can be as point estimates, frequencies, intervals (e.g., lower and upper bounds) or a full parametric probability distribution. Indirect elicitation requires experts to answer questions that relate to their experiences or to respond to scenarios that are presented to them. Their responses can then be converted into the format needed for the analysis.

4.3.1 Direct elicitation

Numeric probabilities

For panellists who are well-versed in dealing with probabilities or where training in probabilistic reasoning has been provided, questions asking for a direct prediction of probability or percentages can be used.

However, many people find it difficult to make judgements about probabilities at the ends of the probability scale (i.e., < 0.2 and > 0.8) (Kahneman & Tversky 1979). This is particularly problematic in risk analyses involving rare hazards that have potentially large consequences. For example, people find it difficult to comprehend the degree of difference between numerical probabilities expressed as

0.005 and 0.0005 despite the ten-fold difference in the likelihood of the event. An alternative approach is to frame the question as a frequency statement (i.e., 5 out of 1,000 and 5 out of 10,000). Research has shown that frequency formats tend to be more easily understood when eliciting subjective probabilities, particularly for low frequency events (Gigerenzer & Hoffrage 1995; Mellers & McGraw 1999). An example format could be:

> *In a forest of n trees, how many (x) would you expect to have the disease?*

The probability can then be calculated as $\frac{x}{n}$.

Intervals

Judgments about intervals are more useful because they contain information about uncertainty that is not provided by point estimates of probabilities or quantities, but are subjectively easier for experts to produce than full distributions. Speirs-Bridge et al., (2010) have developed a stepwise question format to elicit intervals that has been shown to reduce overconfidence by experts. Using this approach, estimates of probability intervals for uncertain events are elicited using a three-question format (Hemming et al., 2018a):

> *Realistically, what do you think is the lowest plausible probability [*event X*] will occur?__*
>
> *Realistically, what do you think is the highest plausible probability that [*event X*] will occur?___*
>
> *Realistically, what is your best estimate for the probability that [*event X*] will occur?___*

The first two questions form an interval that captures the expert's uncertainty about the true value of the probability and the third question prompts them to attempt to estimate the most likely value. The best guess estimate could be interpreted as a measure of central tendency, such as a mean, median or mode, but that is not defined (Hemming et al., 2018a).

A four-step question format is used for quantities and frequencies (Speirs-Bridge et al., 2010):

> *Realistically, what do you think the lowest plausible value for [event X] will be ?___*
>
> *Realistically, what do you think the highest plausible value for [event X] will be ?___*
>
> *Realistically, what is your best guess for [event X]?___*
>
> *How confident are you that your interval, from lowest to highest, could capture the true value of [event X]? Please enter a number between 50% and 100%___*

The wording and order of questions are both important in these formats Hemming et al., (2018a). The words 'plausible' and 'realistic' are used to encourage use of informative bounds rather than defaulting to the limits of the probability range (i.e., 0 and 1). Asking for lower and higher bounds first, encourages consideration of counterfactuals and the evidence for relatively extreme values, and avoids the experts using their best estimate as a point of reference for the bounds (an 'anchoring' bias - Table 4).

In the four-step format, the last question asks the expert to review the interval that they have provided and to evaluate their level of confidence that the true value lies within it. Requiring the experts to self-evaluate their confidence in the interval reduces overconfidence (Speirs-Bridge et al., 2010). When results from this form of elicitation are summarized and presented back to the experts for review, they are standardized to a single level of confidence, usually 80% or 90%, so that the summaries use a consistent scale across all questions. This can be done using linear extrapolation in which:

$$Lower\ standardised\ interval: B - \left((B - L) \times \left({}^{S}\!/_{C}\right)\right)$$

$$Upper\ standardised\ interval: B + \left((U - B) \times \left({}^{S}\!/_{C}\right)\right)$$

where B = best guess, L = lowest estimate, U = upper estimate, S = level of the interval to be standardized to and C = level of confidence indicated by the participant (Hemming et al., 2018a).

4.3.2 Indirect elicitation

Barry & Lin (2010) describe an indirect method that uses constructed scenarios to calibrate estimates of risk. This form of indirect elicitation can be viewed as a form of choice modelling, where experts are asked to compare alternative scenarios and their decision process is modelled to identify important influences on the outcome. Because experts are making judgements about the likelihood of the scenarios, they are usually not required to provide direct, numeric estimates of the probability that they will occur. The experts are presented with a range of scenarios ($i = 1 \ldots . n$) in which the values for a set of risk attributes, $X_1 \ldots X_k$, vary over defined ranges. The experts are asked to make judgements about a specific risk endpoint (y_i). Regression is then used to find the weights (β) that each expert gives (subconsciously) to each risk attribute.

General regression approaches are used to estimate the relationships between the risk attributes, $X_1 \ldots X_k$, and expert predictions of the risk endpoints of interest. For example, consider an event with a probability of occurrence, p_i in which the log odds follow a logistic regression:

$$y_i = log\left(\frac{p_i}{1 - p_i}\right) = \alpha + X_{1,i}\beta_1 + X_{2,i}\beta_2 + \cdots + X_{k,i}\beta_k + \varepsilon$$

such that the $X_1 \ldots \ldots X_k$ are functions of the risk attributes that are thought to contribute to the likelihood that the event will occur, and $\beta_1 \ldots . \beta_k$ are the weights to be estimated from the elicitation that are associated with how much each attribute influences the outcome.

For example, Crombie et al., (2007) modelled the chance of successfully eradicating marine pests under a range of uncertain situational and resourcing conditions. In this instance, the purpose of the study was to relate expert judgements of the cost of eradication (a highly uncertain endpoint) to environmental characteristics of the incursion scenario. Nine variables were used to describe the range of pest traits under consideration, including their size, habit, reproductive output, etc. The situation in which incursions occurred was also described by nine variables that included the area infested, water depth, turbidity, etc. Experts were presented with 21 scenarios that comprised seven pest profiles with three incursion situations for each profile. For each scenario, they were asked to estimate the chance of successfully eradicating the pest (expressed as a % between 0 and 100%) when different levels of funding were available for the operation, ranging from $25,000 to $5,000,000 (for more details, see (Crombie et al., 2007)). A logistic regression was then fitted to each of the 420 responses (20 experts x 21 scenarios) to estimate the level of expenditure required for eradication. These estimates were then aggregated by first using the model to estimate the expenditure required for each scenario to give a probability of eradication of 0.95 by rearranging the equation such that:

$$X_c = \frac{log\left({0.95}/{0.05}\right) - \hat{\alpha}}{\hat{\beta}}$$

where X_c is the estimated expenditure and $\hat{\alpha}$ and $\hat{\beta}$ are the coefficients estimated by the model. A regression model was then fitted to the estimates of costs required to achieve $p = 0.95$ to assess which of the environmental characteristics had the greatest effect on cost. The fitted model had the form:

$$log(X_c) = \alpha + \beta_1 + \beta_2 + \cdots + \beta_k + u + e$$

where X_c is the estimated cost for a probability of eradication of 0.95, α is an intercept term and $\beta_1, \beta_2 \ldots \beta_k$ are coefficients estimated from the data corresponding to the variables describing pests and the incursion setting.

The utility of this approach has also been demonstrated for situations where the risk end-point is an ordering of the relative risk of different scenarios (Barry et al., 2015) or relative to some baseline scenario (Knight et al., 2007) or the probability of occurrence of rare species (O'Leary et al., 2009). For example, where the probability of the event, p_i, is considered low (i.e., close to zero) elicitation of the relative risk of two scenarios, p_1 and p_2, can be modelled as:

$$log\left(\frac{p_1}{p_2}\right) = \alpha + X_{11}\beta_1 + X_{21}\beta_2 + \cdots + X_{k1}\beta_k - (\alpha + X_{12}\beta_1 + X_{22}\beta_2 + \cdots + X_{k2}\beta_k)$$

In this instance, the elicited value is the ratio of relative likelihood, $\frac{p_1}{p_2}$, an endpoint that is intuitively easier to specify than an absolute probability (Barry et al., 2015).

Barry & Lin (2010) note several advantages of scenario-based elicitation methods over other expert elicitation methods:

- it is often intuitively easier for experts to make judgements on scenarios than on individual components of a risk problem
- the elicitation is direct and transparent and is administered just once
- variations between experts can be quantified and incorporated into a risk model
- the statistical modelling approach means that it is relatively easy to calibrate the weights to real world situations.

4.3.3 Reconciling qualitative descriptors of uncertainty as numeric probability ranges

Qualitative risk analyses often use verbal expressions like “possibly,” “probably,” and “likely” as an alternative to numerical probabilities to describe the likelihood of uncertain events or consequences. Research has shown that people prefer to express uncertainty using verbal terms because they find them more intuitive than numerical probabilities (Brun & Teigen 1988; Wallsten et al., 1993) and because they perceive the latter to convey a false sense of precision in situations where there is significant uncertainty (Burgman et al., 2006). Even technical experts sometimes prefer to express likelihoods using words rather than numbers when the events themselves are not well-defined and the uncertainties associated with them are difficult to specify and measure (Budescu & Wallsten 1985; Budescu et al., 2009).

Despite evidence that coarse qualitative descriptors of likelihood compromise the predictive accuracy of risk analyses (Friedman et al., 2018), verbal expressions of likelihood are still widely used in risk analysis and are sometimes recommended practice (EFSA Scientific Committee et al., 2018b) because they:

- Are a simple tool for prioritising uncertainties
- Can aid quantification of uncertainty through use of an approximate probability scale with accompanying qualitative descriptors
- Can describe uncertainties that assessors are unable to include in their quantitative evaluation and
- Can aid the communication of risk. For example, decision-makers or legislation may require or prefer likelihoods to be expressed in qualitative terms.

As an example, Biosecurity New Zealand's Procedures for Risk Analysis allows the likelihood of entry, establishment or exposure of harmful organisms or diseases to be expressed in qualitative verbal terms. Six terms are considered acceptable expressions of likelihood and eight adjectives can be used to qualify it (Table 5). As this example shows, even with some descriptive guidance on interpretation there can be ambiguity in how each of the terms might be used, leading to inconsistency among assessments. This is because the same words mean different things to different people, and because a phrase may take on a different meaning in different contexts (Wallsten et al., 1986; Morgan et al., 2009). For example, the phrase 'very unlikely' may be interpreted quite differently when referring to the possibility of a terrorist attack, or the chance of rain tomorrow. People can conflate the likelihood of an event with its consequences so that the way they interpret verbal expressions for the likelihood of an event can be strongly influenced by the perceived severity of its consequences (Patt & Schrag 2003). Budescu & Wallsten (1985) have shown that while there is some consistency in the way individuals use verbal expressions of likelihood over time, different individuals can have very different interpretations of the same terms. People tend to interpret probability terms more variably at the extremes of the scale and be less sensitive to changes in the middle (Patt & Schrag 2003). They overestimate the probability of rare events and underestimate the probability of frequent events (Kahneman & Tversky 1979). Assumed or informed base rates (i.e., expectations of the long run frequency of an event) alter interpretations of the likelihood terms. Estimates of numeric probabilities tend to be lower for more serious events as people generally assume higher magnitude negative outcomes are less likely to occur (Tversky & Kahneman 1975; Patt & Schrag 2003). Likelihoods terms for neutral and high numeric probabilities tend to be strongly related to base rates whereas interpretations of low terms are much less affected (Wallsten et al., 1986).

Table 5: Terms used to describe likelihood in biosecurity risk analyses. Source: (Biosecurity New Zealand 2006).

Acceptable terms to express likelihood	Definition
Chances	in its plural form chance indicates a probability
Likelihood	probability; the state or fact of being likely
Likely	probable; such as well might happen or be true; to be reasonably expected
Probability	the likelihood of something happening; mathematically it is defined as the extent to which an event is likely to occur, measured by the ratio of the favourable cases to the whole number of cases possible
Probable	May be expected to happen or prove true; likely
Would	To express probability (I guess she would be over 50 by now); past of Will: expressing a wish, ability, capacity, probability or expectation
Adjectives used to qualify likelihood estimates	**Definition**
Average	The usual amount, extent, rate
Extremely	Outermost, furthest from the centre; situated at either end; utmost; the highest or most extreme degree of anything
High	Extending above the normal or average level
Highly	In a high degree
Insignificant	Unimportant; trifling
Low	Less than average, coming below the normal level
Negligible	Not worth considering; insignificant
Significant	Noteworthy; important; consequential
Remote	Slight, faint

To address this ambiguity, organizations have attempted to develop standardized lexicons that map phrases for the likelihood of an event or consequence onto specific numerical ranges of probability. Guidance developed for subjective probability judgements used by working groups on the Intergovernmental Panel on Climate Change Third Assessment Report expresses likelihood using a seven-point scale, ranging from ‘exceptionally unlikely’ to ‘virtually certain’, with numerical probability ranges assigned to each phrase (Table 6). Similarly, Biosecurity Australia uses a six-point scale that ranges from ‘extremely low’ to ‘high’ to map over the same range of probabilities. Guidance issued for US intelligence analysts allows for a seven-point scale with the choice of two sets of nominally equivalent verbal descriptors (Table 6).

Table 6: Guidance on the equivalence of verbal and numerical expressions of likelihood developed for (a) the IPPC Third Assessment, (b) biosecurity risk analyses undertaken by Biosecurity Australia and (c) the US Intelligence community. Source: Moss & Schneider (2000); (Burgman et al., 2010; Friedman et al., 2018)

(a) IPPC guidance		(b) Biosecurity Australia guidance		(c) US National Intelligence Community Directive 203, Analytical Standards		
Verbal term	**Probability range**	**Verbal term**	**Probability range**	**Verbal term**	**Equivalent term**	**Probability range**
Virtually certain	> 0.99	-	-	Almost certain	Nearly certain	95-99%
Very likely	0.90 - 0.99	-	-	Very likely	Highly probable	80-95%
Likely	0.66 - 0.90	High	0.7 - 1.0	Likely	Probable	55-80%
Medium likelihood	0.33 - 0.66	Moderate	0.3 - 0.7	Roughly even chance	Roughly even odds	45-55%
Unlikely	0.10 - 0.33	Low	0.05 - 0.3	Unlikely	Improbable	20-45%
Very unlikely	0.01 - 0.10	Very low	0.001 - 0.05	Very unlikely	Highly improbable	5-20%
Exceptionally unlikely	< 0.01	Extremely low	10^{-6} – 0.001	Almost no chance	Remote	1-5%
		Extremely low	< 10^{-6}			

As these three examples show, the numeric probability ranges specified by organizations for quite similar terms can be very different. They may give emphasis to different parts of the probability distribution to reflect different levels of risk tolerance. For example, the Biosecurity Australia lexicon gives greater emphasis to the lower tail of probabilities, with four separate descriptors for numeric probabilities < 0.3, but just a single term for probabilities > 0.7. This may reflect the greater relevance to biosecurity assessments of events with a less than equal chance of occurrence or could be an effort to improve the consistency of use of terms that describe low probability events. Judgments of risk are strongly influenced by the distribution of small and large probabilities on a response scale (Bilgin & Brenner 2013).

In some guidelines, assessors are also required to provide an indication of the confidence they have in their evaluation of likelihood. For example, the 4th IPPC Assessment Report distinguishes five levels of confidence that can be used to qualify probability assessments (Table 7) whereas the US Intelligence community uses just three: – low, moderate and high (Friedman et al., 2018).

Table 7: Verbal expressions for confidence in an assessment mapped against quantitative ranges in the IPCC 4th Assessment. Source: (IPCC (Intergovernmental Panel on Climate Change) 2005).

Verbal term	Probability range
Very high confidence	At least 9 out of 10 chance
High confidence	About 8 out of 10 chance
Medium confidence	About 5 out of 10 chance
Low confidence	About 2 out of 10 chance
Very low confidence	Less than 1 out of 10 chance

Research has shown that tables such as these that contain numerical guidance for the use of verbal expressions improve consistency in the use of the terms (Budescu et al., 2009) and improve discrimination of likelihoods at the upper and lower ends of the scale (Budescu et al., 2014). People using the scales are better able to calibrate their interpretations of the intended meaning when they are presented with accompanying numerical ranges.

Nevertheless, people may 'de-code' the guidance and interpret the terms in a way that is more commensurate with their beliefs about the event's likelihood (Patt & Schrag 2003) or they may simply ignore the guidance altogether (Budescu et al., 2014). For example, despite the directive that US intelligence analysts use the lexicon described in Table 6, Friedman et al., (2018) reported that the recent Intelligence Community Assessment describing potential Russian interference with the 2016 US presidential election did not use any of the terms to express likelihood, but instead appeared to conflate likelihood with confidence.

Part of the problem with the use of qualitative terms for the communication of risk and uncertainty is that the lexicons are often developed by technical experts (a 'top-down' approach) and do not use the words in same way that other stakeholders might. Recent studies have developed lexicons using people's actual interpretations of the phrases in the context in which they are used; a 'bottom-up' approach that Ho et al., (2015) describe as 'evidence-based'. Ho et al., (2015) developed evidence-based lexicons for the four most used probability phrases in the IPCC guidance- 'very unlikely', 'unlikely', 'likely', 'very likely' - and for the seven-point scale in US National Intelligence Community Directive 203 (Table 6). For each phrase they asked survey participants to read a statement that contained the phrase and to characterize the phrase's numerical meaning by estimating its lower and upper bounds and offering their best estimate for its specific meaning. Two objective approaches were used to associate each phrase with a numerical probability range. The peak value (PV) method reflects the distribution of participants' best estimates of the value of each phrase when it was presented alone, outside of the context of a sentence. The distributions of each phrase's numerical meanings were plotted, and cut-offs between adjacent phrases were determined where the distributions for the terms intersected. The second approach, known as the membership function (MF), modelled the probability phrases as fuzzy subsets within the [0, 1] probability interval, with MFs that describe how well a certain numerical value defines or substitutes for a given phrase. A membership of 0 denotes that the probability value does "not at all" substitute for (define) the phrase, whereas a membership of 1 indicates that a probability value "absolutely" substitutes for the phrase. A value with a membership of 1 is referred to as the peak of the MF, and an intermediate value reflects partial membership. MFs were calculated for each person and term using linear interpolation of the reported lower bound, upper bound and best estimate for each phrase. The optimal cut-off points between adjacent phrases were obtained by identifying the region of values for which the sample membership of a given phrase was higher than all other phrases (Ho et al., 2015).

Table 8: Comparison of evidence-based lexicons with probability ranges specified in guidance for IPCC and US NIC assessments of likelihood. Evidence-based lexicons were derived using two methods: 'Peak value' (PV) and 'Membership functions' (MF). Source: (Ho et al., 2015).

(a) IPPC guidance		(b) Evidence-based lexicon		(c) US NIC Directive		(d) Evidence-based lexicon	
Verbal term	Probability range	PV	MF	Verbal term	Probability range	PV	MF
Virtually certain	> 0.99	-	-	Almost certain	95-99%	80-100%	90-100%
Very likely	0.90 - 0.99	0.65-1.0	0.75-1.0	Very likely	80-95%	75-80%	80-90%
Likely	0.66 - 0.90	0.45-0.65	0.40-0.75	Likely	55-80%	60-75%	50-80%
Medium likelihood	0.33 - 0.66	-	-	Roughly even chance	45-55%	45-60%	40-50%
Unlikely	0.10 - 0.33	0.15-0.45	0.15-0.40	Unlikely	20-45%	25-45%	20-40%
Very unlikely	0.01 - 0.10	0-0.15	0-0.15	Very unlikely	5-20%	15-25%	10-20%
Exceptionally unlikely	< 0.01	-	-	Almost no chance	1-5%	0-15%	0-10%

Studies that have used 'evidence based' approaches to the development of the lexicons report greater consistency in the interpretation of phrases, particularly those in the more extreme ends of the scales (Ho et al., 2015; Wintle et al., 2019). However, Wintle et al., (2019) caution that some of this consistency may be because judgements averaged over multiple people tend to be regressive for terms at the ends of the probability scale and do not capture the precision needed to express extreme numeric probability intervals. This is particularly important for applications that are most concerned about very rare or almost certain events.

Ultimately, there is no universal standard for the expression of likelihood that will suit all settings. The phrases and number of terms used, and the probability ranges associated with them are sensitive to the specific application. Where consistency is required in their use and communication, it is best to tailor the lexicon to the specific needs of the problem and the required level of precision (Ho et al., 2015).

Some general recommendations for deriving and using qualitative descriptors include:

- As much as possible, describe the event in unambiguous language so that judgements of its likelihood are not confounded with uncertainty in how the event will be manifest (Budescu et al., 2009).
- Specify the various sources of uncertainty that underly key events and describe their nature and magnitude, to the degree that this is possible (Budescu et al., 2009)
- Select phrases carefully and, where possible, draw on established lexicons that have been tested and implemented successfully.
- Provide both verbal terms and associated numerical probability ranges to communicate uncertainty.
- In general, the optimal number of categories is a qualitative scale is considered to be about 5 - 7 levels, anchored at certainty and highly uncertain at each end and at an even chance of occurrence in the middle (Renooij & Witteman 1999; Burgman et al., 2006).
- Adjust the width of the numerical ranges to match the uncertainty of the events.
- Where greater precision is required, precise framing of the event along with grounding or calibration of numerical information can greatly increase the ability of assessors to discriminate accurately among 9-12 numerical categories or to specify numeric probabilities (Friedman et al., 2018).
- 'Evidence-based' lexicons improve the translation of verbal phrases to probability ranges. Because these are time-consuming to elicit, they may be most suited to situations where a conversion table will be used repeatedly (Burgman et al., 2006).
- Allow experts to express their degree of confidence (or lack of confidence) in estimates of numerical probabilities.
- As risk analyses typically require numerical information for subsequent calculations the primary use of qualitative expressions of likelihood should be to facilitate more accurate representations that can inform a judgement of numeric probabilities (Burgman et al., 2006; EFSA Scientific Committee et al., 2018b)

4.4 Imprecise probabilities

Imprecise probabilities are expressions of likelihood or uncertainty that are made when it is not possible to specify precise numerical probabilities (Walley 1990). They can take many forms, including an ordering of the relative likelihood of events (e.g., *A* is more likely than *B*), belief functions, possibility measures, or estimates of the upper and lower ranges of belief for a value (Walley 1990). Whereas elicitations of subjective probabilities typically aim to characterize the 'best' belief of expert's about the probability distribution of a variable or its moments, imprecise probability assessments often seek to exclude those probabilities that are incommensurate with the expert's belief (Kriegler et al., 2009). In many cases where there is significant uncertainty about the specific value of a probability it is possible to specify beliefs about the range in which it could occur and where it cannot, often articulated as an upper and a lower bound of possible values. A bounding approach to risk analysis may be necessary when analysts cannot specify (1) precise values for parameter input distributions or point estimates in the risk model (e.g., minimum, maximum, moments, etc.), (2) precise (marginal) probability distributions for some or all of the variables in the risk model, (3) the precise nature of dependencies between variables in the risk model, and (4) the exact structure of the risk model (model uncertainty) (Tucker & Ferson 2003).

Probability Bounds Analysis (PBA) methods combine interval analysis and classical probability theory. They allow quantitative inputs into risk assessment calculations from interval-type bounds on cumulative distribution functions, called "probability boxes", or "*p*-boxes" (Williamson & Downs 1990). A *p*-box consists of a pair of functions that are used to circumscribe an imprecisely known distribution function $F(x)$ that is specified by a pair of upper and lower distribution functions $\bar{F}(x)$ and $\underline{F}(x)$, such that $\underline{F}(x) \leq F(x) \leq \bar{F}(x)$ for all x . In practice, a p-box is often expressed in terms of its inverse functions d and u defined on the interval of probability levels [0,1], where u is the inverse function of the upper bound of the distribution and d is the inverse of the lower bound, such that:

$$d(p) \gg F^{-1}(p) \gg u(p),$$

where p is the probability level (Tucker & Ferson 2003).

Probability bounds can be computed for parametric distribution models, where the distribution family is specified but only interval estimates can be given for the parameters, and for situations where the distribution shape is unknown and cannot be estimated from existing data (nonparametric models). The simplest case of a nonparametric model is where all that can be specified about the distribution function of a variable are its minimum (min) and maximum (max) values. In this case, the cumulative distribution function (CDF) is assumed to be rectangular Figure 4(a). Additional information about the parameters or distribution can be used to refine the box and constrain it to a smaller region. For example, knowledge of the mean, median, mode, etc can be used to constrain the *p*-box to smaller areas through which the true distribution must pass Figure 4(b-e). Where the parameters are also imperfectly known confidence intervals can be used to define them in place of point estimates. In this case, the envelope of all possible underlying distributions can be calculated as the union of bounded probability regions. In this way, PBA has parallels with robust Bayesian analysis where uncertainty about which prior distribution should be used can be addressed by replacing a single precise prior distribution with an entire class of prior distributions. The analysis then examines the variety of outcomes possible across the range of prior distributions (Ferson et al., 2015).

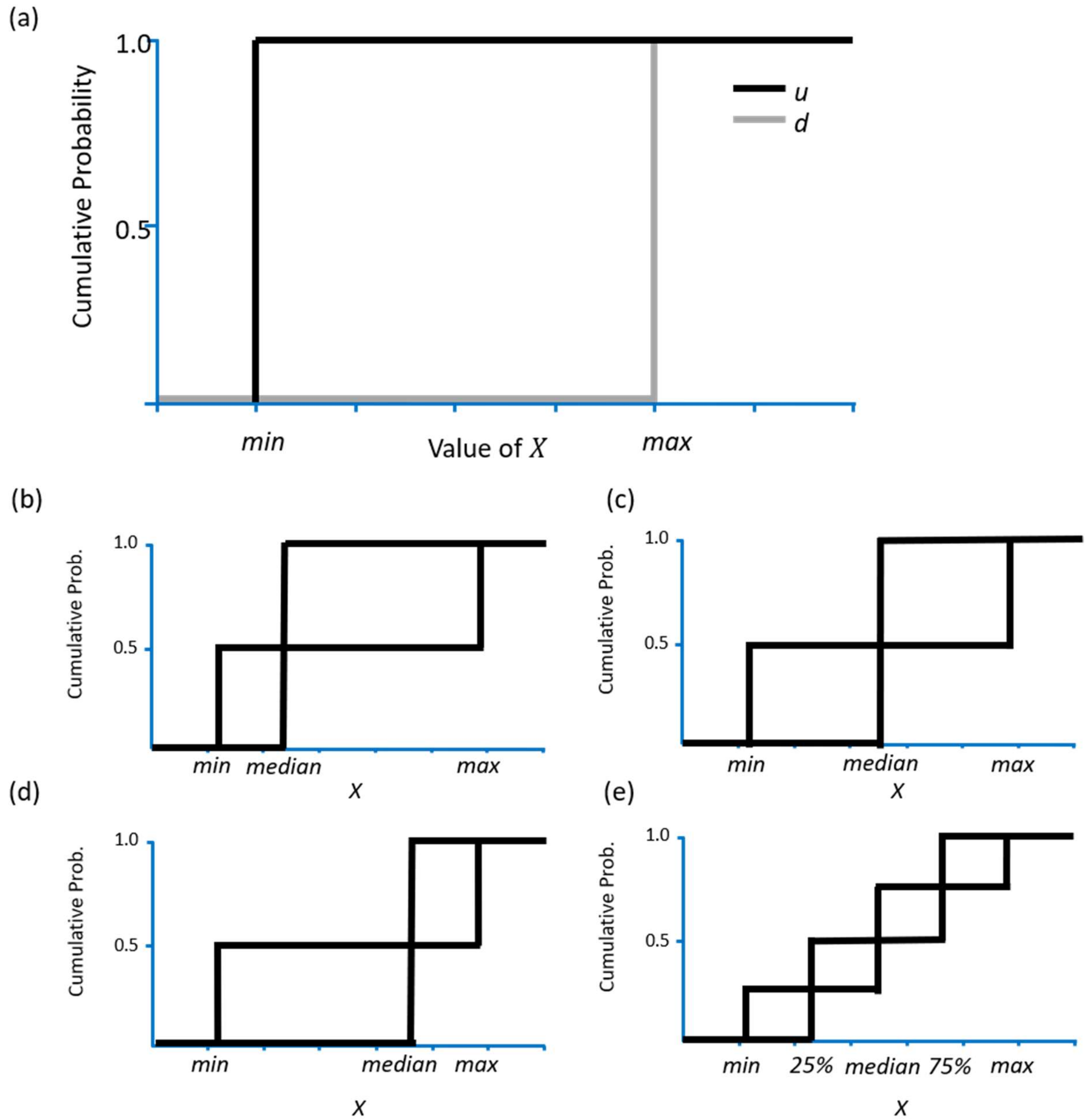

Figure 4: Probability bounds around the unknown distribution function of the random variable X. where the only available information about the values of X are (a) the minimum and maximum, (b, c & d) the minimum, median and maximum, and (e) minimum, median, maximum and 25^{th} and 75^{th} percentiles. (Source: Tucker & Ferson (2003))

Tucker & Ferson (2003) describe the different ways that *p*-boxes can be constructed for a range of situations where data may be limiting or unavailable. Best possible probability boxes have been derived for many different cases that might arise in practice. These include, for instance, cases in which the following sets of information are known (or can be assumed): [minimum, maximum],[minimum, maximum, mean], [minimum, maximum, median], [minimum, maximum, mode], and [sample mean, sample variance]. Qualitative constraints on the shape of the distribution may also be accounted for to tighten the uncertain number (Ferson et al., 2015). An appealing feature of this method is that *p*-boxes can be tailored to the available data in a way that permits rigorous risk calculations without simplifying assumptions about the variable's distribution function (Burgman et al., 2010). Details on numerical methods for computing bounds on the result of addition, subtraction, multiplication and division of random variables when only bounds on the input distributions are given in Williamson & Downs (1990) and Ferson et al., (2015).

A variety of approaches have been proposed to combine *p*-boxes derived from the estimates of multiple experts into a single coherent expression about what is known (or unknown) about the quantity. Ferson et al., (2015) present a detailed review of eight strategies to achieve this: null aggregation, intersection, envelope, Dempster's rule, Bayes' rule, mixture, logarithmic pooling and averaging. Nau (2002) compared the use of 'confidence-weighted lower and upper bounds' and 'risk neutral probabilities' methods to aggregate imprecise probabilities. The reader is referred to these publications for more detail on these methods and an evaluation of their relative merits for PBA.

Despite its appeal in facilitating rigorous quantitative analysis from imprecisely expressed expert judgements, PBA is a relatively new method that has not yet been used widely and is still the subject of some contention among risk analysts (Burgman et al., 2010). There is not yet a large body of work in which it has been applied, particularly in environmental management (EFSA Scientific Committee et al., 2018a). PBA approaches are computationally complex and there is a relatively small base of people with experience and expertise in applying them to risk analyses (Burgman et al., 2010).

Case study: Eliciting imprecise probabilities of climate tipping points

Kriegler et al., (2009) used a PBA approach to explore expert beliefs about the probabilities of triggering five plausible but highly uncertain, large-scale transitions in environmental processes ('tipping points') under different scenarios of climate change. They elicited subjective probability intervals (lower and upper bounds) for the occurrence of identified major transitions possible under global warming from a panel of expert scientists.

Participants were asked three questions about the probability of triggering each event ('B') in the context of each of three global mean temperature projections for climate change. To avoid biases of overconfidence and anchoring, participants were asked two conditioning questions to assess the consistency of their responses. The first asked which of 2 outcomes - ''Triggering'' and ''Not Triggering'' - they judged to be more probable (question (i) below) and the second required them to select the options in a set of verbal statements of uncertainty (based on the IPCC categorization of uncertainty, Table 6) that were incommensurate with their belief of the likelihood of a tipping point occurring (question (ii) below). They were then asked to provide conservative estimates of the lower and upper bounds on the probability of triggering a tipping point taking into account their judgments in (i) and (ii).

i. *The probability of triggering B [is less, is greater, could be more or less] than the probability of not triggering".*

ii. *Consider the following set of statements …. Please select those statements that you definitely believe to be FALSE. Do not tick the statements that you believe to be true or that could possibly be true (i.e., you are uncertain about)."*

Triggering B is. exceptionally unlikely (_),
very unlikely (_),
unlikely (_),
medium likelihood (_),
likely (_),
very likely (_),
virtually certain (_).

iii. *Now tell us the bounds on the range of plausible values of the probability of triggering B. Specify the bounds on this range by real numbers between 0 and 1.*

Participants were also asked about relationships between the five tipping points. For each selected tipping point *B* (chosen under the options in question (1)), participants were asked to consider the statements

iv. *"If A has been triggered this will have a [ ] influence on the probability of B".*
The options to specify the influence of A on the probability of B were:

No influence (_)
Increasing influence (_)
Decreasing influence (_)
Increasing or decreasing influence (_)
I do not know (_)

Three methods were evaluated for aggregating the elicited probabilities: linear pooling, logarithmic pooling and confidence-weighted bounds (Clemen & Winkler 1999; Nau 2002). Linear pooling and the limited-confidence bounds (80% confidence) provided the most convincing pooling rules as they were robust against outliers without masking the imprecision inherent in individual probability intervals and their scatter among each other.

Their results showed that the prospect of triggering major changes in the climate system was not considered remote despite large uncertainty among experts. Even with an overall conservative approach, the aggregated lower probability bounds for triggering major change in the climate system were much higher than probabilities attributed to catastrophic events in other climate assessments.

4.5 Methods for aggregating expert judgements

Three general types of methods are distinguished for aggregating judgements made by groups of experts:

- Behavioural aggregation, whereby individual judgements are aggregated through interaction of the experts within a group to reach a consensus.
- Mathematical aggregation, in which judgements are aggregated using a weighted average to calibrate the experts.
- Mixed methods, where individual judgements are aggregated by moderated feedback avoiding direct interactions in the group and incorporating mathematical aggregation of group scores (EFSA Scientific Committee et al., 2018a).

Commonly used behavioural approaches include the Delphi method, the Nominal Group technique and Cooke's method (Drescher et al., 2013). Each approach consists of multiple rounds of elicitation. The judgements made by individual experts are summarized and circulated following each round so that individuals can adjust their own opinions based on the feedback and information provided (Aspinall 2010). The behavioural approaches differ in whether the judgement made by experts are kept anonymous or discussed face-to-face (Drescher et al., 2013). Anonymity is usually preferred to avoid group biases such as dominance and 'groupthink' (McBride et al., 2012).

Contemporary 'mixed methods' for aggregation do not necessarily seek consensus across the group, but rather use panel settings to resolve ambiguities, promote critical thinking and share knowledge and evidence. The average predictive performance of panellists improves when they are given the opportunity to assess others' judgements and to cross examine their reasoning and data within a facilitated process (Burgman et al., 2011a).

4.5.1 Mathematical aggregation

A wide range of mathematical methods have been proposed for aggregating judgements from multiple experts (Genest & Zidek 1986; Clemen & Winkler 1999; Ranjan & Gneiting 2010).

Linear pooling

Linear pooling is the most commonly used approach. This is simply a weighted linear combination of the expert's point estimates, often their best estimate of the value (Clemen & Winkler 1999; Lichtendahl et al., 2013):

$$p(\theta) = \sum_{i-1}^{n} w_i p_i(\theta)$$

Where n is the number of experts, $p_i(\theta)$ represents expert i's probability distribution, and the weights w_i are non-negative and sum to one.

Linear pooling is numerically simple to implement. It is, however, sensitive to the selection and weighting of experts, so that it is important to include a variety of opinions in the expert panel to provide a balanced (unbiased) outcome (Burgman et al., 2010). It also doesn't perform as well as other aggregation formulae that have a multiplicative structure (Allard et al., 2012).

The weights provide a mechanism to represent the relative quality of estimates elicited from different experts. The simplest case is to assume that the value provided by each expert is roughly the same (i.e., $w_i = \frac{1}{n}$). However, different approaches for determining weights have been proposed (Genest & McConway 1990). One approach is to get each expert to provide a peer rating of other panellists and to use Bayesian updating of the weights as more knowledge is obtained about fellow panellists (Genest & McConway 1990; Regan et al., 2006). Alternatively, a series of preliminary questions for which the outcomes are known or analytical tests can be used to calibrate the skill and knowledge of potential contributors (Aspinall 2010; Burgman et al., 2011b). The determination of weights is, nevertheless, a subjective judgement itself.

Logarithmic pooling

An alternative to linear pooling uses multiplicative averaging, sometimes referred to as a *logarithmic opinion pool*:

$$p(\theta) = k \prod_{i=1}^{n} p_i(\theta)^{w_i}$$

where k is a normalizing constant and the weights, w_i, satisfy some restrictions to ensure that $p(\theta)$ is a probability distribution. Again, the weights are typically restricted to sum to one.

Bayesian pooling

Methods of combining elicited probabilities that are based on a Bayesian updating approach are considered the most appropriate for risk analysis situations but can be difficult to implement (Clemen & Winkler 1999). The idea is that where n experts provide information g_i to a decision maker about an event or quantity of interest θ then Baye's Theorem can be used to update a prior distribution for the event $p(\theta)$ such that:

$$p^* = p(\theta | g_1, \dots, g_n) \propto \frac{p(\theta)L(g_1, \dots, g_n|\theta)}{p(g_1, \dots, g_n)}$$

Where L represents the statistical likelihood function associated with the experts' information and p^* is the posterior distribution.

For example, consider a situation where n experts are asked to estimate the probability p_i that an event θ will occur (i.e., θ = 1). A Bayesian formulation of this problem can be used to express the posterior odds that the event will occur, $q^* = \frac{p^*}{(1-p^*)}$ as:

$$q^* = \frac{p_0}{1-p_0} \prod_{i=1}^{n} \frac{f_{1i}(p_i|q=1)}{f_{0i}(p_i|q=0)}$$

where $f_{1i}(f_{0i})$ represents the probability of expert i giving probability p_i conditional on the occurrence (or non-occurrence) of θ and p_0 is the prior probability. This model assumes that each expert has independent information about p^*.

Genest & Zidek (1986) and Clemen & Winkler (1999) describe a range of other Bayesian models for aggregating point estimates and probability distributions. Newer techniques apply hierarchical models that account for a range of sources of information as well as potential dependence among expert judgments (Albert et al., 2012). Many of these methods are computationally complex as they involve a large number of parameters that need to be fitted.

Log-likelihood aggregation

Satopää et al., (2014) have developed a relatively simple aggregator based on likelihood estimation of the log-odds that is robust to overfitting. The maximum likelihood estimate (MLE) of the geometric mean of the odds is expressed as:

$$\hat{p}_G(a) = \frac{\left[\prod_{i=1}^{n}\left(\frac{p_i}{1-p_i}\right)^{1/n}\right]^a}{1+\left[\prod_{i=1}^{n}\left(\frac{p_i}{1-p_i}\right)^{1/n}\right]^a}$$

where G denotes use of a geometric mean and a is an unknown level of systematic bias ≥ 1. The bias, a, can be estimated from a training set of questions of K binary outcome events for which probabilities are elicited. The maximum likelihood estimator of a is estimated as:

$$\hat{a}_{MLE} = \arg\max_a \prod_{K=1}^{K} \hat{p}_{G,k}(a)^{Z_K}\left(1-\hat{p}_{G,K}(a)\right)^{1-Z_K}$$

4.5.2 Mixed methods of aggregation

Hanea et al., (2017) describe a structured protocol for eliciting and combining expert judgements of uncertain event occurrences using a modified Delphi process. The IDEA protocol has been used for a range of applications and builds upon earlier structured processes for expert elicitation and research on expert judgements to help reduce the influence of biases and to enhance the transparency, accuracy, and defensibility of the expert judgements (Hanea et al., 2018a; Hemming et al., 2018b). It combines a number of the elements of best practice that have been described in earlier sections of this report, including Delphi facilitated workshops (Drescher et al., 2013), frequency format elicitations for probabilities (Gigerenzer & Hoffrage 1995), three- and four-step question formats to elicit plausible intervals for the quantity (Speirs-Bridge et al., 2010), calibration questions to hone thinking and calculate weights (Hanea et al., 2018b), visualization and feed-back of results and mathematical aggregation of judgements (Clemen & Winkler 1999; Ranjan & Gneiting 2010).

The acronym 'IDEA' is derived from the main steps in the process (Figure 5). In the first stage ('Investigate'), experts provide individual estimates of the quantities sought by the analyst in response to questions that are presented to them. They then receive anonymized feedback about the judgements made by other experts for the same questions and, with the assistance of a trained facilitator, discuss the reasoning behind their initial estimates with others. The purpose of the 'Discuss' stage is to establish a shared understanding of the problem by collectively clarifying terms and exchanging information and knowledge. The discussion stage can be done remotely or with the experts in the same room (Hanea et al., 2018a). On conclusion of the discussion, each expert is then asked to revise their individual judgements and to make a second, private estimate ('Estimate'). Estimates provided in this second round are then combined using a mathematical aggregation approach ('Aggregate').

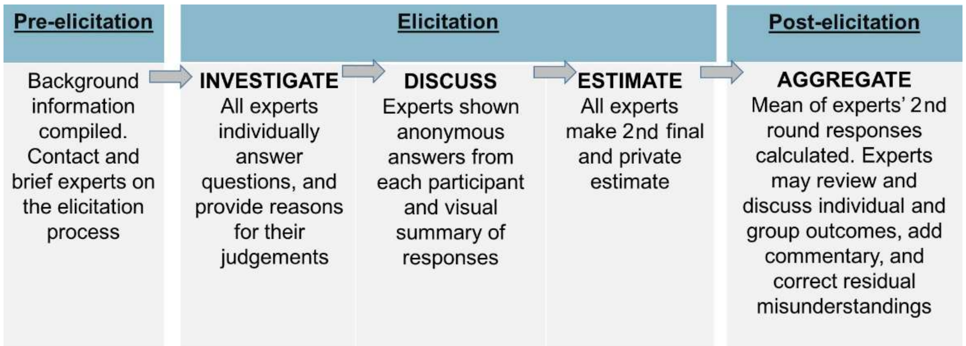

Figure 5: Steps within the IDEA protocol for elicitation of judgements from expert groups. Source: (Hemming et al., 2018a).

Round 1 estimates are usually obtained by an email or internet questionnaire that contains a description of the problem context, the elicitation questions and instructions on how to complete them. The responses are then compiled and summarised anonymously in graphical format to provide feed-back to the participants. The IDEA protocol typically uses quantile aggregation to summarise the elicited values, in which the arithmetic mean of experts' estimates is calculated for the lower, best and upper estimates for each question.

A facilitator is used to guide the discussion phase to ensure there is good deliberation about the results, placing particular emphasis on issues needing clarification and exploring reasons for contrasting results and outliers.

Following the discussion, the experts make a second, anonymous, independent estimate for each question. Once the second-round results have been received the experts' final assessments and the group aggregates are circulated to the group for final review.

Detailed guidance on implementing the IDEA protocol, including methods for selecting experts, question design and implementation are contained in the Supplementary materials provided by Hemming et al., (2018a).

5 Methods for assessing cumulative and indirect risks from multiple stressors

Cumulative risk can be defined as the consequences that may result from the combined effects of multiple activities or stressors and their associated uncertainties. The US EPA defines a stressor as '*a physical, chemical, biological or other entity that can cause an adverse response in a human, other organism or ecosystem*' (U.S. Environmental Protection Agency 2003). Although the terms 'cumulative risk assessment' and 'cumulative impact assessment' (also referred to as 'cumulative effects assessment') are often conflated, the US National Research Council (2009) and National Environmental Justice Advisory Council (2004) recommend maintaining a conceptual distinction between them. They proposed that the former be defined as evaluating the combined effects of an array of stressors to characterize their effects on the environment and human health, taking account of background exposures and vulnerabilities in the system under consideration. Cumulative risk assessment is, therefore, an anticipatory analysis and evaluation of risk in which the focus is on characterizing and, where possible, quantifying the uncertain effects of multiple stressors. 'Cumulative impact assessment' is considered to encompass a somewhat broader range of risk end-points, some quantifiable, and could include retrospective analysis of the incremental effects of actions and stressors (National Research Council 2009). The distinction between the two terms in the U.S.A. reflects their origins in the fields of risk assessment and environmental impact assessment, respectively, and the expectation that outcomes from risk assessment are usually expressed probabilistically, whereas those in environmental impact assessment are typically qualitative in nature (National Environmental Justice Advisory Council 2004). Practically, there are considerable overlaps in approach and method in the two disciplines, but a characteristic of many cumulative impact assessments is that they do not attempt to address uncertainty explicitly in the assessments (Stelzenmüller et al., 2018).

Characterizing the stressors associated with multiple activities and predicting their effects is inherently more complex and uncertain than risk assessments of single activities or events. The stressors themselves may interact so that the nature of the stressor experienced by different components of an ecosystem can be materially different from that experienced when each occurs in isolation. The response of an ecosystem component to a stressor may also be modified by exposure to other existing, contemporaneous or sequential stressors. A corollary is that cumulative risk assessment must consider the potential for more than just aggregate (i.e., additive) effects of multiple activities, but also the possibility of synergistic and antagonistic effects. Temporal and spatial variability in the intensity of the stressors (i.e., level of exposure) and in their combinations add to this complexity so that exposures can be accumulated in system components over time from different sources, and by different routes for multiple activities. In natural and human ecosystems, which are characterized by complex interactions between components of the physical, biological and social environments, the compounding of multiple stressors allows for indirect effects on system components, non-linear responses, and transitions of system states or functionality (Crain et al., 2008; Suding & Hobbs 2009; Selkoe et al., 2015).

As the number of activities and stressors under consideration in the analysis increases and the pathways through which they may have an effect proliferate, so too does the range of values and assets that are potentially at risk and the uncertainties associated with each outcome. Most cumulative risk assessments, therefore, fall into the realm of the Level 3 and Level 4 uncertainties described in Section 2.2.1, where there is a large range of plausible future outcomes, significant gaps in available knowledge and data, a variety of ways of representing the system using models, and a large range of interests and outcomes sought by different stakeholders. Negotiating a path for analysis of these complex problems requires decisions to be made about which dimensions of the problem should be pursued and which can be considered as ancillary information to inform risk management decisions (National Research Council 2009).

Two general approaches have evolved for dealing with this form of risk analysis under uncertainty, each of which is embedded in a participatory framework for decision-making that involves stakeholders and which gives precedence to evaluation of management options as part of the analysis. The first, eschews attempts to predict and rank system responses in a conventional probabilistic framework and, instead, seeks to hedge against uncertain and unwanted outcomes by identifying management options that are robust to a range of uncertainties. These approaches, collectively described as methods for Decision Making under Deep Uncertainty (DMDU), are described in Section 6.2.

The second approach follows a more conventional risk assessment pathway, but endeavours to make the analysis of risks more tractable by prioritising analytical effort to components of the system where detailed analysis will provide the greatest net benefit for evaluating management options (National Research Council 2009). This approach grew out of a series of critical reviews of cumulative risk assessments done using the US EPA Framework for Cumulative Risk Assessment (U.S. Environmental Protection Agency 2003) that recommended:

- *An iterative, Risk-Informed Decision-Making process* (see Section 3.1) in which the analysis is continually adapted and re-evaluated against management objectives as new information on the stressors of interest, system vulnerabilities and management options is obtained.
- *Stakeholder involvement in scoping risk and evaluation of management options* to enhance the transparency of the process and inform stakeholders of changes in the analysis as well as the logic underlying the selection of a risk management option.
- *Consideration of vulnerable* components of the system to improve estimation of risks and ensure that the selected risk management option(s) is adequately protective.
- *Using a tiered approach* to analysis so that resources are focussed on the most important factors contributing to risk.
- *Continued development of risk assessment methods* for evaluating multiple stressors, including approaches for assessing combinations of stressor footprints and evaluating exposure-response relationships in vulnerable populations.
- *Careful consideration of the appropriate spatial and temporal scales of analysis*. The size of the area and time frame under consideration will affect the data needed, model choice, and the utility of the information for comparing risk management options (National Research Council 2009; Gallagher et al., 2015).

A modified version of the US EPA's framework for cumulative risk assessment based on these recommendations is outlined in Table 9. It recognises the need for simple analytical methods to determine when more detailed, quantitative analysis is appropriate and where other forms of analysis are adequate to inform decisions about management options (Sexton 2012). The framework also embraces adaptive learning and risk management as initial analyses reveal more about system components and their vulnerabilities to the stressors.

The first stage of the modified framework involves development and validation of a conceptual model of the system with stakeholders that describes relationships among the stressors and their effects on system components and valued assets. In risk assessment, a conceptual model is a representation of the proposed mechanisms by which an activity or set of activities drive effects on the risk assessment endpoints. They are typically used at the problem framing stage to summarise the cause and effect relationships by which stressors affect the assessment endpoints (Dambacher et al., 2007). A key task in developing a conceptual model for cumulative risk assessment, therefore, is the identification of potential pathways of risks to describe the vulnerability of system components to the stressors.

The second stage uses available data and knowledge to apply a screening tool across the conceptual model to determine the stressors and system components that will be the subject of further analysis. The emphasis is on those that are likely to have the most influence on the objectives for management of the system. These then become the focus of detailed analysis in the third stage. Importantly, this framework recognises that different methodological approaches may be applied at different stages of the analysis and to different system components. These may include a mix of expert judgement, qualitative, semi-quantitative and quantitative risk analyses.

A wide range of methods has been proposed for the assessment of cumulative effects. These have broadly been classified into analytical methods and planning-based approaches depending on whether processes for evaluating management objectives are incorporated in the analysis (Smit & Spaling 1995). Analytical methods include use of simple checklists and matrices to aggregate hazards, spatial analysis and GIS for overlaying stressor footprints with vulnerable ecosystem components and spatial management measures, statistical or mechanistic models of system processes, network and food web models. Planning-based approaches include multi-criteria decision analysis, land suitability evaluations, and scenario-based evaluations of decision alternatives (e.g., linear programming; Section 6). A detailed evaluation of the advantages and limitations of these different methods is beyond the scope of this study and has, in any case, been the subject of earlier reviews (e.g., (Smit & Spaling 1995; European Commission 2001; Gentile & Harwell 2001; National Environmental Justice Advisory Council 2004; Sexton 2012; Stelzenmüller et al., 2018). Our purpose in this section is to highlight novel integrative methods that combine qualitative and quantitative modelling to assess cumulative risks. In keeping with the recommendations in Table 9, these combined approaches are intended to capture whole system behaviour, but to concentrate analytical effort on subcomponents of the system where they will most benefit selection of management actions.

Table 9: The National Research Council's modified version of the stressor-based approach to cumulative risk assessment. Source: National Research Council (2009).

Step	Activities
1	- Develop a conceptual model for the stressors and the way they cause effects emphasising those that would be significant influenced by one or more of the risk management options under evaluation. - Identify the receptors and end points affected by these stressors. - Review the conceptual model and stressors, receptors, and end points of interest with stakeholders in initial planning and scoping
2	- Use available scientific evidence, expert knowledge and screening-level benefit calculations to make an initial determination of which stressors should be included in the assessment. - Review and re-evaluate planning and scoping activities based on feedback from stakeholders and other parties to the decision. - Focus only on stressors that contribute to end points of interest for risk management options and which are either differentially affected by different risk management options or influence the benefits of stressors that are differentially affected.
3	- Evaluate the benefits of different risk-management options with appropriate characterization of uncertainty, including quantification of the effects of individual stressors and bounding calculations of any possible interaction effects.
4	- Conclude the analysis if the results from Step 3 are sufficient to discriminate among risk-management options given other economic, social, and political factors; otherwise, sequentially refine the analysis as needed, taking into account potential interactions among stressors

5.1 Qualitative network models

Qualitative Network Models (QNMs) or 'Loop analysis' (Justus 2006) can provide a useful analytical tool for structuring the analysis of cumulative risk. They provide a simple method for developing a conceptual model of the system collaboratively and for understanding its basic structure and dynamics. QNMs are a simple, graphical representations of system components, their relationships and the pathways through which perturbations may propagate. They are flexible, can incorporate feedback loops and can be used to inform more detailed quantitative sub-component models (Dambacher et al., 2007; Guikema & Aven 2010). Importantly, because they are a simple visual representation of complex systems, QNMs are intuitive and amenable to input from diverse stakeholder groups. Their utility in cumulative risk analysis is that they can be generated and analysed quickly without the need for quantitative data to parameterise them or advanced mathematical expertise.

QNMs portray the components in a system and the relationships between them as signed directed graphs (signed digraphs), which show whether increases in one system component are likely to result in an increase (+) or decrease (-) in other components or leaves them unchanged (0) (Figure 6). In the problem framing stage of risk analysis, they encourage a focus on structural (model) uncertainty (Section 2.2) so that analysts and stakeholders can debate the important features that need to be represented in more detailed analysis. Multiple, alternative signed digraphs can be used to represent different plausible model structures and/or different stressors and management regimes to determine their relative importance in system dynamics (Dambacher et al., 2007).

To develop the models, analysts specify the system components and positive, negative or zero interactions between them, and the dominance of negative and positive feedback in the systems (Dambacher et al., 2007). For example, Figure 6 shows the basic building blocks of a qualitative model of the structure of an ecosystem (Dambacher et al., 2009). The open circles depict system components which could be natural populations, environmental or socio-economic variables. The directed links between them represent direct effects. Solid lines ending in an arrowhead represent a positive direct effect, whereas lines ending in a filled circle denote a negative direct effect. Links connecting a component to itself represent self-effects. In the first model (i), component X and Y have direct negative effects on each other while Y has a positive direct effect on Z. In model (ii), component X enhances the intensity of the interaction between Y and Z (denoted by the dashed line). Model (iii) represents direct effects of X on Y and Z.

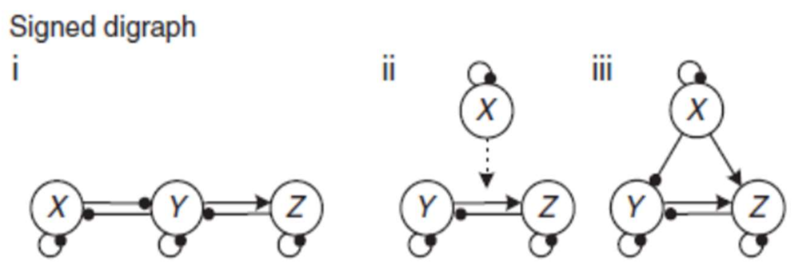

Figure 6: Example signed digraphs of model ecosystems. Source: (Dambacher et al., 2009).

Each signed diagraph can be represented in matrix form, where each element of the matrix a_{ij} represents the direct effect of component j on component i. The pairwise interactions in model (i) of Figure 6 can, therefore, be represented as:

$$\begin{bmatrix} -a_{XX} & -a_{XY} & 0 \\ -a_{YX} & -a_{YY} & -a_{YZ} \\ 0 & a_{ZY} & -a_{ZZ} \end{bmatrix}$$

and in model (iii) as:

$$\begin{bmatrix} -a_{XX} & 0 & 0 \\ -a_{YX} & -a_{YY} & -a_{YZ} \\ a_{ZX} & a_{ZY} & -a_{ZZ} \end{bmatrix}$$

Once the signed digraphs have been constructed, modelling of the system can use either graphical algorithms or matrix algebra to evaluate the stability of the system and to predict the effects of disturbances (Dambacher et al., 2009). The signed digraphs allow examination of indirect interactions and feedback properties of a system. System feedback is formed by the products of links within the system. Negative feedback acts in the opposite direction of an initial change in a component and acts to maintain stability in the system. Perturbations to inputs in the models can be used to examine shifts in the equilibrium state. Predictions for the direction of these shifts are obtained by a summation of all direct and indirect links through the network from the input source to each affected component. This summation of effects can be obtained from a qualitative analysis of the inverse community matrix (Dambacher et al., 2002). An analysis of the system's structure and feedbacks allows its stability to be evaluated and enhances understanding of drivers and constraints on interacting components (Dambacher et al., 2009). Details of the mathematics involved in this qualitative modelling can be found in Dambacher et al., (2002) and Dambacher et al., (2003). Software tools for constructing signed digraphs and for implementing QNMs are available on the R-CRAN repository (Dinno 2018).

Recent extensions to QNMs have included development of Bayesian approaches for comparing alternative qualitative models to examine uncertainty in model structure (Melbourne-Thomas et al., 2012) and using them to incorporate feedback dynamics in semi-quantitative Bayesian Belief Networks (BBNs; (Hosack et al., 2008)). Like QNMs, BBNs are used increasingly for analysis of complex risks in systems where quantitative data are not available for many system components (Pollino et al., 2007). They similarly use a graphical depiction of the conceptual model that captures beliefs about the cause-effect pathways linking stressors to assessment endpoints and, like qualitative network models, are intuitively straight-forward to use with stakeholders. Unlike qualitative modelling, however, the relationships between variables in a BBN are one-way because they represent a conditional probability distribution that describes the relative likelihood of each value of the "child" node (end of the arrow) conditional on every possible combination of values of the "parent" nodes (start of each arrow). The conditional relationships in a BBN are directional (i.e., 'acyclic'; a path traced along its links cannot pass through a variable more than once) so they cannot represent feedback within the system. Hosack et al., (2008) developed a method for embedding a qualitative analysis of sign directed graphs into the conditional probability tables of a BBN. Their approach generates the conditional probability tables needed for a BBN directly from the signed digraph negating the need to parameterize conditional probabilities through expert elicitation or empirical data. It allows feedback behaviour within the framework of a BBN and provides a way to evaluate the effects of uncertainty in model structure. Software for implementing these routines is available in the Supplementary Materials provided by Hosack et al., (2008).

5.2 Integrated methods for cumulative risk assessment

The utility of QNMs in cumulative risk assessment is that they facilitate a graduated approach to the overall analysis, as recommended in Table 9. By this we mean they provide a framework in which a rapid initial analysis can provide direction on subcomponents of the system that are expected to have relatively low risk and those where risks are likely to be greatest and additional analysis is required. Most complex assessments of cumulative risk and impacts will have some ability to utilise mechanistic quantitative models for components of the system under consideration. In other areas, where there is a lack of data or knowledge, qualitative analysis or semi-quantitative analyses may be more appropriate. By providing a conceptual structure of the relationships among system components and stressors, qualitative models facilitate use of a hierarchy of tools, from simple, rapid and low-cost tools to progressively more complex and costly tools to support the prioritisation of management actions. Below, we describe two case studies in which qualitative modelling has been used to structure and organise large-scale assessments of cumulative risk.

5.2.1 Case study 1: Bioregional assessment of risks from coal seam gas and coal mining on water resources.

In 2012 the Australian Government initiated a series of comprehensive bioregional assessments to evaluate potential impacts of coal seam gas and large coal mining developments on freshwater resources. The assessments were designed to provide baseline information on the ecology, hydrology, geology and hydrogeology of a bioregion with explicit assessment of the potential direct, indirect and cumulative impacts of coal seam gas and coal mining development on water resources (Barrett et al., 2013). Risk analysis was an important part of the assessments, which were tasked with providing *'sufficient scientific advice to analyse the level of risk associated with impacts on water-dependent assets, particularly those of high value such as listed threatened species, state-listed wetlands and other important ecological and cultural water features'* (Barrett et al., 2013).

A central component of the impact and risk analysis was a conceptual model of the chain of causation, that related activities associated with the coal resource development within each bioregion or subregion to hydrological response variables (e.g., groundwater drawdown) that could result in changes in the environment and ecology (represented by receptor impact variables, such as vegetation condition). Qualitative network models were used as the method to develop these receptor-impact models at a landscape level. The signed digraphs and their associated community matrices were developed by a multi-disciplinary panel of experts who used their current knowledge of the system to identify the key ecosystem components and dependencies, and their relationships to groundwater and surface water variables (Figure 7). These hydrological variables were then translated into hydrological response variables that could be estimated by quantitative groundwater and surface water hydrological models used to support the assessment.

A key feature of the qualitative models was their ability to explore potential direct, indirect and cumulative impacts of the stress imposed on the ecosystem by coal resource development. Within the signed digraphs, a pathway with one arc can be formally defined as a 'direct effect', whereas a pathway consisting of two or more arcs, involving other intermediate variables, is considered an 'indirect effect'. Cumulative impacts were depicted by changes to two or more potentially interacting hydrological response variables, with concomitant direct and indirect impacts arising from them (Honsack et al., 2018).

The qualitative models were used to assess potential candidate receptor impact variables and hydrological response variables to consider within a receptor impact model. A suite of hydrological scenarios based on plausible combinations of hydrological response variables was developed for each receptor impact model. These scenarios were then presented to a panel of experts as part of an expert elicitation process, and the distribution of receptor impact variable under each scenario elicited. Statistical models were constructed (using generalised linear models) based on the experts' responses to estimate the relationship between the receptor impact variable and the relevant hydrological response variables, and to enable predictions to be made for combinations of hydrological response variables not considered in the elicitation (see Section 4.3.2 for examples of this scenario-based elicitation of prediction models). The statistical models were then used to make probabilistic predictions for changes at specific locations in the landscape, based on the modelled changes in those hydrological response variables at that location in different time periods.

Use of the sign-directed graphs to represent the stressor models provided a conceptual bridge to organise and calibrate other qualitative, semi-quantitative and quantitative analyses. They allowed qualitative predictions of long-term perturbation impacts to variables and provided input into the design of a quantitative receptor impact model that predicted the magnitude of impact in the response variables given changes in the identified hydrological response variables.

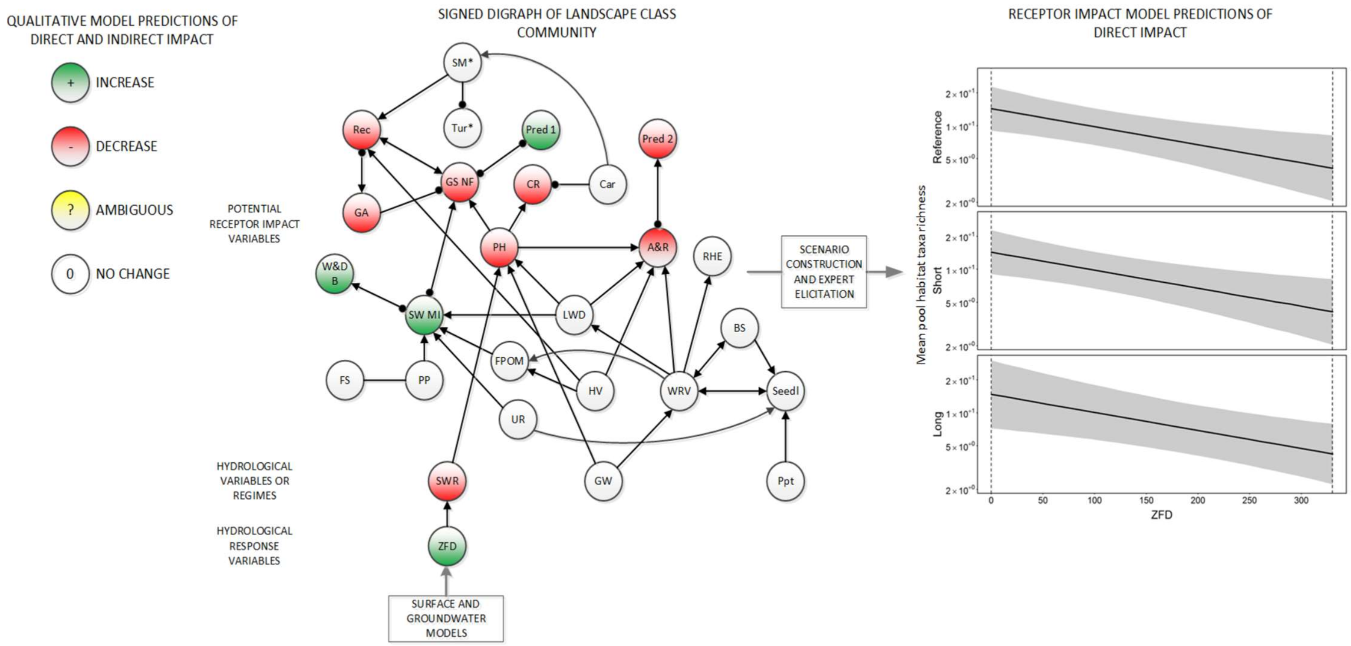

Figure 7: Signed digraph depicting how direct and indirect impacts were assessed qualitatively and quantitatively within the bioregional assessments. In this scenario a decrease in hydrological variable surface water replenishment (SWR) is modelled as an increase in hydrological response variable zero flow days (ZFD; averaged over 30 years) The impact of this direct effect is quantified through the receptor impact model (right-hand panel). The receptor impact model (a generalised linear model developed from scenario-based expert elicitation) predicts how the median value of the receptor impact variable changes in response to changes in the hydrological response variable (HRV; black line in right-hand plots) together with the uncertainty associated with these predictions (grey polygons). Source: Honsack et al., (2018).

5.2.2 Case study 2: Analysis of cumulative risks to marine ecosystems

Dunstan et al., (2015) described a similar, tiered approach to the assessment of cumulative risks and impacts on marine ecosystems that allows for three different levels of analysis (Figure 8). Their framework, based on a model developed to assess ecosystem risks from commercial fishing (Ecological Risk Assessment for the Effects of Fishing, ERAEF; Hobday et al., (2011)), progresses from an initial comprehensive qualitative analysis of risk (Level 1), which they refer to as a 'scale intensity consequence analysis', through a more focused semi-quantitative approach (Level 2), to fully quantitative "model-based" approaches on system sub-components that have been identified in lower level assessments as having high consequences and high uncertainty (Level 3).

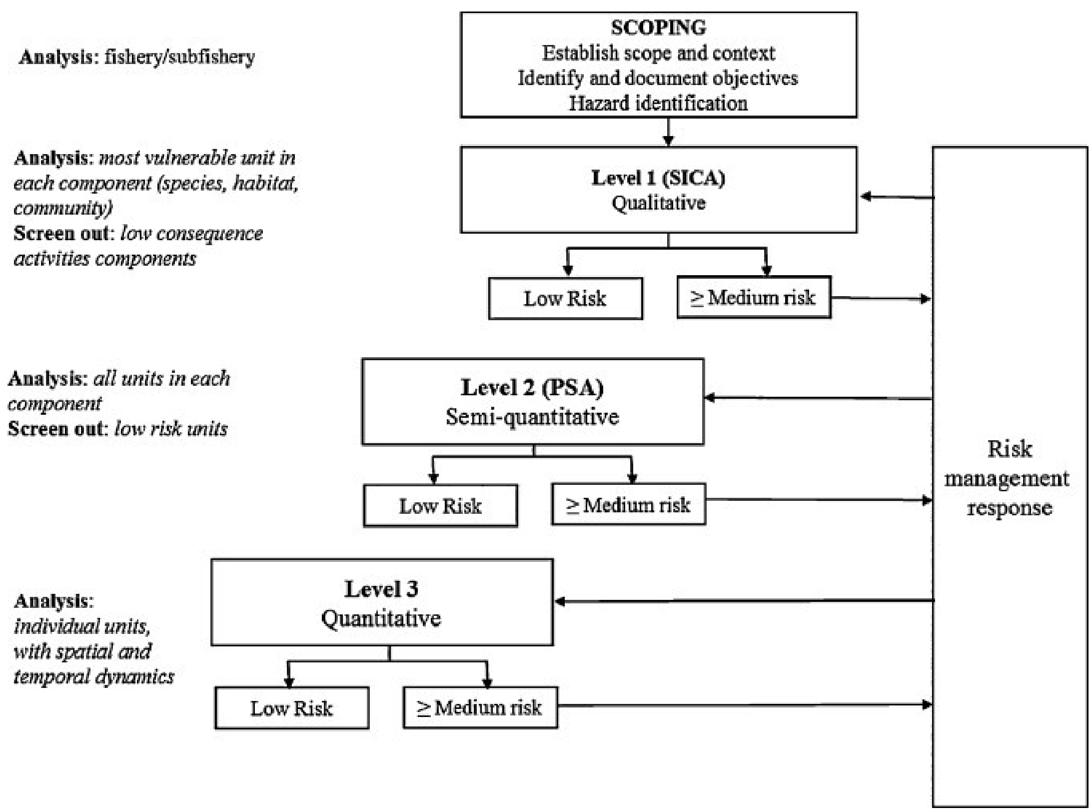

Figure 8: Schematic of the hierarchical ERAEF framework. The focus of analysis at each level is described in italics on the left side of the diagram. Source: (Hobday et al., 2011).

The first level of assessment again develops a general conceptual model of the system, based on the knowledge of experts and stakeholders. It establishes the causal chain of receptor-impact relationships which map identified stressors onto system components and the valued assets that are to be protected. In the original ERAEF model, this was done using an exposure-effects assessment approach in which the range of stressors is identified and a standardized set of scoring guidelines is used to rate the scale and intensity of each stressor (≈ exposure) and its consequences (≈ effect) for a range of assets (Hobday et al., 2011). This screening level assessment allows relatively low consequence effects to be excluded from subsequent analysis and/or management responses designed to mitigate them.

Level 2 analyses utilise qualitative network models (Loop analysis or BBN) that are built from the information obtained during the initial screening level analysis. The qualitative models give a more formal structure to the conceptual model of the system and allow better understanding of sub-components and their relationships. Outputs from the models may be used to identify high risk sub-components that require more considered analysis and to provide inputs into more detailed (and costly) quantitative analyses undertaken at Level 3.

As with other graduated approaches to cumulative risk assessment, a range of quantitative tools may be used at Level 3 to inform the analysis. These are tailored to the resources, data and information available and to the problem under consideration. They can include statistical models, spatial analysis, mechanistic process models, end-to-end ecosystem models (e.g., Atlantis, (Audzijonyte et al., 2019)), and models of intermediate complexity ('MICE' models, (Plagányi et al., 2014)).

At each level in the framework, an analysis occurs that determines the level of risk of stressors to assets. Risks below an agreed threshold are screened from further analysis. This tiered approach of increasingly quantitative analysis, with associated increases in data requirements and analytical costs, means that resources can be efficiently targeted at the system components where there is greatest uncertainty or potential for large consequences.

6 Scenario-based methods for evaluating management alternatives under uncertainty

6.1 Decision analysis

A variety of analytical tools has emerged from the field of decision analysis to evaluate policy choices under uncertainty. Classic decision analysis involves the formal analysis and ranking of decision alternatives against multiple, and often conflicting, decision objectives. Uncertainty is described with well-characterized probabilities and optimal strategies are evaluated based on objectives specified for the policy outcomes (Means et al., 2010).

Unlike conventional probabilistic risk assessment, where the analytical component of the assessment is separated from the evaluation of policy alternatives, decision-analytical methods typically use an analysis that iterates between examining modelled future scenarios and assessing their performance against one or more stated objectives for the system under consideration (Gregory et al., 2005). The unique contribution of decision analysis to risk analysis is the use of an objective or utility function, $u(x)$, to indicate the desirability of each consequence (x) relative to all other consequences. This function is assumed to aggregate the objectives for the problem under consideration and the attitude of the decision-makers to risk (Keeney 1982). The premise is that all decision problems involve subjective judgements about the desirability of the consequences from different decision alternatives and that the likelihood of those consequences and their desirabilities can be estimated separately using probabilities and quantifiable utilities, respectively. The expected utility of each decision alternative can be calculated as the weighted average of all possible outcomes of the decision, with the weights being assigned by the probability that any particular event will occur. In a rational decision context, the decision alternatives with the highest expected utilities should be preferred (Keeney 1982).

Decision analytic methods typically deconstruct complex problems into smaller components (e.g., individual objectives, probability distributions) that can be examined separately. Optimization methods (e.g., mathematical programming) are used to sort through prospective outcomes from a policy option following subjective expected utility theory to achieve the best solution relative to the stated policy objectives.

Most decision-analytic methods for risk and decision analysis are designed to identify ‘optimal’ strategies contingent on a characterization of uncertainty that obeys the axioms of probability theory (Lempert et al., 2006). They begin with a system model that generates outcomes of interest given a choice of strategy and a set of probability distributions over the model’s input parameters to characterize the uncertainties. The analysis seeks the strategy with the optimal expected utility. This approach generates the best possible strategy when the uncertainties are relatively well characterized by single probability distributions (e.g., aleatory uncertainty or natural variability) and the objectives for management can be expressed as some form of utility.

6.1.1 Markov Decision Processes

Markov Decision Processes (MDP) are one of the optimization procedures that are used to model decisions under uncertainty (Puterman 2008). They are useful for sequential decisions that involve trade-offs between strategies, for deciding on optimal actions that can be applied to different parts of a system that may be in different states (Forsell et al., 2011), and for adaptive management problems where there is the opportunity to learn from iterative decisions (Chadès et al., 2017). MDPs define (1) a set of possible states for the system, (2) a set of decision actions that could be implemented, (3) a transition function that represents the probability of the system moving from one state to another given a decision action and (4) a reward or 'utility' function that provides the basis for evaluating the benefits of any decision for the system. In this context, finding the best decisions under uncertainty is an optimisation problem in which the sum of future expected utility is maximised over time (Marescot et al., 2013).

Advances in computing power have increased the functionality and accessibility of MDPs. Extensions have included characterising spatially heterogeneous resources as a network (Graph-based Markov Decision Process), where each spatial domain is represented as a node within a network and dependencies between the domains are denoted as edges between them (Forsell et al., 2011). Methods have also been developed to model hierarchical decisions with different time horizons (Kristensen & Jørgensen 2000; Ge et al., 2010).

MDPs have had some application in biosecurity, public health, and conservation science (Chadès et al., 2011; Marescot et al., 2013; Chadès et al., 2017), but new software tools will likely increase their availability to other researchers (Marescot et al., 2013; Chadès et al., 2014). For example, Ge et al., (2010) used Markov decision processes to assess alternative control strategies for foot-and-mouth animal disease outbreaks. In their analyses, they evaluated the utility of using multi-level, hierarchic Markov processes over a static control approach (i.e., pre-defined strategies with a single decision point). By using dynamic Markov processes, the costs of interventions were more accurately estimated and better guidance was provided to the decision-maker.

6.1.2 Dynamic programming

Dynamic programming is one of a number of decision-analytical methods that can be used to evaluate the effects of management decisions on dynamic ecological and social systems where there is uncertainty in the behaviour of the system and in the effect of the decisions. It is a simulation and optimization technique that is applied most usefully to situations where decisions are made repeatedly through time about the management of a dynamic resource and where there is uncertainty about its future state and the effect of the management upon it. It has been advocated as a way of developing optimal strategies for adaptive management (Hauser & Possingham 2008; Chadès et al., 2017) and for evaluating the effects of management decisions on threshold shifts in the condition of natural systems (Martin et al., 2009).

Dynamic programming uses a mathematical representation of these decision-problems that:

i. simulates different states that a system may exist in and the likelihood of transition among them
ii. considers an optimality condition for all management decisions and,
iii. takes account of how uncertainty affects transition among states and the decisions applied at each stage.

A stylized representation of the method is depicted in Figure 9. The grey arrow represents the timeframe (or states, periods) of the analysis and $1, 2, \dots, t$ are different time intervals such as years, months, weeks or other time metrics. Decision-making is optimized over the entire period of analysis (blue arrow) to achieve an overall, long-term objective (green box).

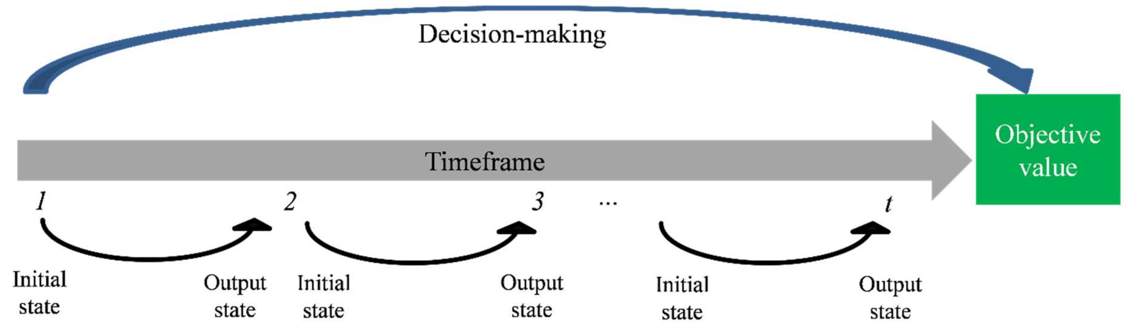

Figure 9: **Schematic representation of dynamic programming.**

In a dynamic programming model it is assumed that the decision maker has a perfect foresight over the entire timeframe under model consideration and that decisions are adjusted to achieve the optimal outcome over the period of analysis. This programming approach is subject to constraints that restrict the optimization level and can include constraints such as initial state variables and resource endowments (Kennedy 1986).

There are six basic steps to implementing a dynamic programming approach:

1. The management objectives for the resource must be defined and agreed upon. This includes defining the time period over which the actions are to be evaluated and the intervals at which decisions are revisited.
2. A set of states must be defined that describe the condition(s) that the resource could be in at each time step (including highly uncertain states).
3. The management actions that can be implemented to change (or maintain) the state of the resource to meet the objective must be specified.
4. A mathematical representation (model) is constructed that describes the system dynamics and how the management actions affect transition from one state to another.
5. A utility function is defined which describes the performance of the system relative to the management objectives. This could be in the form of economic outcomes, levels of a biological population or a measure of social improvement.
6. The final step is determining the strategy or series of actions that optimizes the chances of achieving the overall management objective(s).

A key element in dynamic programming and other optimization techniques is deciding upon an appropriate objective value function that will define the optimality condition for management (i.e., the measure by which the performance of the outcome will be assessed; (Bellman 1954). Table 10 summarises a range of ecological studies that have applied dynamic programming to natural resource decision-making under uncertainty. The summary details the management objectives specified for the resource, the alternate states defined in the modelled system, critical elements of uncertainty and the management strategies that were evaluated. Performance of each strategy was evaluated using a utility measure (or measures), U, intended to quantify the objectives for the resource.

Applications of dynamic programming can be classified into four types: deterministic dynamic programming, dynamic programming stochastic simulation, dynamic stochastic programming (i.e., including uncertainty and risks) and dynamic programming risk models that consider the risk aversion of decision makers. Deterministic approaches to dynamic programming have been used in a number of marine applications (e.g., (Bjørndal et al., 2004)). Dynamic programming models have been applied to the management of fish stocks (Schaefer 1954; Hart 2003; Grafton et al., 2008; Maroto et al., 2012), multiple species interactions (Clark 1985), extraction of marine resources such as oil, gas and minerals (Conrad 2010), and economy-wide impacts of introduced policy and technological changes on marine environments (Asche 1997).

Inclusion of uncertainty and risks provides information on possible various states of nature and allows analysis of impacts on marine environment and resources (living, natural and minerals) in extreme situations. Hence, by considering uncertainty and risks in dynamic programming, it is possible to incorporate different states of nature (e.g., low, mean and high yield levels of fish) and reveal different simulated outcomes (e.g., profits of fisherman under low fish yield, profits of fisherman under high fish yield). Stochastic dynamic programming incorporates variability into the model dynamics. Variation can be parameterised on the basis of historical data, expert elicitation, or through the use of various simulation techniques such as Monte Carlo simulation and Latin Hypercube sampling (Hardaker et al., 2004).

Table 10: Examples of the application of dynamic programming to the management of ecological resources.

Management problem	*Objective*	*System states*	*Management strategy / decision at each time step*	*Critical uncertainty*	*Performance measure* (U)	*Study*
Optimal adaptive harvest strategy for hypothetical population under uncertain dynamics	Maximize the expected number of years in which harvesting takes place over a finite time horizon	Population status: ▪ Collapsed ▪ Vulnerable ▪ Robust	▪ Harvest ▪ Do not harvest	Probability of recovering from a collapsed state to a vulnerable state	Harvest value	Hauser & Possingham (2008)
Optimal allocation of resources to manage spread of invasive pests or diseases	Minimize the number of infected nodes by taking actions to increase the probability of eradication in an infected node.	Node status: ▪ Infected ▪ Susceptible	▪ Manage ▪ Survey ▪ Do nothing	Probability of detection in a node	A function that reflects the number of uninfected (susceptible) nodes relative to the cost of actions to achieve them	Chadès et al., (2011)
Optimal control of a population of wolves in Europe	Maximize the wolf population while providing that the population does not exceed 250 individuals and remains above 50 individuals.	The population size (N_t) at time t	Harvest rate 0-100%	Variability in population growth rates	An index that reflects whether the population size is within the target range specified by the objective	Marescot et al., (2013)
Optimal control of foot-and-mouth disease epidemics	Minimising the total costs of control, including direct expenditures on control and export losses due to a ban on meat sales.	Epidemic status: Infection Index (J_t) that reflects the moving average of the number of newly detected herds that are affected	▪ Implement compulsory EU measures ▪ EU measures + pre-emptive culling on neighbouring farms	Epidemic growth rate	Net present value of the total rewards of a policy	Ge et al., (2010)

Management problem	*Objective*	*System states*	*Management strategy / decision at each time step*	*Critical uncertainty*	*Performance measure (U)*	*Study*
			▪ EU measures + emergency vaccination on neighbouring farms			
Optimal harvest strategy for multiple duck species	Maximize the total harvest over the time horizon	Autumnal population sizes for each species at time t	18 regulatory scenarios that included combinations of bag limits and length of the hunting season	Population recovery rate per species	Average annual (aggregate) harvest; Average season length; Average number of years between a change in regulatory option; Average (aggregate) breeding population size.	Johnson et al., (2019)

Studies that have used a stochastic dynamic model have included applications to identify optimal stock levels, search and information sharing strategies for fishermen, investment, regulation of fishing effort and catch (Clark 1985; Weitzman 2002; Hannesson & Kennedy 2005). For example, Hansen & Jensen (2017) considered ecological, variable economic and structural economic uncertainties in fisheries regulation, since most of the regulators are subject to uncertainty and argue that it is important during the development of pro-price regulation policy often observed in fisheries economic studies. Uncertainty under dynamic context also allowed taking into the account asymmetric information and compliance of fisheries regulation (Jensen et al., 2017).

Li et al., (2014) developed a dynamic programming model that combined a Monte Carlo simulation and mixed integer nonlinear programming to analyse the recovery of spilled oil in offshore areas. They showed that mechanical collection had an important effect on the transport and state of oil. More importantly, they concluded that integrating dynamics and uncertainty in the model facilitated better decisions on the allocation and operation of oil recovery devices. Studies have also investigated the effect of uncertainty in environmental forecasting of offshore conditions and its impact on operations to install and maintain marine infrastructure (Natskår et al., 2015). Link et al., (2012) used dynamic programming to explore the effects of a range of different types of uncertainties on predictions from ecosystem-based fisheries models, including natural and market variability, observation errors, model structure and inadequate communication among stakeholders, decision-makers and scientists. The model outputs provided insight into best practice to address each of these uncertainties in ecosystem models.

Decision-makers make decisions about the best way to achieve desired outcomes based on the current state of a resource and uncertainties about its future state. Their attitudes to risk and to risk aversion often vary depending on the management context. By knowing the attitude of decision makers towards risks it can be possible to develop risk management options that are appropriately conservative or risky. Such attitudes toward risk are not captured in dynamic programming stochastic simulation models. Dynamic programming risk models can, however, be a suitable tool for incorporating risk aversion, as they can consider the behavioural response to risky situations in the objective function of the dynamic programming (Ruszczyński 2010).

Mean-variance analysis is often applied to solve problems of the optimal choice of management strategies in uncertainty and risky settings. In this approach, if activities have the same returns (or variances) but one activity has lower variances (or higher expected return) then this activity is preferred by decision makers (Markowitz 1952). Mean-variance analysis selects the single and optimal activity considering the expected and variance of outcomes, (i.e., the trade-offs between the mean and variance values of outcomes). The mean-variance approach can be extended into the dynamic mean-variance programming model to model the attitude of a decision maker and his/her planning processes through time in a dynamic context (Djanibekov & Villamor 2016).

The objective function used in mean-variance programming is usually maximization of certainty equivalents in terms of profits or other objective value from different activities over time. In this case, the certainty equivalent is the certain profit level that is rated by the decision maker equivalent to an uncertain profit level and includes an annual expected profit value reduced by the annual risk premium. The risk premium is the minimum amount of money required to motivate a decision maker to select a relatively risky activity. Hence, the main characteristics of dynamic mean-variance programming model are consideration of:

- state variables that change from one stage to the next stage,

- transition functions that transform the state variables in one of the stages into the state variable in the next stage,
- a set of possible states of the nature (i.e., distribution of parameters) for each stage, and
- an optimization procedure considering the expected and variance values and risk aversion of decision maker (Djanibekov & Villamor 2016).

To our knowledge there are few studies that have incorporated dynamics and risk aversion in the context of marine studies. For example, Walters (1975) combined a dynamic programming model and stochastic elements, and analysed optimal harvest strategies for fish under environmental variability and uncertain production. He showed that the fish management strategies are sensitive to the mean and variance of catches. Ewald & Wang (2010) analysed the maximization of expected sustainable yields under variance constraints by deriving optimal harvesting strategies for risk-averse decision maker and showed that the optimal fishing effort needs to consider trade-offs in mean and variance of outcomes and is decreasing with the level of risk aversion. Doyen et al., (2007) showed that the ecological and economic risks are reduced when viability strategies are introduced to sustainably manage the marine ecosystem. Allan et al., (2011) analysed investment into the electricity technologies under uncertainty and reported that the lower variance energy technologies can be installed without increasing costs by expanding the marine renewable technologies.

6.2 Decision-Making Under Deep Uncertainty

As described earlier, most risk assessments (Sections 2.1) and scenario evaluations (Section 6.1) analyse and evaluate a subset of the most plausible risk scenarios selected from a larger range of possible scenarios. A single (or relatively few) 'best' risk model is used to represent the system and to simulate the effects of alternate management decisions and their consequences. In this form of analysis, uncertainty is usually incorporated by defining a set of probability distributions for the model input parameters which are then sampled across to understand the range of possible outcomes for a given management option and their likelihoods. These probability distributions are derived from existing data, literature or expert judgement (Section 4). Uncertainty about the probability distributions or about the structure of the model itself is typically addressed by implementing a *sensitivity analysis* to evaluate how the model's outcomes are influenced by the different sources of uncertainty in the inputs (Saltelli 2002).

In situations of deep uncertainty there is not a clear path for choosing which scenarios to model or the appropriate model structure (Kwakkel & Haasnoot 2019). Deep uncertainty occurs when the risk analysts and other experts involved in the risk assessment do not know or cannot agree on the appropriate ways to describe key relationships among: (i) the important influences on the system that affect future outcomes, (ii) the probability distributions used to represent uncertainty or (iii) how to rank or value different predicted outcomes (Section 2.2.1). Indeed, in many situations where there is a lack of knowledge or data about the system of interest it is likely that many different risk models could plausibly represent it. Uncertainty about which specific scenarios to model can also create an impasse in decision-making as different parties dispute the assumptions used to construct and parameterise it (Lempert et al., 2006).

In this section, we describe a range of methods that have been developed to address these situations of Decision-Making Under Deep Uncertainty (DMDU).

6.2.1 Exploratory Modelling and Analysis

'Exploratory Modelling and Analysis' (EMA; or 'Computer Aided Reasoning' – Bankes et al., (2001)) describes a suite of techniques that have been developed to evaluate decision alternatives in complex systems. Conceptually, they have emerged from the field of Assumption-Based Planning (ABP) in which the objective is to prevent a policy from failing by examining its vulnerabilities, particularly the assumptions underlying the policy and the consequences that would occur if they are violated. To deal with potential surprises and to avoid undesirable long-term outcomes, candidate policies are selected on the basis of their performance across a wide range of potential future scenarios, not simply on what is considered the most likely future (Dewar et al., 1993; Marchau et al., 2019a). The objective of ABP is not to choose a single optimal scenario based on a policy decision, but to identify the vulnerabilities in policy choices – things that could go wrong if the assumptions are incorrect - and to evaluate options that best reduce the vulnerabilities and hedge against the possibility of failure (i.e., to make it more robust to uncertainty). Common elements of ABP include:

- identifying important assumptions in a policy setting or plan which if violated could cause its failure,
- identifying vulnerabilities of the policy associated with each assumption (i.e., changes that would occur if the assumption does not hold),
- defining 'signposts' – events or thresholds that indicate changes in a vulnerability of an assumption – and 'shaping actions' that are designed to avert or to cause the failure of an assumption,
- defining 'hedging actions' that are intended to prepare the parties for failure of an important assumption. These are actions that can be taken to achieve the long-term goals even if the assumptions in the plan are not met.
- continuous monitoring to track progress toward the long-term objectives and detect 'signposts' (Dewar et al., 1993; Walker et al., 2013a).

EMA approaches leverage advances in computing power and speed to aid ABP. They are based on the idea that, under conditions of deep uncertainty, where it is unclear which model, parameters and outcomes describes the risk problem most appropriately, an ensemble of plausible models - rather than any single model - best represents the available information about possible futures (Lempert 2002; Lempert et al., 2004; Lempert et al., 2013b). Uncertainty is characterized by the multiple representations of the future rather than by a single set of probability distributions.

EMA approaches utilise computational tools to make large numbers of runs of system models in order to explore the full range of uncertainties and to identify situations in which a plan would fail (Walker et al 2013a). They generate a broad range of plausible future scenarios based on available information, data and the identified policy alternatives (a step referred to as 'case generation' (Section 6.2.2, Lempert et al., (2013b)). Data analysis and visualization tools are then used to explore across the scenarios to obtain insights into how the system would behave under different assumptions and violations of those assumptions (Bankes et al., 2013).

As a simple example, consider a model in which there are *N* parameters each of which could assume a range of values. Each model run with a unique combination of parameter choices represents a single plausible future scenario from the ensemble of possible parameter combinations across *N*-dimensional space (Bankes 1993; Walker et al., 2013a). Large uncertainty can mean that the properties of each model can take a wide range of values across the parameter space so that many combinations of model structure and parameter settings must be examined to describe the full range of plausible outcomes. EMA approaches are, nevertheless, not restricted to dealing just with parametric uncertainty. They can also be utilised where the relationships between model components are unknown but can be approximated by simple functional relationships (Bankes 1993).

Because they attempt to explore the full "space" of possibilities created by uncertainty in the many variables that describe a complex system, implementation of EMA often involves use of simplified, integrated models (variously described as '*meta-models*', or "*fast and simple*" models; (Walker et al., 2013a)) rather than single complex, process-oriented models. Integrated policy models are simple, aggregated models that approximate the behaviour of more complex process models (Haasnoot et al., 2012). To cover the "space" of plausible futures, EMA approaches typically compute very large numbers of model runs (thousands to hundreds of thousands of scenarios). The process of selecting which models to run from the full ensemble is a key element and is driven by the policy question under consideration (Bankes et al., 2013). Data mining and visualization techniques are typically utilized to help parties to the policy decision – analysts, policy makers and stakeholders – generate hypotheses about desirable strategies and to search or sample systematically across the ensemble to identify those strategies and circumstances that achieve the outcomes of interest and which avoid undesirable consequences. (Bankes 1993; Bankes et al., 2013). This can include exploring circumstances in which a policy will perform well or fail, understanding the range of dynamics that the system could plausibly display or identifying circumstances in which the system may exhibit atypical or unexpected behaviour (Kwakkel & Pruyt 2013). Because EMA approaches allow the vulnerabilities of policy alternatives to be explored across a broad range of plausible circumstances they can facilitate the design of strategies that hedge against catastrophic outcomes (Bankes et al., 2013).

EMA approaches are best viewed as methods to "stress-test" a broad range of policy options (Lempert 2019). They incorporate computer aided support to enable participatory decision-making. Computer generated visualizations help users develop propositions about desirable strategies and analytical searches through the ensemble of scenarios can then help test these propositions (Lempert 2002).

Several general approaches and analytical tools have been developed that use EMA. They include:

- Robust Decision Making (RDM) (Groves & Lempert 2007; Lempert & Collins 2007; Lempert 2019)
- Many Objective Robust Decision Making (MORDM) (Kasprzyk et al., 2013; Herman et al., 2015)
- Dynamic Adaptive Policy Pathways (Haasnoot et al., 2013)
- Info-Gap Decision Theory (Ben-Haim 2006; Hemez & Van Buren 2019) and
- Real Options Analysis (Woodward et al., 2014; de Neufville & Smet 2019)

In the following sections, we provide a general overview of Robust Decision Making, Many Objective Robust Decision Making and Dynamic Adaptive Policy Pathways and describe how they have been applied to complex environmental decisions. These approaches have so far had most application in the fields of climate change adaptation, infrastructure planning and flood management. Our aim is to introduce them to the field of risk management in ecological systems. Detailed descriptions on each approach and on Info-Gap Theory and Real Options Analysis can be found in Marchau et al., (2019a) and in the literature cited in each section.

6.2.2 Robust Decision Making (RDM)

Robust Decision Making (RDM) was developed by researchers at the RAND Corporation to enable choice among different management actions in situations of deep uncertainty (Lempert 2002; Lempert et al., 2006; Groves & Lempert 2007). The overall objective of RDM is to identify policies that are 'robust' to uncertainty; strategies that perform well, compared to alternatives, over a wide range of plausible, uncertain future states of the world (Walker et al., 2013a; Lempert 2019). RDM encompasses a variety of approaches that utilize computerized analysis and visualisation to explore policy options and the effects of trade-offs (Lempert & Collins 2007).

There are five general steps (Figure 10).

1. Decision framing

The first stage involves participatory scoping of the risk problem to structure the overall analysis. Decision-makers, planners and stakeholders work together to frame the decision problem, identify objectives for management, actions or policies that could be pursued, and measures for evaluating the performance of those policies.

Lempert et al., (2003) introduced the concept of the "XLRM Matrix" to help structure this scoping phase. The XLRM Matrix is used to document:

- *Exogenous uncertainties* ("X"s) – drivers of the system that are outside the control of the decision-makers and which may influence the success (or failure) of their strategies over the long-term;
- *Policy levers* ("L"s) - the near-term actions that decision-makers can take and which, in combination, comprise the strategies that they wish to explore;
- *Relationships* ("R"s) (generally represented by simulation models) that represent the way that the uncontrolled influences on the system ("X"s) and the management actions ("L"s) relate to each other and to the performance measures ("M"s).
- *Metrics* ("M"s) – measures used to assess performance of the strategies.

For example, stakeholders may express concern that decline in a fishery (an "M") may be associated with increased sediment run-off from deforested land (an "X") that has caused changes in nearshore benthic environments. Options for management could include an increase in riparian planting in surrounding catchments ("L") to mitigate erosion. Stakeholders may also express concern that the costs (an "M") of planting will be greater than the financial returns from the commercial fishery or that there are other potential stressors ("X"s) acting on the coastal environment that contribute to the decline (e.g., climate change, over harvesting, etc). Relationships between sedimentation, replanting and productivity of the fishery ("R") could be described as a series of qualitative functional relationships or by quantitative process models describing erosion under different climate and replanting scenarios and the effects of sediment on fish populations. In either case, the deliberation needs to establish plausible ranges for the uncertainties or assumptions made in defining the problem. These can be derived from existing information, data or from expert elicitation (Section 4).

This scoping stage of RDM requires strong facilitation to ensure that the full range of possible risks and ideas for managing them is captured in the initial XLRM matrix. A key focus of the facilitation process is also deciding on the performance measures ("M") that will be used to rank the desirability, success or failure of various scenarios. These measures embody the long-term outcomes that the stakeholders and other parties to the decision wish to achieve and/or the future states that they wish to avoid. For example, in exploring options for global sustainable development using RDM, Lempert et al., (2003) used a suite of indicators of well-being derived from the United Nations Human Development Index (HDI) and the concept of a Green GDP. In evaluating strategies for future-proofing water supply to London and the Thames Basin, Matrosov et al., (2013) used five metrics for the supply of water: reliability of water supply service, reservoir storage susceptibility, environmental performance, energy consumption and total costs (capital and operating). Shortridge & Guikema (2016) used five performance measures % of years when minimum irrigation demand is met, average annual volume of water delivered to hydropower utilities, % of months where the storage lake is above the minimum acceptable level, average environmental flow requirement, and average flow requirement for the Tis Issat waterfall – to evaluate future scenarios of planned water infrastructure projects in Ethiopia.

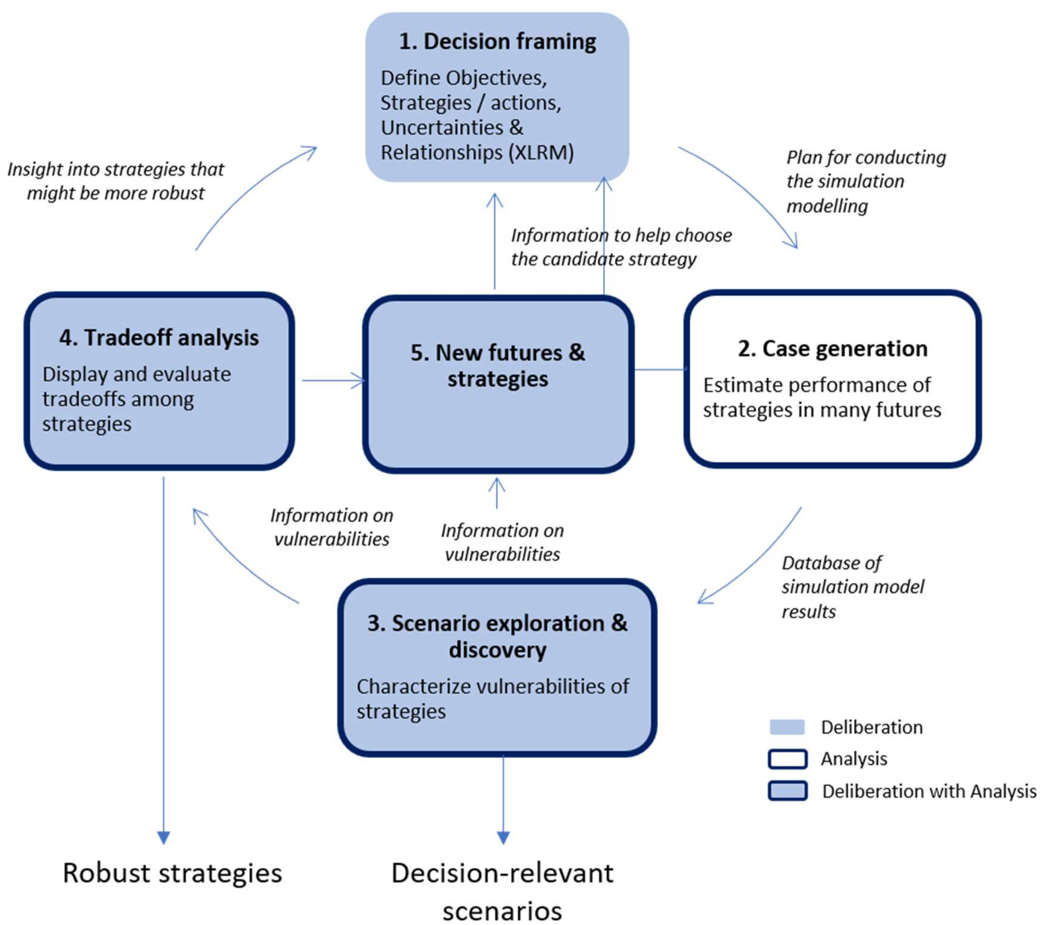

Figure 10: General steps in a Robust Decision-Making process. Steps that involve participatory deliberation are shaded. Source: Lempert (2019)

This participatory scoping step can also be revisited in the RDM process. In this first instance, its focus is to frame the study system by documenting and disaggregating different forms of uncertainty, identifying performance metrics and the initial management actions and relationships among the major influences on system dynamics. After an analysis of current vulnerabilities (Step 3), stakeholders may then structure a dialogue around additional adaptation options to mitigate against vulnerabilities identified in the strategies.

2. Case generation

The second step of RDM – case generation – uses a simulation model (or models) to explore one or more proposed strategies over the full range of the uncertain factors. Unlike conventional risk analysis where the relevant system is characterized by a system model that uses a single "best estimate" set of probability distributions for its input parameters to characterize uncertainty, the purpose here is to generate a very large database of model results across many plausible combinations of the uncertain input variables (Lempert et al., 2006). This database of potential outcomes is then interrogated systematically to identify weaknesses in the strategies (Step 3 - Scenario discovery and exploration).

Case generation can start with one or more actions or policies identified by an agency or decision-body or it could begin with a base-case strategy chosen through traditional expected utility or risk analysis as the most favoured path (Groves & Lempert 2007; Lempert & Groves 2010).

Several approaches can be used to generate cases across the proposed actions and uncertainties. A common method is to use computational experiments to sample systematically across the ranges of uncertain input parameters while holding a candidate policy or parameter constant (Bryant & Lempert 2010; Kwakkel et al., 2016b) or to sample across the policy options (Kwakkel 2017). These experiments can be implemented using different design strategies including Full Factorial sampling, Monte Carlo sampling, Latin Hypercube sampling, and other, more specialized sampling strategies (Morris 1991; Kwakkel 2017) depending on the type of ensemble of models to be explored. In situations where there are many uncertain input variables a Full Factorial sampling can yield a very large number of future scenarios that make this stage very computationally intensive. In these circumstances, Bryant & Lempert (2010) recommended use of Latin Hypercube sampling over other methods as it generates an efficient sample of a model's behaviour across the parameter space without requiring all combinations of the uncertain factors to be included. Latin Hypercube sampling has superior properties for variance reduction relative to other sampling designs.

For example, in applying RDM to analysis of hypothetical flood mitigation strategies in the Rhine Delta of the Netherlands, Kwakkel et al., (2016b) used Latin Hypercube sampling to generate 5000 experimental model outcomes for the base-case "do-nothing" scenario. These were sampled from across the ranges in uncertainty of input variables related to climate change scenarios, potential changes in land use, the uncertain relationship between water levels and the propensity of dykes to fail, the damage functions used in the model and the effects of policy actions. In a similar application – strategies for flood risk mitigation for Ho Chi Minh City, Vietnam - Lempert et al., (2013a) used a combination of Latin Hypercube sampling across five uncertain parameters (rainfall intensity, size and distribution of the human population, poverty rate and population vulnerability) and a Full Factorial design to sample across potential changes in the levels of the Saigon River. They generated 1000 experimental scenarios that formed the basis of their analysis. Molina-Perez (2016) also used a mixed experimental design, comprising both Full Factorial and Latin Hypercube Sampling across uncertainties in 12 climate scenarios and 300 prospective technological scenarios to generate 3,600 different future scenarios to evaluate options for investment and international cooperation in technology development to mitigate climate change.

3. Scenario exploration and discovery

Scenario exploration and discovery is the analytical core of RDM. In this step, statistical data mining algorithms are used to explore the database of model outcomes generated in the previous case generation phase to identify cases that are of interest to decision-makers and stakeholders. These might, for example, be situations where the policy has quite divergent performance in outcomes (Lempert et al., 2006). The aim of the analysis is to identify the combinations of the uncertain input variables that best predict this divergent performance across the range of generated cases and, in particular, the circumstances (i.e., regions of the multi-variate uncertainty space) where a policy performs poorly. The combinations of the uncertain input variables that result in high levels of failure are characterized as the "vulnerabilities" of the policy (Matrosov et al., 2013). They define the conditions under which the proposed policy (or policies) would fail to meet its goals (Lempert 2013). This stage of the analysis provides a quantitative means of identifying those combinations of uncertainties that require attention from decision-makers - where there is a need to hedge against failure - and those that can be more safely disregarded (Groves et al., 2014). Information on a policy's vulnerabilities can inform deliberation with stakeholders and decision makers about necessary modifications to the policy or the development of alternate actions that might reduce these vulnerabilities (return to Step 1). By iterating the case generation and scenario discovery stage (Steps 2 and 3) these alternate strategies can then be evaluated to determine how effective they are in reducing vulnerabilities.

Statistical classification algorithms, such as the Patient Rule Induction Method (PRIM) (Friedman & Fisher 1999), Classification and Regression Trees (CART) (Breiman et al., 1984) and Random Forests (Breiman 2001), are used to find the combinations of values for the uncertain input variables that result in unusually large or small values of the performance measures ('M') or which best distinguish scenarios that variously meet or fail to meet pre-specified performance thresholds (Friedman & Fisher 1999; Lempert 2019).

PRIM is often used in applications of scenario discovery because it is very interactive, allows multiple options for the choice of scenarios, and provides visualizations to assist users choose among the options. Diagnostics implemented in the PRIM routines help guide the selection of scenarios that are relevant to the policy decision (Bryant & Lempert 2010; Dalal et al., 2013). PRIM seeks to optimize three metrics: *density*, the fraction of cases in the data cluster that are of interest (e.g., where the policy fails); *coverage*, the fraction of all cases in the database in which the policy fails that are contained within the data cluster; and *interpretability*, the ease with which users can understand the information conveyed by the scenario. Interpretability is typically measured heuristically as data clusters that have only a small number of uncertain parameters that define it (Kwakkel & Jaxa-Rozen 2016).

A critical step in this stage is defining the (often binary) criteria used to select and classify scenarios of relevance to the decision. The thresholds for measuring success or failure against a performance goal may be specified by the decision-making group or may emerge as cluster sets in analysis of the model outcomes. For example, in evaluating the likely effects of imposing a standard for adoption of renewable energies, Bryant & Lempert (2010) used PRIM to identify circumstances in which the policy that might achieve the renewable energy objectives would result in unacceptably high costs, defined in their application as the cases that exceeded the 90th percentile cost from the database of results. The classification could, alternatively, have been based on a budget threshold (e.g., cases that are within or exceed allocated budgets for the policy) or stakeholder tolerances for public expenditure (i.e., willingness to pay).

In a lake eutrophication example, PRIM was used to identify modelled cases where the policy settings resulted in unacceptably high levels of pollution and the combinations of the uncertain factors that contributed to these cases (Lempert & Collins 2007; Kwakkel 2017). Shortridge & Guikema (2016) defined minimum performance thresholds for each of five metrics they used to evaluate water infrastructure policies under climate change. For measures of the reliability of water delivery for irrigation and hydropower the thresholds were based on design specifications of the plant. For environmental flows and lake levels, the thresholds were set from historic climate conditions and operations.

4. Trade-off analysis

Vulnerabilities of the policies identified in the scenario exploration step provide the basis for evaluating alternative policies or potential modifications to a strategy to reduce its vulnerability. This is sometimes referred to as 'trade-off analysis' or the 'risk management stage' of RDM (Groves et al., 2014)). The trade-off analysis is also designed to be participatory. Visualizations from the analyses undertaken during the case generation and scenario exploration stages enable the parties to the decision to compare the performance of alternative strategies over a wide range of futures.

A variety of different metrics of 'robustness' have been proposed to aid this deliberation (McPhail et al., 2018). They include measures derived from classic decision analytics (e.g., Maximin, Minimax, Hurwicz optimism-pessimism rule, minimax regret, and Laplace's principle of insufficient reason, Table 11) and more contemporary metrics of the density function of performance (e.g., variance and skew) that have been designed to measure the range of performances across the scenarios (Kwakkel et al., 2016a). Choice of policy alternatives can be sensitive to the robustness metric that is used to compare the strategies (Giuliani & Castelletti 2016), since the different measures embody differences in risk tolerance and different meanings of robustness (McPhail et al., 2018). For this reason, Kwakkel et al., (2016a) recommended using a combination of metrics to evaluate the trade-offs between the average performance of a policy across many future states of the world and variation in its performance around the average condition.

McPhail et al., (2018) developed a classification of robustness metrics to provide guidance on their selection and use. The classification is based on (1) whether the metrics measure the absolute performance of a policy or the performance of the policy relative to others or some benchmark standard, and (2) whether the metric provides an actual indication of system performance or is compared to a pre-specified performance threshold. Absolute performance measures may be dictated by regulatory standards or may be specific outcomes that stakeholders want the policy to achieve, such as return on investment, acceptable levels of protection, or other defined policy objectives. Relative performance measures are used where the large uncertainties of the input variables ('X') generate a wide range of plausible futures and there are no well-defined desired outcomes (Lempert 2019).

5. New futures and strategies

Lempert (2019) introduced a fifth stage into the RDM process in which analysts and decisionmakers use the outcomes from the scenario exploration and/or trade-off analysis stages to identify new management strategies that may provide better trade-offs than those initially considered in the modelling. These new alternatives generally incorporate additional policy levers, which can be considered via scenario exploration and trade-off analysis. This step in RDM also provides a natural point of connection with the Dynamic Adaptive Policy Pathways (DAPP) methods (Section 6.2.4) in the consideration of adaptive policies that may incorporate short-term actions and associated contingencies. Different methods can be applied at this stage. Recent applications have used multi-objective robust optimization tools (see Section 6.2.3) to evaluate trade-offs among different strategy objectives. RDM uses both absolute and relative performance measures to compare strategies in the vulnerability and trade-off analyses. Absolute performance measures are useful when decisionmakers are focused on one or more outcomes, such as profit, energy produced, or lives saved. Relative performance measures are often useful when uncertainties create a wide range of outcomes. so decisionmakers seek strategies that perform well compared to alternatives over a wide range of futures. RDM often uses 'regret' to represent relative performance (Table 11).

Hedging against surprises and thresholds

The new futures and strategies stage also provides an opportunity to evaluate the robustness of policies to outcomes that may not be anticipated. For example, Lempert et al., (2003) challenged a group of stakeholders to imagine surprises that represented distinct breaks with current trends or expectations. These included unforeseen technological changes, an unanticipated global pandemic and a shift in the values held by future generations. The imagined surprises were added to the scenario generator and the policy options stress-tested against them.

Lempert & Collins (2007) have also demonstrated the utility of RDM for evaluating policy options in situations where there is potential for nonlinear or threshold responses in natural systems where the conditions necessary to trigger a shift in state into an alternative undesirable condition and the system response are uncertain. By helping to characterize poorly defined uncertainties RDM helps identify strategies where threshold changes are within the realms of possibility and those where they are unlikely. Importantly, the concept of robustness is intended to highlight strategies that keep options open over a wide range of plausible futures. This iterative process, therefore, provides a template for designing and testing robust strategies and characterizing the remaining unquantifiable uncertainties to which they may be vulnerable.

Table 11: Summary of commonly used robustness metrics. Based on (McPhail et al., 2018).

Metric	Description	Calculated on absolute or relative performance values	Relative level of risk aversion	Reference
Maximin	Decision alternatives are ranked on the basis of their worst-case outcomes – the optimal decision is one with the least worst outcome.	Absolute	High	Wald (1950)
Maximax	Decision alternatives are ranked on the basis of their best outcomes and the best of these is selected.	Absolute	Low	Wald (1950)
Hurwicz optimism pessimism rule	Interpolates between selecting the strategy with the best case and the best worst case	Absolute	Moderately low	Hurwiczy (1951)
Laplace's principle of insufficient reason	Assumes equal weighting over all the futures and then selects the strategy that maximizes expected utility	Absolute	Moderate	Simon et al., (1951)
Minimax regret	Selects the strategy which deviates least from the best that could be chosen with perfect information (i.e., least regret)	Relative	High	Savage (1951), Giuliani & Castelletti (2016)
90^{th} percentile minimax regret	Selects the strategies that deviate least from the least regret situation	Relative	Moderately high	Savage (1951)
Mean-variance	Selects the strategy with the highest signal to noise ratio	Absolute	Moderate	Hamarat et al., (2014)
Undesirable deviations	Selects the strategy that minimizes undesirable deviations away from some threshold value	Relative	Moderately high	Kwakkel et al., (2016a)
Percentile-based skewness		Absolute	Moderate	Voudouris et al., (2014), Kwakkel et al., (2016a)
Percentile-based kurtosis		Absolute	Moderate	Voudouris et al., (2014), Kwakkel et al., (2016a)
Starr's domain criterion	Selects the strategy that has highest utility in the most futures.	Absolute	Moderate	Starr (1963), Schneller & Sphicas (1983)

6.2.3 Many-Objective Robust Decision Making (MORDM)

Many-Objective Robust Decision Making uses a similar iterative process to RDM to evaluate management alternatives under uncertainty but extends the steps of generating and evaluating alternative strategies (the 'case generation' step in Figure 10) (Kasprzyk et al. 2013). In RDM this step begins with one or more policy alternatives that have been specified by the decision-makers. These provide the basis for subsequent exploration of the variety of possible futures by systematic variation of uncertain model elements. The initial focus on one or relatively few scenarios is constrained by the ability of decision-makers to simultaneously scrutinize a variety of policy options intensely and by the computationally intensive scenario exploration stage. In most complex planning problems, however, decision makers must consider the likely outcomes of the strategy across a range of objectives for the resource. These typically include objectives that can be at odds with each other, have interdependencies or interact in non-linear ways (Kasprzyk et al., 2013).

MORDM uses a class of problem-solving algorithms (Many-Objective Evolutionary Algorithms - MOEAs) that are designed to search for candidate reference scenarios that simultaneously optimize multiple, conflicting objectives. The aim is to find a set of good compromises (or "trade-offs") among the objectives rather than a single optimal solution (Coello et al., 2007). The advantage of this approach is that the set of initial policy alternatives used in the scenario exploration stage are not constrained by the preconceptions of the decision-makers but represent a careful search of the complex, multi-dimensional design space (Reed et al., 2013; Kwakkel & Haasnoot 2019). The goal is to develop more diverse sets of decision variables that have better objective function performance under extreme conditions (Watson & Kasprzyk 2017). Dynamic visualizations produced during the analyses present decision makers with representations of the trade-offs that are made with different options to support selection of promising scenarios as reference cases (Figure 11, (Reed et al., 2013).

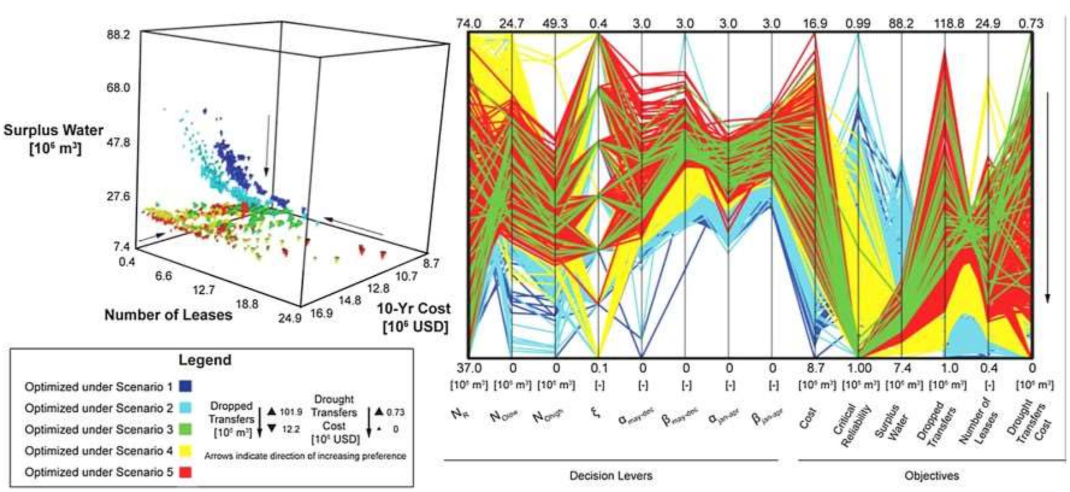

Figure 11: MODRM uses visualizations to facilitate comparison among different scenarios by representing the trade-offs among objectives and decision levers. Left: A 3D glyph plot that uses point shape, size, colour and transparency to communicate up to seven dimensions. Right: Parallel coordinate plots display trade-offs and decision levels optimised under each of five scenarios indicated by the different coloured lines. The example evaluates scenarios for water allocation from multiple market-based supply instruments (Watson & Kasprzyk 2017).

A range of different MOEAs have been developed for use in many-objective optimization (summarised by Reed et al., (2013)). Evolutionary algorithms are an artificial intelligence method that mimics evolutionary processes by iteratively sorting through a large ‘population’ of potential solutions to a problem to select a heuristic that provides a sufficiently good solution under uncertainty. Each new generation is produced by stochastically removing less desired solutions and introducing small random changes. By successively removing less desirable solutions the ‘population’ of solutions converges on the ‘fittest’ state, as represented by the fitness heuristic chosen for the analysis. MOEAs are designed to identify approximately Pareto optimal management decisions. Solutions are Pareto optimal if their performance is not exceeded in any objective by another feasible solution (Figure 12). For example, in a trade-off between the cost and reliability of a system, the optimal trade-off is the population set of least possible costs at every level of reliability (Watson & Kasprzyk 2017).

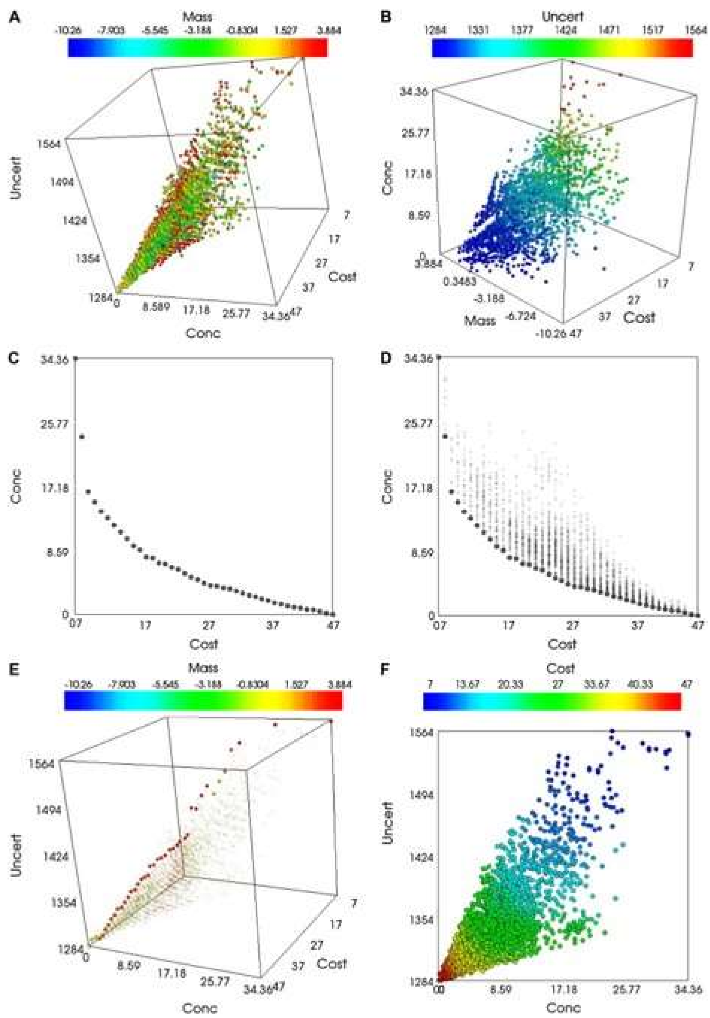

Figure 12: Visualisations of a four-objective Pareto optimal solution. This example, of options for the design of monitoring for groundwater contaminants, has four objectives to optimize: 'Cost' of sampling, error in estimated contaminant concentrations ('Conc'), uncertainty associated with a kriged map of the contaminant plume ('Uncert') and the error in the total mass of contaminants estimated from sampling ('Mass') (Kollat & Reed 2007).

Selection of a small number of candidate scenarios for the scenario exploration and trade-off analysis stages of MORDM from a large set is guided by criteria such as their internal consistency, diversity of outcome indicators, extremeness, and policy relevance (Trutnevyte et al., 2016). Eker & Kwakkel (2018) introduced a more systematic procedure for scenario selection in the search phase that emphasises policy relevance and diversity as the key criteria. 'Diversity' refers to how the scenarios differ from each other in terms of the outcome indicator values, and 'policy relevance' refers to whether the scenarios are undesirable for decision makers. They argued that this approach increased the variety of robustness trade-offs between the objectives, provided more decision options and leads to more robust solutions overall.

6.2.4 Dynamic Adaptive Policy Pathways

Dynamic Adaptive Policy Pathways (DAPP) are a combination of planning approaches that have been developed to support sequential decision-making under uncertainty. Like other methods discussed in this section, the focus of DAPPs is not on identifying the most likely single course of action to reach a particular outcome but rather to identify policies that are robust to failure. DAPP is based on the concept of adaptive planning in which an initial policy is designed and is reassessed repeatedly over time in response to changes in the system and new information (Haasnoot et al., 2013). An adaptive policy needs to be flexible to respond to how uncertain events unfold and should make explicit provision at the outset for monitoring performance so that decisions are not made repeatedly on an *ad hoc* basis.

Walker et al., (2001) described a general process for developing and implementing adaptive policies, which they termed "Dynamic Adaptive Planning" (DAP) (Walker et al., 2013a). In their approach, the problem is framed through initial deliberation and analysis of the system under consideration and long-term objectives and initial policy options are identified. A basic plan for achieving the objectives is developed with attention given to four types of actions that are designed to make the policy robust across a range of plausible futures:

- 'Mitigating actions' to reduce the likely adverse effects ('vulnerabilities') of the policy
- 'Hedging actions' to spread or reduce the uncertain effects of the policy
- 'Seizing actions' to take advantage of likely opportunities to improve the policy and
- 'Shaping actions' which are actions taken to reduce the likelihood that an external condition or event could make the policy fail or enhance its success.

Implementation is guided by a framework for monitoring the performance of the policy over time. This should specify the information that needs to be collected to determine if the policy is being successful ('signposts') and critical points ('triggers') at which additional actions may be needed to modify it. These may include adjustments to the policy in response to specific triggers ('corrective actions'), actions taken to clarify the policy, preserve its benefits, or meet external challenges ('defensive actions'), actions that may take advantage of opportunities to improve the policy ('capitalizing actions'), and reassessment of the plan when the analysis and assumptions that underpin it are no longer valid (Walker et al., 2013a).

Although DAP has many similarities to the concept of adaptive resource management which uses monitoring to reduce uncertainty by learning from and adapting short-term management actions to improve performance (Holling 2004), the key sources of uncertainty being addressed in each approach are different. DAP is intended to address sources of uncertainty that are beyond the control of policymakers whereas adaptive management is focussed more on uncertainties that arise out of the complexity of the system being managed (Kwakkel et al., 2010b). Monitoring systems for DAP, therefore, incorporate what has been termed ‘planned adaptation’ by pre-specifying the responses that will be implemented when specific critical points (‘triggers’) are reached.

DAPP uses the deductive approach of Dynamic Adaptive Pathways by thinking backwards about how a plan might fail and by:

- designing short-term actions to guard against such failures,
- anticipating future actions that might be initiated at a later date to ensure that a plan is still on track to meet its objectives and
- implementing a monitoring system to identify when these actions should be triggered (Haasnoot et al., 2013).

It combines this approach with the use of adaptation trees (a form of decision tree or more aptly a ‘fault tree’) to visualise the sequence of actions that is possible for a policy through time and the range of alternate policies that could be implemented to achieve the long-term objectives. Each branch of the tree (a ‘policy pathway’) consists of a sequence of actions through time, decision points for implementing different actions and critical points when the policy or action is no longer tenable. Adaptation Tipping Points (ATPs) are these critical points (sensu Kwadijk et al., (2010)) at which the current strategy is no longer on course to meet the long-term objectives and an alternative course of action is required. Multiple alternative policy pathways are assembled together in the form of a pathway map that describes the different policy routes (scenarios) that can be taken to achieve the long-term objectives, their vulnerabilities, ‘lock-ins’ (critical points at which changes in the system mean that earlier decisions may severely limit the number of future actions that could be taken), ATPs and the relationships between them.

A characteristic of DAPP is that these maps are presented in the style of metro maps (Figure 13) so that different stakeholders and decisionmakers can use them to identify their preferred pathway to the objectives depending on their values and beliefs and to determine where there is agreement and divergence in opinion about the paths to be taken. Pathways can focus on adapting to changing conditions (‘adaptation pathways’), enabling socio-economic developments (‘development pathways’) or transitioning to a desired future state (‘transition pathways’).

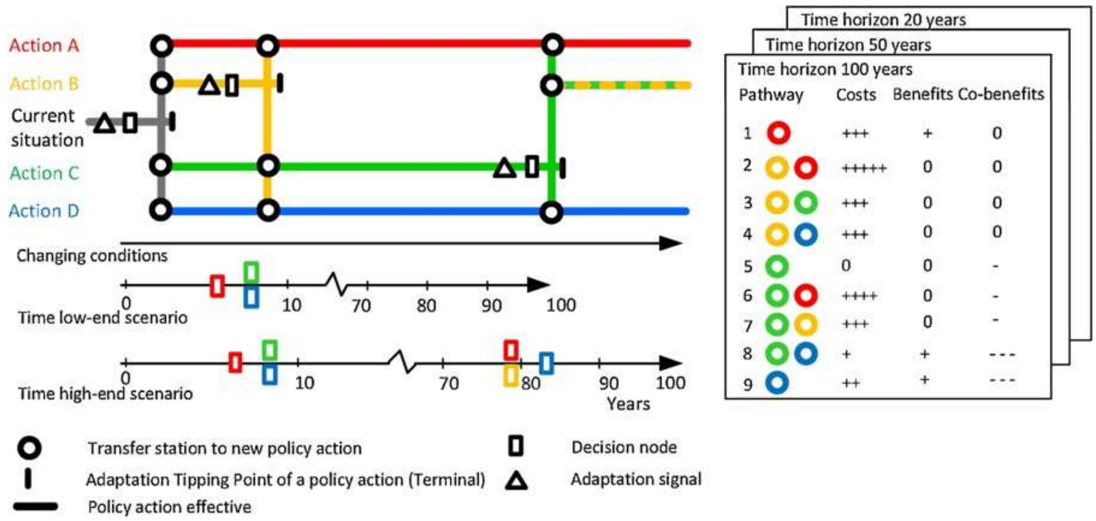

Figure 13: **An Adaptation Pathways Map.** Left: The map displays 9 possible pathways with adaptation signals and decision modes. Right: A scorecard summarizes the costs and benefits of each pathway over different time horizons (Source: (Haasnoot et al., 2013)).

There are seven steps in the DAPP process (Haasnoot et al., 2019) (Figure 14):

1. Problem framing

As with other deliberative approaches to decision-making about complex risks, DAPP is designed to be a participatory process in which the problem is jointly framed by decision-makers, analysts and stakeholders. The initial step, therefore, involves describing the system under consideration, the long-term objectives and the major uncertainties that affect decision-making, including disagreements among the parties. These uncertainties can be changes in external drivers, in the system itself and in how different outcomes are valued by stakeholders. The identified uncertainties are used to generate a range of plausible futures ('scenarios') that describe a desired future end-point or developments through time.

A key focus of the deliberation at this stage is also to identify what success looks like over the long-term and how it can be measured. Indicators and targets for the desired outcomes that can be used to evaluate the performance of management actions and pathways are specified.

2. Assess vulnerabilities and opportunities

The second step resembles the scenario discovery stage of RDM described in Section 6.2.2, where the purpose is to identify influences (vulnerabilities or opportunities) on the policies that would determine their failure or success. In DAPP, this starts with evaluating the current situation against the ensemble of plausible futures (given the identified uncertainties) and to identify the conditions under which the system begins to perform unacceptably (i.e., where it is approaching an ATP).

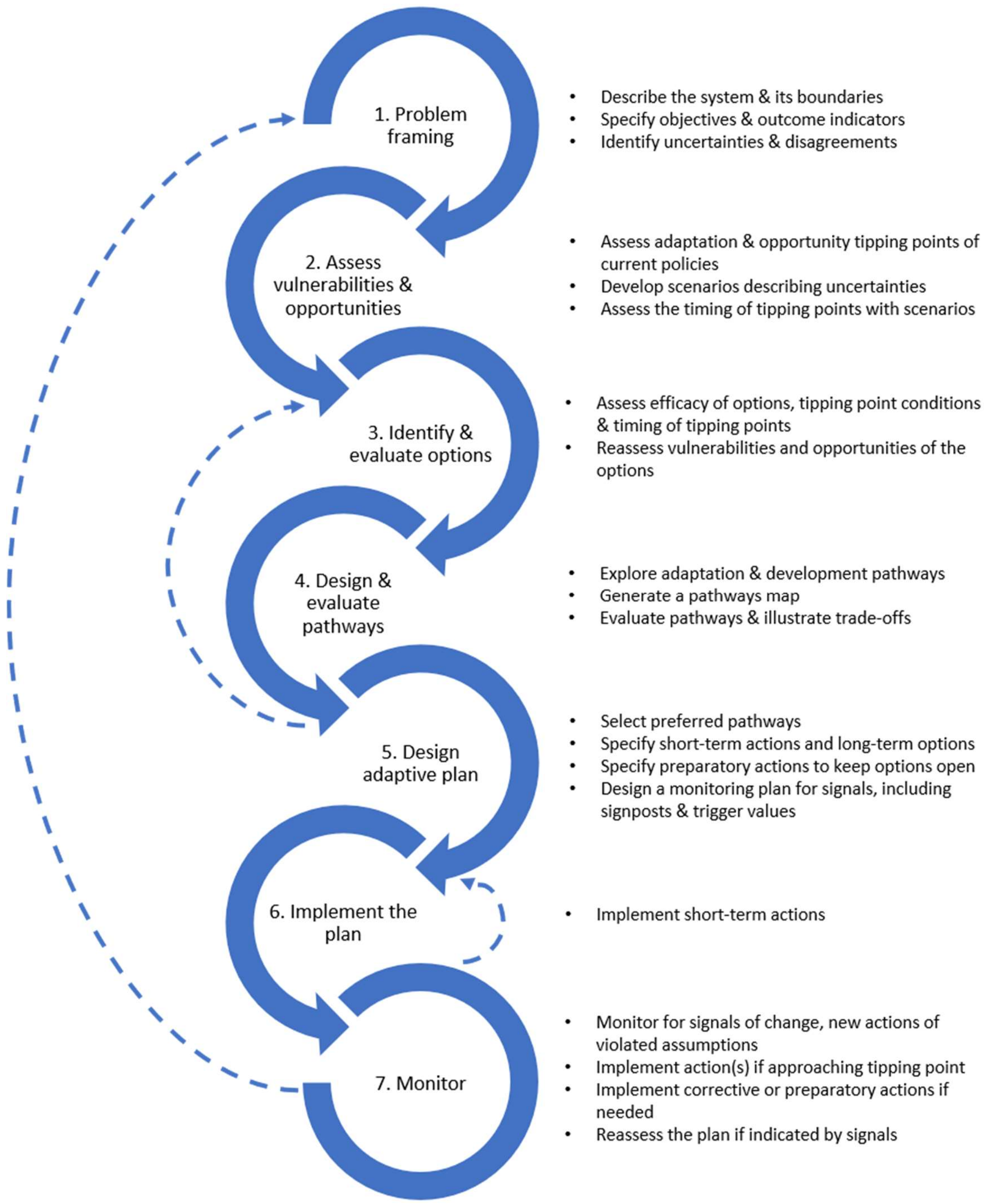

Figure 14: Steps involved in Dynamic Adaptive Pathway Planning.

A range of approaches can be used to identify the ATPs. The approaches chosen depend on the indicators specified and the knowledge and data available for the system. Haasnoot et al., (2019) describe two general approaches. 'Bottom-up' assessments of vulnerability establish unacceptable outcome thresholds and then use scenarios to assess when these tipping points are likely to occur. This can include the use of model-based exploration of the uncertainty space using statistical classification algorithms such as PRIM (Section 6.2.2) to establish failure conditions across an ensemble of scenarios, or they can be specified by expert and stakeholder elicitation. With this approach the aim is to determine how much change in the system can be accommodated within the current strategy before it is unable to meet the objectives.

‘Top-down’ approaches use traditional predictive scenario analysis to determine the timing of ATPs, where the starting point is one or more reference policy scenarios (Kwadijk et al., 2010; Haasnoot et al., 2019).

3. Identify and evaluate options

The third step identifies a rich range of actions that can be taken to address the vulnerabilities identified in steps 1 and 2 or to seize opportunities for improving and strengthening the policies (Haasnoot et al., 2013). They should include a range of the four types of action described as building blocks for DAP, namely: ‘mitigating actions’, ‘hedging actions’, ‘seizing actions’ and ‘shaping actions’. These are then used to assemble the adaptation pathways.

4. Design and evaluate pathways

In this step, a pathway map is constructed (Figure 13) that summarizes the logical policy pathways and actions (and their dependencies) through which the objectives can be reached under changing conditions (Haasnoot et al., 2019). The map identifies ATPs associated with each policy option and specifies the actions that can be taken before an ATP is reached. Decision nodes in the pathways represent when an alternate action needs to be enacted before an ATP is reached to ensure that the pathway remains on track to meet the objectives (Haasnoot et al., 2018).

Adaptation pathways can be assembled from their components and evaluated in different ways. For example, the maps can be developed manually and explored qualitatively using stakeholder focus groups. Alternatively, the assembly and evaluation can be done using a variety of computational approaches (Haasnoot et al., 2013). For example, Lawrence & Haasnoot (2017) used a realistic simulation game to socialise the development of pathway options with stakeholders. The game required teams of stakeholders to develop a water management plan for a stylised river. Participants set a vision, chose policy actions and negotiated these actions with other teams. Simulations of future events allowed participants to get feed-back on their decision choices and to adapt them as needed. This was supported by an associated application (http://pathways.deltares.nl) that assisted with generating alternative pathways and stress-testing them against different future scenarios (Haasnoot et al., 2012; Lawrence et al., 2019).

Kwakkel et al., (2015) used an EMA approach to generate an ensemble of possible futures and candidate pathways for consideration. They then applied a MOEA for robust optimization to assess the robustness of the pathways across multiple, independent, long-term objectives and to identify the set of most promising pathways and adaptation options for consideration.

Trade-offs are identified among the pathways through evaluation of the cumulative costs and benefits of each course of action, potential lock-ins and opportunities, the likely severity of impacts, the level of uncertainty or urgency of actions. Again, the trade-off analysis can be done simply, using a qualitative scorecard approach (e.g., Figure 13; Haasnoot et al., (2013)) or can be supported by more formalized cost-benefit or multi-criteria analyses (Haasnoot et al., 2019).

5. Design adaptive plan

The trade-off analysis is intended to reduce the number of promising and preferred pathways to a manageable set for which initial actions and longer-term contingencies can be identified. Once a basic pathway plan has been constructed that specifies the short-term actions to be taken and longer-term options, it is further enhanced by including adaptive elements and a monitoring system is designed to track its performance. The latter follows the DAP framework and requires identification of signposts and triggers, and their associated *defensive actions*, *corrective actions*, *capitalizing actions* or *reassessment* processes (Kwakkel et al., 2010b).

Haasnoot et al., (2018) described the steps involved in designing a signal monitoring system to support adaptive plans. The signposts or indicators that are tracked to monitor the plan are a critical element. Two general types of signposts are used:

- performance signposts that detect the performance of a system and indicate the extent to which objectives are still being achieved
- environment (context) signposts that monitor (external) changing conditions that jeopardize or provide opportunities for achieving these objectives.

Choice of useful signposts and triggers should be guided by their relevance to the concerns of users (‘salience’), their ability to be observed and measured, their technical credibility, and reliability in giving the right signal (i.e., low rate of false alarms). In a policy context, their perceived fairness to different stakeholders (‘legitimacy’) is also a consideration (Haasnoot et al., 2018).

Where data are available, statistical methods such as trend analysis and signal to noise ratio analyses can be used to resolve trigger points under possible futures. Alternatively, qualitative approaches may be used to elicit expert judgement on the expected rate of change of a signal and the relative timing of tipping points.

6. Implement the plan
7. Monitor

Once the plan is complete, the remaining steps are to implement the initial actions and to establish the monitoring system.

6.2.5 An overview of DMDU methods

Although a range of methods is now available to evaluate management alternatives under deep uncertainty, they have been applied mostly in the fields of climate change adaptation, water resource planning and infrastructure development. There is still limited experience in their application to the management of natural resources and other areas of complex policy choice. There is also limited expertise in their application world-wide.

All the approaches that fall within the DMDU umbrella are designed to examine the vulnerabilities of policy choices by evaluating the situations in which the policies would fail. They do so by combining methods for exploratory modelling, adaptive planning for uncertain future events and through graphical decision-support tools that allows stakeholders and decision-makers to work alongside analysts to evaluate policy options. The ways in which they do this and the planning choices that result from each approach can be quite different. For example, the strengths of DAPPs are that it is intuitively straight-forward, can be implemented using a range of facilitated qualitative or quantitative procedures and uses understandable visualizations of policy alternatives that encourage decision-makers to develop plans that explicitly allow for uncertainty and adaptation over time. Other methods, such as RDM and MORDM leverage the power of computing and statistical algorithms to sort through policy options to identify robust strategies and can be analytically much more challenging.

A key challenge in using EMA approaches is in selecting appropriate strategies for detailed examination ('case generation') and for searching the scenarios to provide useful insights on potential policy choices. While powerful, the analytic complexity and resource requirements of RDM and MODRM can also be an impediment to their more widespread adoption in the short-term (Bhave et al., 2016). For example, Groves et al., (2019) note that their application of RDM required two computer clusters many months to perform all the needed simulations. Other studies have used cloud computing and high-performance computation facilities to support the large ensemble analyses that RDM and MORDM studies require. A range of open-source software tools is now available that can greatly facilitate these analyses (Table 12).

Nevertheless, it remains difficult in many cases to implement a full RDM or MORDM analysis using realistic system models. For this reason, most applications use simplified integrated models to capture system behaviour and which may not, therefore, contain the realism required to explore all options. The cost of developing detailed system models can also put computationally intensive analyses like RDM and MODRM beyond the reach for many decision-makers. Approaches such as expert elicitation and participatory modelling could be used to convert simple conceptual or qualitative system models (Section 5.1) into a meta-model form that might greatly increase their uptake across a range of disciplines (Lempert 2019). An example of this type of approach is presented in the case study described in Section 5.2.1 where a combination of Qualitative Network Models and scenario-based expert elicitation was used to model the effects of changes in hydrology caused by coal seam gas exploitation on a range of response variables.

Decision making for complex system requires an iterative approach that facilitates learning across alternative framings of the problem, and learning about stakeholder preferences and trade-offs, in a collaborative process of discovering what is possible (Herman et al., 2015). Planning for uncertainty requires dynamic policy plans that can be adapted to changing circumstances to achieve long-term policy goals. These two elements – iterative learning and dynamic planning – are present to different degrees in the approaches we have reviewed (Kwakkel & Haasnoot 2019). RDM and MORDM incorporate iterative model-based processes for designing adaptive plans while DAPP provides a systematic structure for how such plans should be organised and monitored. There is, therefore, a case for combining elements of these approaches in risk management for complex problems (Kwakkel et al., 2016b).

Table 12: Available open-source software tools for implementing DMDU approaches.

Software	Analyses incorporated	Location	Reference
Scenario Discovery Toolkit	Scenario generation & discovery Diagnostics toolkit Visualization tools	https://cran.r-project.org/web/packages/sdtoolkit/index.html	Bryant & Lempert (2010)
Exploratory Modelling Workbench	Generation of policy options Scenario discovery Vulnerability analysis Robustness Optimization	https://github.com/quaquel/EMAworkbench	Kwakkel (2017)
Multi-Objective Robust Decision Making (MORDM).	Scenario generation Multi-Objective Evolutionary Algorithms	https://github.com/sibeleker/MORDM---Multi-scenario-search	Eker & Kwakkel (2018)
MOEA framework	Multi-Objective Evolutionary Algorithms Other general-purpose single and multi-objective optimization algorithms.	http://moeaframework.org/	Reed et al., (2013)
Borg Multi-Objective Evolutionary Algorithm (MOEA)	Multi-Objective Evolutionary Algorithms	http://borgmoea.org/	Hadka & Reed (2013)

Software	Analyses incorporated	Location	Reference
Project Platypus	Rhodium – tool for RDM Platypus – library of Multi-Objective Evolutionary Algorithms OpenMODRM – Multi-objective Robust Decision Making PRIM – Scenario discovery J3 – Platform for visualizing and analyzing multi-objective tradeoffs	https://github.com/Project-Platypus	Hadka et al., (2015)
Adaptation Pathways	Generates pathway maps for adaptive policy planning	https://publicwiki.deltares.nl/display/AP/Adaptation+Pathways	Haasnoot et al., (2019)

7 Conclusions

Although principles for the assessment and management of risks have been outlined in numerous guidance documents and technical standards, they are typically not prescriptive of the methods to be used. This recognises the variable contextual setting of risk decisions and the range of techniques that can be applied to them. An expectation is that assessments of risks should incorporate the best available data, technical knowledge and scientifically valid analytical methods to evaluate the prospects of undesirable outcomes. As we outlined in earlier sections of this report, decisions about how the risk problem should be framed and analysed have often fallen solely to technical specialists, but for complex and highly uncertain risk problems where the consequences may be extreme and/or felt by a broad range of interest groups this approach is increasingly untenable. Where the hazards have few precedents and there are limited existing data and knowledge, analytical approaches to assess risk are often necessarily bespoke, with heavy reliance on expert judgement to select the tools, apply them and interpret their outcomes. Extending the peer and expert community for these analyses to include stakeholders and decision makers provides a basis for building trust in the analysis and a collective understanding of the nature of the problem, the limits of knowledge and the rationale for the choice of method. More inclusive evaluation of decisions about risk can broaden perspectives on the problem and enhance the range of options considered for its management.

Our review has described a range of processes for more participatory framing and analysis of complex risk problems (Sections 3, 5, and 6.2) and methods to support more defensible analysis and evaluation of options. The latter included:

- approaches to improve the quality of expert judgements in conventional probabilistic risk assessment when there is a lack of observational data (Section 4),
- ways of integrating qualitative and quantitative assessment methods to analyse cumulative risks from multiple stressors (Section 5),
- approaches to evaluate contrasting management options in the context of risk and uncertainty (Section 6.1), and
- methods for identifying management options that are robust to failure across a range of potential future scenarios (6.2).

The methods included here are not intended to be comprehensive, but rather to represent a range of 'best-practice' tools and approaches that can be applied across a spectrum of risk problems of differing complexity and uncertainty.

The five levels of uncertainty described in Section 2.2.1 (and reproduced in Table 13) provide a useful framework to illustrate when each of these different approaches might be preferred. The levels correspond broadly with a categorization of risk problems used by Aven & Renn (2009b) that is based on the complexity of the problem, the extent of knowledge about the relationships between an event and its potential consequences (components of 'epistemological' and 'model uncertainty', Section 2.2), the level of consensus on the values to be protected and the relevance of outputs from the risk assessments for decision making (aspects of 'decision uncertainty').

Simple risk problems (Level 1 uncertainty) are those where the cause-effect relationship between an event and its consequences are well characterized and residual uncertainties are low. In these circumstances, risks can be modelled from existing data and uncertainties characterized by sensitivity analysis of model parameters. Conventional statistical analysis of existing data or well validated process models are useful tools for predicting the occurrence of the event and its consequences. Probabilities can be used to describe the likelihood of the uncertain alternatives. Because the negative consequences and the circumstances in which they would occur are obvious, there is generally consensus about the values that are to be protected by risk management strategies.

Risk problems with Level 2 uncertainties can also be described adequately using conventional probabilistic risk analysis or scenario-based optimization methods (e.g., decision analytic approaches, Section 6.1). These are problems where it is possible to specify a conceptual model for the relationship between the event(s) and a limited set of possible consequences that are identified based on existing knowledge or data. Statistical or process models can be developed to estimate the likelihood of each outcome. These may be parameterized using available data or through expert elicitation of subjective probabilities or their distributions (Section 4). In this case, the model is used to estimate the probability distributions of the outcomes of interest for these futures. A preferred policy can be chosen based on the outcomes and the associated probabilities of the futures (i.e., based on "expected outcomes" and levels of acceptable risk). Because the conceptual model of the system and the associated analytical model are developed, at least in part, from existing knowledge, there is a need to test its validity against a peer group of experts and stakeholders. This 'epistemological discourse' (*sensu* Renn (2008)) is necessary to review the scope and quality (validity, reliability and relevance) of the information used to develop the analytical framework and is aimed at finding the best approaches for estimating and characterising the risk. Structured protocols for consensus building, eliciting judgements from groups and scenario construction can greatly aid agreement on the conceptual model and the analytical approach (see Sections 4.2 and 4.5). Where it is possible to identify agreed objectives and performance metrics for the management outcomes, scenario-based optimization may be the preferred approach to rank and compare possible future states of the resource (Section 6.1).

Complex risk problems occur when it is difficult to identify and quantify the functional relationships between a range of potential causative agents and their effects (Level 3 uncertainty). This increase in complexity can be accompanied by disagreement among experts about the appropriate way to describe of the system and among stakeholders about the range of values that could potentially be affected. Analysis of complex risks requires 'reflective discourse' (Renn 2008). This involves stakeholders and experts in framing the problem and in evaluating the appropriate level of protection that they would be willing to accept to avoid potentially catastrophic consequences. In these circumstances, tools that can help build conceptual models of the problem under consideration and to prioritise key areas of uncertainty for analysis and key values for protection can be useful. These may include qualitative systems models (Sections 5.1), tiered or semi-quantitative risk analysis (Section 5.2) and scenario-based evaluation of likelihoods and outcomes (Sections 4.3.2, 5.2.1, and 6.1). Assessment of complex risk problems will often require a range of analytical tools to adequately describe different components of the system. It should also be set in an adaptive analytical framework so that outputs can be reviewed by stakeholders and other experts to determine how well they capture system behaviour and the analysis modified to explore a range of plausible scenarios.

In situations of deep uncertainty (Level 4 uncertainty), where a large range of plausible future states for the system can be specified, but it is not possible to estimate or rank their likelihood (Walker et al., 2013b), methods that are aimed at reducing vulnerability to threats are needed. DMDU methods (Section 6.2) that facilitate direct analysis of policy performance under uncertainty are most useful in these situations. Their purpose is to identify options that are adaptable, robust to uncertainty and can withstand surprises (Renn 2008; Marchau et al., 2019a). Because risk problems with Level 4 uncertainty encompass a very wide range of possible future consequences, they are likely to involve a correspondingly large range of interest groups and values that could be affected by policy choices. For this reason, they require the most inclusive form of participation from stakeholders (‘participatory discourse’, *sensu* (Renn 2008)). Aven & Renn (2009b) describe this form of analysis as an innately political process of evaluation in which competing arguments, beliefs and values about the risk problem are openly discussed and management options are evaluated according to how well they are able to achieve equitable outcomes that are robust to uncertainty. The purpose of this form of discourse is to seek mutual understanding of conflicting views and values with a view to eventually reconciling them. As we described in Section 6.2, many DMDU methods are designed as participatory processes that involve iterative discourse and analysis of policy options. Visualization tools help decision-makers and stakeholders evaluate policy performance across a range of future scenarios.

The International Risk Governance Council identified six requirements for building trust in the assessment and management of complex risks (Renn 2008). They include meaningful attempts to:

- involve decision-makers, technical experts and potentially affected groups (if appropriate) in the assessment;
- empower all parties to participate actively and constructively in the discourse;
- co-design the framing of the risk problem with these different groups;
- generate a common understanding of the magnitude of the risk based on expertise of all participants as well as the potential risk management options and to include a range of management options that represent the different interests and values of all parties involved;
- conduct a forum for decision-making that provides equal and fair opportunities for all parties to voice their opinion and to express their preferences; and
- provide a clear connection between the decision-making and policy implementation.

This will require more frequent involvement of skilled facilitators and structured participatory processes in risk assessments to enable framing of the risk problem, inputs to the choice and application of analytical tools and to facilitate evaluation of management options. Skilled facilitation and communication is especially important for more complex assessment tools, such as the DMDU approaches, which are technically challenging to understand and relatively new in their application to risk problems.

Table 13: The spectrum of uncertainty and its relationship to risk analysis and stakeholder participation in risk governance. Adapted from (Aven & Renn 2009b; Kwakkel et al., 2010a; Walker et al., 2013b).

		Levels of uncertainty					
		Level 1	**Level 2**	**Level 3**	**Level 4a**	**Level 4b**	
Description		A single future state	Able to identify multiple alternatives and estimate probabilities (subjective or objective)	Able to identify multiple alternatives and rank order them in terms of perceived likelihood.	Able to identify multiple alternative future states without being able to rank order them in terms of how likely or plausible they are judged to be.	Unable to identify multiple alternatives while admitting the possibility of being surprised	
Nature of the risk	**Complete certainty**	Relatively well understood from precedents.	Knowledge and data adequate to describe the system probabilistically. Some scientific disagreement about how the cause-effect relationships should be described.	Complex cause-effect relationships between events and consequences. A range of potentially affected values that must be considered.	Unresolved complexity and uncertainty in cause-effect relationships. A range of different viewpoints about the relevance of technical information and predictions about risks and their management.	Surprising extreme events that cannot be predicted from our present knowledge, understanding or beliefs.	**Total ignorance**
System model		A single (deterministic) system model	A single (stochastic) system model	A few alternative system models	Many alternative system models	Unknown system model	

	Levels of uncertainty				
	Level 1	**Level 2**	**Level 3**	**Level 4a**	**Level 4b**
Types of analysis	Statistical models Deterministic risk models Sensitivity analysis of model parameters. Cost-benefit analyses	Quantitative risk analysis Bayesian (subjective) probabilities Some stochastic model parameters Tools for seeking consensus among experts (e.g., Delphi, IDEA protocols) Scenario-based optimisation	Quantitative and semi-quantitative risk analysis Scenario-based optimisation	DMDU methods Conflict resolution methods for reaching consensus on risk evaluation and management options	DMDU methods with explicit consideration of surprises
Level of stakeholder participation	Little dispute about values to be protected. Instrumental discourse aimed at finding the most cost-effective measures to make the risk acceptable or at least tolerable.	Epistemological uncertainty addressed through discourse aimed at obtaining the best estimates of the occurrence of events and their consequences	Reflective discourse. Stakeholders are involved in framing the problem and in evaluating the appropriate level of protection that they would be willing to accept to avoid potentially catastrophic consequences. Deliberation reflects on the relative consequences of over- and under-protection.	Participative discourse openly discusses competing arguments, beliefs and values about the risk problem and evaluates how well different management options are able to achieve equitable outcomes that are robust to uncertainty.	Participative discourse openly discusses strategies for reducing vulnerability to unforeseen hazards

8 Acknowledgements

The authors would like to thank Jeremy McKenzie (NIWA) for his constructive review of the draft report. This work was funded by Sustainable Seas National Science Challenge Project 5.1.3 Risk & uncertainty.

9 Glossary of abbreviations and terms

Adaptation Tipping Point	Used in Dynamic Adaptive Policy Pathways to describe a critical point in time when a policy is no longer able to achieve its objectives (i.e., the conditions under which the policy starts to perform unacceptably (sometimes called a 'threshold')) (Kwadijk et al., 2010; Marchau et al., 2019a).
Assumption-Based Planning	A tool for identifying as many of the assumptions underlying policy plans as possible and bringing those assumptions explicitly into the planning process to examine and mitigate their vulnerabilities (Dewar et al., 1993).
Consequence	The outcome of an event that may result from a hazard (often used interchangeably with 'impact'). Consequences may be expressed quantitatively (e.g., monetary value), by category (e.g., high, medium, low) or descriptively in terms of human, environmental, and political/social impacts (Office of the Prime Minister's Chief Science Advisor 2016a).
Decision analysis	A discipline that incorporates aspects of psychology, operations research and economics which is concerned with formalized treatment of complex decisions to achieve optimal outcomes under uncertainty (Keeney 1982).
Decision trees	A sequential graphical representation of decisions and uncertainties that represent all paths the decision maker might follow. Four basic elements are used to construct a decision tree: decision nodes, alternative branches, probability nodes, and outcome branches (sometimes referred to as 'Scenario trees' or 'Event trees').
Deep uncertainty	(a) A situation in which risk analysts do not know or the parties to a decision cannot agree upon (i) the appropriate models to describe interactions among a system's variables, (ii) the probability distributions to represent uncertainty about key parameters in the models, and/or (iii) how to value the desirability of alternative outcomes' (Lempert et al., 2003). (b) A source or sources of uncertainty, the impact of which on the assessment the assessor(s) is not able to quantify (EFSA Scientific Committee et al., 2018a).
Deliberative process	A 'problem solving' discussion in which a group of people with different backgrounds, interests and values receives and exchanges information about an issue, to critically weigh the reasons for and against some proposition, and to come to an agreement that will inform decision making about a course of action (Fearon 1998; Abelson et al., 2003).
Dynamic Adaptive Planning	An approach for designing a plan that explicitly includes provisions for adaptation as conditions change and knowledge is gained (Walker et al., 2019).
Dynamic Adaptive Policy Pathways	An exploratory model-based planning tool that helps design strategies that are adaptive and robust over different scenarios of the future (Haasnoot et al., 2013).
Epistemic uncertainty	Uncertainty due to limitations in knowledge

Epistemological discourse	Deliberation among experts that is aimed at identifying the best available knowledge and information for characterising the risks under consideration (Renn 2008).
Exploratory Modelling and Analysis	The use of computational experiments to analyse complex and uncertain systems. In each experiment, a large number of runs are made over the input space in order to generate an ensemble of runs (Bankes et al., 2013; Marchau et al., 2019a).
Full Factorial design	A sample or experiment in which all possible combinations of the factors and their levels are included.
Hazard	Any source of potential harm, including loss of life or injury, property damage, social and economic disruption or environmental degradation (Office of the Prime Minister's Chief Science Advisor 2016)
Instrumental discourse	Deliberation that is aimed at finding the most cost-effective measures to make the risk acceptable or at least tolerable (Renn 2008)
Latin Hypercube sampling	A statistical method for generating a near-random sample from a multidimensional distribution.
Likelihood	(Risk assessment) - the chance or probability of a specific event occurring.
Linguistic Uncertainty	Imprecise use of language. Arises because language is not exact, including vagueness, context dependence, ambiguity, indeterminacy and underspecificity (Regan et al., 2002).
Many-Objective Optimization	Analyses that consider trade-offs among four or more objectives simultaneously. Also referred to as multi-criteria optimization (Coello et al., 2007; Reed et al., 2013)
Many-Objective Evolutionary Algorithms (MOEAs)	Computational algorithms that seek to simultaneously optimize problems with three or more objectives by simulating the basic principles of the evolutionary process on a set of individuals (solutions), i.e., an evolutionary population, by means of the so-called evolutionary operators (fitness assignment, selection, crossover, mutation and elitism) (Coello et al., 2007).
Model Uncertainty	Bias or imprecision associated with compromises made or lack of adequate knowledge in specifying the structure of a model, including choices of mathematical equation or family of probability distributions. Can also refer to limitations in knowledge affecting the construction of a reasoned argument or qualitative assessment (EFSA Scientific Committee et al., 2018a).
Participatory discourse	Where competing arguments, beliefs and values about the risk problem are openly discussed to find equitable outcomes (Renn 2008).
Pareto optimality	A tool used in the optimization of multiple, conflicting objectives. Pareto optimality defines the 'best' trade-off among conflicting objectives. It is defined as a set of 'non-inferior' solutions in the objective space defining a boundary beyond which none of the objectives can be improved without sacrificing at least one other objective (Coello et al., 2007).

Patient Rule Induction Method (PRIM)	A data mining classification algorithm that is designed to find regions in high-dimensional input space with large (or small) values of a real output variable (Friedman & Fisher 1999).
Probability	(a) A measure for expressing uncertainty about an event. (b) The relative fraction of times an event would occur if the situation were hypothetically ''repeated'' an infinite number of times ('frequentist' probability) (Aven 2010).
Probability Bounds Analysis	A method for combining probability bounds as inputs in order to obtain a probability bound for the output of a deterministic model. It is a special case of the general theory of imprecise probability which provides more ways to obtain partial expressions of uncertainty for the output based on more general partial expressions for inputs. (EFSA Scientific Committee et al., 2018b)
Qualitative risk assessment	An assessment where outputs on the likelihood of the outcome or the magnitude of its consequences are expressed in qualitative terms or ratings, such as 'high', 'medium', 'low' or 'negligible' (Cox Jr. et al., 2005; OIE 2019).
Quantitative risk assessment	An assessment where the outputs of the risk assessment are expressed numerically (OIE 2019).
Reflective discourse	Deliberation among decision-makers, analysts and stakeholders that reflects on the costs and benefits of protective strategies. It seeks to find a consensus on the degree of protection necessary to avoid catastrophic consequences (Renn 2008).
Risk	A combination of the likelihood of occurrence and the magnitude of impact (consequences) of a hazard event on people or things that they value (assets) (Office of the Prime Minister's Chief Science Advisor 2016a).
Risk analysis	The process of characterizing potential consequences of an event (or events) (both harmful and beneficial) and estimating the likelihood that they could occur.
Risk assessment	The process of evaluating the risk of a hazardous event. Risk assessment involves identification and characterization of the hazard(s), estimation of the likelihood/probability of the hazard and analysis of the potential consequences (Office of the Prime Minister's Chief Science Advisor 2016b).
Risk-Informed Decision-Making (RIDM)	A process that uses a set of performance measures, together with other considerations within a deliberative process to "inform" decision-making about risks (NASA 2010; Zio & Pedroni 2012a).
Robust Decision Making (RDM)	An iterative process to evaluate policy decisions under deep uncertainty that combines exploratory model-based methods. RDM characterizes uncertainty using multiple representations of possible future states rather than a single set of probability distributions. It seeks a robust policy solution (i.e., one that performs "well enough" across a broad range of plausible futures), rather than an optimal one (Lempert & Collins 2007).

Sensitivity Analysis	The study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model inputs.
Structured Decision Making (SDM)	An organized, step-wise approach to identifying and evaluating decision alternatives that focuses on engaging stakeholders, experts and decision makers in productive decision-oriented analysis and dialogue and that deals proactively with complexity and judgment in decision making. SDM provides a framework that becomes a decision-focused roadmap for integrating activities related to planning, analysis and consultation (Gregory et al., 2012b).
Subjective (Bayesian) probability	(a) A measure of uncertainty about future events and consequences, seen through the eyes of the assessor and based on some background information and knowledge (Aven 2010). (b) A probability, approximate probability or probability bound obtained by expert judgement (EFSA Scientific Committee et al., 2018a).
Uncertainty	The state, even partial, of deficiency of information related to, understanding or knowledge of, an event, its consequence, or likelihood (ISO 2009).

10 References

Abelson, J., Forest, P.-G., Eyles, J., Smith, P., Martin, E., Gauvin, F.-P. (2003) Deliberations about deliberative methods: issues in the design and evaluation of public participation processes. *Social Science & Medicine*, 57(2): 239-251. https://doi.org/10.1016/S0277-9536(02)00343-X

Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K., Rousseau, J. (2012) Combining expert opinions in prior elicitation. *Bayesian Anal.*, 7(3): 503-532. 10.1214/12-BA717

Allan, G., Eromenko, I., McGregor, P., Swales, K. (2011) The regional electricity generation mix in Scotland: A portfolio selection approach incorporating marine technologies. *Energy Policy*, 39(1): 6-22. https://doi.org/10.1016/j.enpol.2010.08.028

Allard, D., Comunian, A., Renard, P. (2012) Probability aggregation methods in geoscience. *Mathematical Geosciences*, 44(5): 545-581. 10.1007/s11004-012-9396-3

Amendola, A. (2002) Recent paradigms for risk informed decision making. *Safety Science*, 40(1): 17-30. http://dx.doi.org/10.1016/S0925-7535(01)00039-X

Apostolakis, G.E. (2004) How useful is quantitative risk assessment? *Risk Analysis*, 24(3): 515-520. 10.1111/j.0272-4332.2004.00455.x

Asche, F. (1997) Dynamic adjustment in demand equations. *Marine Resource Economics*, 12(3): 221-237. 10.1086/mre.12.3.42629197

Ascough II, J.C., Maier, H.R., Ravalico, J.K., Strudley, M.W. (2008) Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. *Ecological Modelling*, 219(3–4): 383-399. https://doi.org/10.1016/j.ecolmodel.2008.07.015

Aspinall, W. (2010) A route to more tractable expert advice. *Nature*, 463(7279): 294-295. 10.1038/463294a

Audzijonyte, A., Pethybridge, H., Porobic, J., Gorton, R., Kaplan, I., Fulton, E.A. (2019) Atlantis: A spatially explicit end-to-end marine ecosystem model with dynamically integrated physics, ecology and socio-economic modules. *Methods in Ecology and Evolution*, 10(10): 1814-1819. 10.1111/2041-210x.13272

Aven, T. (2008) A semi-quantitative approach to risk analysis, as an alternative to QRAs. *Reliability Engineering & System Safety*, 93(6): 790-797. http://dx.doi.org/10.1016/j.ress.2007.03.025

Aven, T. (2010) On how to define, understand and describe risk. *Reliability Engineering & System Safety*, 95(6): 623-631. http://dx.doi.org/10.1016/j.ress.2010.01.011

Aven, T. (2011a) On different types of uncertainties in the context of the precautionary principle. *Risk Analysis*, 31(10): 1515-1525. 10.1111/j.1539-6924.2011.01612.x

Aven, T. (2011b) On the new ISO guide on risk management terminology. *Reliability Engineering & System Safety*, 96(7): 719-726. https://doi.org/10.1016/j.ress.2010.12.020

Aven, T. (2013) On the meaning of a black swan in a risk context. *Safety Science*, 57: 44-51. http://dx.doi.org/10.1016/j.ssci.2013.01.016

Aven, T., Renn, O. (2009a) On risk defined as an event where the outcome is uncertain. *Journal of risk research*, 12(1): 1-11. 10.1080/13669870802488883

Aven, T., Renn, O. (2009b) The role of quantitative risk assessments for characterizing risk and uncertainty and delineating appropriate risk management options, with special emphasis on terrorism risk. *Risk Analysis*, 29(4): 587-600. 10.1111/j.1539-6924.2008.01175.x

Aven, T., Zio, E. (2011) Some considerations on the treatment of uncertainties in risk assessment for practical decision-making. *Reliability Engineering and System Safety*, 96: 64-74.

Bankes, S. (1993) Exploratory modeling for policy analysis. *Operations Research*, 41(3): 435-449. 10.1287/opre.41.3.435

Bankes, S., Walker, W.E., Kwakkel, J.H. (2013) Exploratory modeling and analysis. In: S.I. Gass & M.C. Fu (Eds). *Encyclopedia of Operations Research and Management Science*. Springer US, Boston, MA: 532-537. 10.1007/978-1-4419-1153-7_314

Bankes, S.C., Lempert, R.J., Popper, S.W. (2001) Computer-Assisted Reasoning. *Computing in Science & Engineering*, 3(2): 71-77. 10.1109/5992.909006

Barrett, D.J., Couch, C.A., Metcalfe, D.J., Lytton, L., Adhikary, d.P., Schmidt, R.K. (2013) Methodology for bioregional assessments of the impacts of coal seam gas and coal mining development on water resources. *A report prepared for the Independent Expert Scientific Committee on Coal Seam Gas and Large Coal Mining Development through the Department of the Environment*: 75.

Barry, S., Lin, X. (2010) Point of Truth Calibration: Putting science into scoring systems. *Australian Centre of Excellence in Risk Analysis (ACERA) Report Number: EP10378, 1 November 2010*. CSIRO, Canberra, Australia: 22.

Barry, S.C., Caley, P., Liu, S., Paini, D.R., Carey, J., Clark, G. (2015) Development of an expert-based model for improved biofouling risk assessment. *CEBRA Project No. 1402A*: 47.

Bellman, R. (1954) The theory of dynamic programming. *Bulletin of the American Mathematical Society*, 60(6): 503-516.

Ben-Haim, Y. (2006) *Info-gap decision theory: decisions under severe uncertainty*. Elsevier.

Bhave, A.G., Conway, D., Dessai, S., Stainforth, D.A. (2016) Barriers and opportunities for robust decision making approaches to support climate change adaptation in the developing world. *Climate Risk Management*, 14: 1-10. https://doi.org/10.1016/j.crm.2016.09.004

Bilgin, B., Brenner, L. (2013) Context affects the interpretation of low but not high numerical probabilities: A hypothesis testing account of subjective probability. *Organizational Behavior and Human Decision Processes*, 121(1): 118-128. https://doi.org/10.1016/j.obhdp.2013.01.004

Bjørndal, T., Lane, D.E., Weintraub, A. (2004) Operational research models and the management of fisheries and aquaculture: A review. *European Journal of Operational Research*, 156(3): 533-540. https://doi.org/10.1016/S0377-2217(03)00107-3

Breiman, L. (2001) Random Forests. *Machine Learning*, 45(1): 5-32. 10.1023/a:1010933404324

Breiman, L., Freidman, J.H., Olshen, R.A., Stone, C.J. (1984) *Classification and Regression Trees*. Wadsworth, Belmont: 358.

Brun, W., Teigen, K.H. (1988) Verbal probabilities: Ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes*, 41(3): 390-404. https://doi.org/10.1016/0749-5978(88)90036-2

Bryant, B.P., Lempert, R.J. (2010) Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting and Social Change*, 77(1): 34-49. http://dx.doi.org/10.1016/j.techfore.2009.08.002

Budescu, D.V., Broomell, S., Por, H.-H. (2009) Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. *Psychological Science*, 20(3): 299-308. doi:10.1111/j.1467-9280.2009.02284.x

Budescu, D.V., Por, H.-H., Broomell, S.B., Smithson, M. (2014) The interpretation of IPCC probabilistic statements around the world. *Nature Climate Change*, 4(6): 508-512. 10.1038/nclimate2194

Budescu, D.V., Wallsten, T.S. (1985) Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes*, 36(3): 391-405. https://doi.org/10.1016/0749-5978(85)90007-X

Burgman, M. (2005) *Risks and decisions for conservation and environmental management*. Cambridge University Press, Cambridge, U.K.,

Burgman, M., Carr, A., Godden, L., Gregory, R., McBride, M., Flander, L., Maguire, L. (2011a) Redefining expertise and improving ecological judgment. *Conservation Letters*, 4(2): 81-87. 10.1111/j.1755-263X.2011.00165.x

Burgman, M., Colyvan, M., Christian, R., Dobiecki, B., Elith, J., Fidler, F., Grant, N., Hauser, C.E., Hayes, K., Jones, S., Korb, K., Kuhnert, P., McBride, M., McCarthy, M., Nicholson, A., Robinson, A., Rout, T., Speirs-Bridge, A., Walshe, T. (2010) Demonstrating risk analysis capabilities. *ACERA Project 0901 - Final Report*: 93.

Burgman, M., Fidler, F., McBride, M., Walshe, T., Wintle, B. (2006) Eliciting Expert Judgments: Literature Review. *ACERA Project 0611, Eliciting Expert Judgments*: 71.

Burgman, M.A. (2015) *Trusting Judgements: How to Get the Best out of Experts*. Cambridge University Press, Cambridge. DOI: 10.1017/CBO9781316282472

Burgman, M.A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., Fidler, F., Rumpff, L., Twardy, C. (2011b) Expert status and performance. *PLoS ONE*, 6(7): e22998. 10.1371/journal.pone.0022998

Chadès, I., Chapron, G., Cros, M.-J., Garcia, F., Sabbadin, R. (2014) MDPtoolbox: a multi-platform toolbox to solve stochastic dynamic programming problems. *Ecography*, 37(9): 916-920. 10.1111/ecog.00888

Chadès, I., Martin, T.G., Nicol, S., Burgman, M.A., Possingham, H.P., Buckley, Y.M. (2011) General rules for managing and surveying networks of pests, diseases, and endangered species. *Proceedings of the National Academy of Sciences*, 108(20): 8323-8328. 10.1073/pnas.1016846108

Chadès, I., Nicol, S., Rout, T.M., Péron, M., Dujardin, Y., Pichancourt, J.-B., Hastings, A., Hauser, C.E. (2017) Optimization methods to solve adaptive management problems. *Theoretical Ecology*, 10(1): 1-20. 10.1007/s12080-016-0313-0

Clark, C.W. (1985) *Bioeconomic Modelling and Fisheries Management*. Wiley Interscience: 291. https://books.google.co.nz/books?id=NCkYAQAAIAAJ

Clemen, R.T., Winkler, R.L. (1999) Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2): 187-203.

Coello, C.A.C., Lamont, G.B., Van Veldhuizen, D.A. (2007) *Evolutionary algorithms for solving multi-objective problems*. Springer.

Conrad, J.M. (2010) *Resource Economics*. Cambridge University Press, New York: 285.

Cooke, R. (1991) *Experts in uncertainty: opinion and subjective probability in science*. Oxford University Press on Demand.

Cox Jr., L.A., Babayev, D., Huber, W. (2005) Some limitations of qualitative risk rating systems. *Risk Analysis*, 25(3): 651-662. doi:10.1111/j.1539-6924.2005.00615.x

Cox, L.A. (2012) Confronting deep uncertainties in risk analysis. *Risk Analysis*, 32(10): 1607-1629. 10.1111/j.1539-6924.2012.01792.x

Crain, C.M., Kroeker, K., Halpern, B.S. (2008) Interactive and cumulative effects of multiple human stressors in marine systems. *Ecology Letters*, 11(12): 1304-1315. 10.1111/j.1461-0248.2008.01253.x

Crombie, J., Knight, E., Barry, S. (2007) Marine pest incursions - a tool to predict the cost of eradication based on expert assessments. *Australian Bureau of Rural Sciences*: 62p.

Dalal, S., Han, B., Lempert, R., Jaycocks, A., Hackbarth, A. (2013) Improving scenario discovery using orthogonal rotations. *Environmental Modelling & Software*, 48: 49-64. https://doi.org/10.1016/j.envsoft.2013.05.013

Dambacher, J.M., Gaughan, D.J., Rochet, M.-J., Rossignol, P.A., Trenkel, V.M. (2009) Qualitative modelling and indicators of exploited ecosystems. *Fish and Fisheries*, 10(3): 305-322. 10.1111/j.1467-2979.2008.00323.x

Dambacher, J.M., Li, H.W., Rossignol, P.A. (2002) Relevance of community structure in assessing indeterminacy of ecological predictions. *Ecology*, 83(5): 1372-1385. 10.1890/0012-9658(2002)083[1372:Rocsia]2.0.Co;2

Dambacher, J.M., Luh, H.-K., Li, H.W., Rossignol, P.A. (2003) Qualitative stability and ambiguity in model ecosystems. *The American Naturalist*, 161(6): 876-888. 10.1086/367590

Dambacher, J.M., Shenton, W., Hayes, K.R., Hart, B.T., Barry, S. (2007) Qualitative modelling and Bayesian network analysis for risk-based biosecurity decision making in complex systems. *ACERA project 06/01*: 55.

de Neufville, R., Smet, K. (2019) Engineering Options Analysis (EOA). In: V.A.W.J. Marchau, W.E. Walker, P.J.T.M. Bloemen & S.W. Popper (Eds). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing, Cham: 117-132. 10.1007/978-3-030-05252-2_2

Dewar, J.A., Builder, C.H., Hix, W.M., Levin, M.H. (1993) Assumption-based planning; a planning tool for very uncertain times. *A report prepared for the United States Army*: 78.

Dinno, A. (2018) A Collection of Tools to Conduct Levins' Loop Analysis. https://www.alexisdinno.com/LoopAnalyst

Djanibekov, U., Villamor, G.B. (2016) Market-based instruments for risk-averse farmers: rubber agroforest conservation in Jambi Province, Indonesia. *Environment and Development Economics*, 22(2): 133-155. 10.1017/S1355770X16000310

Doyen, L., De Lara, M., Ferraris, J., Pelletier, D. (2007) Sustainability of exploited marine ecosystems through protected areas: A viability model and a coral reef case study. *Ecological Modelling*, 208(2): 353-366. https://doi.org/10.1016/j.ecolmodel.2007.06.018

Drescher, M., Perera, A.H., Johnson, C.J., Buse, L.J., Drew, C.A., Burgman, M.A. (2013) Toward rigorous use of expert knowledge in ecological research. *Ecosphere*, 4(7): art83. 10.1890/es12-00415.1

Drouin, M., Parry, G., Lehner, J., Martinez-Guridi, G., LaChance, J., Wheeler, T. (2009) Guidance on the treatment of uncertainties associated with pras in risk-informed decision making - Main report. *NUREG-1855 Vol. 1*: 144.

Dunstan, P., Dambacher, J.M., Bax, N., Smith, T., Fulton, E.A., Hedge, P., Hobday, A.J., Foster, S. (2015) A hierarchical risk assessment framework for ecosystem based management: 24.

EFSA Scientific Committee, Benford, D., Halldorsson, T., Jeger, M.J., Knutsen, H.K., More, S., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Schlatter, J.R., Silano, V., Solecki, R., Turck, D., Younes, M., Craig, P., Hart, A., Von Goetz, N., Koutsoumanis, K., Mortensen, A., Ossendorp, B., Germini, A., Martino, L., Merten, C., Mosbach-Schulz, O., Smith, A., Hardy, A. (2018a) The principles and methods behind EFSA's Guidance on Uncertainty Analysis in Scientific Assessment. *EFSA Journal*, 16(1): e05122. 10.2903/j.efsa.2018.5122

EFSA Scientific Committee, Benford, D., Halldorsson, T., Jeger, M.J., Knutsen, H.K., More, S., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Schlatter, J.R., Silano, V., Solecki, R., Turck, D., Younes, M., Craig, P., Hart, A., Von Goetz, N., Koutsoumanis, K., Mortensen, A., Ossendorp, B., Martino, L., Merten, C., Mosbach-Schulz, O., Hardy, A. (2018b) Guidance on Uncertainty Analysis in Scientific Assessments. *EFSA Journal*, 16(1): e05123. 10.2903/j.efsa.2018.5123

Eker, S., Kwakkel, J.H. (2018) Including robustness considerations in the search phase of Many-Objective Robust Decision Making. *Environmental Modelling & Software*, 105: 201-216. https://doi.org/10.1016/j.envsoft.2018.03.029

European Commission (2001) Guidelines for the assessment of indirect and cumulative impacts as well as impact interactions. *EC DG XI: Environment, Nuclear Safety & Civil Protection*, NE80328/D1/3: 169.

Ewald, C.-O., Wang, W.-K. (2010) Sustainable yields in fisheries: Uncertainty, risk-aversion, and mean-variance analysis. *Natural Resource Modeling*, 23(3): 303-323. doi:10.1111/j.1939-7445.2010.00065.x

Fearon, J.D. (1998) Deliberation as Discussion. In: J. Elster (Ed). *Deliberative Democracy*. Cambridge University Press, Cambridge: 44-68. DOI: 10.1017/CBO9781139175005.004

Ferson, S., Ginzburg, L.R. (1996) Different methods are needed to propagate ignorance and variability. *Reliability Engineering & System Safety*, 54(2): 133-144. http://dx.doi.org/10.1016/S0951-8320(96)00071-3

Ferson, S., Kreinovich, V., Grinzburg, L., Myers, D., Sentz, K. (2015) Constructing probability boxes and Dempster-Shafer structures: 132.

Finnoff, D., Shogren, J.F., Leung, B., Lodge, D. (2007) Take a risk: Preferring prevention over control of biological invaders. *Ecological Economics*, 62(2): 216-222. http://dx.doi.org/10.1016/j.ecolecon.2006.03.025

Finucane, M.L., Alhakami, A., Slovic, P., Johnson, S.M. (2000) The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, 13(1): 1-17. 10.1002/(SICI)1099-0771(200001/03)13:1<1::AID-BDM333>3.0.CO;2-S

Forsell, N., Wikström, P., Garcia, F., Sabbadin, R., Blennow, K., Eriksson, L.O. (2011) Management of the risk of wind damage in forestry: a graph-based Markov decision process approach. *Annals of Operations Research*, 190(1): 57-74. 10.1007/s10479-009-0522-7

Friedman, J.A., Baker, J.D., Mellers, B.A., Tetlock, P.E., Zeckhauser, R. (2018) The value of precision in probability assessment: Evidence from a large-scale geopolitical forecasting tournament. *International Studies Quarterly*, 62(2): 410-422. 10.1093/isq/sqx078

Friedman, J.H., Fisher, N.I. (1999) Bump hunting in high-dimensional data. *Statistics and Computing*, 9(2): 123-143. 10.1023/a:1008894516817

Funtowicz, S.O., Ravetz, J.R. (1994) Uncertainty, complexity and post-normal science. *Environmental Toxicology and Chemistry*, 13(12): 1881-1885. 10.1002/etc.5620131203

Gallagher, S.S., Rice, G.E., Scarano, L.J., Teuschler, L.K., Bollweg, G., Martin, L. (2015) Cumulative risk assessment lessons learned: A review of case studies and issue papers. *Chemosphere*, 120: 697-705. https://doi.org/10.1016/j.chemosphere.2014.10.030

Garthwaite, P.H., Kadane, J.B., O'Hagan, A. (2005) Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470): 680-701. 10.1198/016214505000000105

Ge, L., Mourits, M.C.M., Kristensen, A.R., Huirne, R.B.M. (2010) A modelling approach to support dynamic decision-making in the control of FMD epidemics. *Preventive Veterinary Medicine*, 95(3–4): 167-174. http://dx.doi.org/10.1016/j.prevetmed.2010.04.003

Genest, C., McConway, K.J. (1990) Allocating the weights in the linear opinion pool. *Journal of Forecasting*, 9(1): 53-73. 10.1002/for.3980090106

Genest, C., Zidek, J.V. (1986) Combining probability distributions: A critique and an annotated bibliography. *Statist. Sci.*, 1(1): 114-135. 10.1214/ss/1177013825

Gentile, J.H., Harwell, M.A. (2001) Strategies for Assessing Cumulative Ecological Risks. *Human and Ecological Risk Assessment: An International Journal*, 7(2): 239-246. 10.1080/20018091094358

Gigerenzer, G., Hoffrage, U. (1995) How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4): 684-704. 10.1037/0033-295X.102.4.684

Giuliani, M., Castelletti, A. (2016) Is robustness really robust? How different definitions of robustness impact decision-making under climate change. *Climatic Change*, 135(3): 409-424. 10.1007/s10584-015-1586-9

Goldstein, D.G., Gigerenzer, G. (1999) The recognition heuristic: How ignorance makes us smart. *Simple heuristics that make us smart*. Oxford University Press: 37-58.

Grafton, R.Q., Adamowicz, W., Dupont, D., Nelson, H., Hill, R., Renzetti, S. (2008) Bioeconomics of Fisheries. *Economics of the Environment and Natural Resources*, 35: 95-128.

Gregory, R., Dieckmann, N., Peters, E., Failing, L., Long, G., Tusler, M. (2012a) Deliberative disjunction: Expert and public understanding of outcome uncertainty. *Risk Analysis*, 32(12): 2071-2083. 10.1111/j.1539-6924.2012.01825.x

Gregory, R., Failing, L., Harstone, M., Long, G., McDaniels, T., Ohlson, D. (2012b) *Structured Decision Making: a practical guide to environmental management choices*. John Wiley & Sons.

Gregory, R., Fischhoff, B., McDaniels, T. (2005) Acceptable input: Using decision analysis to guide public policy deliberations. *Decision Analysis*, 2(1): 4-16. 10.1287/deca.1050.0035

Groves, D.G., Fischbach, J.R., Kalra, N., Molina-Perez, E., Yates, D., Purkey, D., Fencl, A., Mehta, V.K., Wright, B., Pyke, G. (2014) Developing Robust Strategies for Climate Change and Other Risks: A Water Utility Framework 118.

Groves, D.G., Lempert, R.J. (2007) A new analytic method for finding policy-relevant scenarios. *Global Environmental Change*, 17(1): 73-85. https://doi.org/10.1016/j.gloenvcha.2006.11.006

Groves, D.G., Molina-Perez, E., Bloom, E., Fischbach, J.R. (2019) Robust Decision Making (RDM): Application to water planning and climate policy. In: V.A.W.J. Marchau, W.E. Walker, P.J.T.M. Bloemen & S.W. Popper (Eds). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing, Cham: 135-163. 10.1007/978-3-030-05252-2_2

Guikema, S.D., Aven, T. (2010) Is ALARP applicable to the management of terrorist risks? *Reliability Engineering & System Safety*, 95(8): 823-827. http://dx.doi.org/10.1016/j.ress.2010.03.007

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 Environmental Change*, 23(2): 485-498. https://doi.org/10.1016/j.gloenvcha.2012.12.006

Haasnoot, M., Middelkoop, H., Offermans, A., Beek, E.v., Deursen, W.P.A.v. (2012) Exploring pathways for sustainable water management in river deltas in a changing environment. *Climatic Change*, 115(3): 795-819. 10.1007/s10584-012-0444-2

Haasnoot, M., van 't Klooster, S., van Alphen, J. (2018) Designing a monitoring system to detect signals to adapt to uncertain climate change. *Global Environmental Change*, 52: 273-285. https://doi.org/10.1016/j.gloenvcha.2018.08.003

Haasnoot, M., Warren, A., Kwakkel, J.H. (2019) Dynamic Adaptive Policy Pathways (DAPP). In: V.A.W.J. Marchau, W.E. Walker, P.J.T.M. Bloemen & S.W. Popper (Eds). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing, Cham: 71-92. 10.1007/978-3-030-05252-2_2

Hadka, D., Herman, J., Reed, P., Keller, K. (2015) OpenMORDM: An open source framework for many-objective robust decision making. *Environmental Modelling & Software*, 74: 114-129. https://doi.org/10.1016/j.envsoft.2015.07.014

Hadka, D., Reed, P. (2013) Borg: An auto-adaptive Many-Objective Evolutionary computing framework. *Evolutionary Computation*, 21(2): 231-259. 10.1162/EVCO_a_00075 %M 22385134

Haimes, Y.Y. (2009) On the complex definition of risk: A systems-based approach. *Risk Analysis*, 29(12): 1647-1654. doi:10.1111/j.1539-6924.2009.01310.x

Hamarat, C., Kwakkel, J.H., Pruyt, E., Loonen, E.T. (2014) An exploratory approach for adaptive policymaking by using multi-objective robust optimization. *Simulation Modelling Practice and Theory*, 46: 25-39. https://doi.org/10.1016/j.simpat.2014.02.008

Hanea, A.M., McBride, M.F., Burgman, M.A., Wintle, B.C. (2018a) Classical meets modern in the IDEA protocol for structured expert judgement. *Journal of risk research*, 21(4): 417-433. 10.1080/13669877.2016.1215346

Hanea, A.M., McBride, M.F., Burgman, M.A., Wintle, B.C. (2018b) The value of performance weights and discussion in aggregated expert judgments. *Risk Analysis*, 38(9): 1781-1794. 10.1111/risa.12992

Hanea, A.M., McBride, M.F., Burgman, M.A., Wintle, B.C., Fidler, F., Flander, L., Twardy, C.R., Manning, B., Mascaro, S. (2017) I nvestigate D iscuss E stimate A ggregate for structured expert judgement. *International Journal of Forecasting*, 33(1): 267-279. https://doi.org/10.1016/j.ijforecast.2016.02.008

Hannesson, R., Kennedy, J. (2005) Landing fees versus fish quotas. *Land Economics*, 81(4): 518-529. 10.3368/le.81.4.518

Hansen, L.G., Jensen, F. (2017) Regulating fisheries under uncertainty. *Resource and Energy Economics*, 50: 164-177. https://doi.org/10.1016/j.reseneeco.2017.08.001

Hardaker, J.B., Huirne, R.B.M., Anderson, J.R., Lien, G. (2004) *Coping with Risk in Agriculture*. CABI Publishing, Oxfordshire, United Kingdom: 332.

Hart, R. (2003) Dynamic pollution control—time lags and optimal restoration of marine ecosystems. *Ecological Economics*, 47(1): 79-93. https://doi.org/10.1016/j.ecolecon.2002.09.002

Hauser, C.E., Possingham, H.P. (2008) Experimental or precautionary? Adaptive management over a range of time horizons. *Journal of Applied Ecology*, 45(1): 72-81. 10.1111/j.1365-2664.2007.01395.x

Hayes, K.R. (2011) Uncertainty and uncertainty analysis methods - Report Number: EP102467. *Issues in quantitative and qualitative risk modeling with application to import risk assessment ACERA project (0705)*: 131.

Hemez, F.M., Van Buren, K.L. (2019) Info-Gap (IG): Robust design of a mechanical latch. In: V.A.W.J. Marchau, W.E. Walker, P.J.T.M. Bloemen & S.W. Popper (Eds). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing, Cham: 201-222. 10.1007/978-3-030-05252-2_10

Hemming, V., Burgman, M.A., Hanea, A.M., McBride, M.F., Wintle, B.C. (2018a) A practical guide to structured expert elicitation using the IDEA protocol. *Methods in Ecology and Evolution*, 9(1): 169-180. 10.1111/2041-210x.12857

Hemming, V., Walshe, T.V., Hanea, A.M., Fidler, F., Burgman, M.A. (2018b) Eliciting improved quantitative judgements using the IDEA protocol: A case study in natural resource management. *PLoS ONE*, 13(6): e0198468. 10.1371/journal.pone.0198468

Herman, J.D., Reed, P.M., Zeff, H.B., Characklis, G.W. (2015) How should robustness be defined for water systems planning under change? *Journal of Water Resources Planning and Management*, 141(10): 04015012. doi:10.1061/(ASCE)WR.1943-5452.0000509

Ho, E.H., Budescu, D.V., Dhami, M.K., Mandel, D.R. (2015) Improving the communication of uncertainty in climate science and intelligence analysis. *Behavioral Science & Policy*, 1(2): 43-55.

Hobday, A.J., Smith, A.D.M., Stobutzki, I.C., Bulman, C., Daley, R., Dambacher, J.M., Deng, R.A., Dowdney, J., Fuller, M., Furlani, D., Griffiths, S.P., Johnson, D., Kenyon, R., Knuckey, I.A., Ling, S.D., Pitcher, R., Sainsbury, K.J., Sporcic, M., Smith, T., Turnbull, C., Walker, T.I., Wayte, S.E., Webb, H., Williams, A., Wise, B.S., Zhou, S. (2011) Ecological risk assessment for the effects of fishing. *Fisheries Research*, 108(2): 372-384. http://dx.doi.org/10.1016/j.fishres.2011.01.013

Holling, C.S. (2004) *Adaptive Environmental Assessment and Management*. Blackburn Press: 377.

Honsack, G.R., Ickowicz, A., Hayes, K.R., Dambacher, J.M., Barry, S.A., Henderson, B.L. (2018) Receptor impact modelling. Submethodology M08 from the Bioregional Assessment Technical Programme: 89. http://data.bioregionalassessments.gov.au/submethodology/M08

Horlick-Jones, T. (1998) Meaning and contextualisation in risk assessment. *Reliability Engineering & System Safety*, 59(1): 79-89. https://doi.org/10.1016/S0951-8320(97)00122-1

Hosack, G.R., Hayes, K.R., Dambacher, J.M. (2008) Assessing model structure uncertainty through an analysis of system feedback and Bayesian networks. *Ecological Applications*, 18(4): 1070-1082. 10.1890/07-0482.1

Hurwiczy, L. (1951) Optimality Criteria for Decision Making Under Ignorance. *Cowles Commission Discussion Paper*: 16.

IPCC (Intergovernmental Panel on Climate Change) (2005) Guidance Notes for Lead Authors of the IPCC Fourth Assessment Report on Addressing Uncertainties: 4.

ISO (2009) Guide 73. Risk management — Vocabulary. International Organization for Standardization, Switzerland: 15.

ISO (2018) ISO 31000:2018 Risk Management - Principles and Guidelines. International Organization for Standardization, Switzerland.

Jasanoff, S. (1993) Bridging the two cultures of risk analysis. *Risk Analysis*, 13(2): 123-129.

Jensen, F., Frost, H., Abildtrup, J. (2017) Fisheries regulation: A survey of the literature on uncertainty, compliance behavior and asymmetric information. *Marine Policy*, 81: 167-178. https://doi.org/10.1016/j.marpol.2017.03.028

Johnson, F.A., Zimmerman, G.S., Huang, M.T., Padding, P.I., Balkcom, G.D., Runge, M.C., Devers, P.K. (2019) Multi-species duck harvesting using dynamic programming and multi-criteria decision analysis. *Journal of Applied Ecology*, 0(ja). 10.1111/1365-2664.13377

Justus, J. (2006) Loop analysis and qualitative modeling: limitations and merits. *Biology and Philosophy*, 21(5): 647-666. 10.1007/s10539-006-9050-x

Kahneman, D. (2011) *Thinking, Fast and Slow*. Penguin Books Ltd, London, England: 499.

Kahneman, D., Tversky, A. (1979) Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2): 263-291. 10.1142/9789814417358_0006

Kahneman, D., Tversky, A. (1982) Variants of uncertainty. *Cognition*, 11: 143-157.

Kahneman, D., Tversky, A. (1984) Choices, frames and values. *American Psychologist*, 39(4): 341-350.

Kaplan, S., Garrick, B.J. (1981) On the quantitative definition of risk. *Risk Analysis*, 1(1): 11-27. 10.1111/j.1539-6924.1981.tb01350.x

Kaplan, S., Haimes, Y.Y., Garrick, B.J. (2001) Fitting hierarchical holographic modeling into the Theory of Scenario Structuring and a resulting refinement to the quantitative definition of risk. *Risk Analysis*, 21(5): 807-807. 10.1111/0272-4332.215153

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

Keeney, R.L. (1982) Decision Analysis: An overview. *Operations Research*, 30(5): 803-838. 10.1287/opre.30.5.803

Kennedy, J.O. (1986) *Dynamic programming: applications to agriculture and natural resources*. Elsevier Applied Science Publishers Ltd., Essex, United Kingdom: 341.

Kinzig, A., Starrett, D. (2003) Coping with uncertainty: a call for a new science-policy forum. *AMBIO: A Journal of the Human Environment*, 32(5): 330-335. 10.1579/0044-7447-32.5.330

Knight, E., Barry, S., Summerson, R., Cameron, S., Darbyshire, R. (2007) Designated Exchange Areas Project – Providing informed decisions on the discharge of Ballast Water in Australia (Phase 2): 93.

Knight, F.H. (1921) *Risk, uncertainty and profit*. Houghton Mifflin Company, The Riverside Press, Cambridge, United Kingdom.

Kollat, J.B., Reed, P. (2007) A framework for Visually Interactive Decision-making and Design using Evolutionary Multi-objective Optimization (VIDEO). *Environmental Modelling & Software*, 22(12): 1691-1704. https://doi.org/10.1016/j.envsoft.2007.02.001

Kriegler, E., Hall, J.W., Held, H., Dawson, R., Schellnhuber, H.J. (2009) Imprecise probability assessment of tipping points in the climate system. *Proceedings of the National Academy of Sciences*, 106(13): 5041-5046. 10.1073/pnas.0809117106

Kristensen, A.R., Jørgensen, E. (2000) Multi-level hierarchic Markov processes as a framework for herd management support. *Annals of Operations Research*, 94(1): 69-89. 10.1023/a:1018921201113

Kuhnert, P., Barry, S. (2010) Bayesian Learning and Synthesis through the Elicitation of Risk: BLASTER. *Australian Centre of Excellence for Risk Analysis (ACERA) Report No. EP102289*: 26.

Kuhnert, P., Martin, T.G., Griffiths, S.P. (2010) A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecology Letters*, 13: 900-914.

Kwadijk, J.C.J., Haasnoot, M., Mulder, J.P.M., Hoogvliet, M.M.C., Jeuken, A.B.M., van der Krogt, R.A.A., van Oostrom, N.G.C., Schelfhout, H.A., van Velzen, E.H., van Waveren, H., de Wit, M.J.M. (2010) Using adaptation tipping points to prepare for climate change and sea level rise: a case study in the Netherlands. *Wiley Interdisciplinary Reviews: Climate Change*, 1(5): 729-740. doi:10.1002/wcc.64

Kwakkel, J.H. (2017) The Exploratory Modeling Workbench: An open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. *Environmental Modelling & Software*, 96: 239-250. https://doi.org/10.1016/j.envsoft.2017.06.054

Kwakkel, J.H., Eker, S., Pruyt, E. (2016a) How robust is a robust policy? Comparing alternative robustness metrics for robust decision-making. In: M. Doumpos, C. Zopounidis & E. Grigoroudis (Eds). *Robustness Analysis in Decision Aiding, Optimization, and Analytics*. Springer International Publishing, Cham: 221-237. 10.1007/978-3-319-33121-8_10

Kwakkel, J.H., Haasnoot, M. (2019) Supporting DMDU: A Taxonomy of Approaches and Tools. In: V.A.W.J. Marchau, W.E. Walker, P.J.T.M. Bloemen & S.W. Popper (Eds). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing, Cham: 355-374. 10.1007/978-3-030-05252-2_15

Kwakkel, J.H., Haasnoot, M., Walker, W.E. (2015) Developing dynamic adaptive policy pathways: a computer-assisted approach for developing adaptive strategies for a deeply uncertain world. *Climatic Change*, 132(3): 373-386. 10.1007/s10584-014-1210-4

Kwakkel, J.H., Haasnoot, M., Walker, W.E. (2016b) Comparing Robust Decision-Making and Dynamic Adaptive Policy Pathways for model-based decision support under deep uncertainty. *Environmental Modelling & Software*, 86: 168-183. https://doi.org/10.1016/j.envsoft.2016.09.017

Kwakkel, J.H., Jaxa-Rozen, M. (2016) Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes. *Environmental Modelling & Software*, 79: 311-321. https://doi.org/10.1016/j.envsoft.2015.11.020

Kwakkel, J.H., Pruyt, E. (2013) Exploratory Modeling and Analysis, an approach for model-based foresight under deep uncertainty. *Technological Forecasting and Social Change*, 80(3): 419-431. https://doi.org/10.1016/j.techfore.2012.10.005

Kwakkel, J.H., Walker, W.E., Marchau, V.A. (2010a) Classifying and communicating uncertainties in model-based policy analysis. *International journal of technology, policy and management*, 10(4): 299-315.

Kwakkel, J.H., Walker, W.E., Marchau, V.A.W.J. (2010b) Adaptive airport strategic planning. *European Journal of Transport and Infrastructure Research*, 10(3). 10.18757/ejtir.2010.10.3.2891

Lawrence, J., Haasnoot, M. (2017) What it took to catalyse uptake of dynamic adaptive pathways planning to address climate change uncertainty. *Environmental Science & Policy*, 68: 47-57. https://doi.org/10.1016/j.envsci.2016.12.003

Lawrence, J., Haasnoot, M., McKim, L., Atapattu, D., Campbell, G., Stroombergen, A. (2019) Dynamic Adaptive Policy Pathways (DAPP): From Theory to Practice. In: V.A.W.J. Marchau, W.E. Walker, P.J.T.M. Bloemen & S.W. Popper (Eds). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing, Cham: 187-199. 10.1007/978-3-030-05252-2_2

Leitch, M. (2010) ISO 31000:2009—The New International Standard on Risk Management. *Risk Analysis*, 30(6): 887-892. 10.1111/j.1539-6924.2010.01397.x

Lempert, R. (2013) Scenarios that illuminate vulnerabilities and robust responses. *Climatic Change*, 117(4): 627-646. 10.1007/s10584-012-0574-6

Lempert, R., Kalra, N., Peyraud, S., Mao, Z., Tan, S.B., Cira, D., Lotsch, A. (2013a) *Ensuring robust flood risk management in Ho Chi Minh City*. The World Bank.

Lempert, R., Nakicenovic, N., Sarewitz, D., Schlesinger, M. (2004) Characterizing climate-change uncertainties for decision-makers. An editorial essay. *Climatic Change*, 65(1): 1-9. 10.1023/B:CLIM.0000037561.75281.b3

Lempert, R.J. (2002) A new decision sciences for complex systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3): 7309-7313. 10.1073/pnas.082081699

Lempert, R.J. (2019) Robust Decision Making (RDM). In: V.A.W.J. Marchau, W.E. Walker, P.J.T.M. Bloemen & S.W. Popper (Eds). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing, Cham: 23-51. 10.1007/978-3-030-05252-2_2

Lempert, R.J., Collins, M.T. (2007) Managing the risk of uncertain threshold responses: Comparison of robust, optimum, and precautionary approaches. *Risk Analysis*, 27(4): 1009-1026. doi:10.1111/j.1539-6924.2007.00940.x

Lempert, R.J., Groves, D.G. (2010) Identifying and evaluating robust adaptive policy responses to climate change for water management agencies in the American west. *Technological Forecasting and Social Change*, 77(6): 960-974. http://dx.doi.org/10.1016/j.techfore.2010.04.007

Lempert, R.J., Groves, D.G., Popper, S.W., Bankes, S.C. (2006) A general, analytic method for generating robust strategies and narrative scenarios. *Management Science*, 52(4): 514-528. 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: 187. https://www.rand.org/pubs/monograph_reports/MR1626.html

Lempert, R.J., Popper, S.W., Groves, D.G., Nidhi, K., Fischbach, J.R., Bankes, S.C., Bryant, B.P., Collins, M.T., Keller, K., Hackbarth, A., Dixon, L., LaTourrette, T., Reville, R.T., Hall, J.W., Mijere, C., McInerney, D.J. (2013b) Making good decisions without predictions:

Robust decision making for planning under deep uncertainty. https://www.rand.org/pubs/research_briefs/RB9701.html

Levin, P.S., Fogarty, M.J., Murawski, S.A., Fluharty, D. (2009) Integrated ecosystem assessments: Developing the scientific basis for ecosystem-based management of the ocean. *PLOS Biology*, 7(1): e1000014. 10.1371/journal.pbio.1000014

Li, P., Chen, B., Zhang, B., Jing, L., Zheng, J. (2014) Monte Carlo simulation-based dynamic mixed integer nonlinear programming for supporting oil recovery and devices allocation during offshore oil spill responses. *Ocean & Coastal Management*, 89: 58-70. https://doi.org/10.1016/j.ocecoaman.2013.12.006

Lichtendahl, K.C.J., Grushka-Cockayne, Y., Winkler, R.L. (2013) Is it better to average probabilities or quantiles? *Management Science*, 59(7): 1594-1611. 10.1287/mnsc.1120.1667

Link, J.S., Ihde, T.F., Harvey, C.J., Gaichas, S.K., Field, J.C., Brodziak, J.K.T., Townsend, H.M., Peterman, R.M. (2012) Dealing with uncertainty in ecosystem models: The paradox of use for living marine resource management. *Progress in Oceanography*, 102: 102-114. https://doi.org/10.1016/j.pocean.2012.03.008

Marchau, V.A.W.J., Walker, W.E., Bloemen, P.J.T.M., Popper, S.W. (2019a) *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing, Cham, Switzerland: 405. 10.1007/978-3-030-05252-2_1

Marchau, V.A.W.J., Walker, W.E., Bloemen, P.J.T.M., Popper, S.W. (2019b) Introduction. In: V.A.W.J. Marchau, W.E. Walker, P.J.T.M. Bloemen & S.W. Popper (Eds). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing, Cham: 1-20. 10.1007/978-3-030-05252-2_1

Marescot, L., Chapron, G., Chadès, I., Fackler, P.L., Duchamp, C., Marboutin, E., Gimenez, O. (2013) Complex decisions made simple: a primer on stochastic dynamic programming. *Methods in Ecology and Evolution*, 4(9): 872-884. 10.1111/2041-210X.12082

Markowitz, H. (1952) Portfolio selection. *The journal of finance*, 7(1): 77-91. doi:10.1111/j.1540-6261.1952.tb01525.x

Maroto, J.M., Moran, M., Sandal, L.K., Steinshamn, S.I. (2012) Potential collapse in fisheries with increasing returns and stock-dependent costs. *Marine Resource Economics*, 27(1): 43-63. 10.5950/0738-1360-27.1.43

Martin, J., Runge, M.C., Nichols, J.D., Lubow, B.C., Kendall, W.L. (2009) Structured decision making as a conceptual framework to identify thresholds for conservation and management. *Ecological Applications*, 19(5): 1079-1090. 10.1890/08-0255.1

Martin, T.G., Burgman, M.A., Fidler, F., Kuhnert, P.M., Low-Choy, S., McBride, M., Mengersen, K. (2012) Eliciting expert knowledge in conservation science. *Conservation Biology*, 26: 29-38.

Matrosov, E.S., Woods, A.M., Harou, J.J. (2013) Robust Decision Making and Info-Gap Decision Theory for water resource system planning. *Journal of Hydrology*, 494: 43-58. https://doi.org/10.1016/j.jhydrol.2013.03.006

McBride, M.F., Garnett, S.T., Szabo, J.K., Burbidge, A.H., Butchart, S.H.M., Christidis, L., Dutson, G., Ford, H.A., Loyn, R.H., Watson, D.M., Burgman, M.A. (2012) Structured elicitation of expert judgments for threatened species assessment: a case study on a continental scale using email. *Methods in Ecology and Evolution*, 3(5): 906-920. 10.1111/j.2041-210X.2012.00221.x

McPhail, C., Maier, H.R., Kwakkel, J.H., Giuliani, M., Castelletti, A., Westra, S. (2018) Robustness metrics: How are they calculated, when should they be used and why do they give different results? *Earth's Future*, 6(2): 169-191. 10.1002/2017ef000649

Means, E.I., Laugier, M., Daw, J., Kaatz, L., Waage, M. (2010) Decision Support Planning Methods: Incorporating Climate Change Uncertainties into Water Planning *Water Utility Climate Alliance White Paper*: 102.

Melbourne-Thomas, J., Wotherspoon, S., Raymond, B., Constable, A. (2012) Comprehensive evaluation of model uncertainty in qualitative network analyses. *Ecological Monographs*, 82(4): 505-519. 10.1890/12-0207.1

Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S.E., Ungar, L., Bishop, M.M., Horowitz, M., Merkle, E., Tetlock, P. (2015a) The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of Experimental Psychology: Applied*, 21(1): 1-14. 10.1037/xap0000040

Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., Chen, E., Baker, J., Hou, Y., Horowitz, M., Ungar, L., Tetlock, P. (2015b) Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspectives on Psychological Science*, 10(3): 267-281. 10.1177/1745691615577794

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S.E., Moore, D., Atanasov, P., Swift, S.A., Murray, T., Stone, E., Tetlock, P.E. (2014) Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science*, 25(5): 1106-1115. 10.1177/0956797614524255

Mellers, B.A., McGraw, A.P. (1999) How to improve Bayesian reasoning: Comment on Gigerenzer and Hoffrage (1995). *Psychological Review*, 106(2): 417-424. 10.1037/0033-295X.106.2.417

Meyer, M.A., Booker, J.M. (2001) *Eliciting and analyzing expert judgment: a practical guide*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, USA: 491.

Molina-Perez, E. (2016) Directed international technological change and climate policy: New methods for identifying robust policies under conditions of deep uncertainty: 175. https://www.rand.org/pubs/rgs_dissertations/RGSD369.html

Morgan, M.G., Dowlatabadi, H., Henrion, M., Keith, D., Lempert, R., McBrid, S., Small, M., Wilbanks, T. (2009) Best practice approaches for characterizing, communicating, and incorporating scientific uncertainty in decisionmaking: 96.

Morris, M.D. (1991) Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2): 161-174. 10.1080/00401706.1991.10484804

Moss, R., Schneider, S.H. (2000) Uncertainties in the IPCC TAR: Recommendations to lead authors for more consistent assessment and reporting. In: R. Pachauri, T. Taniguchi & K. Tanaka (Eds). *Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC*. Intergovernmental Panel on Climate Change, Geneva, Switzerland: 33-51.

NASA (2010) Risk-Informed Decision Making Handbook. *NASA/SP-2010-576 Version 1.0*: 128.

NASA (2016) Standard for Models and Simulations. *NASA Technical Standard NASA-STD-7009A*: 72.

National Environmental Justice Advisory Council (2004) Ensuring risk reduction in communities with multiple stressors: Environmental justice and cumulative risks/impacts https://www.epa.gov/sites/production/files/2015-02/documents/nejac-cum-risk-rpt-122104.pdf

National Research Council (1982) *Risk and Decision Making: Perspectives and Research*. The National Academies Press, Washington, DC: 75. doi:10.17226/775

National Research Council (2009) Science and Decisions: Advancing Risk Assessment: 403.

Natskår, A., Moan, T., Alvær, P.Ø. (2015) Uncertainty in forecasted environmental conditions for reliability analyses of marine operations. *Ocean Engineering*, 108: 636-647. https://doi.org/10.1016/j.oceaneng.2015.08.034

Nau, R.F. (2002) The aggregation of imprecise probabilities. *Journal of Statistical Planning and Inference*, 105(1): 265-282. https://doi.org/10.1016/S0378-3758(01)00213-0

O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakow, T. (2006) *Uncertain judgements: eliciting experts' probabilities*. John Wiley & Sons.

O'Leary, R.A., Choy, S.L., Murray, J.V., Kynn, M., Denham, R., Martin, T.G., Mengersen, K. (2009) Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby Petrogale penicillata. *Environmetrics*, 20(4): 379-398. 10.1002/env.935

Office of the Prime Minister’s Chief Science Advisor (2016a) Making decisions in the face of uncertainty: Understanding risk. Part 1: 38.

Office of the Prime Minister’s Chief Science Advisor (2016b) Making decisions in the face of uncertainty: Understanding risk. Part 2: 27.

OIE (2019) Glossary. *Terrestrial Animal Health Code*. World Organization for Animal Health (OIE), Paris, France: i-x.

Oil & Gas UK (2014) Guidance on Risk Related Decision Making. *Issue 2*: 25.

Patt, A., Dessai, S. (2005) Communicating uncertainty: lessons learned and suggestions for climate change assessment. *Comptes Rendus Geoscience*, 337(4): 425-441. https://doi.org/10.1016/j.crte.2004.10.004

Patt, A.G., Schrag, D.P. (2003) Using specific language to describe risk and probability. *Climatic Change*, 61(1): 17-30. 10.1023/A:1026314523443

Plagányi, É.E., Punt, A.E., Hillary, R., Morello, E.B., Thébaud, O., Hutton, T., Pillans, R.D., Thorson, J.T., Fulton, E.A., Smith, A.D.M., Smith, F., Bayliss, P., Haywood, M., Lyne, V., Rothlisberg, P.C. (2014) Multispecies fisheries management and conservation: tactical applications using models of intermediate complexity. *Fish and Fisheries*, 15(1): 1-22. 10.1111/j.1467-2979.2012.00488.x

Pollino, C.A., Woodberry, O., Nicholson, A., Korb, K., Hart, B.T. (2007) Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling & Software*, 22(8): 1140-1152. https://doi.org/10.1016/j.envsoft.2006.03.006

Purdy, G. (2010) ISO 31000:2009—Setting a New Standard for Risk Management. *Risk Analysis*, 30(6): 881-886. 10.1111/j.1539-6924.2010.01442.x

Puterman, M.L. (2008) *Markov Decision Processes: Discrete Stochastic Programming*. John Wiley & Sons Inc., New York: 649.

Ranjan, R., Gneiting, T. (2010) Combining probability forecasts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1): 71-91. 10.1111/j.1467-9868.2009.00726.x

Reed, P.M., Hadka, D., Herman, J.D., Kasprzyk, J.R., Kollat, J.B. (2013) Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in Water Resources*, 51: 438-456. https://doi.org/10.1016/j.advwatres.2012.01.005

Regan, H.M., Colyvan, M., Burgman, M.A. (2002) A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications*, 12(2): 618-628. 10.1890/1051-0761(2002)012[0618:ATATOU]2.0.CO;2

Regan, H.M., Colyvan, M., Markovchick-Nicholls, L. (2006) A formal model for consensus and negotiation in environmental management. *Journal of Environmental Management*, 80(2): 167-176. https://doi.org/10.1016/j.jenvman.2005.09.004

Renn, O. (2008) White Paper on Risk Governance: Toward an Integrative Framework. In: O. Renn & K.D. Walker (Eds). *Global Risk Governance: Concept and Practice Using the IRGC Framework*. Springer Netherlands, Dordrecht: 3-73. 10.1007/978-1-4020-6799-0_1

Renooij, S. (2001) Probability elicitation for belief networks: issues to consider. *The Knowledge Engineering Review*, 16(03): 255-269. doi:10.1017/S0269888901000145

Renooij, S., Witteman, C. (1999) Talking probabilities: communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning*, 22(3): 169-194. https://doi.org/10.1016/S0888-613X(99)00027-4

Rosa, E.A. (2003) The logical structure of the social amplification of risk framework (SARF); metatheoretical foundations and policy implications. In: N. Pidgeon, R.E. Kasperson & P. Slovic (Eds). *The social amplification of risk*. Cambridge University Press, Cambridge, U.K.: 47-79.

Ruszczyński, A. (2010) Risk-averse dynamic programming for Markov decision processes. *Mathematical Programming*, 125(2): 235-261. 10.1007/s10107-010-0393-3

Saltelli, A. (2002) Sensitivity analysis for importance assessment. *Risk Analysis*, 22(3): 579-590. 10.1111/0272-4332.00040

Satopää, V.A., Baron, J., Foster, D.P., Mellers, B.A., Tetlock, P.E., Ungar, L.H. (2014) Combining multiple probability predictions using a simple logit model. *International Journal of Forecasting*, 30(2): 344-356. http://dx.doi.org/10.1016/j.ijforecast.2013.09.009

Savage, L.J. (1951) The theory of statistical decision. *Journal of the American Statistical Association*, 46(253): 55-67. 10.1080/01621459.1951.10500768

Schaefer, M.B. (1954) Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Inter-American Tropical Tuna Commission Bulletin*, 1(2): 23-56.

Schneller, G.O., Sphicas, G.P. (1983) Decision making under uncertainty: Starr's Domain criterion. *Theory and Decision*, 15(4): 321-336. 10.1007/BF00162111

Schoemaker, P.J., Tetlock, P.E. (2016) Superforecasting: How to upgrade your company's judgment. *Harvard Business Review*, 94: 72-78.

Schweder, T., Hjort, N.L. (2016) *Confidence, Likelihood, Probability*. Cambridge University Press, Cambridge, United Kingdom: 500.

Selkoe, K.A., Blenckner, T., Caldwell, M.R., Crowder, L.B., Erickson, A.L., Essington, T.E., Estes, J.A., Fujita, R.M., Halpern, B.S., Hunsicker, M.E., Kappel, C.V., Kelly, R.P., Kittinger, J.N., Levin, P.S., Lynham, J.M., Mach, M.E., Martone, R.G., Mease, L.A., Salomon, A.K., Samhouri, J.F., Scarborough, C., Stier, A.C., White, C., Zedler, J. (2015) Principles for managing marine ecosystems prone to tipping points. *Ecosystem Health and Sustainability*, 1(5): 1-18. 10.1890/EHS14-0024.1

Sexton, K. (2012) Cumulative risk assessment: An overview of methodological approaches for evaluating combined health effects from exposure to multiple environmental stressors. *International Journal of Environmental Research and Public Health*, 9(2): 370-390. https://www.mdpi.com/1660-4601/9/2/370

Shortridge, J.E., Guikema, S.D. (2016) Scenario discovery with multiple criteria: An evaluation of the robust decision-making framework for climate change adaptation. *Risk Analysis*, 36(12): 2298-2312. 10.1111/risa.12582

Simon, P., de Laplace, M., Truscott, F. (1951) *A philosophical essay on probabilities*. Dover Publications New York.

Slovic, P. (1999) Trust, emotion, sex, politics, and science: surveying the risk-assessment battlefield. *Risk Analysis*, 19(4): 689-701. 10.1111/j.1539-6924.1999.tb00439.x

Slovic, P., Monahan, J., MacGregor, D.G. (2000) Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, 24(3): 271-296. 10.1023/A:1005595519944

Smit, B., Spaling, H. (1995) Methods for cumulative effects assessment. *Environmental impact assessment review*, 15(1): 81-106. http://dx.doi.org/10.1016/0195-9255(94)00027-X

Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., Burgman, M. (2010) Reducing overconfidence in the interval judgments of experts. *Risk Analysis*, 30(3): 512-523. DOI: 10.1111/j.1539-6924.2009.01337.x

Standards Australia, Standards New Zealand (2009) Joint Australian New Zealand International Standard - Risk management - Principles and guideline. Standards Australia / Standards New Zealand,, New Zealand: 24.

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

Stelzenmüller, V., Coll, M., Mazaris, A.D., Giakoumi, S., Katsanevakis, S., Portman, M.E., Degen, R., Mackelworth, P., Gimpel, A., Albano, P.G., Almpanidou, V., Claudet, J., Essl, F., Evagelopoulos, T., Heymans, J.J., Genov, T., Kark, S., Micheli, F., Pennino, M.G., Rilov, G., Rumes, B., Steenbeek, J., Ojaveer, H. (2018) A risk-based approach to cumulative effect assessments for marine management. *Science of the Total Environment*, 612: 1132-1140. https://doi.org/10.1016/j.scitotenv.2017.08.289

Suding, K.N., Hobbs, R.J. (2009) Threshold models in restoration and conservation: a developing framework. *Trends in Ecology & Evolution*, 24(5): 271-279. https://doi.org/10.1016/j.tree.2008.11.012

Sutherland, W.J., Burgman, M.A. (2015) Use experts wisely. *Nature*, 526: 317-318. doi:10.1038/526317a

Taleb, N.N. (2007) *The Black Swan - The Impact of the Highly Improbable*. Random House Publishing, United Kingdom.

Tetlock, P., Gardner, D. (2016) *Superforecasting - The Art and Science of Prediction*. Random House Books, London, U.K.: 340.

Trutnevyte, E., Guivarch, C., Lempert, R., Strachan, N. (2016) Reinvigorating the scenario technique to expand uncertainty consideration. *Climatic Change*, 135(3): 373-379. 10.1007/s10584-015-1585-x

Tucker, W.T., Ferson, S. (2003) Probability bounds analysis in environmental risk assessments 63. https://www.researchgate.net/profile/Scott_Ferson/publication/240192153_Probability_Bounds_Analysis_in_Environmental_Risk_Assessment/links/0deec531998e77ba79000000/Probability-Bounds-Analysis-in-Environmental-Risk-Assessment.pdf

Tversky, A., Kahneman, D. (1973) Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2): 207-232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., Kahneman, D. (1975) Judgment under uncertainty: Heuristics and biases. *Utility, probability, and human decision making*. Springer: 141-162.

Tversky, A., Kahneman, D. (1982) Judgement under uncertainty - heuristics and biases. In: D. Kahneman, P. Slovic & A. Tversky (Eds). *Judgement Under Uncertainty - Heuristics and biases*. Cambridge University Press, Cambridge, United Kingdom: 3-20.

U.S. Environmental Protection Agency (1992) Framework For Ecological Risk Assessment. *EPA/630/R-92/001* 41.

U.S. Environmental Protection Agency (2003) Framework for Cumulative Risk Assessment. *EPA/630/P-02/001F*: 109.

Voudouris, V., Matsumoto, K.i., Sedgwick, J., Rigby, R., Stasinopoulos, D., Jefferson, M. (2014) Exploring the production of natural gas through the lenses of the ACEGES model. *Energy Policy*, 64: 124-133. https://doi.org/10.1016/j.enpol.2013.08.053

Wald, A. (1950) *Statistical decision functions*. Wiley, Oxford, England: ix, 179-ix, 179.

Walker, W., Haasnoot, M., Kwakkel, J. (2013a) Adapt or perish: A review of planning approaches for adaptation under deep uncertainty. *Sustainability*, 5(3): 955. http://www.mdpi.com/2071-1050/5/3/955

Walker, W.E., Harremoës, P., Rotmans, J., van der Sluijs, J.P., van Asselt, M.B.A., Janssen, P., Krayer von Krauss, M.P. (2003) Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1): 5-17. 10.1076/iaij.4.1.5.16466

Walker, W.E., Lempert, R.J., Kwakkel, J.H. (2013b) Deep Uncertainty. In: S.I. Gass & M.C. Fu (Eds). *Encyclopedia of Operations Research and Management Science*. Springer US, Boston, MA: 395-402. 10.1007/978-1-4419-1153-7_1140

Walker, W.E., Marchau, V.A.W.J., Kwakkel, J.H. (2019) Dynamic Adaptive Planning (DAP). In: V.A.W.J. Marchau, W.E. Walker, P.J.T.M. Bloemen & S.W. Popper (Eds). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing, Cham: 52-69. 10.1007/978-3-030-05252-2_2

Walker, W.E., Rahman, S.A., Cave, J. (2001) Adaptive policies, policy analysis, and policy-making. *European Journal of Operational Research*, 128(2): 282-289. https://doi.org/10.1016/S0377-2217(00)00071-0

Walley, P. (1990) *Statistical Reasoning with Imprecise Probabilities* Chapman & Hall/ CRC: 720.

Wallsten, T.S., Budescu, D.V., Zwick, R., Kemp, S.M. (1993) Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society*, 31(2): 135-138. 10.3758/BF03334162

Wallsten, T.S., Fillenbaum, S., Cox, J.A. (1986) Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language*, 25(5): 571-587. https://doi.org/10.1016/0749-596X(86)90012-4

Walters, C.J. (1975) Optimal harvest strategies for salmon in relation to environmental variability and uncertain production parameters. *Journal of the Fisheries Research Board of Canada*, 32(10): 1777-1784. 10.1139/f75-211

Watson, A.A., Kasprzyk, J.R. (2017) Incorporating deeply uncertain factors into the many objective search process. *Environmental Modelling & Software*, 89: 159-171. https://doi.org/10.1016/j.envsoft.2016.12.001

Weitzman, M.L. (2002) Landing fees vs harvest quotas with uncertain fish stocks. *Journal of Environmental Economics and Management*, 43(2): 325-338. https://doi.org/10.1006/jeem.2000.1181

Williamson, R.C., Downs, T. (1990) Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*, 4(2): 89-158. https://doi.org/10.1016/0888-613X(90)90022-T

Wintle, B.C., Fraser, H., Wills, B.C., Nicholson, A.E., Fidler, F. (2019) Verbal probabilities: Very likely to be somewhat more confusing than numbers. *PLoS ONE*, 14(4): e0213522. 10.1371/journal.pone.0213522

Woodward, M., Kapelan, Z., Gouldby, B. (2014) Adaptive Flood Risk Management Under Climate Change Uncertainty Using Real Options and Optimization. *Risk Analysis*, 34(1): 75-92. doi:10.1111/risa.12088

Zio, E., Pedroni, N. (2012a) Overview of risk-informed decision-making processes. *Number 2012-10 of the Cahiers de la Sécurité Industrielle*: 48.

Zio, E., Pedroni, N. (2012b) Uncertainty characterization in risk analysis for decision-making practice. *Number 2012-07 of the Cahiers de la Sécurité Industrielle*: 39.