JNCC Report
No: 490

Review of Case Studies and Recommendations for the Inclusion of Expert Judgement in Marine Biodiversity Status Assessments

Barnard, S. & Boyes, S.J.

August 2013

ISSN 0963 8901

For further information please contact:

Joint Nature Conservation Committee
Monkstone House
City Road
Peterborough PE1 1JY
www.jncc.defra.gov.uk

This report should be cited as:

Barnard, S. & Boyes, S.J. (2013) Review of Case Studies and Recommendations for the Inclusion of Expert Judgement in Marine Biodiversity Status Assessments - A report for the Joint Nature Conservation Committee, JNCC Report No.490 by the Institute of Estuarine and Coastal Studies, University of Hull.

UNIVERSITY OF Hull
Institute of Estuarine and Coastal Studies

Summary

Assessments of the status of marine biodiversity in the UK can be made at a range of geographic scales and can focus on different aspects of biodiversity. For example, geographic scales may vary from site-specific (such as a Marine Conservation Zone reporting under the Marine and Coastal Access Act 2009) to the full extent of the North-East Atlantic maritime area (for reporting under the 1992 OSPAR Convention). Similarly, the particular aspects of biodiversity that are reported on can range from specific features of conservation importance within a designated Marine Conservation Zone (e.g. Ocean quahog, a type of shellfish) to 'all biological diversity' (e.g. under the 1993 Convention on Biological Diversity).

In recent years there has been an increasing shift in conservation and natural resource management away from a focus on single species, to an emphasis on ecosystem-level management. Therefore assessments of ecosystem condition are increasingly needed to support evaluation of management effectiveness, and to inform resource-management decisions. While these ecosystem-level policy goals reflect a growing scientific consensus on the importance of multi-species interactions and system-level processes in the marine environment, there is no simple direct measurement that can be made to evaluate the condition of a whole ecosystem, although a variety of types of evidence can be drawn upon to help inform an assessment.

In response to the demand for ecosystem status evaluations, expert judgement which follows formal systematic processes can be applied to obtain quantitative assessments of scientific phenomena and thereby marine biodiversity status. Such 'expert judgement', in its formal definition (US EPA, 2009), necessarily excludes personal or social values and preferences.

This expert judgement often takes the form of a *vulnerability assessment* where information on a benthic habitat's sensitivity to known pressures is combined with evidence of exposure to activities exerting those pressures in order to derive an indication of the benthic habitat's vulnerability. In the UK this approach has been used to support the development of *conservation objectives* (for example for offshore Special Areas of Conservation or for recommended Marine Conservation Zones; Natural England & JNCC, 2011) and to inform assessments of benthic habitats (for example Charting Progress 2; UK Marine Monitoring and Assessment Strategy, 2010).

The use of expert judgement should not preclude the delivery of robust and transparent status assessments for marine biodiversity, although standards and guidelines are required alongside more-explicit descriptions of uncertainty to allow policy makers, scientists and stakeholders to both understand and have confidence in the reliability and integrity of marine biodiversity status assessments.

Recent marine biodiversity assessments – including those underpinning Charting Progress 2 (UK Marine Monitoring and Assessment Strategy, 2010), assessments of offshore Annex I reef and sandbank features for Article 17 reporting under the Habitats Directive (Vaughan *et al*., in prep.) and for the 2010 OSPAR Quality Status Report (OSPAR, 2010) – have demonstrated a number of limitations with the present methods for undertaking such assessments. In turn this has highlighted the need for a more robust and efficient assessment process which is centred on data, regularly repeated, and harmonised with, and responsive to, emerging requirements. However, whilst work is underway to address these points, the next cycle of reporting will (particularly with reference to status assessments of benthic habitats under the 2008 Marine Strategy Framework Directive) continue to rely

strongly on expert judgement, alongside limited monitoring data and information on the distribution of benthic habitats and pressures.

The use of expert judgement to support decision making is not unusual and it has always played a large role in science and engineering. Increasingly expert judgement is being recognised as just another type of scientific data (Goossens *et al*., 2008), and it can be argued that the use of expert judgement in relation to technical problems is not only unavoidable, but is also desirable (Keeny & von Winterfeldt, 1989). Even a cursory glance at the subject areas shows that, aside from conservation and natural resource management, expert judgement is used across a wide range of fields including, for example, nuclear safety (Thorne & Williams, 1992; Simola *et al.*, 2005); the effects of nanoparticles ($PM_{2.5}$; particulate matter 2.5μm in diameter and smaller) on public health (Kandlikar *et al.*, 2007; Roman *et al.*, 2008); aircraft engineering (Peng *et al*., 2011) and air traffic control (Nunes & Kirlik, 2005); drug legalisation (MacCoun, 2011); economics (Braun & Yaniv, 1992); pharmacoeconomics (Evans & Crawford, 2000); and software production (Jorgensen, 2004).

This report examines the potential use of expert judgement as a tool or approach within marine biodiversity status assessments. Initially, four studies from outside of the field of biodiversity assessment are reviewed, with the intention of demonstrating the potential value of expert judgement and to identify any key learning points that may subsequently be applicable to studies relating to biodiversity assessments. These four studies cover: environmental health impact assessment (Knol *et al*., 2010); the risk assessment of nanotechnology-enabled food products (Flari *et al*., 2011); the response of the Atlantic meridional overturning circulation to climate change (Zickfeld *et al.,* 2007); and structural reliability relating to engineering in the nuclear energy sector (Simola *et al.*, 2005). These particular four studies have been selected as they are relatively recent, relate to a number of different disciplines and make use of different underlying methodologies. From this brief review, a number of learning points are identified. These include, but are not limited to, suggestions that:

- there should be a transparent and structured process to help plan and guide elicitation work, and to provide an audit trail of decision-making in the light of expert opinion;
- the selection of experts is important and should include an appropriately wide range of views and expertise including not only technical specialists, but also practitioners and stakeholders, that is, those with a broader perspective;
- opinions from different experts can be weighted; and
- techniques other than simple requests for quantitative data can be used to help capture information.

Later in the report, a series of thirteen case studies that involve the use of expert judgement in biodiversity assessments (within either terrestrial or marine environments) are critically reviewed, and the strengths, weaknesses, opportunities and threats of each case study (SWOT analysis) is appraised. The selected studies cover a range of subjects and geographic areas (e.g. assessment of threats to coral reef systems at the global scale; assessment of grassland status in the USA-Mexico borderlands; identification and prioritisation of terrestrial conservation areas in the Cape Floristic Region of South Africa; and comparisons of coastal benthic macrofauna community condition assessment between the USA and Europe).

A number of elements from these studies are identified as potentially being applicable to the development of an appropriate framework or methodology for UK marine biodiversity assessments. From these, a series of 11 recommendations outlining good practice for the use of expert judgement are set out, and specific methods and potential approaches are briefly discussed. These recommendations cover the selection of experts (e.g. studies should seek to engage with at least ten relevant experts from the governmental bodies, academia and the private sector); the detail of the elicitation process (e.g. a series guidelines for consideration when developing a process that will employ expert judgement, and the application of post-hoc analyses to check for bias in experts' responses); and methods for recording judgement (e.g. the use of interactive reporting systems in workshop environments, and the application of novel paradigms (such as 'fuzzy logic') to assist in assigning and recording the degree of uncertainty in judgements).

These recommendations are set into a generic framework for applying expert judgement to marine benthic habitat vulnerability assessments. A five stage process to assess the vulnerability (and likely condition) of benthic habitats, and the contribution of expert judgement methods to each stage of this process is discussed. The role that expert judgement can play in benthic habitat vulnerability assessments is clearly demonstrated and the effective deployment of a range of approaches and techniques (including elements of good practice identified and discussed earlier in the report) is outlined.

Contents

1 Introduction

1.1 Expert Judgement and Decision Making

Expert judgement (also referred to, *inter alia*, as expert elicitation, best professional judgement or best professional opinion) emerged from the growing field of decision theory in the 1950s as a technique for quantifying uncertainty and estimating otherwise unobtainable values to support analytical decision making (US EPA, 2009). However, both professionals and non-professionals can provide expert judgement, and the term 'expert judgement' is the preferred term used within this report.

A wide range of activities may fall under the term 'expert judgement'; but the US EPA White Paper on the use of expert elicitation helpfully restricts the term to '*a formal systematic process to obtain quantitative judgements on scientific questions (to the exclusion of personal or social values and preferences)*'. In acknowledgment of the fact that it is often the process of elicitation of judgement that is of interest, the term 'elicitation of expert judgement' has also been adopted within this report. Expert judgement is however, intended to be synonymous with the alternative terms (as above) that are used in the quoted literature.

Decision making, as a response to a given question or problem, forms part of a process loop linking it to problem solving (see Figure 1). Typically, inputs are used to inform a problem solving process which provides one or more possible solutions to the problem or question. Decision making subsequently identifies a solution or answer to carry forward as an output. Additionally, a feedback 'review' loop may be applied to ensure that the selected solution provides an adequate or appropriate solution to the problem. Expert judgement can be applied at both the problem solving and decision making levels of this linked process and, rather than restrict the consideration of expert judgement solely to decision making in its narrowest sense, it is at both of these levels that its use is being considered in this report.

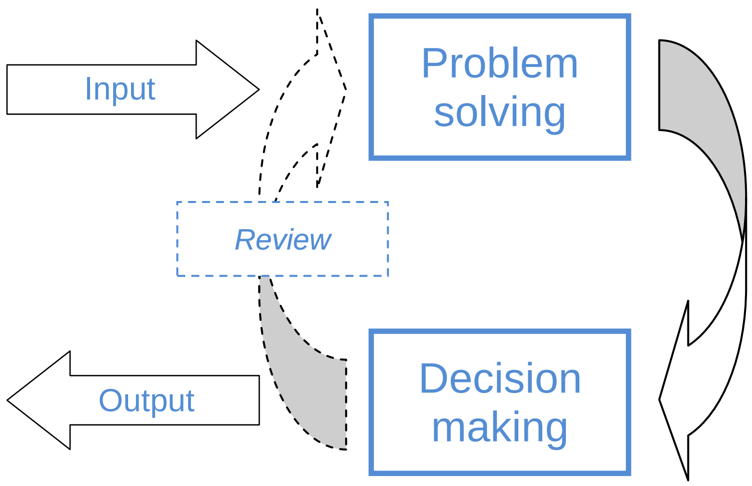


Figure1. Answering questions: problem solving and decision making

1.2 Objectives

The Joint Nature Conservation Committee (JNCC) commissioned the Institute of Estuarine and Coastal Studies (IECS) at the University of Hull to review and evaluate case studies that make use of expert judgement and to recommend good practice, including key principles and methods that should be adopted when including expert judgement in marine biodiversity status assessments. This work was undertaken through a series of investigations culminating in this final report. The objectives of this report are to:

1. Examine the use of expert judgement to support decision making processes;
2. Critically review and assess the strengths and weaknesses of selected case studies that have used expert judgement in biodiversity assessments;
3. Provide recommendations on good practice when using expert judgement in biodiversity assessments; and
4. Test the application of these good practice recommendations in a benthic habitat vulnerability assessment.

2 Application of Expert Judgement Across Disciplines

2.1 Overview of Case Studies

Guidelines developed by the Centre for Evidence-Based Conservation at Bangor University (CEBC, 2009) were used to conduct literature searches (for more details of this approach see Appendix 1). A literature search on the ScienceDirect website (www.sciencedirect.com) – searching any field for the terms 'expert' and 'elicitation' or 'judgement' or 'judgment', and with the results limited to articles published in journals since 2002 – produced over 42,000 matches, clearly demonstrating that this is a methodology that is widely applied.

Expert judgement, where the elicited judgements of more than one expert are brought together, has always played a large role in science and engineering and, increasingly, is being recognised as just another type of scientific data (Goossens *et al*., 2008). It can be argued that the use of expert judgement in relation to technical problems is not only unavoidable, but is also desirable (e.g. Keeny & von Winterfeldt, 1989). Keeny and von Winterfeldt suggest that experts are sought to work on complex problems precisely because of their expertise, not because they are able to avoid the use of judgement. However, it should be considered from the outset that important insight and valuable opinion can come from individuals who may not be immediately identified as 'experts', such as local practitioners and informed stakeholders (Burgman *et al.*, 2011a). This contribution has been recognised across a number of studies including, for example, Oliver *et al.* (2012) who described the use of local knowledge from dairy farmers in the development of a decision support system to determine the impact of farming practices on environmental quality in the Taw catchment (Devon, UK) valley, and the Net Gain Marine Conservation Zone (MCZ) project which made use of 'expert judgement' from informed stakeholders to contribute to the quality assurance of initial vulnerability assessments made for habitat and species Features of Conservation Importance (FOCI) within recommended MCZs (Net Gain, 2011).

Even a cursory glance at the subject areas shows that, aside from conservation and natural resource management, expert judgement is used across a wide range of fields including, for example, nuclear safety (Thorne & Williams, 1992; Simola *et al.*, 2005); the effects of nanoparticles ($PM_{2.5}$; particulate matter 2.5μm in diameter and smaller) on public health (Kandlikar *et al.*, 2007; Roman *et al.*, 2008); aircraft engineering (Peng *et al*. 2011) and air traffic control (Nunes & Kirlik, 2005); drug legalisation (MacCoun, 2011); economics (Braun & Yaniv, 1992); pharmacoeconomics (Evans & Crawford, 2000); and software production (Jorgensen, 2004). A number of such studies, incorporating the use of expert judgement, are listed within Appendix 2, demonstrating its application across a number of wide-ranging fields. From this list, four references were selected for review to demonstrate and explore the use of expert judgement across a number of differing fields:

1. *Application of formal expert judgement to the evaluation of structural reliability* (Simola *et al.,* 2005) working in the field of engineering/nuclear safety.
2. *Expert judgements on the response of the Atlantic meridional overturning circulation to climate change* (Zickfeld *et al.,* 2007);
3. *The use of expert elicitation in environmental health impact assessment: a seven step procedure* (Knol *et al*., 2010);
4. *Expert judgement based multi-criteria decision model to address uncertainties in risk assessment of nanotechnology-enabled food products* (Flari *et al*., 2011);

These four studies were chosen as they are relatively recent, they each related to different disciplines and each made use of different methodologies. Each study was reviewed, looking at why expert judgement was used, the inherent benefits and limitations of its application, how it affected or influenced decision making and its potential application to marine biodiversity assessments. A summary of the findings are presented in Table 1 with the full supporting discussion presented in Appendix 3.

Table 1. Summary of key points from selected studies

	Engineering/nuclear safety (Simola *et al.*, 2005)	**Effects of climate change on Atlantic currents (Zickfeld *et al.*, 2007)**	**Environmental health impact assessment (Knol *et al.*, 2010)**	**Risk assessment of nanotechnology-enabled food products (Flari *et al.*, 2011)**
Why was expert judgement used?	Limitations and inconsistencies in the underlying knowledge base. There are several degradation mechanisms for safety-critical components in the nuclear industry for which no validated structural reliability tools are available.	Limitations and inconsistencies in the underlying knowledge base. There is no single method for a definitive prediction, and expert consensus allows these lines of evidence to be drawn together.	Limitations and inconsistencies in the underlying knowledge base. Limited available knowledge can be presented and used to inform policy development before conclusive scientific evidence becomes available.	Limitations in the underlying knowledge base regarding the interactions of nanoparticles and potential human health impacts, coupled with high levels of uncertainty.
What were the benefits that were realised through using expert judgement?	Structured expert judgement seen as the optimum means of obtaining numerical (quantitative) estimates for structural reliability issues.	Individual expert judgement allowed synthesis of several disparate lines of evidence and quantification of uncertainty. Expert judgement allowed the assessment of quantitative estimates for a number of parameters (relating to factors that may influence the Atlantic currents) which would not otherwise be easily derived.	Outputs could help prioritise research agendas – this information would not otherwise be available and research agendas would remain uninformed. Expert judgement can help gain insight on issues that might otherwise be ignored.	Levels of uncertainty can only be effectively assessed by expert judgement. Participants at a subsequent workshop agreed that the approach was of significant value in aiding the assessment of safety regarding nanotechnology-enabled food products.
What limitations were apparent due to using expert judgement?	It was not possible to completely fulfil all of the (nine) requirements that were postulated as underpinning expert selection and the study used a relatively small group of experts (four). Subsequently, even though participating experts may be familiar with most of the concepts involved, individuals may lack specific knowledge in key areas.	Adopted process seen as being lengthy with high requirement for manpower resource.	Potential cost implications.	The work was extensive, involving a number of ‘successive, lengthy steps’. Quality of judgement (arising from weakness in methods used to elicit information from experts) may affect quality of subsequent decision-making tools. Incomplete knowledge regarding nanotechnology-enabled products limited effective elicitation (and application) of expert knowledge.

	Engineering/nuclear safety (Simola *et al.*, 2005)	**Effects of climate change on Atlantic currents (Zickfeld *et al.*, 2007)**	**Environmental health impact assessment (Knol *et al.*, 2010)**	**Risk assessment of nanotechnology-enabled food products (Flari *et al.*, 2011)**
How did the use or application of expert judgement affect or influence decision making?	Differences in opinion across the experts were noted, but there was agreement on some major themes directing subsequent modelling approaches and providing consensus on where leaks or failures would be possible.	Allowed dominant physical processes responsible for determining Atlantic current strength to be identified, so facilitating an informed approach regarding subsequent research priorities.	Expert elicitation provided useful means of gaining insight where information was limited or inconclusive. Provided a relatively quick and inexpensive substitute for time- or money-consuming research.	Subsequent use of a multi-criteria decision modelling (MCDM) approach allowed the development of decision support tool for assessing the potential risk of novel products based on their profile against 10 defined characteristics.
What lessons are applicable to, or might inform, the use of expert judgement in marine biodiversity assessment?	Overall, the process should be reliable, transparent and robust. As high levels of uncertainty are typical, the process should allow experts to state their true opinions without influence or the risk of being pre-judged by the analyst/assessor.	Graphical outputs used to show degree of (dis)agreement between experts. 'Poker chips' approach used to elicit judgements (see section 2.24 herein). Use of alternative elicitation methods to check consistency in judgements. Definition of baseline terms important at outset in order to reduce ambiguity – can be done as preliminary exercise. Briefing/training material useful in advance of elicitation sessions; experts encouraged to make use of material (notes, papers, etc.) during elicitation sessions.	Group elicitation considered to counter tendency of different experts of responding to differing understanding (interpretation) of questions. Use of structured process for eliciting expert judgement. Experts selected from range of disciplines to help stimulate discussion (also noted possibility of restricting participation to one expert per organisation). Importance of assessing consistency is highlighted. Under-specificity (allowing excessive room for individual interpretation of questions or scenarios) should be addressed by ensuring adequate detail is provided to all experts involved. Briefing/training material useful in advance of elicitation sessions. Two-stage elicitation process may be useful – with an initial judgement elicitation being followed by a debate and then by second judgement elicitation. Consideration given to retaining anonymity of experts.	Use of structured process for eliciting expert judgement. Derived decision-making tools can be used to inform level of precaution that should be adopted (rather than applying blanket 'precautionary principle' approach. Expert judgement elicited remotely via bespoke interactive software (allowing participation from across the global scientific community). Training of experts before elicitation – especially regarding identification and treatment of uncertainty, and its subsequent management. Inclusion of generalists within the expert panels thought to be useful. Briefing/training material useful in advance of elicitation sessions.

2.2 Lessons Learnt

Learning points that can be taken from the four case studies are discussed below, with particular reference to their potential application to marine biodiversity assessments.

2.2.1 General

The case studies reviewed show that the use of expert judgement in relation to technical problems is often unavoidable, and is also sometimes desirable (e.g. Keeny & von Winterfeldt, 1989). Knol *et al*. (2010) noted that a formalised process for employing expert judgement can serve as a way of presenting limited available knowledge in order to inform policies which have to be made before conclusive scientific evidence becomes available. In addition, they noted that the transparency and reproducibility, and most likely also the quality of the elicited information, increases when the expert elicitation is carried out according to a systematic protocol.

2.2.2 Process

Learning point: having a transparent and structured process is important both to help plan and guide elicitation work, and also to provide an audit trail of decision making in the light of expert opinion. A number of authors refer to the use of a predefined structured process for obtaining and applying expert judgement.

Both Simola *et al.* (2005) and Knol *et al.* (2010) present seven step processes for use in eliciting expert judgement. Simola *et al. (2005)*, for example, provide the following:

1. Identification and selection of issues about which judgements of experts should be made;
2. Identification and selection of experts;
3. Training of experts and definition of variables to be elicited;
4. Execution of individual work by experts;
5. Elicitation of expert judgements
6. Analysis and aggregation of results and, in case of disagreement, attempt to resolve differences; and
7. Documentation of results, including expert reasoning in support of their judgement.

Simola *et al.* (2005) suggest that, even though other references might have presented a slightly different list of phases, essentially the underlying procedure is similar. This process is briefly compared to that described in more detail by Knol *et al*. (2010) within Section 4 of this report.

2.2.3 Choice of Experts

Learning point: experts should cover an appropriately wide range of views and expertise, with generalists (individuals with broad experience and knowledge) as well as nominated technical experts, involved.

Regarding the process of expert selection, Knol *et al.* (2010) reported that experts used in the ultrafine particle elicitation (Knol *et al*., 2009) were invited from a range of disciplines specifically to help stimulate interdisciplinary discussion. Similarly, elicitation work on *Campylobacter* transmission (van der Fels-Klerx *et al*., 2005, again summarised by Knol *et al.*, 2010) actively sought to engage experts from different disciplines while also restricting participation to one expert per organisation (a restriction that was also employed in the ultrafine particle elicitation work that they reported on).

Based on experience from their initial case study, Simola *et al.* (2005) made a couple of recommendations regarding expert training, suggesting that it would be beneficial to give examples to demonstrate the identification and treatment of uncertain parameters, as well as provide guidance on the propagation and management of uncertainties. Although referring specifically to the field of structural reliability (engineering) they also highlight the importance of including a generalist alongside 'normative experts' (i.e. individuals with a theoretical and conceptual knowledge of probability and practical experience in the elicitation of judgements from individuals) within the discussions to inform or elicit expert judgement.

2.2.4 **Elicitation of Judgements**

Learning point: elicitation can seek to obtain the same information by different methods to check internal consistency.

Consideration of the design of the elicitation protocol raised several learning points, one of which concerns methods to assess internal consistency. Experts can be asked to make their judgement in two or more different ways (as was done in genetically modified (GM) crop elicitation work; Krayer von Krauss *et al*. 2004, reported by Knol *et al.*, 2010). Where experts are internally consistent, alternative approaches should lead to (roughly) the same outputs. For example, in the study reported by Knol *et al*. (2010) experts were asked to rank the relative sensitivity of the risk assessment conclusion to each of a range of sources of uncertainty; they were first asked to prioritise the uncertainties on a simple 0-1 scale and then subsequently asked to allocate poker chips to uncertainty sources which indicated how much they would be willing to invest to completely eliminate that uncertainty.

Learning point: agreement on definitions of key terms and concepts ahead of the main elicitation exercise can help to clarify what is being asked for and reduce uncertainty and 'noise' in experts' responses.

Baseline definitions may be defined at the start of the elicitation process. For example Zickfeld *et al.* (2007) acknowledge that there may be ambiguity around the meaning of Atlantic Meridional Overturning Circulation (AMOC) and so started interviews by suggesting a definition (which was, fortunately, accepted by all experts for the course of their interviews).

Problems relating to potential under-specificity (which occurs when too few details being provided, allowing too much room for interpretation) were addressed in the *Campylobacter* elicitation work summarised by Knol *et al.* (2010) by purposefully ensuring that adequate detail about broiler-chicken processing was provided to enable each of the experts to refine their judgements accordingly.

Alternatively, rather than specifying (and fixing) the detail at the outset, it is possible to open up the wording of questions for discussion and agreement during the elicitation session. This process may increase common understanding and improve the experts' approval of the questions. Whilst it is particularly appropriate in an interdisciplinary expert elicitation (since semantics or perspectives may differ between disciplines), modification of questions is not recommended for a series of individual elicitation sessions (such as personal interviews) as it would tend to diminish the inter-expert comparability.

Learning point: some degree of training or familiarisation ahead of the elicitation itself may help to reduce uncertainty and improve the quality of information that is provided.

Regarding the preparation of the elicitation session, the organisers of the ultrafine particle expert workshop (as reported in Knol *et al.*, 2010) ensured that full briefing material was available in their pre-workshop briefing document. In the *Campylobacter* elicitation work, a specific group-training session was held prior to the interviews, providing an opportunity to discuss the documentation and to train the experts in estimating probabilities. This approach was also taken by Simola *et al.* (2005), who provided a training session for the experts involved in the elicitation exercise to address issues arising from an apparent lack of specific knowledge about expert judgement techniques (relating to, for example, how to express subjective assessment of probability; or how to deal with bias relating to their judgements). This training session focussed on:

- Familiarising the participants with the expert judgement process and motivating them to provide formal judgements;
- Providing training on probability issues and on providing probability judgements;
- Informing participants on possible cognitive biases inherent in expert judgement and possible techniques to reduce or remove bias; and
- Solving simple exercises, as a means of providing experience of the process.

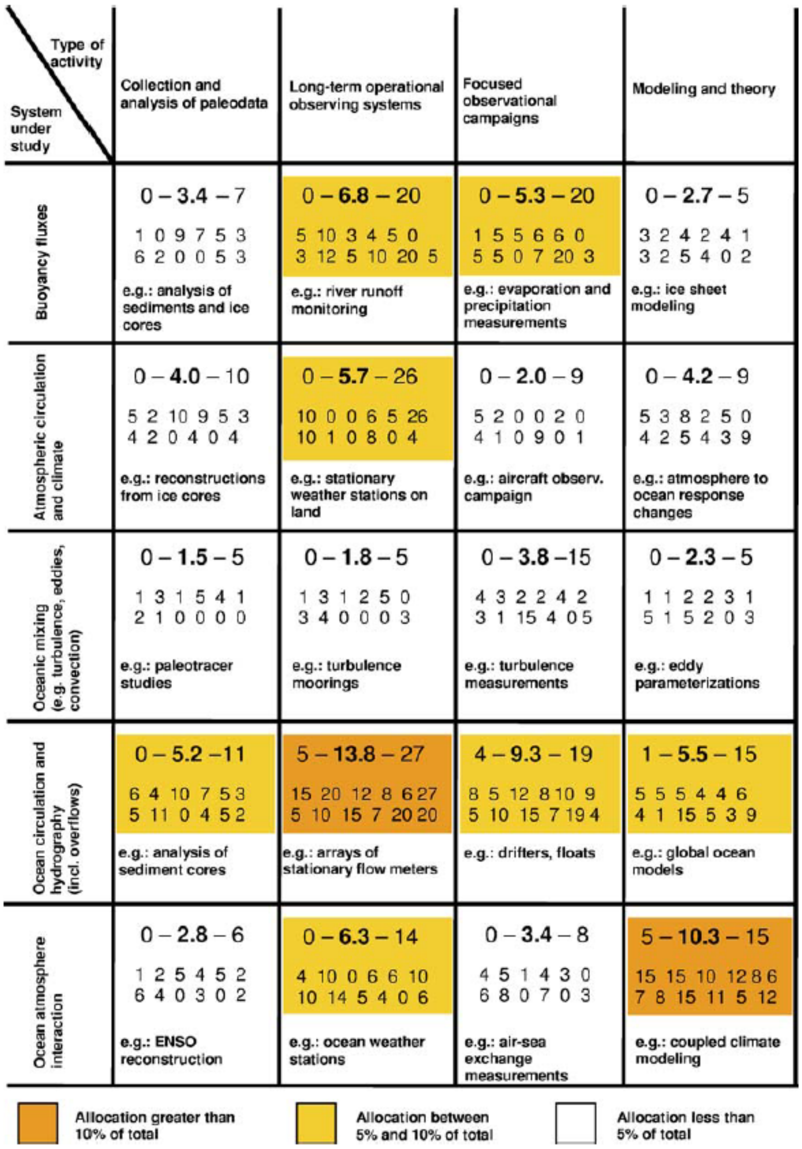
Zickfeld *et al.* (2007) also invested in pre-elicitation training. A couple of weeks before interviews participants were presented with the protocol to allow them to prepare themselves on specific topics. Also, the ‘interactive’ referencing of materials (literature, notes, simulation results, etc.) during the course of the interviews was encouraged.

Learning point: splitting the elicitation process to more than one single session allows participants to reflect on their judgements and refine their opinions in the light of information from the other experts.

Elicitation of expert judgement regarding ultrafine particles (as reported in Knol *et al.*, 2010) involved a two round process. In the first round, experts provided initial judgements on an individual basis. This was followed by a group discussion and then a second (and final) individual set of judgements. Changes in judgements between rounds were effectively motivated by the experts themselves (through their own discussions). The observed changes were considered to be mostly the result of contemplating new arguments or of a modified (and more harmonised) interpretation of the question, rather than being due to 'peer pressure' or ‘anchoring effects’ (a tendency to focus on one piece of information).

Learning point: techniques other than simple requests for quantitative data can be used to help capture additional information.

Zickfeld *et al.* (2007) present the use of ‘scoring chips’ as a means of allocating importance across each of a range of pre-defined topics. For example, each expert was allocated 50 ‘relative importance’ chips (each representing 2% of the total cumulative importance). A ‘playing board’ was constructed, presenting each of a range of climate science research areas as individual ‘cells’. The experts provided judgement on relative importance of each research area by allocating their chips across the cells of the playing board. Scores were combined (across all experts) and presented as shown in Figure 2.



Type of activity / System under study	Collection and analysis of paleodata	Long-term operational observing systems	Focused observational campaigns	Modeling and theory
Buoyancy fluxes	0 – **3.4** – 7 1 0 9 7 5 3 6 2 0 0 5 3 e.g.: analysis of sediments and ice cores	0 – **6.8** – 20 5 10 3 4 5 0 3 12 5 10 20 5 e.g.: river runoff monitoring	0 – **5.3** – 20 1 5 5 6 6 0 5 5 0 7 20 3 e.g.: evaporation and precipitation measurements	0 – **2.7** – 5 3 2 4 2 4 1 3 2 5 4 0 2 e.g.: ice sheet modeling
Atmospheric circulation and climate	0 – **4.0** – 10 5 2 10 9 5 3 4 2 0 4 0 4 e.g.: reconstructions from ice cores	0 – **5.7** – 26 10 0 0 6 5 26 10 1 0 8 0 4 e.g.: stationary weather stations on land	0 – **2.0** – 9 5 2 0 0 2 0 4 1 0 9 0 1 e.g.: aircraft observ. campaign	0 – **4.2** – 9 5 3 8 2 5 0 4 2 5 4 3 9 e.g.: atmosphere to ocean response changes
Oceanic mixing (e.g. turbulence, eddies, convection)	0 – **1.5** – 5 1 3 1 5 4 1 2 1 0 0 0 0 e.g.: paleotracer studies	0 – **1.8** – 5 1 3 1 2 5 0 3 4 0 0 0 3 e.g.: turbulence moorings	0 – **3.8** –15 4 3 2 2 4 2 3 1 15 4 0 5 e.g.: turbulence measurements	0 – **2.3** – 5 1 1 2 2 3 1 5 1 5 2 0 3 e.g.: eddy parameterizations
Ocean circulation and hydrography (incl. overflows)	0 – **5.2** –11 6 4 10 7 5 3 5 11 0 4 5 2 e.g.: analysis of sediment cores	5 – **13.8** – 27 15 20 12 8 6 27 5 10 15 7 20 20 e.g.: arrays of stationary flow meters	4 – **9.3** – 19 8 5 12 8 10 9 5 10 15 7 19 4 e.g.: drifters, floats	1 – **5.5** – 15 5 5 5 4 4 6 4 1 15 5 3 9 e.g.: global ocean models
Ocean atmosphere interaction	0 – **2.8** – 6 1 2 5 4 5 2 6 4 0 3 0 2 e.g.: ENSO reconstruction	0 – **6.3** – 14 4 10 0 6 6 10 10 14 5 4 0 6 e.g.: ocean weather stations	0 – **3.4** – 8 4 5 1 4 3 0 6 8 0 7 0 3 e.g.: air-sea exchange measurements	5 – **10.3** – 15 15 15 10 12 8 6 7 8 15 11 5 12 e.g.: coupled climate modeling

Figure 2. Example of tabular/graphical representation of elicited expert judgement on relative importance based on allocation of scoring chips (reproduced from Zickfeld *et al.*, 2007)

Learning point: consideration should be given to weighting the opinions from different experts.

Knol *et al.* (2010) highlighted several beneficial approaches within the aggregation and reporting stage of their proposed seven-step process. For example, where value judgements are to be combined into one final estimate, assessors have the choice of various methods to apply weighting to the experts' individual estimates. Although the most obvious is to simply apply equal weighting across all experts, a number of more complex aggregation processes may be applied (which usually involve valuing the judgements of some experts more than those of others, based, for example, on estimates of their individual performance or experience).

Learning point: anonymity can be both advantageous (for example in allowing individuals to give a more honest or candid response) and problematic (for example in reducing the apparent transparency of the process).

In terms of reporting the proceedings in a suitably transparent manner, a listing of all participating experts can be provided in an acknowledgement statement or, alternatively, the experts could be invited to co-author a results paper. The *Campylobacter*-elicitation work reported by Knol *et al.* (2010) allowed the experts to retain their anonymity having announced, prior to the elicitation interviews, that names of experts were not to be mentioned anywhere in the resultant manuscript. This latter approach was purposefully selected in order to make the experts feel more comfortable in providing their judgements that could be in conflict with, for example, ideas of other leading experts (or, indeed, their own previous beliefs). Such complete anonymity can be sensible when a high level of diversity in assigned values exists.

Flari *et al.* (2011) succinctly raise the issues of reliability, transparency and robustness. The issue of choosing reliable experts to provide answers to specific questions has been addressed numerous times in the past. Historically, processes to obtain expert judgement have been *ad hoc* and hard to reproduce, particularly when consensus was reached by means of group discussions. In cases requiring expert opinion, high levels of uncertainty are typical, so the elicitation process should be transparent and must allow experts to state their true opinions without being:

(i) influenced by other participants and/or stakeholders; or
(ii) pre-judged by the risk analyst/assessor.

3 Use of Expert Judgement in Biodiversity Assessments

3.1 Introduction

As highlighted in the preceding section, there is evidence of the widespread use of expert judgement across a range of disciplines. More specifically, expert knowledge is used widely in the science and practice of conservation – often as a consequence of the complexity of problems, the relative lack of data, and the imminent nature of many conservation decisions (Martin *et al*., 2011). To help identify the strengths and opportunities associated with the use of expert judgement, SWOT (Strengths, Weaknesses, Threats and Opportunities) analyses were undertaken on a series of case studies that incorporated the expert judgement within biodiversity assessments. Subsequently the same case studies were critically reviewed with the aim of understanding and developing the application of expert judgement to environmental disciplines (and, more specifically, to marine biodiversity).

3.2 Methodology

3.2.1 Case Study Selection

A range of recent studies[1] that make use of expert judgement to produce or support biodiversity assessments were identified. These included studies from both the terrestrial and marine environments. From this initial 'long-list' (see Appendix 4) a shortlist of 13 case studies were selected for critical review (see Table 2).

Table 2. Selected case studies

Citation
BASHARI, H., SMITH, C. & BOSCH, O.J.H. (2009) Developing decision support tools for rangeland management by combining state and transition models and Bayesian belief networks, *Agricultural Systems*, **99**, 23–34.
CERTAIN, G., SKARPAAS, O., BJERKE, J-W., FRAMSTAD, E., LINDHOLM, M., NILSEN, J-E., NORDERHAUG, A., OUG, E., PEDERSEN, H-C., SCHARTAU, A-K., van der MEEREN, G.I., ASLAKSEN, I., ENGEN, S., GARNÅSJORDET, P-A., KVALØY, P., LILLEGÅRD, M., YOCCOZ, N.G. & NYBØ, S. (2011) The Nature Index: A general framework for synthesizing knowledge on the state of biodiversity. *PLoS ONE*, **6**(4): e18930. doi:10.1371/journal.pone.0018930. Available at: http://www.plosone.org/article/info:doi/10.1371/journal.pone.0018930 [accessed October 2012].
COWLING, R.M., PRESSEY, R.L., SIMS-CASTLEY, R. le ROUX, A., BAARD, E., BURGERS, C.J. & PALMER, G. (2003) The expert or the algorithm? - comparison of priority conservation areas in the Cape Floristic Region identified by park managers and reserve selection software, *Biological Conservation*, **112**, 147–67.
ENQUIST, C.A.F. & GORI, D.F. (2008) Application of an expert system approach for assessing grassland status in the U.S.-Mexico borderlands: implications for conservation and management, *Natural Areas Journal*, **28**(4), 414–28.
HALPERN, B.S., SELKOE, K.A., MICHELI, F. & KAPPEL, C.V. (2007) Evaluating and ranking the vulnerability of global marine ecosystems to anthropogenic threats, *Conservation Biology*, **21**(5), 1301–15.
KLEYPAS, J.A. & EAKIN, C.M. (2007) Scientists' perceptions of threats to coral reefs: results of a survey of coral reef researchers, *Bulletin of Marine Science*, **80**(2), 419-36.

[1] Although the review focussed on studies published in the last ten years, relevant studies from before 2003 would have been included had they been considered to be of value.

Citation
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**, 906–20.
MPA Monitoring Enterprise (2012) *A framework for advancing the use of expert judgments of ecosystem condition*. Available at: http://monitoringenterprise.org/pdf/Expert_Judgment_Draft_Framework.pdf [accessed October 2012].
MUXIKA, I., BORJA, A. & BALD, J. (2007) Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive, *Marine Pollution Bulletin*, **55**, 16–29.
TECK, S.J., HALPERN, B.S., KAPPEL, C.V., MICHELI, F., SELKOE, K.A., CRAIN, C.M., MARTONE, R., SHEARER, C., ARVAI, J., FISCHHOFF, B., MURRAY, G., NESLO, R. & COOKE, R. (2010) Using expert judgment to estimate marine ecosystem vulnerability in the California Current, *Ecological Applications*, **20**, 1402–16.
TEIXEIRA, H., BORJA, A. WEISBERG, S.B., RANASINGHE, J.A., CADIEN, D.B., DAUER, D.M., DAUVIN, J-C., DEGRAER, S., DIAZ, R.J., GRÉMARE, A., KARAKASSIS, I., LLANSÓ, R.J., LOVELL, L.L., MARQUES, J.C., MONTAGNE, D.E., OCCHIPINTI-AMBROGIM, A., ROSENBERG, R., SARDÁ, R., SCHAFFNER, L.C. & VELARDE, R.G. (2010) Assessing coastal benthic macrofauna community condition using best professional judgement – developing consensus across North America and Europe, *Marine Pollution Bulletin*, **60**, 589-600.
UK Marine Monitoring and Assessment Strategy (2010) *Charting Progress 2 Healthy and Biological Diverse Seas Feeder report*. (Eds. Frost, M. & Hawkridge, J). Published by Department for Environment Food and Rural Affairs on behalf of UKMMAS. 682pp. Available at: http://chartingprogress.defra.gov.uk/healthy-and-biologically-diverse-seas-feeder-report [accessed October 2012].
WARD, T.J. (2011) *SOE 2011: National marine condition assessment – decision model and workshops*. Report prepared for the Australian Government Department of Sustainability, Environment, Water, Population and Communities on behalf of the State of the Environment 2011 Committee. Canberra: DSEWPaC, 2011.

Although initially identified as potential case studies, two further works (Okey *et al*., 2012; and Zacharias & Gregr, 2005) were, following initial review, dropped from this report (see Appendix 4 for further detail).

3.2.2 **Review Methodology**

SWOT analyses were used to identify the main strengths and opportunities afforded by the range of approaches to applying expert judgement used in the selected case studies, within the context of their weaknesses and any relevant threats. Whilst classic SWOT analyses are routinely used in the business or marketing environment, their use as a descriptive and analytical tool outside of these environments requires the process to be modified slightly. In undertaking the SWOT analyses for the selected case studies as part of this study, the following points were taken into consideration:

Strengths:

- What are the key strengths of the approaches used or highlighted in the paper (transparency of process, ability to derive useful data from disparate expert sources, etc.)?
- What aspects of the methodology set this study apart from others?
- Are unique (or unusual) resources being drawn upon (and, if so, what are they)?
- What factors would be central in leading to the described approach being adopted or transferred to further work in the future?

Weaknesses:

- What (if any) are the weaknesses in the approach(es) used and how might they have been improved?
- What apparent pitfalls could have been avoided in the approach?

Opportunities:

- What opportunities are highlighted in the paper?
- Does the paper highlight or represent any particular trends in approaches?
- N.B. opportunities can come from such things as:
  - Changes in relation to ‘new’ technology or to needs
  - Changes in government policy or societal desire

Threats:

- Are the requirements for studies based on expert judgement, or the mainstream understanding of the value of expert judgement, likely to change such that aspects of the approach(es) used in the study would no longer be applicable (and, if so, how)?
- Could the validity of the approach(es) used in the study be undermined (or their applicability otherwise reduced) in the future?

The outputs from the SWOT analyses were used to provide an indication of the ‘success’ of each study (or of their constituent methodologies) and to assist in identifying those approaches and techniques that should be considered further.

In addition, a subsequent detailed analysis of the performance and attributes of the selected case studies considered several separate elements which, together, effectively covered a number of broad facets:

- Practical aspects:
  - practicality of method(s) (applicability and ease of use);
  - transparency and comprehensibility of method(s); and
  - size/source of expert panel.

- Statistical and methodological aspects:
  - principles used to apply expert judgement to the assessment process (e.g. methods for amalgamating/combining individual judgements);
  - robustness of the method(s) (ability to reduce bias, reduce inconsistency and assure scientific quality);
  - ability to quantify associated confidence;
  - repeatability of the method(s); and
  - validation of outputs after applying the method(s).

- Opportunities or insights:

  - scale of application (methods might have been acceptable in their original form, but may not transfer to a different scale (wider or narrower) of application);
  - opportunity for modification (methods may have some shortcomings but could be readily adapted or improved to better meet the needs of marine biodiversity assessment in the UK).

3.3 Case Study Analyses

The following text provides the SWOT analyses together with further detailed analysis of each of the 13 selected case studies under consideration, including commentary on their practical aspects, statistical and methodological aspects and opportunities or insights.

3.3.1 Developing Decision Support Tools for Rangeland Management by Combining State and Transition Models and Bayesian Belief Networks (Bashari *et al.*, 2009)

The paper is based on the development of a decision support system that can be applied to rangeland management, specifically cleared Ironbark-spotted gum woodland in south-east Queensland, Australia. Bashari *et al.* (2009) demonstrate an approach which links a ‘state and transition model’ (STM) with a ‘Bayesian belief network’ (BBN) in order to provide a relatively simple and dynamic model that is able to accommodate uncertainty, and support scenario-, diagnostic- and sensitivity- analyses.

SWOT Analysis

Strengths Flexibility - allows use of subjective probabilities (expert judgement) to replace measured values; Captures advantages of both STMs (simple, graphical) and BBNs (allows use of experiential knowledge where empirical data is scarce; can accommodate uncertainty).	**Weaknesses** Where probabilities within the models are based mainly on expert opinion, limits of human judgement become important – reliability may be questioned and opinions may be subject to bias; Some users may see complexity in the STM as being a weakness – may lead to ‘over-parameterisation’.
Opportunities Use of BBN appears to be a growing field; Improved understanding of probabilities can be readily incorporated into an updated BBN making the approach suitable for adaptive management.	**Threats** BBN may be mistrusted by some people – may be viewed as overcomplicated.

Practical Aspects

Bashari *et al.* (2009) present a system to combine two approaches: state and transition models (STMs) and Bayesian belief networks (BBNs). The reference provides a useful insight into potential approaches, although only limited expert elicitation is presented. The premise of their paper is that decision support tools for rangeland management can be developed by combining STMs and BBNs. These new tools would retain the benefits of STMs (such as diagrammatic, low cost, flexible, and suited to participatory development with managers) whilst providing scenario analysis capabilities, adaptive management capabilities, and the ability to accommodate uncertainty.

Whilst the approach may superficially appear complex, the inclusion of a clear graphical element within the overall model structure helps facilitate communication and acceptance by managers, policy makers or the public.

Expert judgement regarding pasture dynamics within the study area was elicited through two workshops. One workshop involved livestock owners, whilst the second involved rangeland scientists. No information is provided regarding numbers of experts or the processes used to elicit, collate or process opinion, but the potential value of obtaining expert judgement from more than one source should be noted.

Statistical and methodological aspects

Bashari *et al.* (2009) held two workshops to elicit experiential knowledge of pasture dynamics within the study area, with participants being asked to review vegetation state definitions, possible transitions and their causes (as taken from previously published STMs). The authors describe how the STM was converted into an influence diagram (the directed acyclic graph; the graphical component of a BBN) and how the relationships between the nodes within the influence diagram were subsequently defined using Conditional Probability Tables (CPTs). CPTs store the probabilities of outcomes under particular scenarios and these probabilities allow uncertainty and variability to be accommodated in model predictions. As measured probabilities for transitions can be obtained only from long-term studies (and these were absent in the system studied by Bashari *et al.,* 2009) subjective probability estimates were obtained from rangeland scientists (as expert judgement) and used in their place. The process adopted does not, however, allow for the reliability of elicited probability estimates to be directly recorded.

Although databases, scientific literature, and other models can be used to determine the probabilities of transitions in a combined STM and BBN model, the study by Bashari *et al.* (2009) indicates that the knowledge and practical experience of experts are often the only available sources of data.

Opportunities or insights

Bashari *et al.* (2009) consider the combination of STMs and BBNs, identifying the benefits of both individual approaches. Because of their graphical and descriptive nature, STMs are excellent tools for communicating knowledge about rangeland dynamics between scientists, managers, and policy makers. However, because they are essentially descriptive diagrams, STMs have limited predictive capability (which has restricted their practical application in scenario analysis), and their coarse handling of uncertainty represents a further shortcoming. Bayesian belief networks (BBNs) (also known as belief networks, causal nets, causal probabilistic networks, probabilistic cause effect models, and graphical probability networks) are graphical models consisting of nodes (boxes) and links (arrows) that represent system variables and their cause-and-effect relationships. BBNs consist of a qualitative part (a directed acyclic graph, or cause-and-effect diagram, composed of a series of linked nodes) and an associated quantitative part (a set of conditional probabilities quantifying the strength of dependencies between the variables represented in the acyclic graph). BBNs are becoming an increasingly popular modelling tool, particularly in ecology and environmental management. This is largely because they can be used in a predictive capacity and also, because they use probabilities to quantify relationships between model variables, they explicitly allow uncertainty and variability to be accommodated in model predictions.

3.3.2 **The Nature Index: A General Framework for Synthesizing Knowledge on the State of Biodiversity (Certain *et al.*, 2011)**

Certain *et al.* (2011) developed a framework to produce an index, the Nature Index (NI), representing a synthesis of information on biodiversity from across a wide range of sources. Major ecosystems are represented by a series of individual biodiversity indicators which are amalgamated into a single index. The NI is closely allied to the Dutch Natural Capital Index (ten Brink and Tekelenburg, 2002) and to the South African Biological Intactness Index (Scholes and Biggs, 2005) but benefits from some important conceptual differences. For example, the NI does not assume that the importance of a major ecosystem is proportional to its area, and also allows the combination of several different types of reference state rather than relying on an assumed relationship with a covariate. Implementation of the NI framework is detailed for Norway, covering nine major habitat groups (four coastal and marine, two wetland and freshwater and three terrestrial).

SWOT Analysis

Strengths	**Weaknesses**
Allows the synthesis of expert judgement together with both monitoring-based and model-based estimates of biodiversity status, producing a simple set of measures that can be easily communicated; Allows for three forms of uncertainty to be considered: numerical uncertainty, data source uncertainty and uncertainty due to lack of knowledge; Provides three alternative models for scaling indicators onto a common 0-1 range, accounting for different ways of interpreting observed values relative to expected values under reference conditions.	Major ecosystems are represented by one or more biodiversity indicators but the distribution of indicators across ecosystems is not discussed; Choice of indicators may affect outcome of assessment.
Opportunities	**Threats**
The approach described allows for the incorporation of forecasting and scenario testing.	The production of a single composite index may mask large variation in biodiversity status; The method (particularly scaling of indicators etc.) may potentially be difficult to understand.

Practical Aspects

The Nature Index (NI) approach described by Certain *et al.* (2011) is not an approach based on expert judgement, but is rather a framework for combining information on biodiversity and from a number of sources, allowing it to be synthesised to produce simple metrics that can be used to inform environmental managers, policy makers and the public. As applied to the Norwegian environment (Certain and Skarpaas, 2010) the framework considered 130 indicators (covering nine major habitat groups: four coast and marine; two wetland and freshwater; and three terrestrial) and involved input from more than 120 experts (most of them being either the recognised national expert on their indicator, or one of several such experts for the same indicator).

Experts used a website to enter observed values for each indicator, by municipality and by date. Experts also entered values for the reference states of each indicator in each municipality. Lower (25%) and upper (75%) quartiles for each observed indicator value were also submitted by the experts as a measure of numerical uncertainty.

Statistical and Methodological Aspects

The NI framework presented by Certain *et al.* (2011) makes use of reference state indicator values, which are used in conjunction with one of three possible models to scale the observed value of each indicator such that all scaled indicator values are then directly comparable. Estimation of reference state indicator values was undertaken by expert judgement, with the expert in charge of the indicator in question being responsible for identifying the critical values in each of the three ‘scaling models’:

- The ‘optimal’ scaling model implicitly assumed that any departure (positive or negative) from the reference state would result in a degradation of the state of the major ecosystem related to the indicator;
- The ‘minimal’ scaling model was applied when the reference state refers to a low, precautionary level. When scaling the indicator for the minimal model, it was assumed that a deteriorated state for the indicator corresponds to a decrease below the reference level, and that any value above this reference level corresponded to an optimal situation; and
- The ‘maximal’ scaling model was applied when the reference state refers to a maximal value above which detrimental effects on ecosystems are observed.

In addition, as part of the expert elicitation each expert was asked to provide 50% confidence intervals, reflecting both measurement error and natural variability, for each estimate and for reference state values (Certain and Skarpaas, 2010).

Opportunities or Insights

The full NI, as presented by Certain *et al.* (2011), represents a framework that can be applied repeatedly across several reporting periods, allowing changes in biodiversity to be captured and reported. By providing a consistency of approach, the NI facilitates the subsequent comparability of outputs between years.

The approach for normalising or scaling data outlined by Certain *et al.* (2011), using three possible models: the optimum, minimum and maximum models as outlined earlier, represents a simple approach that could be adapted for use elsewhere.

3.3.3 A Comparison of Priority Conservation Areas in the Cape Floristic Region Identified by Park Managers and Reserve Selection Software (Cowling *et al.*, 2003)

Working in a South African terrestrial environment (the Cape Floristic Region), Cowling *et al.* (2003) present an assessment of priority areas for nature conservation as identified by expert judgement.

The expert judgement elicitation process used to identify priority areas was not particularly detailed. A list of 29 candidate areas was compiled using contributions from reserve managers working for three conservation agencies who were involved in the process of expanding the Cape Floristic Region’s conservation system. The list was compiled over a relatively short time period and did not incorporate all of the expert knowledge that resides in each of the conservation agencies. Two workshops were held to enable managers to provide and refine justifications for identifying the candidate areas.

The study concluded that the priority areas identified in the candidate ‘wish list’ fell short of achieving many of the biodiversity targets that had previously been formulated for system conservation planning in the region. The authors note that closer interaction between the outcomes of a systematic approach (such as maps of irreplaceability and vulnerability) and the priority areas identified by managers (i.e. expert judgement outputs) is likely to greatly improve the prospects for the successful implementation of new reserves, and that expert judgement should be integrated into all stages of a systematic conservation planning process.

SWOT Analysis

Strengths	**Weaknesses**
Approach used to elicit information (production of hand-drawn maps) relatively simple.	Expert judgement elicitation process was limited and not particularly robust (e.g. selection of potential new reserve areas were based on a ‘wish list’ approach, with limited structure to guide experts.
Opportunities	**Threats**
Suggests that even where systematic, algorithm-based processes exist for conservation planning, expert judgement should be integrated into all stages of the planning process.	No threats identified.

Practical Aspects

In practical terms, the elicitation of expert judgement described by Cowling *et al.* (2003) was relatively simple. Contributions to a ‘wishlist’ totalling 29 priority areas for nature conservation were elicited from reserve managers working for the three agencies that were in the process of expanding the (South Africa) Cape Floristic Region’s conservation system. In addition to understanding the practical constraints of reserve implementation and management, most managers in the Cape Floristic Region have a sound knowledge of biodiversity issues and many have specialist knowledge of particular taxa. Managers were each asked to indicate the location of candidate areas on a map; the identified areas were then digitised. The list of areas was compiled over a very short period (managers had, at most, a few weeks to identify and recommend priority conservation areas). Following a one-day workshop, at which managers were able to provide supporting justifications for their identified wishlist areas, the areas identified by experts were assessed against sites identified using systematic, algorithm-based reserve-selection software. In particular, the sites were assessed for the extent to which they achieved targets for biodiversity pattern and process over and above the existing conservation system, as well as the extent to which they incorporated priority areas identified in terms of conservation value and vulnerability to processes that threaten biodiversity. Although the authors do not directly state how many experts were contacted for contributions, 19 managers and officials (from two of the three agencies in the region) attended the workshop.

Statistical and Methodological Aspects

The approach adopted by Cowling *et al.* (2003), in their comparison of expert judgement methods with established systematic approaches, made no allowance for accounting for associated confidence in judgements. Whilst there is not a great deal of clarity over methods used to elicit expert judgement, the process as described does not appear to have been particularly robust or well-defined. The approach does, however, incorporate a form of validation, with expert judgement being compared to the outputs from reserve selection software.

Opportunities or Insights

Cowling *et al.* (2003) noted that the candidate list configuration was strongly determined by pragmatic considerations (including reserve consolidation and boundary rationalisation, and alignment with pre-existing initiatives) as well as by the sample of managers used (with bias being introduced through, for example, individual preferences for particular regions and/or knowledge about specific biodiversity features).

Overall, the list reflected a desire to improve the efficiency of management and to facilitate the rapid inclusion of new conservation areas through the expansion of existing reserves into adjacent areas where land tenure is sympathetic to conservation, that would capture a specific subset of biodiversity pattern, and where certain biodiversity processes would be accommodated. However, rather than emphasize the dichotomy between expert and systematic approaches, Cowling *et al.* (2003) conclude that conservation planners should instead look to integrate them. In particular, priority conservation areas identified by experts should be carefully considered against the backdrop of the outcomes of associated systematic conservation planning.

3.3.4 Application of an Expert System Approach for Assessing Grassland Status in the U.S.-Mexico Borderlands: Implications for Conservation and Management (Enquist & Gori, 2008)

Enquist & Gori developed a framework for producing maps detailing the spatial extent of five grassland classes across the arid and semi-arid grasslands of the Apache Highlands ecoregion of the US/Mexico borderlands based on the input from a range of management specialists (expert judgement). They describe an extensive ground-truthing exercise that was completed to verify data for the US grasslands, and this was used to determine the accuracy rate of the expert judgement designations.

SWOT Analysis

Strengths	**Weaknesses**
The study employed expert judgement of grassland type/status engaging a wide range of experts (24 individuals from across 10 federal and state agencies, institutions and non-governmental organisations); Expert judgement was verified by follow-up field visits to provide an indication of their accuracy rate (initially determined to be 71%, based on an error matrix of omissions and commissions developed from randomised field visits, and later estimated at 77% based on field visits plus long term monitoring data); Follow-up study covering other areas of Arizona produced accuracy rate of 74%.	Some detail appears lacking (e.g. methods for integrating spatial assessments by individual experts into a single combined assessment); Elicitation of expert judgement was restricted to identifying the likely occurrence of each of five classes of grassland vegetation (i.e. the elicitation was effectively based on a closed set of possible responses); Is an expert judgement accuracy of <80% high enough?
Opportunities	**Threats**
The paper supports the use of expert judgement in scientific studies as an appropriately rigorous method; Provides some potentially transferable procedures for testing accuracy in expert judgement.	Although subsequent work by the authors using the same approach found high accuracy rates (74%), these were lower that the ‘headline’ values presented for the main study (77%) suggesting that there may be a risk that the true underlying accuracy of the elicited expert judgement is lower than that suggested by the verification process.

Practical Aspects

In deriving expert opinion on grassland status in the US/Mexico borderlands, Enquist & Gori (2008) interviewed 24 range management specialists. Each was asked to draw polygons on hard copy base maps to show the distribution of five standardised grassland classes or condition types. Experts were selected from ten federal and state agencies, institutions and non-governmental organisations (although, with the exception of commenting that all experts had considerable knowledge of grassland extent and conditions within their local jurisdiction, no indication of how the panel of 24 experts were identified is provided). The regional base maps used, which contained elevation contours and location data, were at scales between

1:100,000 and 1:125,000. Subsequently, a series of 17 field surveys were undertaken to provide an independent 'validation' dataset covering a total of 190 points throughout the study area; this validation dataset was then used to verify the expert data.

Statistical and Methodological Aspects

A hands-on approach was taken in Enquist & Gori's (2008) study, where hard-copy maps of grassland extent were provided by 24 experts and then digitised into a GIS. These GIS layers were overlaid with a hexagonal grid (50ha cell size) and, where there were discrepancies in grassland categories of individual cells between expert-derived categories and field verification data, expert judgements were modified. The individual experts' layers were then combined to a single GIS layer (although no methodology for this combination process is described). Whilst the process as described does not incorporate measures of uncertainty or confidence, Enquist and Gori do undertake an assessment of the validity of the experts' views. Based on an error matrix of omissions and commissions developed from randomised field visits, experts' accuracy rate was initially determined to be 71%. Subsequent incorporation of long term monitoring data into the validation dataset increased this accuracy rate to 77%, whereas a follow-up study, using the same approach but covering other areas of Arizona, produced an accuracy rate of 74%.

Opportunities or Insights

Whilst noting that many researchers do not view the use of expert opinion in scientific studies as a sufficiently rigorous method, Enquist & Gori (2008) believe that their use of such an approach was warranted given the absence of viable alternatives for meeting their specific objectives (e.g. the development of a rapid and cost-effective assessment of habitat status for immediate conservation planning and management). They comment that expert-based methods have been employed with success in other studies reported in the literature, and conclude that they were able to demonstrate a rapid, regional assessment approach to evaluating habitat extent and condition at a regional scale that is both straightforward in its application and instructive in its ability to produce meaningful results for conservation planning and management in a time and cost effective manner.

3.3.5 Evaluating and Ranking the Vulnerability of Global Marine Ecosystems to Anthropogenic Threats (Halpern *et al.*, 2007)

The aim of the work by Halpern *et al.* (2007) was to quantify the actual or potential impacts of 38 distinct anthropogenic threats on 23 marine ecosystems (as identified in workshops involving academic, non-governmental and agency scientists from around the world). The paper presents a table of vulnerability scores indicating the relative importance of 38 threats to each of 23 distinct marine and coastal ecosystems. Expert elicitation was web-based and focused on marine and coastal ecosystems at the global scale (i.e. it was not restricted to any one particular area but sought views on the each of the 23 identified ecosystem types). Views were elicited on ratings for five independent vulnerability criteria (scale, frequency, functional impact, resistance and recovery time) in relation to each of the 38 identified threats.

SWOT Analysis

Strengths	Weaknesses
The web-based approach facilitates the effective integration of input at a global scale from a number of geographically dispersed experts and provides a record of expert decisions that can be referred back to; Uncertainty data was embodied within the expert judgement responses that were provided.	All five individual vulnerability criteria were assumed to be equally important when combined to derive an overall vulnerability score for each threat: ecosystem combination; Although the elicitation requested relatively simple information, the remote (web-based) approach removes the ability to clarify requests and may lead to bias in the responses provided; Whilst there was a defined process for identifying experts, subsequent agreement of experts to participate may have biased the overall spatial coverage (and hence consideration of specific ecosystem types). Supplementary literature-based inputs were used to address this bias where possible by producing an additional supplementary 'expert' view.
Opportunities	**Threats**
The use of expert judgement to help focus future effort is highlighted – i.e. the outputs themselves are not necessarily the final product but instead are a means of identifying priorities for further work (e.g. by identifying critical threats and the incidence of information or knowledge gaps); The approach can be readily transported to other management or conservation priority-setting tasks related to the identification of key threats or priority ecosystems.	If a similar approach to identifying experts were to be used on a more targeted elicitation it may be challenged as being too 'passive' – a more 'active' selection of experts may be seen as providing more robust and less biased outputs.

Practical Aspects

Halpern *et al.* (2007) addressed three critical questions:

- What are the most important current threats within and across marine ecosystems?
- Which marine ecosystems are most vulnerable to human activities?
- Which factors drive differences in ecosystem susceptibility, and is it possible to quantify those differences?

Experts were identified by searching the Web of Science for literature on each threat-ecosystem combination and all authors with listed email addresses were contacted about the survey. Authors contacted in this way were also requested to pass the invitation on to other experts (in one or more ecosystems) to provide opinion on these matters. A total of 370 experts were invited to take part in the survey, of which 135 submitted responses to an on-line survey, providing estimates of the severity of each of 38 potential threats to any one or more of 23 previously identified marine ecosystems.

Statistical and Methodological Aspects

In their consideration of 874 threat-ecosystem combinations (38 threats affecting 23 marine ecosystems) Halpern *et al.* (2007) obtained expert assessments of the impact of threats expressed as 'vulnerability scores', determined by each ecosystem's vulnerability to each threat.

Vulnerability scores were based on the integration of six separate factors (vulnerability criteria; see Table 3) relating to each threat:ecosystem pairing, these being: spatial scale (the average scale at which the threat affects the ecosystem); frequency (how often discrete threat events occur in a given ecosystem); functional impact (the degree to which a threat affects only selected species within an ecosystem or impacts upon the entire ecosystem); resistance (the average tendency of a species, trophic level, community or ecosystem to resist changing its 'natural' state in response to a threat); and recovery (the average time required for the affected species, trophic level, community or ecosystem to return to its initial state), together with an indication of the overall level of certainty in the judgements made.

For each threat:ecosystem pairing, the five vulnerability measures and the certainty measure were combined into a single weighted-average vulnerability score representing (in relative terms) how vulnerable a given ecosystem is to a given threat[2].

The approach used by Halpern *et al.* (2007) meant that, inevitably, there was a small sample size for some of the marine ecosystems under consideration. Whilst the approach made no allowance for perceived confidence in responses, a key strength was that it could be adapted for use in almost any management or conservation priority-setting effort tasked with identifying key threats or priority ecosystems. It is noted by the authors however that the approach would be less suitable if species (and not habitats) were to be the focus of management.

In terms of the results obtained, Halpern *et al.* (2007) found that the variance amongst responses was fairly high even for ecosystems with large survey sample sizes, a result attributed to four sources:

- Valid differences in perspectives and knowledge base;
- Regional differences in ecosystem response to threats;
- The paucity of data on some threat: ecosystem combinations (forcing experts to rely on extrapolation or intuition); and
- Respondent error due to misunderstanding of instructions.

Additionally, an assessment of the consistency was made by comparing the similarity between the top three threats to marine ecosystems calculated from the expert survey (average ecosystem vulnerability scores) and the most frequently stated threats by respondents to the survey on how threats affect each marine ecosystem[3].

This revealed a surprisingly large inconsistency between the lists of top three threats derived from the survey versus the volunteered responses, both across all respondents and for each individual expert. The authors conclude that this emphasises the need for conducting quantitative and transparent threat analyses.

[2] The authors noted that this weighted average gives greater importance to values with higher certainty (and presumably higher precision), but may lower weighted scores for poorly studied threat-ecosystem combinations. Additionally, it may be sensitive to the scale of the categories used for each vulnerability measure.

[3] Whereby respondents stated what they believed were the top three threats *without* evaluating them by the five measures of ecosystem vulnerability.

Table 3. Ranking systems for each vulnerability measure used to assess how threats affect marine ecosystems (after Halpern *et al.*, 2007)

Vulnerability measure & category	Rank	Descriptive notes	Example
Scale (km^2)			
No threat	0		
<1	1		Anchor damage
1–10	2		Single trawl drag
10–100	3		Sediment run-off from deforestation
100–1,000	4		Land-based pollution from run-off of large rivers
1,000–10,000	5		An invasive species
>10,000	6		Sea surface temperature change
Frequency			
Never occurs	0		
Rare	1	Infrequent enough to affect long-term dynamics of a given population or location	Large oil spill
Occasional	2	Frequent but irregular in nature	Toxic algal blooms
Annual or regular	3	Frequent and often seasonal or periodic in nature	Runoff events due to seasonal rains
Persistent	4	More or less constant year-round, lasting through multiple years or decades	Persistent hypoxic zones
Functional impact			
No impact	0		
Species (single or multiple)	1	One or more species in a single or different trophic levels	Ship strikes on whales
Single trophic level	2	Multiple species affected; entire trophic level changes	Overharvest of multiple species within the same trophic guild
>1 trophic level	3	Multiple species affected; multiple trophic levels change	Overharvest of key species from multiple trophic guilds
Entire community	4	Cascading effect that alter the entire ecosystem	Ocean temperature increase and fatal bleaching of coral reefs
Resistance			
No impact	0		
High	1	No significant change in biomass, structure, or diversity until extreme threat levels	Trawling on soft-sediment communities
Medium	2	Moderate intensities or frequencies of a threat lead to change	Effects of industrial pollution run-off on coastal species
Low	3	Slightest occurrence of a threat causes a change, or all-or-nothing threats	Blast fishing in coral reefs
Recovery time (yrs)			
No impact	0		
<1	1		Kelp recovery after disturbance
1–10	2		Short-lived species recovery from episodic toxic pollution
10–100	3		Long-lived species recovery from overfishing
>100	4		Deep-sea coral recovery after trawl damage
Certainty			
None	0		
Low	1	Very little or no empirical work exists	
Medium	2	Some empirical work exists or expert has some personal experience	
High	3	Body of empirical work exists or the expert has direct personal experience	
Very high	4	Extensive empirical work exists or the expert has extensive personal experience	

Opportunities or Insights

In discussion, Halpern *et al*. (2007) suggest that the results of their study may provide guidance on where the greatest information gaps and research needs exist (i.e. those threats and ecosystems that received the lowest certainty scores in the elicitation exercise) and which are therefore in need of more research. For example, soft-bottom ecosystems generally had the lowest certainty scores and so might be construed as being understudied.

3.3.6 Scientists' Perceptions of Threats to Coral Reefs: Results of a Survey of Coral Reef Researchers (Kleypas & Eakin, 2007)

Kleypas & Eakin used expert opinion to identify the major threats facing coral reef ecosystems around the world. The survey that underpinned their study was primarily web-based, with the elicitation of expert judgement consisting of a set list of questions that were presented in a fixed order and answered via an interactive website. Responses from 286 participants at the 10th International Coral Reef Symposium and members of the International Society for Reef Studies were used to rank the relative severity of 39 threats to coral reefs in the region with which they were most familiar by indicating whether they believed the level of each of the 39 threats to be high, medium or low.

SWOT Analysis

Strengths	**Weaknesses**
Very large group of 'experts' contributed (responses received from 286 participants across 41 different countries); Web-based approach reduced effort required to undertake elicitation and subsequent collation of outputs and provided record (audit trial) of responses.	Level of expertise was potentially very variable between participants.
Opportunities	**Threats**
The web-based approach used in the study would appear to help facilitate participation.	Basis for selection of 'experts' (participation at the 10th International Coral Reef Symposium plus members of the International Society for Reef Studies) was not particularly rigorous.

Practical Aspects

A form of 'self selection' was used to identify experts for participation in the study by Kleypas & Eakin (2007) who emailed their web-based questionnaire to delegates attending the 10th International Coral Reef Symposium. This approach provided responses from 286 'experts' spread across 41 countries (20% of attendees and nearly 50% of the countries represented at the symposium). Whilst generating a reasonable response rate this approach needs to account for bias within responses.

Statistical and Methodological Aspects

Kleypas & Eakin (2007) asked participants in their study to rank the relative severity (high, medium, low) of each of 39 threats to coral reefs in the region with which they were most familiar. After elicitation, an overall score for each threat was calculated using the average response (assigning a score of 3 to the 'high' responses; 2 to 'medium' responses' and 1 to 'low' responses). No account was taken of uncertainty.

Opportunities or Insights

Although Kleypas & Eakin (2007) demonstrated a means of generating a large participant group composed of workers in the field of study (by engaging with delegates registered for an international symposium) this process does not particularly lend itself to developing an 'expert' view and probably is better framed as a means of identifying an 'informed opinion'.

3.3.7 Structured Elicitation of Expert Judgements for Threatened Species Assessment: a Case Study on a Continental Scale Using Email (McBride *et al.*, 2012)

McBride *et al.* (2012) reported on the evaluation of extinction risk for nine Australian bird taxa. Sixteen experts independently provided judgements regarding up to125 parameters for each taxon. Initial outputs were collated and circulated for information. Experts were subsequently provided with each others' values and were able to discuss any differences in opinion via email before reassessing and confirming their own views (a version of the 'Delphi method', whereby initial estimates are made by each participant who then, subsequently, receives *anonymous* feedback regarding the estimates of the other participants. On the basis of this information they then make a second, (potentially) revised estimate. The estimate and feedback stages continue for a set number of rounds or until a pre-specified level of agreement between participants is reached).

Two forms of elicitation question were used: one elicited lowest and highest plausible estimates, best estimates and probabilities that the true values were contained within the upper and lower bounds; the second elicited yes/no answers and a degree of credibility in the answer provided. The analysis makes use of 'fuzzy number procedures' (e.g. Zadeh, 1965; Kaufmann & Gupta, 1985) to derive best estimates of uncertain quantitative values.

SWOT Analysis

Strengths	Weaknesses
Discussion as part of the Delphi process was done via email, which reduced dominance and gave less confident and articulate people a better opportunity to contribute meaningfully; no one's voice or view was able to dominate, and discussion was frank and balanced; The study employed a four-point question format to elicit quantitative data – mitigates for potential 'over-confidence' effects in responses.	With no group gathering, the ability to discuss topics face-to-face was lost. For contentious issues it was not possible to give and receive behavioural cues during debate, which might otherwise assist in reducing misunderstandings and conveying greater nuance; The effort involved in the elicitation and subsequent analysis was high (e.g. setting up and testing the process, allowing for experts to contribute and following up non-respondents).
Opportunities	**Threats**
The approach of using email groups to undertake the elicitation process is highly flexible and allows individuals to participate at their convenience with ready access to outside resources whilst removing the need for a number of participants to be assembled simultaneously; Uses 'fuzzy number procedures' to help handle uncertainty in the expert judgements.	No threats identified.

Practical Aspects

McBride *et al.* (2012) applied a modified Delphi approach, incorporating facilitator-assisted discussion, to undertake an assessment of threatened Australian birds via email. The expert panel, who were responsible for completing the full elicitation process and for providing assessments for multiple species, consisted of 16 ornithologists identified by their track record, experience, knowledge of the birds of particular regions or specialist skills (taxonomy, IUCN Red Listing, particular bird taxa). All panellists had published extensively on Australian birds and were selected from what were believed to be a relatively small pool of people (<100) with similarly high levels of experience. An additional group of 12 taxon specialists provided assessments for their specialty taxon and informed the expert panel of their views on the parameters. Taxon specialists were identified by their association with interest groups and scientific societies, or by their relevant publications. The overall process of elicitation and compilation took nearly six months to complete.

Statistical and Methodological Aspects

The work by McBride *et al.* (2012) is of particular note in that it seeks to capture and account for uncertainty in experts' opinions through the use of 'triangular fuzzy numbers' (see Figure 3). Where expert judgement on quantitative values (quantities and percentages) was requested during the elicitation process a four point procedure was used (see, for example, Speirs-Bridge *et al.* 2010; Burgman *et al.* 2011a):

1. What is the lowest the value could be? (α);
2. What is the highest the value could be? (β);
3. What is your best estimate - the most likely value? (γ); and
4. How confident are you that the interval you provided contains the truth (in the range of 50–100%)? (ρ)

For binary (yes ⁄ no) questions a two-step procedure was used:

1. Is the statement true or false? (ι); and
2. How sure are you that your answer is correct (provide an answer in the range of 50-100%)? (p)

Each participant refined their opinions through a subsequent Delphi process before values were averaged to provide a set of aggregate estimates for each parameter. These aggregated responses were then used to produce triangular fuzzy numbers (see Figure 3).

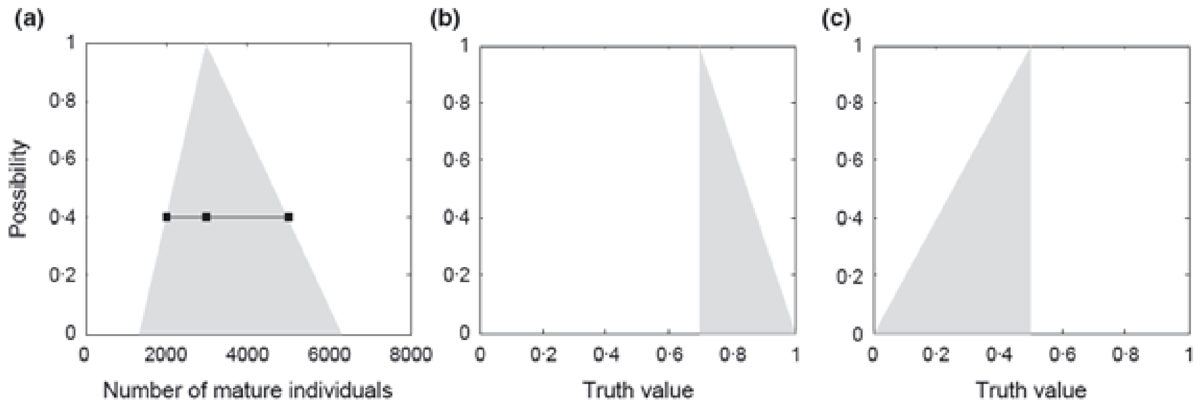


Figure 3. Triangular fuzzy numbers for (a) quantitative and (b, c) categorical questions[4] (After McBride *et al.*, 2012).

[4] The y-axis measures the possibility level, which corresponds inversely to the confidence with which it is believed the true value lies within the bounds of the fuzzy number at that level of possibility. For example, in (a), the black square markers show the expert's assessment of the lower (α), upper (β) and best estimate (γ) values for 'number of mature individuals in the population' plotted on the

The McBride *et al*. paper goes on to discuss several potential sources of bias in the expert judgement estimates (anchoring, dominance, overconfidence, framing affects, availability bias, language-based misunderstanding, inter-expert variability and convergence; see also Table 10, below).

Opportunities or Insights

McBride *et al.* (2012) suggest that email discussions have a number of advantages over face-to-face workshops or telephone conferences (e.g. by removing restrictions on both the number and location of experts that are potentially engaged). However, whilst group emails also made discussions transparent, the overall reduction in potential for discussion (e.g. with discussion tending to be stilted, subject to time delays between questions and answers and, because only a small number of issues can be dealt with in any single email, narrowly focused) may well be a key drawback to the use of email.

Also, facilitation was difficult and panellists were often left to compare their own answers with those of others, and judge for themselves whether they should adjust their responses. The authors suggest that group workshops may be superior where interactive discussion is required on a large number of issues, and conclude that facilitated group workshops employing similar techniques for structured elicitation should remain the tool of choice for assessments in geographically confined areas where panellists can gather without the excessive costs of long-distance travel.

3.3.8 A Framework for Advancing the Use of Expert Judgements of Ecosystem Condition (MPA Monitoring Enterprise, 2012)

The MPA Monitoring Enterprise report, ‘Advancing the use of expert judgements of ecosystem condition’, defines a draft framework and principles to guide the use of expert judgement in assessing ecosystem condition. Arising primarily from a workshop bringing together decision-makers and experts, the guidelines that form the basis of this framework seek to identify and evaluate the process steps, external drivers and communications tools that can contribute credibility and legitimacy to ecosystem condition assessments produced by expert judgement processes.

The development of the framework was also informed by a review of academic literature and reports focused on ecosystem condition, as well as semi-structured interviews with academics and other practitioners who have been involved in expert judgement processes. An appendix to the report provides summaries of five case studies (four US and one Australian).

possibility scale at 0.4, the inverse of the expert’s stated confidence level (ρ) of 60%. Linear extrapolation is used to determine the absolute minimum and maximum values at which the bounds of the fuzzy number bounds cross the x-axis (1333 and 6333 respectively). In (b), the panellist’s ‘false’ response with assigned confidence (ρ) of 50% is represented by the fuzzy number [0, 0.5, 0.5]. In (c), the panellist’s ‘true’ response with assigned confidence (ρ) of 70% is represented as the fuzzy number [0.7, 0.7, 1].

SWOT Analysis

Strengths	**Weaknesses**
Provides clear guidance to inform development of expert judgement processes; Guidance is based on literature reviews and expert judgement practitioners/participants.	The framework provides generic guidance – no specific/novel techniques or methodologies are showcased.
Opportunities	**Threats**
The establishment and reporting of a quasi-formalised process lends credibility to the adoption of expert judgement in other studies.	There is a small but finite risk that the framework as presented may not easily accommodate specific assessment problems.

Practical Aspects

The 2012 report from the MPA Monitoring Enterprise (MPA Monitoring Enterprise, 2012) did not report on a study *per se* but rather provided a draft framework for advancing the use of expert judgement. The report was based on a workshop that had involved decision-makers, managers and scientists, and which identified and evaluated the process steps, external drivers and communications tools that can contribute credibility and legitimacy to ecosystem condition assessments produced by expert judgement processes. A review of academic literature and reports relating to ecosystem condition was used along with further input from academics and others involved in expert judgement processes to help inform development of the framework. The report concludes with a brief synopsis of five sets of case studies:

- NOAA National Marine Sanctuaries (NMS) - Sanctuary Condition Reports;
- Great Barrier Reef Marine Park Authority (GBRMPA) – Outlook Report;
- Puget Sound Partnership (PSP) - State of the Sound (SOS) Reports;
- Chesapeake Bay Program (CBP) – various reports; and
- Santa Monica Bay Restoration Commission (SMBRC) - State of the Bay (SOTB) Reports.

Statistical and Methodological Aspects

The MPA Monitoring Enterprise (2012) report provides a set of guidelines for designing a process to elicit expert judgement (reproduced in summary in the boxes below and discussed in more detail in Section 4.2.1):

Key guidelines for defining the scope and desired outcomes of an expert judgement process

Defining the function of expert judgement

- Clearly articulate how the results will be used (e.g. to inform management decisions, to educate the public)
- Define the goals and the broad context for the process
- Ground the process in a well-defined question, so the experts understand what they are being asked to evaluate

Defining the role of the expert

- Identify the role of experts and the expert judgement process (e.g. to make decisions, provide recommendations, synthesize data)
- Define additional ancillary roles for the experts (e.g. lending credibility to results, achieving buy-in among stakeholders)

Defining the desired end point

- Represent a diversity of opinions; consensus is not necessarily needed
- Acknowledge when there is uncertainty in conclusions

Key guidelines for selecting the experts

Selecting experts
- In consideration of the scope of the question, define the criteria for selecting experts
- Select experts using an open and transparent process

Identifying the types of knowledge needed
- Ensure that a diversity of perspectives is represented within a given discipline
- Consider different perspectives taken by different disciplines concerned with the topic.

Key guidelines for soliciting and working with expert judgement

Deciding on appropriate methods for eliciting expert judgement
- Consider the appropriateness of quantitative versus qualitative assessments
- Identify and select elicitation processes that employ individual assessments and/or group discussions
- Identify whether consensus is the goal, and structure workshop discussions appropriately
- Define a mechanism to handle disagreements

Identifying the source of information that will form the basis of the assessment
- Identify the information available and the criteria for inclusion (e.g. peer reviewed, state funded data)
- Clearly define a role for non-traditional sources of information (e.g. traditional ecological knowledge, local knowledge of fishing activities)

Reviewing the results of the assessment
- Include participant review as an early review step to ensure that their views are accurately represented
- Employ a process for external review that builds credibility and legitimacy employing a combination of a formal peer-review and less formal review by key stakeholder groups as appropriate

Key guidelines for communicating results

Developing an appropriate communications strategy
- Identify the targeted audience(s) as that will inform all aspects of a communications strategy
- Tailor communications for different audiences, but ensure message consistency

Identifying important characteristics of a reporting tool
- Design and employ reporting tools that display information in a way that is intuitive/accessible
- Engage communications experts, or members of the target audience, early on in the assessment process
- Display results in a way that reflects multiple types of uncertainty and variation.
- Articulate, alongside the reporting tool, information about the process, including evidence used, assumptions, caveats, data gaps, uncertainty

Deciding on authorship of a report (the report includes more than the ecosystem condition assessment)
- Identify authors and their role

Key guidelines for developing a process for broader engagement

Identifying the groups/communities that should be brought in to the process

- Design a process to engage other groups (beyond the experts) likely to be interested in the results of the assessment (stakeholders, managers, decision-makers)
- Find and engage 'critical friends'

Identifying at what stage other groups should be engaged and potential mechanisms for engagement

- Consult decision-makers early in the process to ensure salience in situations where the results will be used to inform management
- Consider the broader social and political context in which the results will land when timing stakeholder engagement
- Identify and employ mechanisms for most effectively engaging with other groups (e.g. workshops, briefings)

Opportunities or Insights

The 2012 report by the MPA Monitoring Enterprise provides a useful framework for expert engagement that could easily be adapted and expanded upon to provide a robust and transparent system for future use.

3.3.9 **Using Historical Data, Expert Judgement and Multivariate Analysis in Assessing Reference Conditions and Benthic Ecological Status, According to the European Water Framework Directive (Muxika *et al.*, 2007)**

Muxika *et al.* (2007) present an expert judgement based means of assessing ecological quality status under the European Water Framework Directive (WFD) based on benthic community metrics. They use the AMBI system (the AZTI[5] Marine Biological Index), species richness and diversity together with factor analysis and discriminant analysis to produce a single objective tool for classifying the ecological quality of coastal and estuarine sites in the Basque region of north-eastern Spain.

After developing the new classification tool the authors subsequently employed expert judgement to provide validation of the results. The level of agreement between methods was established, with the importance of misclassification being taken into account (i.e. slight differences in classification being of less concern than greater differences).

[5] AZTI-Tecnalia, is a Spanish Technological Centre specialised in Marine and Food Research, is a private non-profit organization, whose objective is the social development and the improvement of competitiveness in its area of influence by means of technological Research and Innovation

SWOT Analysis

Strengths	**Weaknesses**
Provides an insight into validation metrics (i.e. assessment of the levels of agreement in classification derived using different methods).	Only uses expert judgement as part of a validation process.
Opportunities	**Threats**
Expert judgement seen as high quality method that provides the 'right' answers to classification assessments i.e. validated by direct monitoring data.	As expert judgement is used only for validation there is the implication that, whilst it provides good quality outputs, the expert judgement process is too cumbersome or involves too complex a methodology for it to be used routinely as a validation technique, and where data are available there are alternative (cheaper) methods that can be developed as viable alternatives.

Practical Aspects

Expert judgement did not play a central role in the work by Muxika *et al.* (2007) which instead made use of an earlier assessment of the ecological quality status (EcoQS) of 49 sampling stations on the Basque coast (Northern Spain) undertaken using an objective model-based approach employing benthic community metrics (the AZTI Marine Biological Index, species richness and diversity) together with factor analysis and discriminant analysis. However, expert judgement was used subsequently to represent an independent baseline against which model results could be compared and validated (see Prior *et al.*, 2004). For this aspect of expert judgement the sample composition from each of the sampling stations, together with associated structural parameters (density, biomass, richness, Shannon Wiener index and AMBI, the AZTI Marine Biological Index), was sent to three experts who assessed the quality of each of the samples and allocated each sample to one of the five WFD levels, based upon their own experience. This was a subjective process in which each of the experts evaluated the general quality of the area, based upon their knowledge of the locations and the knowledge of the benthic communities' composition and structure.

Muxika *et al.* compared methods (modelling/discriminant analysis and expert judgement) using a Kappa analysis[6] (Cohen, 1960; Landis & Koch, 1977), the level of agreement between the two methods was measured using the equivalence table from Monserud & Leemans (1992). As the importance of 'misclassification' (or disagreement between methods) is not the same between close categories (e.g. between 'High' and 'Good', or 'Poor' and 'Bad') as between categories that lie further apart (e.g. between 'High' and 'Moderate', or 'High' and 'Bad'), the authors applied Fleiss-Cohen weights[7] to the analysis (Fleiss & Cohen, 1973). After taking expert judgement into account the authors re-classified the status of only five of the 49 sampling sites; Kappa analysis showed an almost perfect agreement between the two methodologies (modelling / discriminant analysis and expert judgement).

[6] Cohen's Kappa coefficient (Cohen, 1960) is a statistical measure of agreement between raters for qualitative (categorical) items

[7] Whilst a simple Kappa analysis considers only whether different raters (e.g. different experts) agree or disagree on their classification or grouping, Fleiss-Cohen weights (Fleiss & Cohen, 1973) are used to take the degree of agreement into consideration, attaching greater importance to near-disagreements

Statistical and Methodological Aspects

The study by Muxika *et al.* (2007) employed expert judgement as a means of validating the outputs of an alternative assessment (statistical) methodology. However, this case study does not provide any further relevant information.

Opportunities or Insights

No information relevant to this section was apparent in the work presented by Muxika *et al*. (2007).

3.3.10 Using Expert Judgement to Estimate Marine Ecosystem Vulnerability in the California Current (Teck *et al.*, 2010)

Teck *et al.* (2010) built on the earlier work by Halpern *et al.* (2007) by addressing the issue of deriving weightings to reflect the relative importance of each of five vulnerability criteria. These weightings were used in the production of an overall vulnerability score for a series of stressor: ecosystem combinations. In total, 53 stressors (threats) associated with human activities and 19 marine ecosystems were considered. The work was based in the California Current region of the eastern Pacific and involved input from 107 experts. Elicitation was undertaken by completion of a survey, either by hand, phone, online or in-person interview.

The study found that, when judging the relative vulnerability of ecosystems to stressors, experts primarily considered two criteria: (i) the ecosystem's resistance to the stressor; and (ii) the number of species or trophic levels affected.

SWOT Analysis

Strengths	**Weaknesses**
Builds on, and improves, a previous documented approach; Survey was tested on a sample group of experts with feedback being used to revise the elicitation process ahead of the main elicitation process; Robust process for identifying pool of experts; Weights derived by defined process, using probabilistic inversion methods.	Co-ordination of the elicitation with a large number of experts may be problematic.
Opportunities	**Threats**
The study stands distinct from many other (Delphi-based) processes in its use of values for decision rules (criteria) and the assessment of the relative importance of those criteria (weights) that were both explicit and quantified (rather than implicit and qualitative).	No threats identified.

Practical Aspects

The scale of the work by Teck *et al.* (2010), which involved 107 respondents, would suggest that the approach may not be transported to other areas without some significant resource implications. In total 240 confirmed potential expert respondents had been identified from an original long list of 525 which was based on scientific experts with personal experience in marine science, conservation, management or policy within the California Current and affiliated primarily with academic institutions, governmental agencies, non-governmental organisation or private-sector consultancy.

Statistical and Methodological Aspects

The vulnerability model used by Teck *et al.* treats vulnerability as the weighted sum of five independently assessed vulnerability criteria, and the details of the mathematical manipulations are described in their paper, but summarised as:

$$\text{Vulnerability (stressor } i\text{, ecosystem } j\text{)} = \sum_{k=1,\ldots,5} W_k S^j_{i,k}$$

where $S^j_{i;k}$ is the value of stressor *i* on criterion *k* in ecosystem *j*, and W_k is the weight assigned to criterion *k*. The weights are normalized so that they sum to unity and are assumed to be the same for all ecosystems and stressors under consideration (an assumption that allows for a single model to be applied to all ecosystem–stressor combinations, in turn allowing for direct comparison among them).

There are many mathematical models exist for combining the weightings to create a single value (e.g. linear, logarithmic, polynomial). However, because environmental vulnerability is expected to be monotonic for all criteria (i.e. higher values denote greater impacts) it can be reasonably approximated by a simple linear model.

Opportunities or Insights

The scale of the work reported by Teck *et al.* (2010) is large (in terms of the numbers of participants) although the elicitation was reasonably focussed as regards the geographic range being considered. The study was unusual in that it tested the approach by undertaking a ‘pre-survey’ involving seven experts to test and confirm the questionnaire system being employed (the experts used in the pre-test did not participate in the subsequent judgement elicitation). They recognise that vulnerability is an abstract concept and defining it at an ecosystem-level scale adds further complexity to the concept. The use of a multi-criteria decision model (MCDM) revealed that experts primarily used percentage change (resistance) and trophic impact when evaluating ecosystem vulnerability to stressors, despite vulnerability also being thought to be a function of exposure and not just a surrogate for sensitivity. Using the MCDM approach allowed Teck *et al.* (2010) to assess the rank importance of each of the 53 stressors considered to be estimated. It may be possible to employ a similar approach to identify the key stressors that should be considered in UK waters.

3.3.11 **Assessing Coastal Benthic Macrofauna Community Condition Using Best Professional Judgement – Developing Consensus Across North America and Europe (Teixeira *et al*., 2010)**

Teixeira *et al.* (2010) tested whether there was agreement between independent judgements by US and European experts regarding coastal benthic macrophyte community condition (as sampled within four regions in the US and Europe).

Their study addressed two objectives: an evaluation of whether best professional judgement assessments were independent of the home regions of the experts; and whether the level of agreement among experts was sufficient to establish a universal benthic assessment scale for the four regions that could be used subsequently to intercalibrate benthic indices and assessment methodologies across habitat boundaries.

Sixteen benthic experts from four geographic regions were provided with species-abundance data for twelve sites from each region and asked to determine the condition of the benthos at each site. The four regions included the Atlantic and Pacific coasts of the US, and the Atlantic and Mediterranean coasts of Europe. A total of 16 benthic ecology experts, whose experience in benthic monitoring ranged from 16 to 38 years, were identified as participants. Of the 16, nine were from academic institutions; four from municipalities that implement benthic monitoring programs to assess the effect of discharge outfalls; two from non-profit research organisations; and one from a private consulting firm.

No systematic difference in assessments due to experts' region of origin was observed, with site rankings being highly correlated among experts, regardless of whether they were assessing samples from their home region. There was also good agreement on condition category, though agreement was better on the extremes of the disturbance gradient.

SWOT Analysis

Strengths The approach employed to elicit expert judgement was straightforward (each expert being presented with three tasks: a ranking exercise; a categorisation exercise; and an assessment of what criteria they had used to evaluate benthic community condition and the relative importance of these); The size of the expert panel (16 experts) was manageable.	**Weaknesses** No weaknesses identified.
Opportunities Demonstrates the apparent absence of regional bias in expert judgement; Supports the use of expert judgement in establishing common scaling principles to allow benthic indices (typically developed independently, at relatively local scales, by habitat) to be integrated over larger geographic scales.	**Threats** No threats identified.

Practical Aspects

The procedures used by Teixeira *et al.* (2010) do not readily translate to other expert judgement applications as they were specifically concerned with having experts judge the condition of a number of benthic samples. Judgement of condition status was based on species-abundance data for each of 48 samples, together with limited habitat data (region, salinity, depth, and percentage of fine-grained sediments) sufficient to establish an expectation for what kinds of organisms should occur there under undisturbed conditions. The study was more concerned with the levels of agreement between experts than the absolute status of each sample. Nevertheless it provided easily understood and transparent results that can be easily communicated to stakeholders and/or the public at large, and makes a valuable contribution by demonstrating the apparent absence of regional bias in expert judgement. The study employed 16 experts from across the four regions participating in their study (European 'Atlantic' and 'Mediterranean' regions, and US East Coast and West Coast regions). The panel of experts was made up of academics, monitoring practitioners and consultants. In addition to assessing the relative condition of each of the 48 samples and ranking the samples both within and across the four regions, the experts were asked to assign each site to one of four condition classes (from 'unaffected' through to 'severely affected') and to identify the criteria they used to evaluate the samples. Of the eight criteria available, six were used by more than half of the experts, with the other two used by only two experts. The three most widely used criteria were 'Dominance by tolerant taxa', 'Presence of sensitive taxa', and 'Biodiversity number of taxa measures' although these were not equally important to experts from different regions.

Statistical and Methodological Aspects

The study presented by Teixeira *et al.* (2010) introduces a methodology to assess the level of agreement (in this instance, regarding condition categories) among the experts, employing Kappa analysis (Cohen, 1960; Landis & Koch, 1977) by establishing moderate, good, very good, and almost perfect levels of agreement using the equivalence table of Monserud & Leemans (1992). As misclassifications between distant categories (for example between 'unaffected' and 'affected', or between 'unaffected' and 'severely affected') are more important than misclassifications between closer categories (e.g. 'affected' and ''severely affected') Teixeira *et al.* applied Fleiss–Cohen weights to the data (Fleiss & Cohen, 1973).

Opportunities or Insights

As noted earlier, the study by Teixeira *et al.* (2010) is valuable as it demonstrates the possibility of employing expert judgement using experts from spatially distinct regions[8].

3.3.12 Charting Progress 2 Healthy and Biological Diverse Seas Feeder Report (UK Marine Monitoring and Assessment Strategy, 2010)

The Healthy and Biologically Diverse Seas Evidence Group (HBDSEG) is one of four evidence groups set up by the United Kingdom Marine Monitoring and Assessment Strategy (UKMMAS) community. The UKMMAS community was set up in response to a recommendation in Charting Progress (the first assessment of the UK Seas, published in 2005) to provide a more-coordinated approach to the assessment and monitoring of the state of the UK marine environment. Each evidence group has a broad membership across the academic and research communities as well as experts in government agencies and non-governmental organisations. HBDSEG was tasked with producing a 'Feeder Report' assessing all the evidence available under its remit which could be used as source material for the biodiversity chapters in the main Charting Progress 2 report. The Charting Progress 2 report builds on the 2005 Charting Progress report and seeks to show the extent to which the UK Government and the Devolved Administrations are making progress towards their vision of achieving clean, healthy, safe, productive and biologically diverse oceans and seas.

The Charting Progress 2 Healthy and Biological Diverse Seas Feeder report (UK Marine Monitoring and Assessment Strategy, 2010) includes assessments that range from those based on expert judgement (due to insufficient data and/or lack of appropriate assessment methodology) to those based on more-formal assessment protocols. The assessment protocol was usually determined by the amount and type of data available, relating to both the ecosystem component being assessed as well as the associated pressures. For certain components and pressures, data are available over wide geographical areas and temporal scales, whilst for others this is not the case. Expert judgement was applied across a number of components of the report including:

- Benthic habitats;
- Seals and turtles;
- Cetaceans; and
- Marine birds.

[8] However, it should be noted that in similar studies other authors have found lower levels of agreement. For example Thomson *et al*. (2012) found only low agreement between experts using best professional judgement to assess benthic condition in the San Francisco Estuary and Delta.

SWOT Analysis

Strengths	**Weaknesses**
Use of expert judgement is declared up-front as being necessary due to insufficient data and/or lack of appropriate assessment methodology; Combines expert judgement with assessments based on more formal assessment protocols. For benthic habitats, the expert judgement assessment process was based on the peer reviewed work of Robinson *et al.* (2008).	Limited information on expert judgement elicitation process; Only one four-day workshop used to elicit expert judgement – no suggestion of subsequent validation.
Opportunities	**Threats**
Whilst highlighting the need for scientifically robust monitoring programmes the report nevertheless makes good use of expert judgement and presents a good precedent for its use in national reporting.	Suggestion that confidence in assessments based on 'expert judgement' is generally low. Together with the suggestion that it is not possible to produce accurate statements on trends in ecosystem components (or pressures affecting them) without information from scientifically robust monitoring programmes facilitating robust, data-led approaches to assessments, this sentiment may undermine adoption or acceptance of expert judgement methods in the future.

Practical Aspects

Benthic habitat assessments underpinning the Charting Progress 2 Feeder Report on Healthy and Biologically Diverse Seas (UK Marine Monitoring and Assessment Strategy, 2010) were based on the best-available information on current and historical habitat distribution and extent, and the spatial distribution and variability in intensity of pressures. Benthic habitat maps derived from survey data cover just 10% of the UK continental shelf. Where this information was unavailable, or there was an interaction between pressures, expert judgement was applied to complete the assessments[9]. The use of expert judgement is generally reflected in a low confidence ranking in the assessment[10].

Separate assessments were made for six broad habitat categories: intertidal rock; intertidal sediments; subtidal rock; shallow subtidal sediments; shelf subtidal sediments, and deep-sea habitats. Most of these were based on expert judgement, considering the relationship between habitats and pressures, and drawing upon limited evidence from monitoring studies and research. The assessment of broad habitats was completed at a four-day workshop (November, 2008) and was based on the methodology described in Robinson *et al.* (2008). The expert panel comprised representatives from the environmental, conservation and fisheries agencies, together with members of the academic community and independent consultants. Six expert groups were established, one for each of the six broad habitat types, and each group assessed their habitat against each individual pressure for each regional sea. In each case, the current status of the habitat was judged relative to an expert view regarding likely former natural conditions (i.e. condition in the absence of human pressures),

9 No available method was able to incorporate the wide range of indicators in use across different benthic habitats; consequently the assessment largely relied on expert judgement.
10 Confidence was rated as 'low' when limited or no supporting data were available and assessments were based largely on expert judgement. Confidence was rated as 'high' in all other circumstances. The assessment against each pressure was supported by an audit trail, documenting the sources of the information used. Full details of the methodology used can be found in Robinson *et al.* (2008). The assessment process did not involve any kind of statistical analysis of trends, largely due to the lack of available data that could be used for such a process.

the trends over the past ten years, and the prospects for the next two decades. The results for each broad habitat were aggregated, using a set of specific rules, to provide an assessment of the overall status of the habitat in each regional sea.

Similar to the assessment of benthic habitats, the Charting Progress 2 Feeder Report on Healthy and Biologically Diverse Seas based the assessment of cetaceans on expert judgement, using mainly the 2007 Favourable Conservation Status (FCS) assessments of all cetacean species occurring in UK waters. Similarly, it based assessments for marine birds (including the magnitude of the impacts of pressures from human activities on seabirds and waterbirds, encompassing 22 pressures across eight broad pressure themes) on expert judgement, supported by published evidence where possible.

Statistical and Methodological Aspects

A simple method of aggregating multiple pressures was used within the Charting Progress 2 Healthy and Biological Diverse Seas Feeder report (UK Marine Monitoring and Assessment Strategy, 2010) to arrive at an overall assessment of the status of the habitat. The report notes that it was not possible to apply anything other than a basic approach and expert judgement to cumulative pressures and impacts, and identifies this as an area for future method development. However, the process was supported by a comprehensive assessment audit trail and a series of confidence statements regarding the expert judgements made.

Opportunities or Insights

As might be anticipated, the Charting Progress 2 Feeder Report on Healthy and Biologically Diverse Seas (UK Marine Monitoring and Assessment Strategy, 2010) identifies numerous areas where further research is needed. However, there are some common themes throughout the report: the need to fill gaps in monitoring (in terms of geographical and temporal coverage, as well as the parameters being measured) and the need for better integration between monitoring programmes being the most frequently occurring. The report notes that it is not possible to produce accurate statements on trends in ecosystem components or the pressures affecting them without having information from scientifically robust monitoring programmes facilitating robust, data-led approaches to assessments rather than the current reliance on expert judgement.

3.3.13 **SOE 2011: National Marine Condition Assessment – Decision Model and Workshops (Ward, 2011)**

Ward produced the 2011 Australia State of the Environment (SoE) decision model and workshop reports for marine condition assessment over Australia's five marine regions.

The approach that was adopted followed an iterative process of opinion updating (the Closure Method) which is closely allied to the Delphi Method. At a series of workshops (each covering a different Australian marine region: South-west; North-west; North; East; and South-east) all scores and comments on the state of the environment produced by the participating experts were entered to an on-screen spreadsheet. In this way, all recorded information could be checked and verified by participants in real time. The full draft spreadsheet of raw data was circulated to all participants by email after the workshop, for subsequent checking and further verification. Where appropriate, further information, references and so on could be included and returned to the co-ordinator.

Participants were selected on the basis of consultation with agencies and universities, by reference to earlier national workshop activities, and in consultation with the Marine Division of the Department of Sustainability, Environment, Water, Population and Communities (SEWPaC).

At each workshop views were elicited on a range of aspects ('components') relating to the marine environment: Biodiversity; Quality of habitats; Species populations; Ecological processes; Ecosystem health; Physical and chemical processes; Pests and diseases; and Pressures. Participants contributed to group discussions and provided scores that best represented their judgement about the condition and trends of each component. Scoring was on a scale from 0 (poor or low) to 10 (very good or high). For each component in each region, three scores were provided. These related to consideration of:

- the worst 10% of sites;
- most of the sites; and
- the best 10% of sites.

The baseline (point of reference) for each assessment was taken as being the condition that would have existed prior to the changes in type and intensity of use/exploitation that accompanied settlement of Australia by Europeans.

Participants also assigned the median judgement to a quartile (equivalent to one of four reporting bands) as well as, for estimates of condition, trend, and importance of factors affecting the environment, recording their confidence in the evidence base they used to make their judgements. Participants were advised that this uncertainty should cover all aspects of the evidence base (including: technical quality/robustness; spatial and taxonomic coverage; process uncertainty; all forms of model uncertainty; and access to appropriate levels of detail). Participants were requested to directly assign the confidence surrounding their estimate as: high (adequate high quality evidence and high consensus); moderate (limited or low quality evidence or limited consensus); or low (evidence and consensus too low to make an assessment).

Scores were used to develop 'radar plots' showing, for each aspect considered, the estimated state or condition in each of the five marine regions.

SWOT Analysis

Strengths	Weaknesses
Incorporated relative estimates from participants (experts) regarding uncertainty in the knowledge base for each component; Although assessments for the five different regions considered were made across three workshops, cross-workshop coherency of approaches and scoring assumptions was facilitated by several key experts who attended all workshops, and the full assessment process (including the workshops) was moderated by the SoE Committee and the SEWPAC SoE Team to ensure consistency of scoring approaches and findings across the workshops. Closure Method produces defined outputs that lend themselves to simple graphical representation.	Equal weighting applied across all scores/gradings (assumed by the author to be a reflection of the diversity and expert standing of the participants in the assessment process); Because of the timescales involved, the choice of baseline (conditions pre-European settlement) may result in some ambiguous judgements.
Opportunities	**Threats**
No opportunities identified.	No threats identified.

Practical Aspects

Participants in the Australian SoE reporting described by Ward (2011) were selected on the basis of consultation with agencies and universities, by reference to earlier national workshop activities, and in consultation with the Marine Division of the Department of Sustainability, Environment, Water, Population and Communities (SEWPaC).

Statistical and Methodological Aspects

The SoE reporting described by Ward (2011) notes that there was significant up-front investment in the elicitation work (in terms of preparation and execution of several workshops). The level of effort invested in the elicitation work would be justified if the framework and process were to be ‘recycled’ (e.g. for subsequent rounds of SoE reporting) in which case results (and their interpretation) would be directly comparable between rounds. Their approach incorporated a measure of confidence to accompany the judgements but did not go as far as to provide ranges around estimates. No post-elicitation validation was undertaken although opportunity was provided for participants to review and modify their input following the workshops.

Opportunities or Insights

On similar lines, Ward’s (2011) methodology was applied to relatively large marine areas (the five marine areas around Australia) and would transfer to other geographic areas. However, the outputs as presented are relatively low in resolution. The summary statistics presented by Ward are useful, clear and concise; similar outputs would readily facilitate the transparent communication of findings of other studies to a non-expert audience.

3.4 Summary

The case studies reviewed above are summarised in Table 4. They demonstrate that a number of approaches to the use of expert judgement have been applied in relation to biodiversity assessments. Whilst no one study can be isolated as an example that may be directly re-applied to a UK marine biodiversity assessment scenario, a number of elements can be identified that may contribute to the development of an appropriate framework or methodology to address such a need (see Section 4 of this report).

Table 4. Summary of analysis of case studies

	Case study:												
	Bashari *et al*., 2009	Certain *et al*., 2011	Cowling *et al*., 2003	Enquist & Gori, 2008	Halpern *et al*., 2007	Kleypas & Eakin, 2007	McBride *et al*., 2012	MPA Monitoring Enterprise, 2012	Muxika *et al*., 2007	Teck *et al*., 2010	Teixeira *et al*., 2010	UK MAAS, 2010	Ward, 2011
Spatial coverage	Regional Queensland, Australia	National Norway	Regional Cape Floristic Region, South Africa	Regional US/Mexico border	Global	Global	National Australia	n/a	Regional Basque coast, North Spain	Regional US Pacific coast	(Global) US & Europe	National UK (11 regional seas)	National Australia (five marine regions)
Environment	Terrestrial	General	Terrestrial	Terrestrial	Marine	Marine	Terrestrial	Marine	Marine	Marine	Marine	Marine	Marine
Topic of expert judgement	Rangeland progression: potential development of cleared ironbark-spotted gum woodland	Habitat status: Assessment of nine major Norwegian habitat types x4 coast & marine; x2 wetland & freshwater; x3 terrestrial	Identification of potential conservation sites	Grassland: delineation of extent of five defined grassland condition types	Vulnerability assessment: vulnerability of 23 defined ecosystems to each of 20 defined threats	Threat evaluation: relative importance of 39 defined threats to coral ecosystems	Extinction risk: evaluation of (up to) 125 defined parameters relating to each of nine bird taxa	n/a	Site quality (relative to WFD): Evaluation of samples from 49 sites (based on benthic community and structural data)	Marine ecosystem impacts: evaluation of potential impacts of 53 defined stressors on 19 defined marine ecosystems	Benthic condition assessment: relative quality of 48 benthic samples	Status of six broad marine habitat categories: assessment of the habitat status in each of 11 UK regional seas	Assessment of eight aspects of ecosystem health across each of five SoE regions
Expert panel size	n/a	>120	≥19	24	135	286	16	n/a	3	107	16	n/a	n/a

	Case study:												
	Bashari *et al.*, 2009	**Certain *et al.*, 2011**	**Cowling *et al.*, 2003**	**Enquist & Gori, 2008**	**Halpern *et al.*, 2007**	**Kleypas & Eakin, 2007**	**McBride *et al.*, 2012**	**MPA Monitoring Enterprise, 2012**	**Muxika *et al.*, 2007**	**Teck *et al.*, 2010**	**Teixeira *et al.*, 2010**	**UK MAAS, 2010**	**Ward, 2011**
Source of expertise	Livestock owners & rangeland scientists	'Recognised national experts'	Park managers	Experts from federal/state agencies, institutes, etc	Scientists (authors of relevant publications listed in Web of Science)	Delegates to international symposium and members of relevant international society	Experts (identified by track record, experience, local knowledge, specialist skills)	Suggests selection criteria are pre-defined and an open and transparent selection process is used	n/a	525 potential respondents invited – initial identification through web-based searches on specific key words, authors' own knowledge and suggestions from previously identified experts	Benthic ecology experts	Experts from environment fisheries and conservation agencies, academia and private sector consultancy	Experts identified through consultation with agencies and universities, through earlier national workshops and in consultation with government
Principal means of expert judgement elicitation	Workshops	Web-based	Email (?) and workshop	Face-to-face interview	Web-based	Web-based	Email and telephone	Suggest initial email elicitation followed by workshop to discuss and derive final conclusion	Email (?) - direct evaluation of sample data by each expert	Survey – phone, online or face-to-face	Email - data for sample evaluation sent to each expert	Four-day workshop	Series of five (regional) workshops
Practicality of (expert judgement) method(s): subjective assessment of whether the methods are simple and easily repeated	Moderate: simple expert selection process coupled with workshop-based expert elicitation	Moderate: web-based approach increases practicality	Moderate: simple methodology to elicit expert judgement but approach elicits only very subjective views	High: described good process for eliciting geospatial information combined with assessments of ecological status	Moderate: web-based approach enables wide group of experts to be gathered but initial identification process may have implication regarding its overall repeatability	Moderate: web-based approach enables wide group of experts to be gathered but initial identification process may have implication regarding its overall repeatability	Moderate: overall process was time consuming (the overall process taking six months to complete for just 16 experts)	n/a	Low: expert judgement only used to validate alternative approaches	Moderate: relatively high manpower investment required	High: expert judgement was easily elicited given the prior identification of the experts being used	Moderate: Approach appears reasonable although information on workshop structure not available	High: relatively basic expert selection process coupled with workshop approach and development of simple consensus or aggregated view

	Case study:												
	Bashari *et al.*, 2009	**Certain *et al.*, 2011**	**Cowling *et al.*, 2003**	**Enquist & Gori, 2008**	**Halpern *et al.*, 2007**	**Kleypas & Eakin, 2007**	**McBride *et al.*, 2012**	**MPA Monitoring Enterprise, 2012**	**Muxika *et al.*, 2007**	**Teck *et al.*, 2010**	**Teixeira *et al.*, 2010**	**UK MAAS, 2010**	**Ward, 2011**
Transparency and comprehensibility: subjective assessment of the degree to which the (expert judgement) process could be readily communicated to decision-makers, general public, etc.	High: process, although technical and involving statistical modelling software, is easy to communicate (esp. due to graphic nature of underlying concepts)	Moderate: overall process quite involved	High: process used for, and outputs from, the expert elicitation, were very simple and therefore is easy to communicate	High: approach is easily understood and can be used to generate outputs that are easy to communicate	High: both the approach and the outputs can be easily described and are relatively easy to communicate	High: easily understood	Moderate: asked relatively simple questions and ultimately provided simple outputs although some of the methods used (e.g. fuzzy triangle numbers) may need to be presented in a sympathetic manner	n/a	Moderate: use of expert judgement to validate alternative (statistical) assessment methods easily explained but only of moderate transferable value	High: clear process producing simple outputs (relative vulnerability of different ecosystems to stressors)	High: information elicited from identified experts was relatively simple and the outputs could be easily explained	Moderate: overall process quite involved although underlying concepts can be presented graphically so making overall approach much easier to communicate	High: relatively simple approach
Robustness: subjective assessment of the degree to which the (expert judgement) approach is able to reduce bias and inconsistency, and assure scientific quality	High: use of BBN allows for production of probabilistic outputs	Moderate: application of weightings (for specific indicators, ecosystems or spatial units) allows for certain indicators, ecosystems or localities to impart greater or lesser bias to outputs	Low: elicitation and use of expert judgement is not undertaken in a particularly sophisticated manner; no account made of between-expert differences or of how opinions were combined	Moderate: no information is provided on how expert judgements are combined	Moderate: some: between-expert variability in judgement is discussed but not accounted for	Low: although demographics of experts (respondents) are discussed, other than directing respondents to 'the region with which they were most familiar', any resultant bias was not accounted for	High: elicitation process was designed to offset a range of predictable biases[11] (albeit with varying levels of success)	n/a	Low: expert judgement taken as being used to revise classifications derived using alternative methods; success of this approach reliant on experts selected (no details on this aspect were provided)	High: undertakes an assessment of potential bias amongst respondents	Moderate: inconsistency was assessed (in that the central theme was one of attempting to generate consensus); other than in the choice of experts (and no significant information is provided on this) there was no attempt to address bias	Moderate: approach is seen as best option given current knowledge base, although information on approaches used to elicit expert judgement was lacking	Moderate: information lost in the process (e.g. simple aggregated outputs from each of the five regional workshops are presented; detailed information from the experts is not presented alongside the main assessments)

11 Including anchoring, dominance, overconfidence, framing effects, availability bias, language-based misunderstanding, between-expert variability and convergence

	Case study:												
	Bashari *et al.*, 2009	**Certain *et al.*, 2011**	**Cowling *et al.*, 2003**	**Enquist & Gori, 2008**	**Halpern *et al.*, 2007**	**Kleypas & Eakin, 2007**	**McBride *et al.*, 2012**	**MPA Monitoring Enterprise, 2012**	**Muxika *et al.*, 2007**	**Teck *et al.*, 2010**	**Teixeira *et al.*, 2010**	**UK MAAS, 2010**	**Ward, 2011**
Do the methods used allow the quantification of confidence associated with the expert judgements that are elicited?	No: although variability is effectively modelled through subsequent application of a BBN	Yes: experts requested to record 50%ile range around mean estimates	No	No	Yes: experts provide measure of certainty for each response	No	Yes: uses triangular fuzzy numbers to capture uncertainty	Highlights need for process to acknowledge where there is uncertainty in experts' conclusions	No	No	No: not directly, but considers between-expert variability	No: although includes an audit trail that includes confidence statements regarding judgements made	Yes: a measure of confidence is presented alongside aggregated (consensus) expert views, but confidence is not incorporated in anything other than a semi-quantitative manner
Do the methods include some validation of outputs?	Yes: BBN model behaviour tested using scenario and sensitivity analysis	No	Yes: expert judgement is compared with algorithm-based outputs	Yes: field assessment is used to validate judgements	Yes: level of consistency was assessed between top threats as derived by direct questioning and by process used to elicit, collate and combine expert opinion	No	No: other than a comparison of outputs before and after group discussion to reconfirm or modify judgements	n/a	No: expert judgement is not validated	No	No: other than assessment of level of agreement between independent experts	No	No
Is scale of application similar to, or applicable to, that for UK (MSFD) marine benthic habitat assessments	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

	Case study:												
	Bashari *et al.*, 2009	**Certain *et al.*, 2011**	**Cowling *et al.*, 2003**	**Enquist & Gori, 2008**	**Halpern *et al.*, 2007**	**Kleypas & Eakin, 2007**	**McBride *et al.*, 2012**	**MPA Monitoring Enterprise, 2012**	**Muxika *et al.*, 2007**	**Teck *et al.*, 2010**	**Teixeira *et al.*, 2010**	**UK MAAS, 2010**	**Ward, 2011**
Do any of the underlying methodologies present possible opportunities for modification and incorporation to other paradigms?	Yes: BBN approach	Yes: use of scaling model approach to deriving 'common scale' quantitative estimates	Yes: use of weightings within model development post-elicitation	No	No	No	Yes: fuzzy numbers present a simple and graphic means to capture uncertainty; bias reduction (and assessment techniques) may be useful	Yes: presents useful process guidelines	No: alternative approaches discussed are reliant on sampling data and are not of relevance regarding potential application to data-poor environments	Yes: including derivation and use of weightings for individual vulnerability criteria (spatial scale, frequency, trophic impact, percentage change and recovery time)	No: although the good levels of agreement seen between experts provides ancillary support to the use of expert judgement in other studies	Yes: applicable to benthic habitat assessment under MSFD	No: with exception of the use of live consensus scoring in workshops)

4 Recommendations of Good Practice When Using Expert Judgement

4.1 Introduction

The review of 13 selected case studies showed that expert judgement has been successfully applied to a number of aspects of biodiversity assessment (see Section 3 of this report). Whilst no one study has been identified as an example that may be directly and fully re-applied to UK benthic habitat assessments (such as for reporting under the Marine Strategy Framework Directive, MSFD), a number of methods and approaches that either involve or relate to expert judgement have been identified as being of potential value as regards to their contribution to a framework or process to address such assessments. These methods and approaches, which may relate either to obtaining or applying expert judgement, are summarised in Table 5. Following on from the SWOT analyses and review of selected case studies using expert judgement in biodiversity assessment, as reported in Section 3 of this report, recommendations have been developed outlining what might constitute good practice when using expert judgement.

Table 5. Expert judgement methods and approaches potentially applicable to the process of undertaking marine biodiversity assessments

Role	Method/approach
Obtaining expert judgement:	Approaches for identifying experts (including numbers involved and background/credentials)
	Use of web-based approaches for elicitation
	Use of 'live' consensus development in workshop environments
	Use of fuzzy logic to capture uncertainty
	Use of map-based approaches (linked with GIS) to capture geospatial information from experts
Applying expert judgement:	Potential for use of Bayesian belief networks (BBNs) in developing quantitative (probabilistic) links between activities and the pressures they can cause, and between pressures and the impacts they may have on sensitive receptors such as species or habitats
	Use of scaling techniques to 'normalise' values[12]
	The need to account for bias in experts' responses
	Generation and use of weights

Each of these methods and approaches relating to the elicitation and application of expert judgement is discussed below, outlining good practice where possible and clarifying each aspect's potential role within marine biodiversity assessments. Where appropriate, principles that could be adopted for marine biodiversity status assessments, and the method(s) that could be used to implement these principles, are outlined. In cases where alternative methods are available an attempt is made to identify a preferred method.

[12] Whilst the use of scaling techniques, as discussed by Certain *et al*. (2011), may have value for specific applications and should be seen as a useful technique that can incorporate expert judgement, its specific value within a framework for biodiversity assessment in UK marine waters is not immediately clear. Consequently, scaling methods are not discussed further within the scope of this report.

At this stage it is not intended that all of these methods or approaches should be used in all cases where expert judgement is to be applied (Section 5 will examine what techniques should be applied to different specific stages or phases within a marine habitat biodiversity assessment process). Rather, they should be considered as possible tools within an ‘assessment toolbox’.

4.2 Obtaining Expert Judgement

4.2.1 Process Overview

The MPA Monitoring Enterprise (2012) report provides a set of guidelines for designing a process to elicit expert judgement. Studies from other, unrelated, fields of research have produced similar guidance. For example, Simola *et al.* (2005) and Knol *et al.* (2010) (working on structural integrity issues and public health respectively) present alternative schemata typical of an expert judgement process. Referring specifically to their own process description, Simola *et al.* (2005) suggest that, although other references may present slightly different lists of phases, the underlying procedure is essentially similar. By way of example, the general overlaps or commonalities between the three processes are shown in Table 6.

Table 6. Comparison of general guidelines for expert judgement processes

<table>
<tr><th>MPA Monitoring Enterprise (2012)</th><th>Simola et al. (2005)</th><th>Knol et al. (2010)</th></tr>
<tr><td rowspan="2">Step 1 - Define the scope and desired outcomes of an expert judgement process</td><td rowspan="2">Step 1 - Identification and selection of issues about which judgements of experts should be made</td><td>Step 1 - Characterisation of uncertainties</td></tr>
<tr><td>Step 2 - Scope and format of the elicitation</td></tr>
<tr><td>Step 2 - Select the experts</td><td>Step 2 - Identification and selection of experts</td><td>Step 3 - Selection of experts</td></tr>
<tr><td rowspan="5">Step 3 - Solicit, and work with, expert judgement</td><td rowspan="2">Step 3 - Training of experts and definition of variables to be elicited</td><td>Step 4 - Design of the elicitation protocol</td></tr>
<tr><td>Step 5 - Preparation of the elicitation session</td></tr>
<tr><td>Step 4 - Individual work of experts</td><td rowspan="2">Step 6 - Elicitation of expert judgements</td></tr>
<tr><td>Step 5 - Elicitation</td></tr>
<tr><td>Step 6 - Analysis and aggregation of results and, in case of disagreement, attempt to resolve differences</td><td rowspan="3">Step 7 - Possible aggregation and reporting</td></tr>
<tr><td>Step 4 - Communicate results</td><td rowspan="2">Step 7 - Documentation of results, including expert reasoning in support of their judgement</td></tr>
<tr><td>Step 5 - Develop a process for broader engagement</td></tr>
</table>

In general terms the procedures outlined in Table 6 are not too dissimilar; differences are largely due to the emphasis on, and apparent detail within, each of the stages. Especially considering that these process descriptions are for guidance only it is not expected that any one of the procedures is likely to have significant advantages over the others. However, as the procedure presented by the MPA Monitoring Enterprise (2012) relates specifically to marine ecosystem condition assessment and is supported by quite detailed text (given as Appendix 2 of the MPA Monitoring Enterprise (2012) report) this particular process is considered further in the context of the current review, and has been used to provide the basis for a modified process, as described in Table 7.

As well as taking due consideration of the main elements of Table 7, consideration should be given to confirming that any resultant process works successfully. Teck *et al*. (2010), for example, tested a preliminary draft of the survey instrument developed for their work (a structured questionnaire) on a separate group of experts. Whilst the outputs from this test elicitation were not used in subsequent formal analyses or interpretation, feedback and discussion was used to improve the overall process.

Table 7. Key guidelines for consideration under different stages of a process involving the use of expert judgement (main substantive elements to left of table; supplementary or ancillary elements to right of table)

Defining the scope and desired outcomes of an expert judgement process	
Defining the function of expert judgement	
· Clearly articulate how the results will be used (e.g. to inform management decisions, to educate the public) · Define the goals and the broad context for the process · Ground the process in a well defined question, so the experts understand what they are being asked to evaluate	
Defining the role of the expert	
· Identify the role of experts and the expert judgement process (e.g. to make decisions, provide recommendations, synthesize data)	· Define additional ancillary roles for the experts (e.g. lending credibility to results, achieving buy-in among stakeholders)
Defining the desired end point	
· Represent a diversity of opinions; consensus is not necessarily needed · Acknowledge when there is uncertainty in conclusions	
Selecting the experts	
Selecting experts	
· Define the criteria for selecting experts, with consideration of the scope of the question · Select experts using an open and transparent process	
Identifying the types of knowledge needed	
· Ensure that a diversity of perspectives is represented within a given discipline	· Consider different perspectives taken by different disciplines concerned with the topic

Soliciting and working with expert judgement	
Deciding on appropriate methods for eliciting expert judgement	
· Consider the appropriateness of quantitative versus qualitative assessments · Identify and select elicitation processes that employ individual assessments and/or group discussions as appropriate · Identify whether consensus is the goal, and structure workshop discussions appropriately · Consider potential sources of bias and structure methods to minimise these as far as possible · Provide mechanism for capturing level of confidence in opinions or assessments	· Define a mechanism to handle disagreements
Identifying the source of information that will form the basis of the assessment	
· Identify the information available and the criteria for inclusion (e.g. peer reviewed, state-funded data)	· Clearly define a role for non-traditional sources of information (e.g. traditional ecological knowledge, local knowledge of fishing activities)
Reviewing the results of the assessment	
· Include participant review as an early review step to ensure that their views are accurately represented	· Employ a process for external review that builds credibility and legitimacy employing a combination of a formal peer-review and less formal review by key stakeholder groups as appropriate
Communicating results	
Developing an appropriate communications strategy	
	· Identify the targeted audience(s) as that will inform all aspects of a communications strategy · Tailor communications for different audiences, but ensure message consistency
Identifying important characteristics of a reporting tool	
· Design and employ reporting tools that display information in a way that is intuitive/accessible · Engage communications experts, or members of the target audience, early on in the assessment process · Display results in a way that reflects multiple types of uncertainty and variation · Articulate, alongside the reporting tool, information about the process, including evidence used, assumptions, caveats, data gaps, uncertainty	
Deciding on authorship of a report (the report includes more than the ecosystem condition assessment)	
	· Identify authors and their role
Developing a process for broader engagement	
Identifying the groups/communities that should be brought in to the process	
· Find and engage 'critical friends'	· Design a process to engage other groups (beyond the experts) likely to be interested in the results of the assessment (stakeholders, managers, decision-makers)
Identifying at what stage other groups should be engaged and potential mechanisms for engagement	
· Consult decision-makers early in the process to ensure salience in situations where the results will be used to inform management	· Consider the broader social and political context in which the results will land when timing stakeholder engagement · Identify and employ mechanisms for most effectively engaging with other groups (e.g. workshops, briefings)

Recommendation 1.
The guidelines outlined in Table 7 should be considered when developing a process that will employ expert judgement. Whilst not all elements within the table will be applicable in all instances, and some have greater importance than others, they could usefully form a checklist against which the development of a robust expert judgement process may be set.

As noted, the detailed design of the elicitation process should be carefully considered with respect to minimising the potential impact of judgemental bias (see Section 4.3.2 for further information).

Having completed the design of the elicitation process, it should be tested in a 'dry run' using a small group of experts (ideally individuals whom it is not intended to involve subsequently in the formal elicitation process). Critical evaluation of the process should be used to improve the overall approach before its final use.

4.2.2 **Approaches for Identifying Experts (Including Numbers Involved and Background/Credentials)**

How Many Experts Do You Need to Make a Judgement?

The selected case studies made use of expert judgement from a wide range of expert 'panel' sizes, for example ranging from just three individuals (in the study reported by Muxika *et al*., 2007) up to 286 experts (in the study by Kleypas & Eakin, 2007). There was no clear suggestion in the reviewed case studies regarding what might be considered to be an appropriate panel size.

Whilst, in general terms, a panel comprising more, rather than fewer, experts might be seen as being the ideal, this needs to be balanced against practical constraints, not least of which is the availability of appropriate experts who are able to demonstrate an adequate level of expertise. Equally, after the number of experts has passed a certain level, one might anticipate diminishing returns as the further addition of experts to the pool simply 'dilutes' the data without adding any new information (Gehris, 2008). The fundamental question is one of how many experts are sufficient to ensure reasonable confidence that the addition of more experts will not substantially change the results.

Simulation studies reported by Gehris (2008) have suggested that there is a benefit in adding experts up to ten. This view is supported by the theoretical analysis of weighted arithmetic average combination methods (which, due to both their ease and simplicity, are commonly used to combine individual judgements).

Additionally, the number of experts will affect project costs (e.g. through an increase in organisational administration) and lead time (e.g. as the input from more individuals will have to be coordinated and, ultimately, more data will need to be analysed). Consequently, project resources will tend to dictate an upper limit of the number of experts that can be involved.

Recommendation 2.
It is suggested that, given the views of Gehris (2008), studies using expert judgement should seek to engage with at least ten experts. Where there is a distribution of experts across a relatively wide geographic region or across a number of institutions/backgrounds then consideration should be given to involving more individuals. However, even in such cases, an upper limit of 20 or so may be appropriate.

Identification of Experts, Including Expertise and Credentials

Expert judgements are a necessary part of environmental management. Typically, experts are defined by their qualifications, track record, professional standing, and experience (Burgman *et al*., 2011a). However, it is reasonable to seek clarity regarding the range of criteria that should be applied to the selection of experts. A person's formal training and technical knowledge (known as their 'substantive' expertise; Stern & Fineberg, 1996; Walton 1997) are often contrasted with the knowledge of people with no formal training (known as 'lay' knowledge). Expert judgements are attractive when time and resources are stretched, and are especially important where existing data are inadequate, circumstances are unique, or extrapolations are required for novel, future and uncertain situations (Burgman *et al*., 2011a).

Clearly, such experts should have some knowledge of the topic under consideration that is in some way over and above that of an informed layperson (i.e. they should have some 'expert' credentials). This expert knowledge may be the result of training, research, and skills (Burgman *et al*. 2011a), but should not exclude certain individuals, such as local practitioners and informed stakeholders, who although lacking a formal background may nevertheless be able to contribute important insight and valuable opinion (Burgman *et al*., 2011b).

Beyond that rather simplistic qualification, further recommendation becomes less straightforward with clear suggestions from the literature being hard to identify. The selected case studies included work that made use of opinion or judgement from experts identified according to a number of widely varying criteria. Some of these approaches identified a large number of experts, for example: authors of relevant publications (Halpern *et al*., 2007); delegates to a relevant international symposium and related international society (Kleypas & Eakin, 2007). Other reported approaches were more specific and gave rise to smaller panels of experts, for example: experts from federal/state agencies, institutions, etc. (Enquist & Gori, 2008).

The review by Burgman *et al.* (2011a) recommended a set of general prescriptions for the selection of experts:

(1) Identify core expertise requirements and the pool of potential experts, including lay expertise.
(2) Create objective selection criteria and clear rules for engaging experts, stratify the pool of experts and select participants transparently based on the strata.
(3) Evaluate the social and scientific context of the problem.
(4) Identify potential conflicts of interest and motivational biases and control bias by "balancing" the composition of expert groups, with respect to the issue at hand (especially if the pool of experts is small).
(5) Test expertise, relevant to the issues.
(6) Provide opportunities for stakeholders to cross examine all expert opinions.
(7) Train experts and provide routine, systematic, relevant feedback on their performance.

Pragmatically, whatever set of criteria are decided upon should be clear and transparent. They need to be general enough to encompass a reasonably wide set of experts but specific enough to ensure that, whilst the level of expertise is kept high, the numbers of potential experts returned is manageable. As noted above, it is recommended that the overall number of experts should generally be in the range of 10-20. In the context of UK marine biodiversity assessments it is suggested that this could be achieved by applying the type of selection criteria employed by Enquist & Gori (2008); McBride *et al* (2012); Aish *et al*. (2010) and by Ward (2011) (see Table 8).

Table 8. Source of expert opinion for selected studies

Citation	Source of expertise
Enquist & Gori (2008)	Experts selected from federal/state agencies, institutes, etc.
McBride *et al*. (2012)	Experts identified by track record, experience, local knowledge, specialist skills
Aish *et al*. (2010)	Experts selected from environment, fisheries and conservation agencies, academia and private sector consultancy
Ward (2011)	Experts identified through consultation with agencies and universities, through earlier national workshops and in consultation with government

Recommendation 3.
In the context of the application of expert judgement to UK marine biodiversity assessments, appropriate experts should generally be selected from the Statutory Nature Conservation Bodies (SNCBs), academia and private sector consultancy. The process of identification of experts from amongst these potential sources should be undertaken through consultation with agencies and universities, by reference to earlier workshops and in consultation with government; and should seek to identify experts with appropriate track records, experience and (where relevant) local knowledge and other specialist skills. A formal, transparent process for the definition and selection of those with relevant expertise should be adopted.

4.2.3 Use of Web-Based Approaches for Elicitation

A number of studies have made use of the internet as a ‘virtual meeting space’ for eliciting expert judgement (e.g. Halpern *et al*., 2007; Kleypas & Eakin, 2007, Certain *et al*., 2011;). The use of the internet confers a number of advantages, including cost, speed and audience penetration. It is particularly valuable where the intended expert audience is large, as the administration of alternative elicitation techniques (group meetings or face-to-face meetings) would be likely to prove difficult and time consuming.

The internet can be used in two distinct ways:

- As a means of questionnaire dissemination (i.e. delivery and collation by email); or
- As a direct surrogate for a face-to-face meeting, where questions are presented in the form of a web page[13].

Both approaches were evident amongst the selected case studies, and these types of web-based approaches tended to be employed where the number of participants was relatively high.

Whilst a web-based approach may be used as a surrogate for a face-to-face meeting, it may still be desirable, or even necessary, for the experts who are engaged to hold additional discussions over certain aspects of the elicitation. Such discussions can be held via email or by video-conferencing, maintaining the advantages of the ‘virtual’ meeting (e.g. reducing costs and increasing the potential pool of experts that are able to contribute) although

[13] A number of commercial web-based ‘survey’ or ‘questionnaire’ applications are available that could be used to good effect in eliciting individual-based expert judgement. Without recommending any particular product, possible options include:

- Survey Monkey - http://www.surveymonkey.com/
- Smart Survey - http://www.smart-survey.co.uk/
- Checkbox - http://www.checkbox.com/
- Dot Survey - http://www.dotsurvey.com/

discussions under such conditions can be stilted (e.g. McBride *et al*., 2012) and, as such can be hard to lead or facilitate and may be subject to domination by one or more individuals. Other disadvantages might include technical problems (including time delays) and time zone synchronisation issues.

Given the suggestion that the numbers of experts involved in an elicitation exercise can be relatively low (i.e. in the range of 10-20) then the benefit of using a 'virtual meeting space' is lost, and the practical benefits associated with face-to-face meetings (workshops) become more important. For example, McBride *et al*. (2012) recognised that, although the adoption of an email format reduced dominance (giving less confident and articulate participants a better opportunity to provide meaningful contributions with no single views dominating) and was able to provide a forum for open and balanced discussion, the absence of any ability to give or receive behavioural cues (as would be present within a face-to-face setting and which would be likely to convey greater nuance and tone to discussion and to assist in the reduction of misunderstandings) is an important drawback.

The value of a workshop environment was underlined by McBride *et al*. (2012) who concluded that, where panellists can gather without excessive long-distance travel, facilitated group workshops should be the tool of choice for expert elicitation where interactive discussion is likely to be necessary or desirable.

Recommendation 4.
Whilst the value of web-based approaches under certain scenarios is recognised, it is recommended that for expert judgement elicitation in the context of UK marine biodiversity assessment, where possible, workshop-based environments are employed instead.

4.2.4 **Use of Fuzzy Logic to Capture Uncertainty**

Although it is possible to restrict an elicitation process to record only 'best' estimates from experts, this assumes (or implies) that they each have a perfect knowledge of the system under consideration. In reality, each expert would have a degree of uncertainty relating to his or her judgements. It is therefore important, whenever experts are asked for quantitative estimates, to record some indication of the level of uncertainty associated with their responses. Whilst this can be done directly, e.g. by asking experts to quote a confidence interval around their estimates, such approaches can appear unfriendly to the participants. One useful alternative, which allows for uncertainty to be recorded whilst avoiding the need for it to be formally represented by a statistical distribution, is that presented by fuzzy numbers. McBride *et al.* (2012), for example, provide a framework for capturing and accounting for uncertainty in experts' opinions through the use of triangular fuzzy numbers.

The use of fuzzy logic (or fuzzy set theory) is primarily applicable to situations where quantitative values are being sought from experts. Developed in the 1960s as a means to model the uncertainty of natural language, fuzzy logic is a branch of conventional (Boolean) logic that embraces the concept of partial truth, i.e. truth values between 'completely true' and 'completely false' (Zadeh, 1965).

In their study, McBride *et al*. (2012) elicited expert judgement on quantitative values (quantities and percentages) through a four point procedure whereby experts were asked to give an estimate of the number of mature individuals of a given target species within a particular area. They were asked to provide:

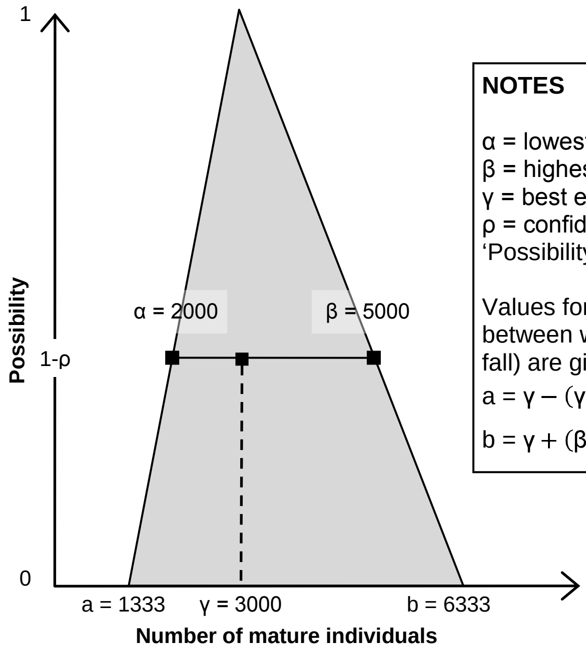
1. Their opinion on what the lowest value could be? (α);

2. Their opinion on what the highest value could be? (β);

3. Their best estimate of the most likely value (γ); and

4. An indication of their confidence that the interval they provided contains the truth (in the range of 50–100%)? (ρ)

These values (γ, α, β and ρ) were used to construct a fuzzy triangular number ([a,0], [γ,1], [b,0]); see Figure 4 for an example.



NOTES

α = lowest likely value
β = highest likely value
γ = best estimate
ρ = confidence that range (α to β) will contain the true value
'Possibility' = 1 - ρ

Values for a and b (the absolute lower and upper bound between which 100% of all estimates might be expected to fall) are given by linear extrapolation from α and β, such that:
$a = \gamma - (\gamma - \alpha)(\frac{1}{\rho})$; and
$b = \gamma + (\beta - \gamma)(\frac{1}{\rho})$

Figure 4. Quantitative triangular fuzzy numbers

A similar approach was used for binary (yes/no) questions, where a two-step procedure was used and experts were asked to indicate:

1. Whether they believed the statement to be true or false? (ı); and

2. How sure they were that their answer was correct (providing an answer in the range of 50–100%)? (p)

Each participant was able to refine their opinions through a modified Delphi process[14] (where each of the participants was able to review the reasoning behind the selection of values by the other experts before subsequently reconsidering their own values and submitting final selections). In presenting information at this stage the researchers 'standardised' the data using the triangular fuzzy numbers to provide 'best estimate' and 'min-max ranges' at a 0.8 level of confidence (i.e. they used the fuzzy triangle numbers to produce standardised 80% intervals around the best estimates).

---

[14] NB The approach was described by McBride *et al.* as being 'modified' as the review/feedback stage of the Delphi process was undertaken remotely, via email. A classic Delphi process would make use of a face-to-face meeting to provide this review/feedback information.

In the study by McBride *et al*. values for the 'most-likely' estimate and for the standardised upper and lower bounds from the elicitation process (i.e. after the submission of final, reconsidered vales) were compiled and averaged across the participating experts to provide a set of aggregate estimates for each parameter.

An alternative approach would be to 'de-fuzzify' or 'crisp' each fuzzy triangle number to produce a series of single values, representing each expert's opinion. Defuzzifying or crisping triangular numbers can be done in several ways but a simple, common and useful technique is by the 'centre of gravity (CoG) method' (e.g. Yager and Filev, 1993) - the 'crisp' output being given by the x-axis value of the CoG.

For example, taking the fuzzy triangle number shown in Figure 4, the CoG lies at:

$$[\frac{a+\gamma+b}{3}, \frac{0+1+0}{3}]$$

$$= [\frac{1333+3000+6333}{3}, \frac{0+1+0}{3}]$$

= [3555.3, 0.3]

Hence, the best overall estimate for the 'Number of mature individuals' from this example is 3555.

An alternative option for applying a fuzzy number approach is provided below. Table 9 shows hypothetical values as elicited from five experts. In each case, the elicited values of γ, α, β and ρ are used to construct fuzzy triangle numbers (Figure 5). The values for a and b are calculated (as shown in Table 9) and these are used, together with the original elicited value for γ, to estimate the x-coordinate of the CoG for each triangle.

Table 9. Example values from five experts

	'At least'	'At most'	Confidence		Best estimate	'Likelihood'	Minimum	Maximum	CoG
Expert	**α**	**β**	**%**	**ρ**	**γ**	**1-ρ**	**a**	**b**	**x**
1	1300	3500	60	0.6	2200	0.4	700	4367	2422
2	1800	2500	55	0.55	2100	0.45	1555	2827	2161
3	1850	2800	50	0.5	2300	0.5	1400	3300	2333
4	1400	2100	66	0.66	1900	0.34	11425	2203	1748
5	2250	3200	70	0.7	2500	0.3	2143	3500	2714
Average									**2276**

Either the mean or the median value from the range of separate best estimates for the 'true' value can then be taken forward as the definitive output from the elicitation. Additional value might be gained by presenting the range of best estimates as a box and whisker plot (e.g. Figure 6). Such plots provide a simple visual representation of the distribution of data (Massart *et al.*, 2005), showing the median value and the interquartile range (i.e. the 25th percentile to the 75th percentile) together with an indication of the wider spread of data (e.g. from the 25th percentile *minus* 1.5 x the interquartile range up to the 75th percentilee *plus* 1.5 x the interquartile range) and any additional outlying data points.

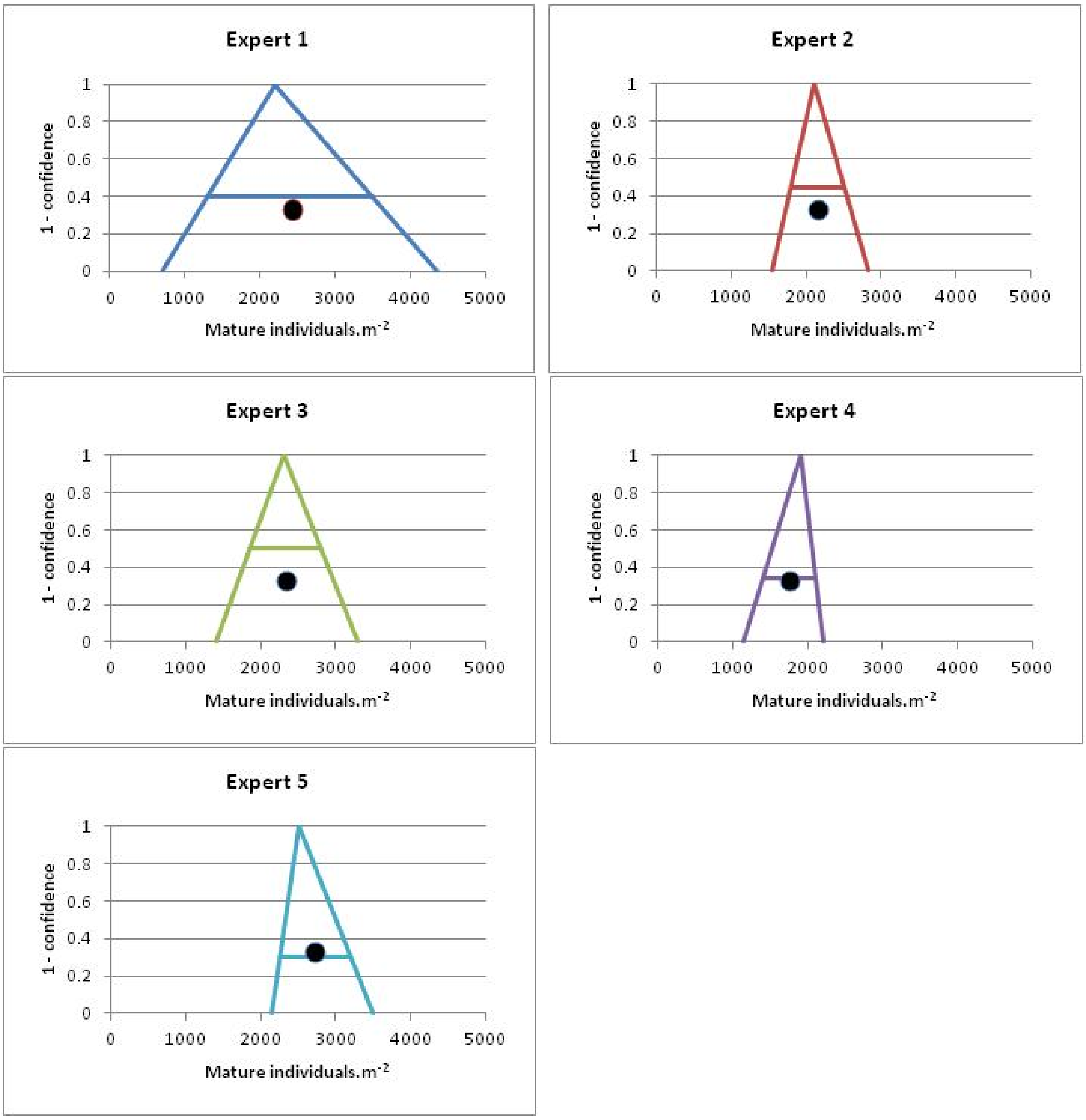


Figure 5. **Examples of constructed triangular fuzzy numbers using data from Table 9 (black markers indicate the calculated positions of centre of gravity) – see text for details**

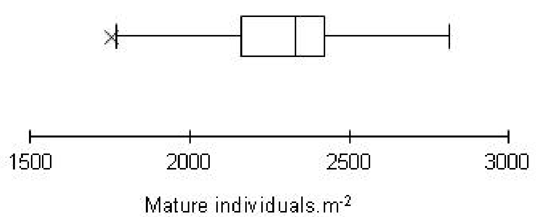


Figure 6. **Box and whisker plot showing distrubution of de-fuzzified estimates from across experts (see text for details)**

Recommendation 5.
The adoption of fuzzy triangular numbers to capture and use uncertainty information, employing a simple 'centre of gravity' (CoG) approach to defuzzification, should be considered within a framework for eliciting expert judgement in support of marine biodiversity assessment. The approach is valuable not just for quantitative values but also for categorical questions with the same CoG methodology potentially applying to averaged fuzzy values to derive crisp values for simple yes/no type responses.

4.2.5 Use of 'Live' Consensus Development in Workshop Environments

Where workshop environments are used to elicit expert opinion, the information produced by experts can be used in two different ways. Information from individual experts can be taken away as a series of independent assessments and collated by the team undertaking the study, or it can be refined by participants within a workshop environment to develop a consensus view. Both approaches have their benefits: the former providing useful information on, for example, between-expert variability; the latter providing a mechanism for deriving single 'best estimates' that can be confirmed or supported by the experts present.

The former approach is supported by the classic Delphi process, whereby initial estimates are made by each participant who then, subsequently, receives anonymous feedback regarding the estimates of the other participants. On the basis of this information they then make a second, (potentially) revised estimate. The estimate and feedback stages continue for a set number of rounds or until a pre-specified level of agreement between participants is reached. The Delphi process is well-established in ecology (McBride *et al*., 2012) and has the advantage over single stage questionnaires and unstructured groups of allowing experts to revise their judgements in the light of others in the group. At the same time the anonymity afforded to participants can alleviate some of the more obvious pressures that may result from unstructured group discussion settings (e.g. dominance, where less experienced group members conform to the views of the senior members of the group; or 'groupthink', where groups become more concerned with achieving consensus than with carefully considering their individual judgements). The McBride *et al*. (2012) report suggests that, to achieve improvements in accuracy from round to round, experts should be provided with rationales to accompany the feedback they receive about the responses from other group members; in the absence of these rationales, their responses will tend to converge only towards a majority position. The incorporation of discussion into the feedback stage of the elicitation is one natural and effective means for providing such rationales.

The latter approach, where consensus views are being sought, requires an alternative process to be adopted, such as that employed in the Australian SoE assessment workshops reported by Ward (2011). In developing consensus views, Ward made use of a dedicated meeting facilitator who entered all scores/comments to an on-screen spreadsheet such that all information could be checked and verified by participants in real time. In addition, all participants were circulated the full raw data draft spreadsheet by email after the workshop, for subsequent checking, further verification and the addition of further information or references. Meeting facilitators need to, as far as possible, maintain a position of neutrality and independence as regards the views being discussed. Notwithstanding this, there is considerable merit in facilitators being well informed regarding the topics being discussed. The successful use of such 'informed and independent facilitators' in workshop environments has been demonstrated recently in large-scale stakeholder engagement environments (e.g. Net Gain, 2011).

The use of interactive, on-screen tools has application to the classic Delphi process as well, as it allows for the effective reporting of rationales behind the decision-making process adopted by different participants.

Recommendation 6.
Where workshops are used, it is recommended that consideration be given to using interactive, on-screen tools for recording and discussing expert judgements. In addition, although information in the selected case studies is generally lacking regarding the arrangements made for expert judgement workshops, it is recommended that the choice of an 'informed' independent facilitator can be important to achieving a successful workshop.

4.2.6 Use of Map-Based Approaches (Linked With GIS) to Capture Geospatial Information From Experts

A simple approach to obtaining expert opinion on the geospatial extent of a specific feature or pressure is to ask individuals to outline the distribution on a map (or map overlay). Relevant orientation or interpretive markers (such as height or depth contours, infrastructure, etc.) can be provided in addition to standard grid markers. This approach was used by Enquist & Gori (2008) in their elicitation of information of the extent of certain grassland types in the USA–Mexico borderlines.

However, where geospatial information is being captured or verified through the application of expert judgement, there is merit in attempting to identify not only the spatial extent of (for example) features or pressures, but also the confidence with which it can be delineated. Consider an area (or polygon) that represents the spatial distribution of a particular feature. The data for this polygon can be readily stored in a Geographic Information System (GIS), but there should be some allowance made for possible uncertainty regarding the feature boundaries. In effect, the polygon boundary could be viewed as being fuzzy. Alternatively, and given separate assessments of the spatial extent of the feature by different experts, a probability of feature presence/absence can be built up. For example, Figure 7 shows a stylised polygon denoting the spatial extent of a hypothetical feature. Although we have a distinct boundary to the feature this is, in effect, simply an indication of the 'best guess' as to where the boundary really lies. For areas within the polygon we can be more certain that the feature is truly present, whereas the converse holds true for areas outside of the polygon.

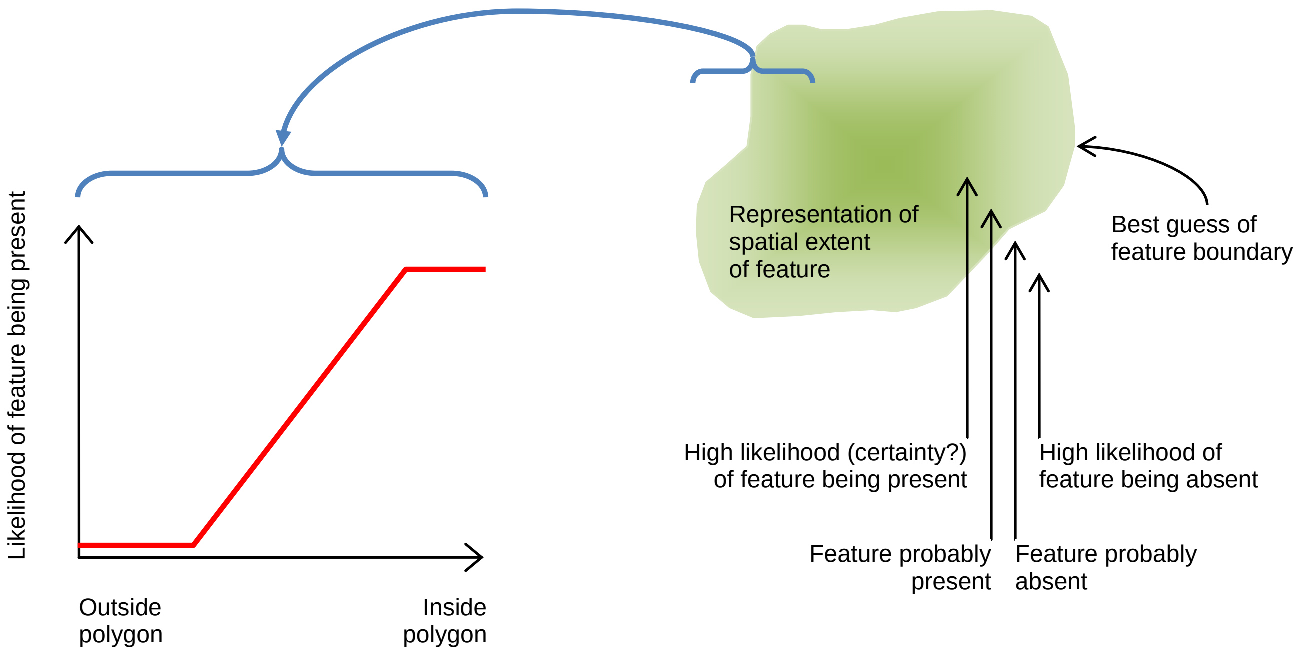


Figure 7. Boundary uncertainty for polygon data within a GIS

Consequently the subsequent interpretation and use of expert opinion on geospatial distributions, although relatively straightforward to elicit, needs to be considered carefully. Initially, information derived from expert judgement regarding the distribution of, for example, features or pressures should be digitised and stored in a suitable GIS in raster format[15].

In terms of digitising the elicited views, the distributions as suggested by each individual expert can be recorded as a simple presence/absence (coded for example as 1 or 0 within the raster framework) or, where experts have been asked to grade their response to include (for example) 'possible' as well as 'definite presence' and 'absence' a graded coding system can be employed (e.g. 2, 1, 0 respectively). When each respondent's distribution is recorded as a separate raster data layer it is possible to use the GIS database management system to combine them to a single layer – with data values relating to the mean score plus a measure of variability (e.g. standard deviation from the mean, usually represented as the Greek sigma, σ).

Whilst the mean values can be used to provide a 'likelihood' surface for the experts' composite view of the feature distribution (which would be, for example, in the range 0-1 or 0-2, dependent on the initial scoring approach used), the associated 'confidence' layer (the measure of between-expert' variability) is available, if necessary, for use as a weighting factor in future analyses.

[15] There are three basic types of data model used for storing geospatial data within a GIS: vector; raster; and image.

- **Vector** storage uses vectors (directional lines) to represent geographic features. Vector data is characterised by the use of sequential points in Cartesian space (i.e. X-Y coordinates, or vertices) to define each linear segment or 'arc'.
- **Raster** data models incorporate a grid-cell data structure where the geographic area is divided into cells identified by row and column. Whilst the grid-cell structure is most commonly square, other tessellations (such as regular hexagons) can be used. Although the term 'raster' implies a regularly spaced grid, other tessellated data structures do exist in grid based GIS systems. For example, the application of quadtree data structures, where the plane is recursively subdivided into four quadrants so that some areas have a higher resultant spatial resolution, has found some acceptance as an alternative raster data model.
- **Images** (typically pictures or photographs of the landscape, e.g. satellite imagery) use techniques very similar to those applied to raster data although the approach lacks the internal formats required for analysis and modelling of the data.

Whilst vector-based systems are often constrained by the capabilities and language of the underlying relational database management system, grid-cells can be handled as two-dimensional arrays in computer encoding and many analytical operations are easy to program. Consequently the use of raster data structures allow for more sophisticated mathematical modelling processes.
To facilitate subsequent analysis, geospatial data related to marine biodiversity assessment should, as far as practicable, be stored in raster format. To maintain appropriate spatial resolution of the raster data consideration should be given to using quadtree-based gridding as an alternative data model to regular gridding, although this aspect of the GIS specification should be discussed with the GIS developer when the system is being designed.

Recommendation 7.
Where expert opinion of the geospatial extent of features is being sought and analysed, potentially useful information regarding the level of agreement between experts (in effect a form of confidence in the estimates) should be retained by digitising each expert's set of opinions separately. Subsequent processing within the GIS will allow a composite view of the experts' opinions to be built up together with a single associated confidence layer.

Data should be stored to a GIS using a raster data model, with gridding set to a resolution appropriate to the features being mapped. Differential resolutions can be used (e.g. quadtree data structures) as an alternative to regular gridding (e.g. square or hexagonal cells).

4.3 Applying Expert Judgement

4.3.1 **Potential for Use of Bayesian Belief Networks**

Bayesian belief networks (BBNs; also known as belief networks, causal nets, causal probabilistic networks, probabilistic cause-and-effect models, and graphical probability networks) are graphical models incorporating nodes (boxes) and links (arrows) that represent system variables and their cause-and-effect relationships. BBNs consist of a qualitative part (a directed acyclic graph, or cause-and-effect diagram, composed of a series of linked nodes) and an associated quantitative part (a set of conditional probabilities quantifying the strength of dependencies between the variables represented in the acyclic graph).

Expert judgement has recently been used in the UK to inform and develop both a 'Sensitivity Matrix' (a pressures/features matrix, as produced by ABPMer and MarLIN) and an associated activities/pressures matrix (produced by the SNCBs). The sensitivity matrix identifies the sensitivity of given marine features to given pressures, whilst the activities/pressures matrix identifies the pressures that may arise from (human) marine activities. This information is invaluable as it brings clarity to, for example, discussions over assessments of features' status or potential management implications (see, for example, Natural England & JNCC, 2011).

However, when being considered as part of biodiversity vulnerability assessments by the SNCBs, the relationship between marine (human) activities and resultant pressures, and the impacts of these pressures on marine features, is handled in a deterministic way (where a particular given input will always produce the same output) (see, for example, Natural England & JNCC, 2011). Presenting a possible alternative paradigm, BBNs can be used in a predictive capacity; because they use probabilities to quantify relationships between model variables therefore they explicitly allow uncertainty and variability to be accommodated in model predictions. It is due to these features that they are increasingly becoming a popular modelling tool, particularly in ecology and environmental management.

It is possible to describe the activity/pressure/feature sensitivity relationships currently being considered by the SNCBs in terms of an influence diagram (a directed acyclic graph; the graphical component of a BBN). Figure 8 presents a simplified example. In this diagram, the occurrence of an activity (activity *i*) is linked to the generation of *j* pressures (1 to *j*; each of which is recorded at high, medium or low intensity). The impact of each of the pressures is then linked to a specific feature (feature *k*).

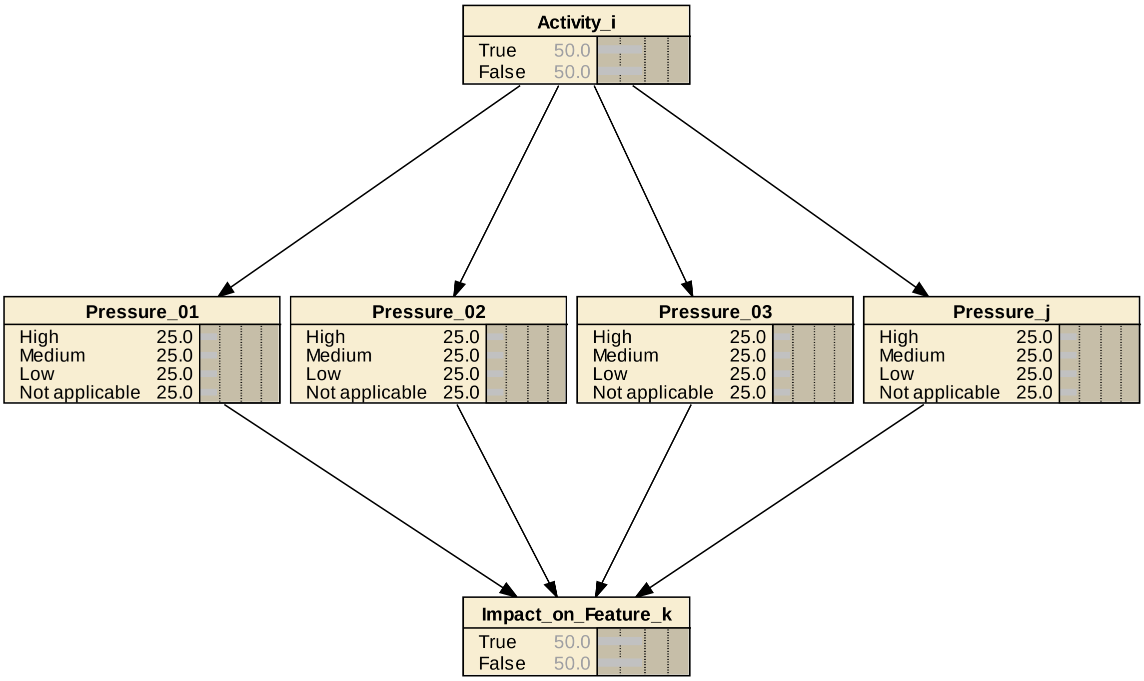


Figure 8. Simplified example of Bayesian belief network linking 'Activities' to 'Pressures' to 'Impacts' (see text for further detail)

The scope of the BBN can potentially be expanded to include the full range of activities and features, and may provide an effective way of representing combined effects.

The issue is how to describe the relationships between the nodes within the influence diagram. Although they can be defined using Conditional Probability Tables (CPTs), which store the probabilities of outcomes under particular scenarios (it is these probabilities that allow uncertainty and variability to be accommodated in model predictions), measured probabilities for the relationships (or transitions) between nodes can generally only be obtained from long-term studies. However, an alternative approach to deriving these probabilities is to use a series of observations (or 'cases') to inform the BBN, which will then identify the set of probabilities that best match the observations supplied. It is possible that these cases can, at their simplest, replicate expert judgement.

Without a greater understanding of the availability of existing information on these relationships (derived, for example, from expert judgement) – and without a more comprehensive grounding in the development of BBN models – it is not possible to provide more detail on their direct and immediate application it the context of marine biodiversity assessment in the UK. However, it is recommended that the potential role of BBN methodologies should be considered further.

Recommendation 8.
Further consideration should be given to the potential for using a BBN approach to derive a better probabilistic interpretation of the associations that exist between marine (human) activities and resultant pressures, and the impacts of these pressures on marine features.

4.3.2 The Need to Account for Potential Bias in Experts' Responses

Geographic Bias

The selection of experts should, to some extent, address the potential issues of geographical bias (experts should, as far as possible, be selected such that their knowledge and understanding is distributed evenly across the study area).

The case study presented by Teixeira *et al.* (2010) demonstrated an apparent absence of regional bias in expert judgement (their study employed 16 experts from across four geographic regions: European 'Atlantic', 'Mediterranean', and US East Coast and West Coast). After establishing levels of agreement using the equivalence table from Monserud & Leemans (1992), the authors applied Kappa analysis (e.g. Cohen, 1960; Landis & Koch, 1977) to assess condition category assignment among experts. As the importance of 'misclassification' would not be the same between close categories (e.g. between 'High' and 'Good', or 'Poor' and 'Bad') as between categories that lie further apart (e.g. between 'High' and 'Moderate', or 'High' and 'Bad'), Teixeira *et al.* (2010) applied Fleiss-Cohen weightings (Fleiss & Cohen, 1973) to their analyses.

Recommendation 9.
Where the selection of experts might introduce a geographic bias and this might affect the validity of any outputs, post-hoc analysis should be considered (e.g. using Kappa analysis) to assess whether geographic bias is actually in evidence. Where necessary, equivalence tables should be used to support his process and, if appropriate, Fleiss-Cohen weighting applied to account for the differential importance of mis-categorisation across different possible categories.

Response Bias

Care should be taken not to introduce bias into expert elicitation. It may be that those people with the most appropriate expertise to make the judgements may also have a stake in the outcome of the decision. Where this is the case, there is always the danger that their rating judgements may (perhaps unconsciously) be influenced by factors other than simply the performance of the options on the criterion being assessed. Ideally, such judgements should come from individuals who are both expert and impartial (DCLG, 2009). McBride *et al*. (2012) present a total of 11 different types of subjective response bias, a summary of which is presented as Table 10.

The design of the elicitation process should be undertaken in such a way as to, as far as possible, reduce the potential bias from each of these sources. This will entail careful review of the elicitation 'script' or questionnaire and the workshop design and process (if applicable). Bias may be mitigated by setting tasks that allow for deliberate practice, including unambiguous feedback, and phrasing questions for experts in such a way that they are aligned with an expert's knowledge. Several authors provide more extensive advice on managing elicitation bias (Meyer & Booker, 1991; O'Hagan *et al*., 2006; Kynn, 2008; Low-Choy *et al*., 2009).

Table 10. Subjective biases commonly encountered in expert elicitation (taken from McBride *et al*., 2012)

Bias type	Description	Illustration
Individual biases		
Anchoring	Final estimates are influenced by an initial salient estimate, either generated by the individual or supplied by the environment	People give a higher estimate of the length of the Mississippi River if asked whether it is longer or shorter than 5000 miles, than if asked whether it is longer or shorter than 200 miles
Anchoring & adjustment	Insufficient adjustment of judgements from an initial anchor, known to be incorrect but closely related to the true value	People's estimates of the boiling point of vodka are biased towards the self-generated anchor of the boiling point of water
Availability bias	People's judgements are influenced more heavily by the experiences or evidence that most easily come to mind	Tornadoes are judged as more frequent killers than asthma, even though the latter is 20 times more likely
Confirmation bias	People search for or interpret information (consciously or unconsciously) in a way that accords with their prior beliefs	Scientists may judge research reports that agree with their prior beliefs to be of higher quality than those that disagree
Framing	Individuals draw different conclusions from the same information, depending on how that information is presented	Presenting probabilities as natural frequencies (e.g. 6 subpopulations out of 10) helps people reason with probabilities and reduce biases such as overconfidence
Overconfidence	The tendency for people to have greater confidence in their judgements than is warranted by their level of knowledge	People frequently provide 90% confidence intervals that contain the truth on average only 50% of the time
Group biases		
Dominance	Social pressures induce group members to conform to the beliefs of a senior or forceful member of the group	Groups spend more of their time addressing the ideas of high-status members than they do exploring ideas put forward by lower-status members
Egocentrism	Individuals tend to give more weight to their own opinions than to the opinions of others than is warranted	Individuals attribute weightings of on average 20–30% to advisor opinions in revising their judgements, when higher weightings would have been optimal
Groupthink	When groups become more concerned with achieving concurrence among their members than in arriving at carefully considered decisions	The invasion of North Korea and the Bay of Pigs invasion have been attributed to decision makers becoming more concerned with retaining group approval than making good decisions
Halo effects	When the perception of an attribute for an individual or object is influenced by the perception of another attribute or attributes	Attractive people are ascribed more intelligence than those who are less attractive
Polarisation	The group position following discussion is more extreme than the initial stance of any individual group members	Punitive damages awarded by juries tend to be higher than the median award decided on by members prior to deliberation

Recommendation 10.
When developing an elicitation process, care should be taken to avoid generating situations where such bias is unwittingly introduced. The list of types of potential bias produced by McBride *et al*. (2012, Table 5 therein) can be used as a checklist for the development of the process but the listing should, in any case, be considered in any review of outputs.

4.3.3 Generation and Use of Weightings

Zickfeld *et al.* (2007) presented the use of 'scoring chips' as a means of allocating importance across each of a range of pre-defined topics. For example, each expert was allocated 50 'relative importance' chips (each representing 2% of the total cumulative importance). A 'playing board' was constructed, presenting each of a range of climate science research areas as individual 'cells'. The experts provided judgement on relative importance of each research area by allocating their chips across the cells of the playing board. Scores were combined (across all experts) and used to define the relative importance of each topic.

It is not difficult to see how such a process could be adapted to provide indicative values for weightings, for example with 'relative importance' chips being allocated across different stressors in the marine environment to better describe their significance regarding the overall vulnerability of a given habitat or species.

An alternative technique, making use of statistical modelling, was reported by Teck *et al*. (2010) whose work on assessing marine ecosystem vulnerability has some parallels with work completed for the MB0102 project in the UK (ABPmer Ltd, 2011). Where MB0102 sought to examine the relationships between marine activities and the pressures that they generate, and the sensitivity of marine features (both species and habitats) to those pressures, Teck *et al*. (2010) sought to evaluate the relative vulnerability of a series of marine ecosystems to a range of stressors associated with human activities. Vulnerability of each feature to each stressor was assessed by experts and scored on five criteria: spatial extent; frequency; trophic impact; percentage change; and recovery time.

Rather than assume that the five criteria were equally important, their study generated weightings for each criterion based on a multi-criteria decision model. The expert panel made a separate assessment of 30 hypothetical scenarios, each of which had hypothetical but realistic values for the five vulnerability criteria, identifying in rank order the top five scenarios that they would judge to have the largest negative impact at the ecosystem level. Using a probabilistic inversion technique within the multi-criteria decision model allowed the authors to identify values for weightings that reflected the importance of each criterion in the experts' decision making[16]. To properly test the relationship between criteria scores and the judged rankings requires that scenarios with both low and high values for each criterion be included. Consequently, criteria values for the 30 hypothetical scenarios were chosen deliberately to capture the full range of possible combinations.

Recommendation 11.
In any process that involves the combination of expert judgement values across a number of criteria or fields, it is recommended that consideration be given to deriving appropriate weightings for the values. Such requirements should be embodied within the design of the elicitation process. Several techniques are available including the allocation of relative importance and the use of probabilistic inversion modelling methods.

The need for and (where appropriate) the selection of specific weighting methods should be considered as part of the process design; both will need to reflect data types, elicitation methods and subsequent use of data.

[16] Operationally, and more accurately, the probabilistic inversion method was used to find a distribution for a function that maps onto the target distribution for the set of five vulnerability weightings.

4.4 Conclusions

The following provides a high level summary of the recommendations of good practice. (Note: these have been reordered to better follow their likely application to a biodiversity assessment process.)

4.4.1 Identification of Experts

Recommendation 2.

- It is suggested that studies using expert judgement should seek to engage with at least ten experts. Where there is a distribution of experts across a relatively wide geographic region or across a number of institutions/backgrounds then consideration should be given to involving more individuals. However, there is perhaps less importance to be attached to increasing the number of participants significantly and an upper limit of 20 may be appropriate.

Recommendation 3.

- In the context of the application of expert judgement to UK marine biodiversity assessments, appropriate experts should generally be selected from the SNCBs, academia and private sector consultancy. The process of identification of experts from amongst these potential sources should be undertaken through consultation with agencies and universities, by reference to earlier workshops and in consultation with government, and should seek to identify experts with appropriate track records, experience and (where relevant) local knowledge and other specialist skills.

4.4.2 Detail of Elicitation Process (Format of Elicitation and Consideration of Bias)

Recommendation 1.

- Guidelines (shown in Table 7) are available for consideration when developing a process that will employ expert judgement. Whilst not all elements within the guidelines will be applicable in all instances, and some have more importance than others, they could usefully form a checklist against which the development of a robust expert judgement process may be set.
- Having completed the design of the process the elicitation process should be tested beforehand using a small group of experts (ideally individuals whom it is not intended to involve subsequently in the formal elicitation process). Critical evaluation of the process should be used to improve the overall approach before its final use.

Recommendation 4.

- Whilst the value of web-based approaches under certain scenarios is recognised, it is recommended that for expert judgement elicitation in the context of UK marine biodiversity assessment, where possible, workshop-based environments are employed instead.

Recommendation 10.

- When developing an elicitation process care should be taken to avoid generating situations where response bias is unwittingly introduced. The list of types of potential bias can be used as a checklist for the development of the process, but the listing should, in any case, be considered in any review of outputs.

Recommendation 9.

- Where the selection of experts might introduce a geographic bias and this might affect the validity of any outputs, post-hoc analysis should be considered (e.g. using Kappa analysis) to assess whether geographic bias is actually in evidence. Where necessary, equivalence tables should be used to support this process and, if appropriate, Fleiss-Cohen weighting applied to account for the differential importance of mis-categorisation across different possible categories.

4.4.3 Recording Judgements

Recommendation 6.

- Where workshops are used, it is recommended that consideration be given to using interactive, on-screen tools for recording and discussing expert judgements. In addition, although information in the selected case studies is generally lacking regarding the arrangements made for expert judgement workshops, it is recommended that the choice of an ‘informed’ independent facilitator can be important to achieving a successful workshop.

Recommendation 5.

- The adoption of fuzzy triangular numbers to capture and use uncertainty information relating to individual judgements, employing a simple centre of gravity (CoG) approach to defuzzification, should be considered within a framework for eliciting expert judgement in support of marine biodiversity assessment. The approach is valuable not just for quantitative values but also for categorical questions with the same CoG methodology potentially applying to averaged fuzzy values to derive crisp values for simple yes/no type responses.

Recommendation 7.

- Geospatial data should be stored to a GIS using a raster data model, with gridding set to a resolution appropriate to the features being mapped. Differential resolutions can be used (e.g. quadtree data structures) as an alternative to regular gridding (e.g. square or hexagonal cells).
- Where expert opinion of the geospatial extent of features is being sought and analysed, potentially useful information regarding the level of agreement between experts (in effect a form of confidence in the estimates) should be retained by digitising each expert’s set of opinions separately. Subsequent processing within the GIS will allow a composite view of the experts’ opinions to be built up together with a single associated confidence layer.

4.4.4 Use of Bayesian Belief Network (BBN) Approaches

Recommendation 8.

- Further consideration should be given to the potential for using a BBN approach to derive a better probabilistic interpretation of the associations that exist between marine (human) activities and resultant pressures, and the impacts of these pressures on marine features. In part, the options for this will be determined by the availability of suitable information sets to inform (initialise) the network.

4.4.5 Weighting

Recommendation 11.

- In any process that involves the combination of expert judgement values across a number of criteria or fields it is recommended that consideration be given to deriving appropriate weightings for the values. Such requirements should be embodied within the design of the elicitation process. Several techniques are available including the allocation of relative importance and the use of probabilistic inversion modelling methods.

4.4.6 Further Considerations

Although all of the aspects outlined above are important within a wider application of expert judgement to a biodiversity assessment, certain elements do not necessarily relate directly to the process of a marine benthic habitat vulnerability assessment as described in more detail below. In particular, consideration of weighting and the use of fuzzy logic relate more to the direct elicitation of quantitative information.

Nevertheless, whilst these elements do not necessarily sit directly within a vulnerability assessment process, methods and approaches that relate to weighting and the use of fuzzy logic should not be totally disregarded as they provide a useful insight into the potential use and application of expert judgement.

5 Application of Good Practice for Marine Biodiversity Status Assessments

Following on from the development of recommendations as to what constitutes good practice (including key principles and methods) when using expert judgement, this section provides an overview of how expert judgement could be applied when undertaking a vulnerability assessment of a benthic habitat. In so doing, consideration is given to the process that might be used to undertake a vulnerability assessment, and a five stage process is described.

5.1 Vulnerability Assessment Methods

5.1.1 **Recent Approaches**

In the absence of more specific information, habitat vulnerability has been used as an indicator of habitat condition in a number of assessments of conservation features in the UK marine environment. A number of flowcharts, or process diagrams, are available to describe these assessments, and are presented below.

For example, Connor & Enserink (2009) present a framework used to undertake the Quality Status Report 2010 (QSR 2010) which reported on the environmental quality of the five regions in the OSPAR[17] maritime area (see Figure 9). The process described by Connor & Enserink was, in turn, developed from earlier work presented by Robinson *et al*. (2008), who developed a marine assessment and monitoring framework under commission from JNCC for application by UKMMAS and OSPAR on the assessment of pressures.

The first comprehensive assessment of benthic habitats throughout the UK was undertaken as part of Charting Progress 2, a comprehensive report on the state of the UK seas published by the UK Marine Monitoring and Assessment community (comprising over 40 member organisations). One of the four Feeder Reports that provides a detailed evidence base for the ‘healthy and biologically diverse seas’ section of *Charting Progress 2* includes a detailed framework for assessing the condition of benthic habitat features (Aish *et al.*, 2010) (see Figure 10).

Both of the above processes were considered within the development of a method for assessing the ‘structure and function’ parameter that forms part of the proposed Conservation Status assessment for offshore Annex I reef and sandbank features for Article 17 reporting under the Habitats Directive (Vaughan *et al.*, in prep.). This latter approach (shown as Figure 11) together with its explanatory text represents the most detailed process of the three considered here.

[17] OSPAR manages and administrates the Oslo and Paris Conventions, guiding international cooperation on the protection of the marine environment of the North-East Atlantic.

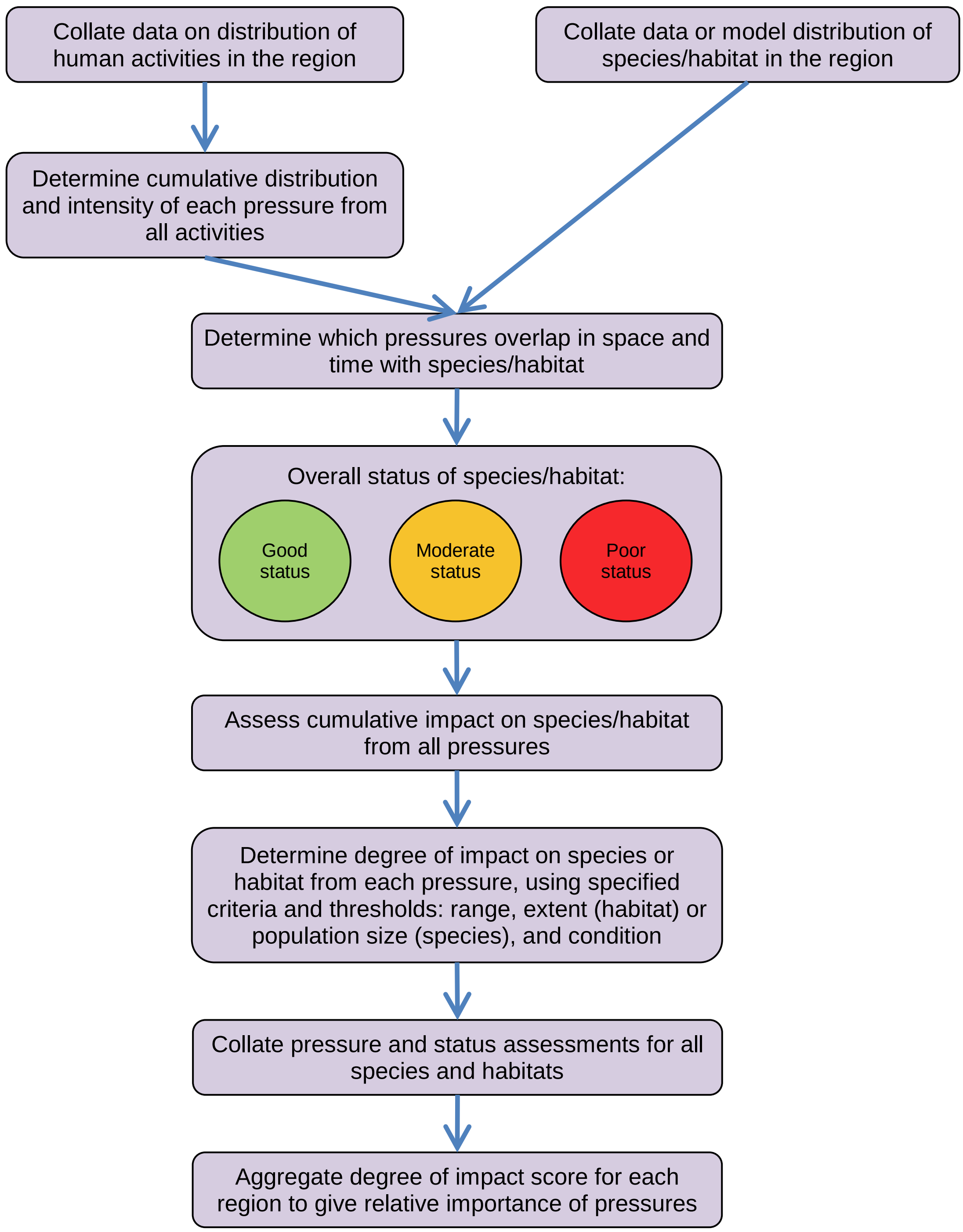


Figure 9. ‘Utrecht’ assessment process flow diagram – based on Robinson *et al*., 2008 (reproduced from Connor & Enserink, 2009)

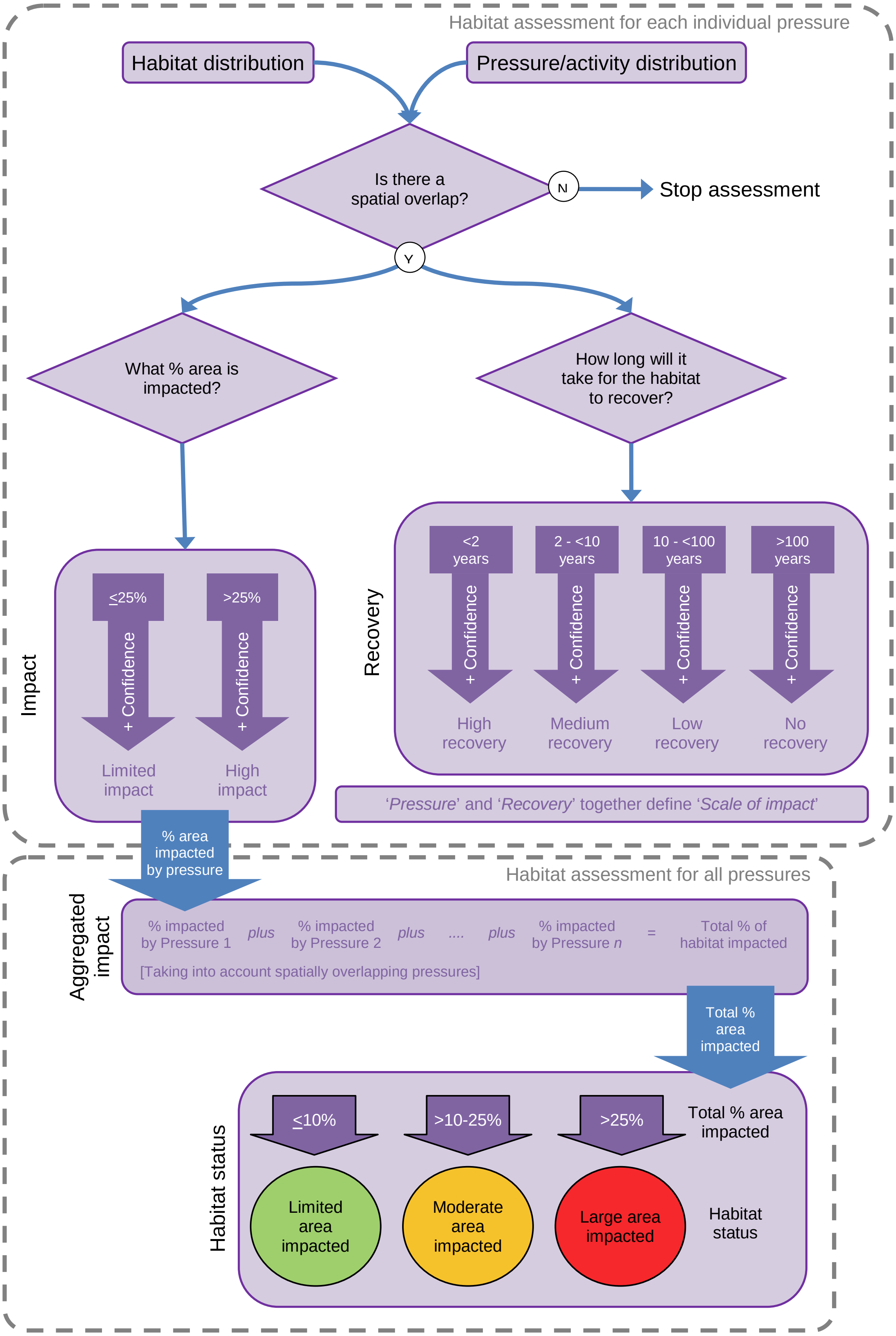


Figure 10. Schematic overview of process used to derive broad habitat status assessments for Charting Progress 2 (reproduced from Aish *et al.*, 2010)

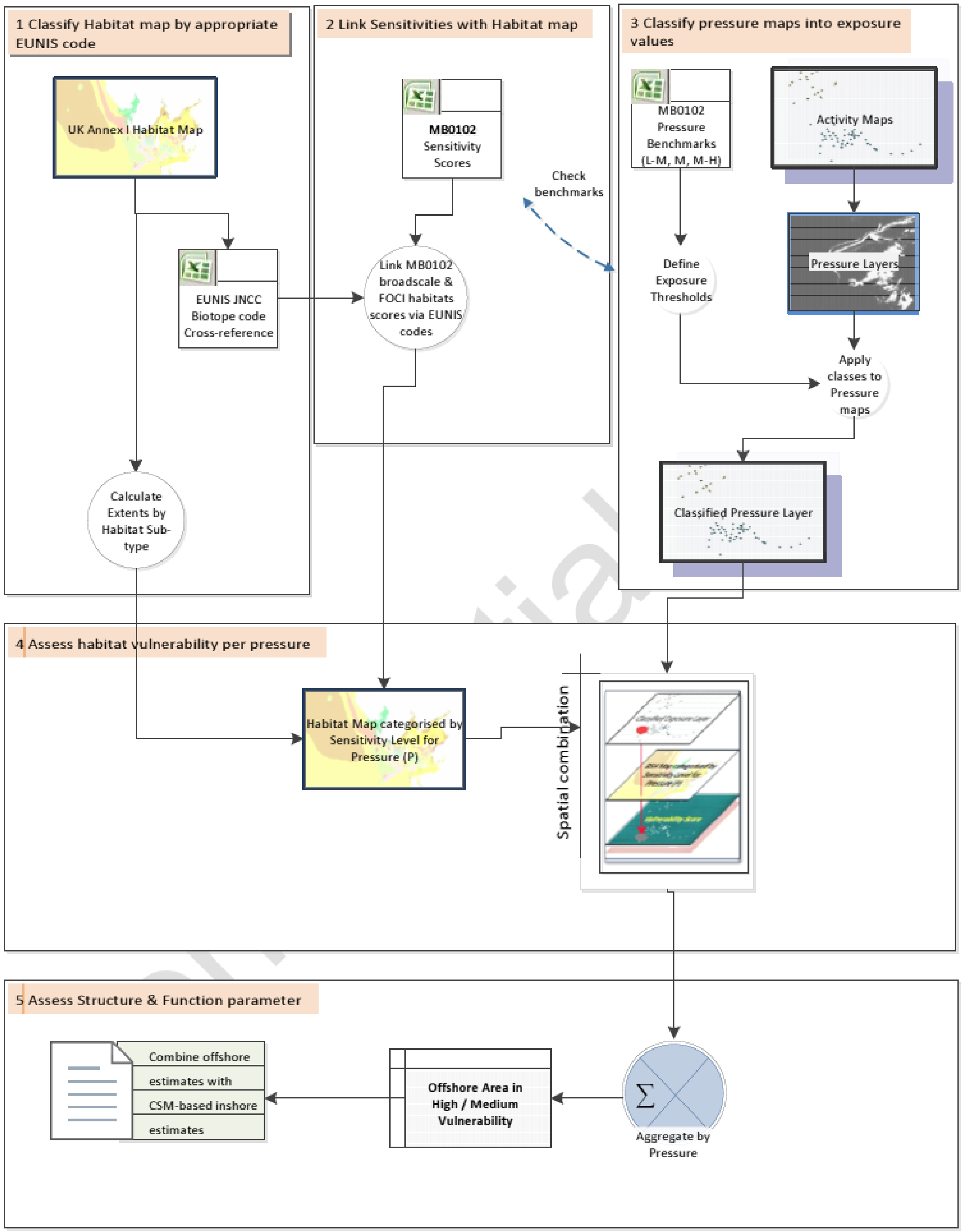


Figure 11. Overview of structures and functions assessment process for offshore Annex I habitats (reproduced from Vaughan *et al.*, in prep.) [CSM – Common Standards Monitoring; EUNIS – European Nature Information System; FOCI – Features of Conservation]

5.2 Modified Framework

The approach described by Vaughan *et al.* (in prep.) has been used as the basis for the development of a modified (and more generic) vulnerability assessment process for marine benthic habitats, comprising five stages. These stages are shown diagrammatically in Figure 12 and described below:

Stage 1 Classification of habitat layers by EUNIS code

Having identified the appropriate habitat maps (GIS layers), all habitats or biotopes should be classified to the appropriate scale (e.g. EUNIS level 3), producing a new GIS habitat layer classified using standard EUNIS biotope codes.

Stage 2 Linking habitat layers to sensitivity information

For each series of pre-identified anthropogenic pressures, EUNIS habitat codes should be linked to sensitivity scores (e.g. as generated by MB0102[18]) to the habitat types identified in Stage 1. Each habitat type should be assigned a sensitivity score (Not Sensitive (NS), Low (L), Medium (M) or High (H)). The resulting GIS layer effectively represents the 'sensitivity layer' for a given pressure.

Stage 3 Production of pressure layers and classification into exposure values

Activity layers (e.g. as produced under MB0106[19]) should be used to derive pressure layers. This could be done by deriving one layer per pressure; each developed considering the footprint, intensity, frequency and duration of the human activities that can give rise to the pressure and classified into potential exposure (L, M or H) according to agreed benchmarks.

Stage 4 Assessment of habitat vulnerability

Following the process described by Vaughan *et al.*, the habitat sensitivity at each location (derived in Stage 2) and the pressure exposure (as derived in Stage 3) should be used to derive a vulnerability score for each pressure under consideration (e.g. see Table 11, below). These vulnerability scores should be stored within a GIS as a (pressure-specific) vulnerability layer.

18 i.e. the relevant scores from within the sensitivity matrix developed as Task 3 of the MB0102 Defra-led contract (Tillin *et al*., 2010) which presents the sensitivity of (each of 41 MCZ/MPA) conservation features to each of 40 human pressures.

19 Objective 1 of the MB0106 Defra-led contract (Lee *et al.*, 2009) which includes the provision of geo-database containing standardised layers showing the distribution of specified activities.

Table 11. Example of derivation of vulnerability class from sensitivity and exposure

	Relative sensitivity of the feature:			
Relative exposure of the feature:	**High**	**Moderate**	**Low**	**None**
High	High vulnerability	High vulnerability	Moderate vulnerability	No known vulnerability
Medium	High vulnerability	Moderate vulnerability	Low vulnerability	No known vulnerability
Low	Moderate vulnerability	Low vulnerability	Low vulnerability	No known vulnerability
Unknown	Vulnerable but not quantified	Vulnerable but not quantified	Vulnerable but not quantified	No known vulnerability
None	No known vulnerability	No known vulnerability	No known vulnerability	No known vulnerability

The individual vulnerability layers, one for each pressure under consideration, should then be aggregated across pressures; the highest (pressure-specific) vulnerability score at each location is selected to give an ‘overall vulnerability’ score for that position.

Stage 5 Assessment of habitat condition

Subsequently, the ‘overall vulnerability’ score is used to determine the likely condition of the habitat at each location using the relationships shown in Table 12. For example, where moderate or high overall vulnerability scores have been assigned the habitat is deemed likely to be in unfavourable condition.

Table 12. Derivation of likely condition from vulnerability

Overall vulnerability score	Assumed condition
High vulnerability	Likely unfavourable
Moderate vulnerability	Likely unfavourable
Low vulnerability	Likely favourable
Vulnerable but not quantified	Unknown
No known vulnerability	Likely favourable

The amount of each marine benthic habitat that is likely to be in unfavourable condition is then assessed. This should be done by summing those areas with Moderate or High overall vulnerability scores.

A judgement should then be made on the overall likely habitat condition, according to the percentage of the total area of habitat that is likely to be in unfavourable condition, using agreed targets derived from expert judgement or opinion[20].

The above five-stage process is shown schematically in Figure 12 (below) which also indicates (through the use of grey-shaded boxes) where in the process different applications of expert judgement may potentially be used. Subsequently, Table 13 provides a précis of the process indicating, for each constituent step, those that are straightforward ‘procedural’ steps (not involving expert judgement) and ‘expert judgement’ steps where such judgements would be applied. Table 13 also indicates the inputs (sources of information) that are required at each step and the outputs that are produced. Inputs that would be documented or recorded prior to the step being undertaken are shown separately from those that would be derived (or potentially derived) through the application of expert judgement.

[20] e.g. Vaughan *et al.*(in prep.) described the structures and functions parameter of each habitat on the basis of the following rules:

- where less than 5% of the total area of a habitat was likely to be in unfavourable condition the structures and functions parameter of the habitat was deemed to be ‘Favourable’;
- where between 5 and 25% of the area of a habitat was likely to be in unfavourable condition the structures and functions parameter of the habitat was deemed to be ‘Unfavourable-Inadequate’; and
- where more than 25% of a habitat was likely to be in unfavourable condition the structures and functions parameter of the habitat was deemed to be ‘Unfavourable-Bad’.

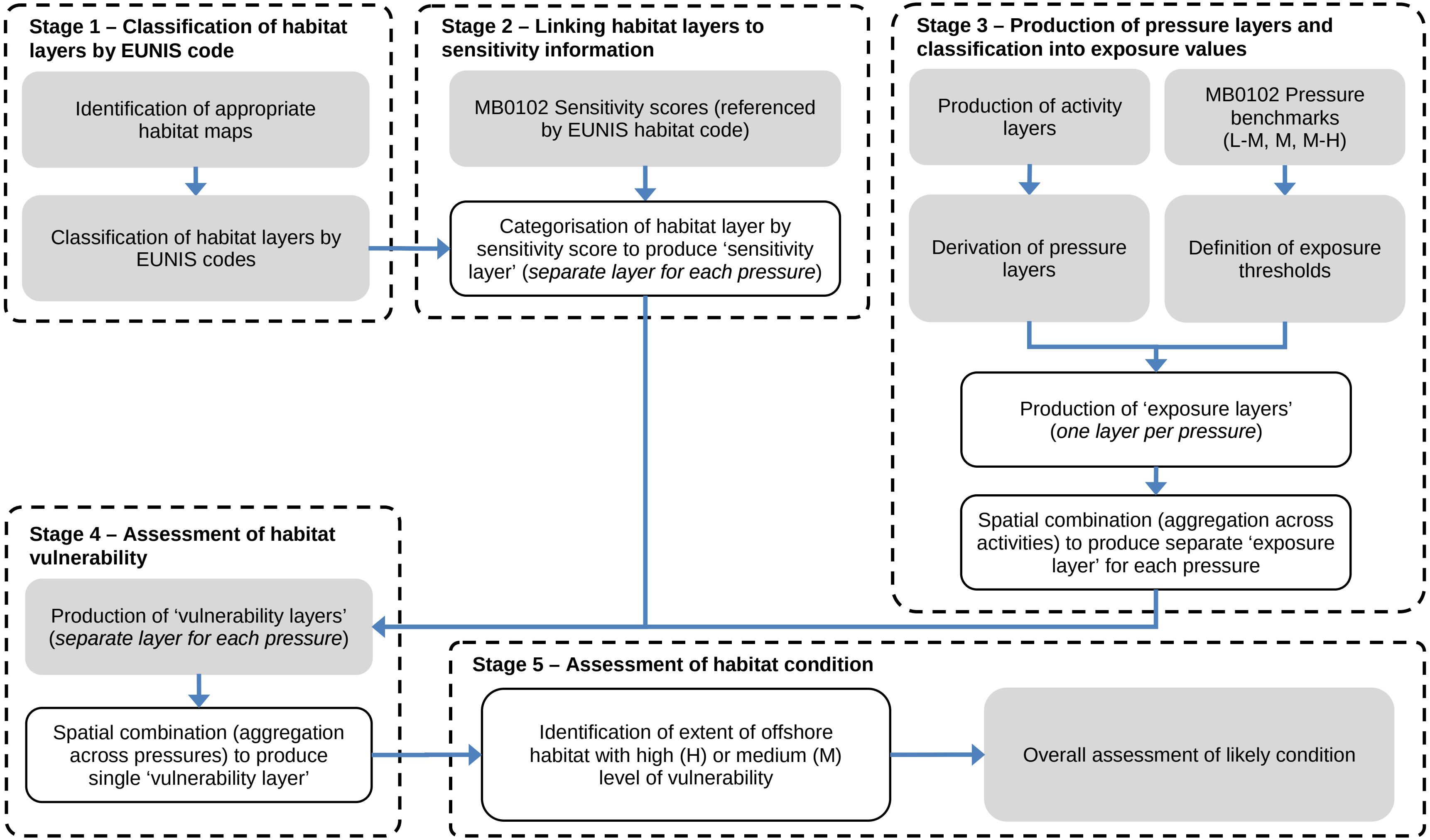


Figure 12. Overview of generic vulnerability (condition) assessment process for benthic habitats – see text for details (adapted from Vaughan *et al.*, in prep.) (grey boxes show possible integration of expert judgement)

Table 13: Role of expert judgement within marine biodiversity assessment indicating 'procedural' and 'expert judgement' steps and potential application of expert judgement to provide or support information sources

Assessment stage	Steps		Information sources/inputs		Outputs
	Procedural	Expert Judgement	Documented/recorded	Expert Judgement	
Stage 1 Classification of habitat layers by EUNIS code	· Appropriate habitat map(s) identified	· Habitats classified to appropriate EUNIS code		· Identification of relevant habitat data sources (e.g. survey data; modelled data; and habitat maps explicitly based on expert judgement such as local or lay knowledge) · Habitat classification (e.g. how to assign codes to ambiguous habitats)	Spatial extent of habitat types referenced by EUNIS code
Stage 2 Linking habitat layers to sensitivity information	· For each pressure, sensitivity scores linked to identified habitat types (via EUNIS codes)		· Spatial extent of habitat types (from Stage 1)	· Tabulated sensitivity scores for each habitat/pressure combination	Pressure-specific habitat sensitivity layers (recorded as: Not Sensitive, or Low, Medium or High Sensitivity)
Stage 3 Production of pressure layers and classification into exposure values	· Spatial extent (footprint) of human activities used to infer spatial extent of potential pressures[21]			· Activity layers · Linkage between activities and pressures	Pressure layers (one layer for each activity-pressure combination)
	· Spatial extent and frequency of each activity used to derive level of exposure to each potential pressure			· Frequency and extent of each activity · Benchmarks expressing degree of exposure to pressures as a function of the spatial extent and frequency of each human activity under consideration	Pressure-specific 'exposure' layers for each activity under consideration (recorded as: No Exposure, Exposure Not Known, or Low, Moderate or High Exposure)
	· Separate exposure layers for each activity-pressure combination aggregated across activities to derive single 'exposure layer' for each pressure (with maximum level of exposure from across all activities taken forward)		· Pressure-specific 'exposure' layers for each activity under consideration		Pressure-specific 'exposure' layers (recorded as: No Exposure; Exposure Not Known, or Low, Moderate or High Exposure)

21 N.B. actual pressure may be more spatially restricted locally (and consequently more intense) than suggested by some relatively low resolution activity data layers

Assessment stage	Steps		Information sources/inputs		Outputs
	Procedural	Expert Judgement	Documented/recorded	Expert Judgement	
Stage 4 Assessment of habitat vulnerability	· For each pressure, habitat sensitivity and exposure are used to derive an indication of relative habitat vulnerability (by fixed, tabulated relationship)		· Pressure-specific habitat sensitivity (from Stage 2) and exposure (from Stage 3) layers	· Expression of habitat vulnerability as a function of sensitivity and exposure	Pressure-specific habitat vulnerability layers (recorded as: Vulnerable but Not Quantified, or High, Moderate, Low, or No Known Vulnerability)
		· Aggregation of individual (pressure-specific) vulnerability layers[22] (maximum vulnerability across all pressures used to set overall vulnerability)	· Individual (pressure-specific) vulnerability layers		Overall habitat vulnerability layer (recorded as: Vulnerable but Not Quantified, or High, Moderate, Low, or No Known Vulnerability)
Stage 5 Assessment of habitat condition	· The likely condition of each habitat is assessed on the basis of the overall percentage of habitat with moderate or high overall vulnerability, expressed as percentage of total area of habitat present		· Overall habitat vulnerability (from Stage 4)	· Tabulated relationship between percentage of total area of habitat with moderate or high vulnerability and habitat condition target expressed as a percentage of the total area of habitat present	Assessment of condition for each habitat under consideration

[22] Expert judgement potentially applied to consideration of 'in-combination' effects

5.3 Habitat Vulnerability Assessment and the Application of Expert Judgement

Both Figure 12 and Table 13 indicate where expert judgement might be integrated into the overall vulnerability assessment process. As can be seen, expert judgement is potentially applied at each of the five stages described. The practical application of different methods, approaches or considerations regarding expert judgement to each of the different steps within the stages in the process, as identified in Table 13, is summarised in Table 14.

Table 14 was produced as a *prima facie* overview of how the different aspects of the application judgement (e.g. source and identification of experts, use of facilitated workshops) are likely to relate to the different stages within the generic vulnerability assessment process. The relationships shown are high level and, in general, no attempt has been made to quantify the relative importance of any one aspect against any particular stage of the assessment process. The information presented in Table 14, together with an overview of the application of these methods, approaches or considerations across the different elements within the benthic habitat vulnerability assessment process, is presented below (Section 5.3.1).

Table 14. Summary overview of how different aspects of expert judgement should be considered in relation to key points within biodiversity assessment process – see text for detail

	Aspects of expert judgement:									
		Detail of elicitation process (format of elicitation and consideration of bias):					Recording judgements:			
Elements within overall assessment process where expert judgement may potentially be employed (see Table 13):	Experts: source and identification	Elicitation process checklist	Use of facilitated workshops	Elicitation process dry run	Consideration of the minimisation of judgement bias	*Post hoc* checking for geographic bias	Interactive recording & discussion within workshops	Adaptive resolution for underlying GIS raster models	Use of 'confidence' layers in GIS	BBN approach with regard to activity/ pressure associations, and impacts
Identification of relevant habitat map(s)	ü							ü		
Classification of habitats by EUNIS habitat codes	ü	ü	ü	ü	ü		ü	ü		ü
Application of sensitivity scores for habitat/pressure combinations (MB0102) (initial derivation of scores)	ü	ü	ü	ü	ü	ü	ü		ü	ü
Production of activity layers	ü	(ü)	ü	(ü)	ü		(ü)	ü	ü	
Linkage between activities and pressures	ü		ü	ü	ü					ü
Assessment of activity frequency	ü	ü	ü	ü	ü			ü	ü	
Exposure to pressures expressed as a function of spatial extent and frequency of each activity under consideration	ü	ü	ü		ü		ü			
Expression of habitat vulnerability as a function of sensitivity and exposure	ü		ü				ü			
Relationship between % of habitat in unfavourable condition and implied habitat condition	ü		ü				ü			

(ü) – May not be fully applicable; production of activity layers may largely be a data confirmation process so not all aspects of elicitation necessarily apply (although expert judgement may be used to derive activity layers in some instances).

5.3.1 Experts: Source and Identification

The choice of experts is clearly central to the use of expert judgement, and is important at each stage of the assessment process. Published studies employing expert judgement have used a variety of methods to identify the experts they use.

However, given that only a relatively small number of experts are likely to be involved (e.g. 10-20; see Section 4, Recommendation 2), more-extensive methods for engaging experts (e.g. mail shots or engaging with experts during a conference organised by others) can be discounted.

Experts should generally be selected from the Statutory Nature Conservation Bodies (SNCBs), academia and private sector consultancies and their selection should be undertaken through consultation with agencies and universities, by reference to earlier workshops, and in consultation with government (e.g. Defra). The process should seek to identify experts with appropriate track records, experience and (where relevant) local knowledge and other specialist skills (Section 4, Recommendation 3). A formal, transparent process for the definition and selection of those with relevant expertise should be adopted and would form part of the audit trail for the work being undertaken.

Both the identification of suitable data layers and, where required, the application of additional data layers (to better improve the accuracy of habitat classification) can be seen as areas where expert judgement is applied. As a minimum, the process behind the selection of those individuals who are best placed to undertake these steps, both to identify relevant data (or additional data) and to develop a process for the application of additional data layers, should be considered.

5.3.2 Elicitation Process Checklist

Table 6 (presented within Section 4.2.1) presents a comparison of generalised processes for the elicitation of expert judgement, whilst Table 7 (also presented within Section 4.2.1) provides a more-detailed set of guidelines for the design of an elicitation process. In turn, more detail on the elements within Table 7 can be found in the detailed reproduction from Knol *et al*. (2010) which is presented as Annex 1 of Appendix 3. Whilst not all elements within Table 7 will be applicable in all instances, and some have greater importance than others, they should be used to form a checklist against which the development of a robust expert judgement process may be set.

5.3.3 Use of Facilitated Workshops

A facilitated workshop is likely to provide the optimum environment for eliciting expert judgement in the context of UK marine biodiversity assessments (see Section 4, Recommendation 4) and should be the ‘tool of choice’ (McBride *et al*., 2012). Once experts have been identified, they should be given a reasonable period of notification to ensure their availability. Careful selection of the facilitator (it is suggested that an informed but independent individual be selected for this role), and of the workshop design, is important.

Regarding the overall five-stage generic assessment process discussed above, it is likely that a range of separate tasks need to be undertaken. For example, the identification of appropriate habitat data sources and classification to standard EUNIS codes may need to be undertaken separately to the assessment of overall habitat condition as a function of the percentage of habitat deemed to be in unfavourable condition. Consequently, it is not necessarily the case that the same group of experts is the most appropriate for all tasks. However, there may be instances where the same group of experts can be used to complete a number of separate tasks and this situation could be used to help make better use of experts' time and contributions.

Simola *et al*. (2005) considered the use of training ahead of formal workshops to:

- Familiarise participants with the expert judgement process and to motivate them to provide formal judgements;
- Provide training on probability issues and on providing probability judgements;
- Inform participants on possible cognitive biases an expert judgement and possible debiasing techniques; and
- Solve simple exercises, as a means of providing experience of the process.

As indicated in Table 14, the process of deriving judgement on the likely sensitivity of habitat to pressures should be undertaken through facilitated workshops. The complexity and extent of the topics under discussion (MB0102 considered 32 broad-scale habitats, 40 habitat features of conservation importance and 36 species of conservation importance) would suggest that serious consideration should be given to carry out a trial run of the process in advance of the actual exercise, to help develop the final format of the elicitation exercise (see below). When designing the process the generalised elicitation process guide that is reproduced as Table 7 should be used to help guide development. As noted below, it would also be important to design the process so as to minimise the different types of judgement bias that may be in evidence, and where possible consideration should be given to examining the potential occurrence of geographic bias through appropriate post-hoc tests. As above, the identification of appropriate experts to participate in the workshop(s) should involve ensuring that an appropriate spread and depth of expertise is available. Methods used to identify participants should be recorded to form part of an 'audit trail' for the overall elicitation process.

Whilst the classification of habitats to a standardised level, such as EUNIS level 3, can be largely automated, certain biotopes may not be fully described within the principle GIS habitat layers (an issue that, in order to be addressed requires the application of a degree of expert judgement). For example the extent of biogenic reef within a 'mixed' area classified as bedrock/biogenic reef as described in the UK Annex I habitat maps may not be known. In such cases, expert judgement and additional data sources (such as EUSeaMap or specific individual point records) may be used to infer habitat type and could be undertaken in a workshop environment.

5.3.4 **Elicitation Process Trial**

It is recommended that workshop procedures are tested ahead of elicitation by undertaking a trial using a different set of experts to those that are invited to participate in the formal workshops. Such a process is effectively good practice for large-scale workshop scenarios where there is a great deal of investment just in ensuring that the optimum group of experts are able to attend. The use of a 'dry-run' should be considered for most workshop-based elicitation exercises, although its adoption for smaller-scale applications (e.g. for looking at the 'rules' for expressing habitat vulnerability as a function of sensitivity and exposure) is not so important.

When planning a trial, attention should be paid not just to the structure and balance of the workshop, but also to the tools that are used (including the questions that are being asked, and any practical methods that are being used to elicit, record or collate views and judgements).

5.3.5 Consideration of the Minimisation of Judgement Bias

A range of potential sources of bias has been described earlier (described by McBride *et al*., 2012, and presented as Table 10) and include bias at both the individual and group level. When developing a workshop plan, the facilitator(s) should be aware of these potential sources of bias and, as far as possible, alter the design of the process to eliminate them.

5.3.6 Post-hoc Checking for Geographic Bias

Given that experts involved in potential work to develop or undertake assessments for UK waters are likely to be from a restricted (i.e. UK) origin, there is perhaps little danger of there being systematic geographic bias inherent in their judgements. There is, however, the possibility that there may be some small-scale geographic influence involved (e.g. possible systematic differences in view arising from differences in experience such as familiarity with Atlantic versus North Sea environments). It may be possible to detect such effects through suitable post-hoc testing of outputs (e.g. using Kappa analysis).

5.3.7 Interactive Recording and Discussion Within Workshops

Where possible, discussion and consensus should be recorded ‘live’ within workshops. This does not mean that proceedings need to be audiovisually recorded, but that all participants are able to interact with each other in open discussion and that comments or views are somehow recorded in a way that subsequently provides an audit trail. Where appropriate, the use of ‘live’ on-screen tools for capturing discussion at the time of origin (e.g. projecting a spreadsheet from an active laptop computer to a screen, and having it edited by a secretary as agreed by a facillitator) can support debate and allow individuals to better follow the process.

5.3.8 Adaptive Resolution for Underlying GIS Raster Models

Current practice regarding GIS/habitat work within the SNCBs is generally to store habitat data within a GIS as polygon or point data, not raster. Where a raster model is used instead it is typical to use a regularly spaced grid within the underlying GIS data models. However, other tessellated data structures are available for grid-based GIS systems (for example, quadtree data structures, where the plane is recursively subdivided into quadrants so that some areas have a higher resultant spatial resolution). The use of such adaptive scaling methods may be relevant for certain GIS layers, allowing greater spatial resolution to be preserved for specific areas, and allowing for a more realistic application of data sources such as point records.

5.3.9 Use of Information on ‘Confidence’

The elicitation process for the MB0102 Sensitivity Matrix included a measure of the level of confidence associated with each assessment of the sensitivity of each habitat to each pressure. As a minimum, this data should be held within the GIS model derived within the second stage of the generic five stage habitat assessment process outlined earlier, providing a series of ‘confidence’ layers each of which would correspond to a (pressure) sensitivity layer.

Similarly it is possible to derive confidence scores for activities, and hence pressures. For example, the pressure layers produced by Eastwood *et al*. (2007) were assigned confidence ratings which related to whether pressure extents were based on data or on the estimates of location and extent of underlying human activities. Three cases were considered: both location and extent estimated (least confidence); location known, but extent estimated (medium confidence); and both location and extent known (highest confidence). With the exception of extraction, the spatial extent of pressures was based on estimates from known locations of the activities these being assigned ‘medium’ confidence rating. Abrasion pressures were assigned the lowest confidence rating, because both location and spatial extent were estimated based on the analysis of Vessel Monitoring System (VMS) data.

Where new judgements are being elicited, data such as between-expert variability in judgement scores could potentially be used to develop a confidence layer (where higher variability implies a lower level of agreement and hence an implicit lower confidence in the judgement scores).

In either case, the incorporation of such layers into the GIS database management system allows for uncertainty or confidence to be used, for example, to weight subsequent analyses. Additionally, where spatial data have associated confidence scores, these can be provided to the user as a separate inset map.

Where qualitative or semi-quantitative outputs are being presented cartographically, it is possible to provide a small inset map showing (for example in shades of red, amber and green) how confidence or uncertainty values are distributed spatially. However, where fully quantitative data are being elicited, there is perhaps more of an opportunity to make use of associated information regarding uncertainty to develop confidence intervals around estimated values derived by expert judgement.

5.3.10 BBN Approach with Regard to Activity/Pressure Associations, and Impacts

Wooldridge (2003) notes that BBNs are particularly useful for making probabilistic inference about modelled systems (the ‘model domain’) that are characterised by inherent complexity and uncertainty. In addition to being able to deal with problems whose complexity cannot be feasibly modelled by other approaches, BBNs offer many advantages over other methods for dealing with uncertainty, and limited data including:

- *Merging different types of information*: BBNs can combine both subjective information (e.g. expert opinion) and quantitative information (e.g. monitoring data, modelling results). The flexible nature of BBNs also means that new information can easily be incorporated as it becomes available.
- *Formal structuring of our understanding*: BBNs are helpful for challenging experts to articulate what they know about the model domain, and to knit those influences into dependency networks. The visual nature of BBNs facilitates easy transfer of understanding about key linkages. As subjective expert opinions (hypotheses) are made explicit in the formal structure of the network, they can be challenged and revised, and can also be directly evaluated (potentially with process-based models) to determine whether results are robust.
- *Modular design*: Given their network structuring, BBNs successfully capture the notion of modularity i.e. a complex system is built by combining simpler parts. A BBN can be started small, with limited knowledge about a domain, and grown (with additional variables being added) as new knowledge is acquired.
- *Informed decision-making before scientific knowledge is complete*: The representation of a model domain through the use of a BBN means that you don't need complete knowledge about the particular instance of the world it is being applied to. Because uncertainty in particular linkages can be acknowledged in the probabilistic dependency relationships, the models are not necessarily limited by the

mechanistic detail of existing information or understanding. Consequently, BBNs can facilitate informed decision-making before scientific understanding is complete.

- *Predictions are amenable to risk analysis*: BBNs express predicted outcomes as likelihoods, which can form the basis for risk analysis and which provide a sound basis for adopting rational decisions based on a precautionary (risk-averse) attitude.
- *Future scenario testing*: BBNs provide an ideal framework to test the most 'likely' consequence of future events or scenarios. This contributes to 'future memory', and an understanding of '*what will happen when* …'. The ability of BBNs to perform bidirectional reasoning also provides an excellent diagnostic tool for identifying the most likely causes of specific outcomes.

The schemata available to model or represent the inter-relationships of activities, resultant pressures and subsequent impacts on conservation features within the marine environment are currently limited. For example, the outputs produced under MB0102 relate sensitivities to pressures for each of a series of habitats or other conservation features. Although additional work was undertaken to link activities to the pressures that they can cause, the source of the pressure (i.e. the causative activity) was not explicitly considered when the degree of sensitivity was assessed (for example different levels of abrasion being caused by different types of fishing gear – although in some instances this may be adequately accounted for in the associated benchmarking work). Such differences, whether real or perceived, may have an effect on judgements made regarding impacts.

There are clear benefits to be gained from re-developing the work undertaken on MB0102 and setting it within a more realistic framework; ideally one that permits a more intuitive use of uncertainty. In this context, it is recommended that further work is commissioned to examine the potential application of BBN methods and approaches to the development of sensitivity matrices.

5.4 Other Considerations

In addition to the foregoing, there are two further areas that could usefully be considered when developing expert judgement based approaches to support marine biodiversity assessments: fuzzy logic and weighting.

5.4.1 **Fuzzy logic**

The use of fuzzy logic techniques as a way of capturing uncertainty on quantitative topics in workshop environments should be considered during the workshop design process. As described in Section 4.2.4, fuzzy triangle numbers provide a straightforward means of capturing information from individuals and, whilst perhaps most obviously suited to quantitative estimation, can be used for 'true/false' assessments.

5.4.2 **Weighting**

As noted by French (2011) a key issue is not just how to assimilate the advice of one expert, but also how to draw together and learn from the advice of several experts, particularly when there is conflict in their views. Accordingly (and as noted earlier) the use of weighting is an important method to consider when eliciting expert judgement and using the outputs. However, the approach relates principally to quantitative procedures and not the qualitative judgements that are typical of the process of marine habitat assessment outlined here.

Nevertheless, the use of weighting techniques (including the allocation of relative importance amongst expert judgements and the use of probabilistic inversion modelling methods) should be considered alongside the wider list of considerations when developing an expert judgement elicitation process.

5.5 Concluding Comments

The application of expert judgement, or the use of information previously derived through the elicitation of experts, is in evidence at each stage of a benthic habitat vulnerability assessment processes.

Some applications can be relatively straightforward; for example, describing the relationship between the percentage of a habitat that is vulnerable to prevailing pressures (i.e. the percentage that is likely to be in unfavourable condition) and the overall habitat condition. Other applications, such as the development of 'sensitivity matrices' where the sensitivity of each of a series of habitats or features to each of a series of pressures is estimated, are more fundamental to the overall assessment process and are the areas where most attention on the appropriate elicitation and application of expert judgement should perhaps be placed.

Unsurprisingly, some aspects of the application of expert judgement (including the choice and selection of experts) cut across the full spectrum of expert judgement use. Other aspects, such as the need to check for geographic bias in elicited responses, are likely to be restricted to discrete steps within an overall assessment framework.

Other valuable tools and approaches (including approaches to weighting, the use of fuzzy logic, the application of Bayesian belief network models and the incorporation of uncertainty into expert judgement-derived GIS layers) may not sit within an assessment framework as it is currently envisaged. Nevertheless their applicability and potential value should be considered when designing an elicitation process.

6 Conclusions

6.1 Introduction

This report covers four main areas of investigation. Initially, by looking at the application of expert judgement across widely differing fields, it examines the use of expert judgement as a means of supporting decision making processes. It then narrows the focus onto examples where expert judgement has been used in biodiversity assessments, critically reviewing and assessing the strengths and weaknesses of a series of selected case studies. Following from this, a series of recommendations on good practice when using expert judgement in biodiversity assessments are drawn out. Finally, these recommendations are set into a revised generic framework for applying expert judgement to marine benthic habitat vulnerability assessments.

6.2 Summary of Findings and Reviews

There is a shift underway in conservation and natural resource management from a historical focus on single species to an emphasis on ecosystem-level protection. As a consequence, assessments of ecosystem condition are increasingly called for to support the evaluation of management effectiveness and to inform resource management decisions.

Expert judgement is already used in the UK to assess marine biodiversity status, often in the form of vulnerability assessments (where information on a benthic habitat’s sensitivity to a series of known pressures is combined with a range of evidence of exposure to activities exerting those pressures, in order to derive an indication of the benthic habitat’s vulnerability). Vulnerability assessments (or similar) are likely to be used in the future to help inform benthic habitat assessments of Good Environmental Status under the MSFD, whilst expert judgement is also increasingly being used in the development of conservation objectives for offshore Special Areas of Conservation and for recommended Marine Conservation Zones. As straightforward measurements of ecosystem condition are beyond the reach of current science, managers are likely to be increasingly reliant on expert judgements.

In addition to conservation and natural resource management, expert judgement is used across a wide range of fields. This includes, for example: nuclear safety; the effects of nanoparticles on public health; aircraft engineering; air traffic control; drug legalisation; economics; pharmacoeconomics; and software production. From across these diverse fields, four studies were selected for review, with the intention of demonstrating the potential value of expert judgement and to identify any key learning points that may subsequently ‘translate’ to studies relating to biodiversity assessments. These four studies covered: environmental health impact assessment; the risk assessment of nanotechnology-enabled food products; the response of the Atlantic meridional overturning circulation to climate change; and structural reliability relating to engineering in the nuclear energy sector. These particular four studies were selected as they are relatively recent, relate to a number of different disciplines and make use of different underlying methodologies.

From this limited and brief review a number of learning points were identified:

- a transparent and structured process is important (helping to plan and guide elicitation work, and providing an audit trail of decision-making in the light of expert opinion);
- experts should cover an appropriately wide range of views and expertise, with generalists, as well as nominated experts, present;
- elicitation can seek to obtain the same information by different methods to check internal consistency;
- agreement on definitions of key terms and concepts ahead of the main elicitation exercise can help to clarify what is being asked for and reduce uncertainty and 'noise' in experts' responses;
- training or familiarisation ahead of the elicitation itself may help reduce uncertainty and improve the quality of information that is provided on the day;
- splitting the elicitation process to more than one session allows participants to reflect on their judgements and refine their opinions in the light of information from the other experts;
- techniques other than simple requests for quantitative data can be used to help capture additional information;
- opinions from different experts can be weighted;
- anonymity can be both advantageous (for example in allowing individuals to give a more honest or candid response) and problematic (for example in reducing the apparent transparency of the process).

Subsequently, a range of recent studies that made use of expert judgement to produce or support biodiversity assessments (within both terrestrial and marine environments) were identified, and a shortlist of 13 case studies were selected for SWOT analysis and critical review. The case studies demonstrated that a number of approaches to the use of expert judgement have been successfully applied in relation to biodiversity assessments. From these it was possible to identify a number of elements that may contribute to the development of an appropriate framework or methodology to address a UK marine biodiversity assessment scenario (such as for reporting under the MSFD).

Examination of these elements in more detail led to the production of eleven recommendations outlining good practice for the use of expert judgement, along with a brief discussion of specific methods and potential approaches. The recommendations included the following elements:

- Recommendations 2 & 3: guidance on the approaches for identifying experts (including numbers involved and background/credentials);
- Recommendations 1, 4, 9 & 10: detail on the elicitation process (the design of the workshop or web based interview and its facilitation, the need to account for bias in experts' responses and any geographic bias which may occur);
- Recommendations 5, 6 & 7: the recording of judgements during the elicitation process, including the use of interactive tools and the use of map-based approaches (linked with GIS) to capture geospatial information from experts;
- Recommendation 8: the use of Bayesian Belief Network (BBN) approaches for expert elicitation; and
- Recommendation 11: weighting as a means to account for bias in experts' responses.

The report then puts these recommendations (and other earlier findings) properly into context by considering where and how expert judgement might be employed to support a process to undertake a vulnerability assessment of a benthic habitat. A five stage process is described which is itself a refinement of other recent approaches to marine biodiversity assessment, especially that employed by the MPA Monitoring Enterprise (2012). The contribution that expert judgement makes is described at each stage of the five stage process and includes:

- Stage 1: The classification of habitat layers by EUNIS code (expert judgement applied in identification of habitat maps and applying EUNIS codes);
- Stage 2: Linking habitat layers to sensitivity information (expert judgement used when applying sensitivity scores);
- Stage 3: Production of pressure layers and classification into exposure values (expert judgement used for the production of activity and pressure layers and defining exposure threshold values);
- Stage 4: Assessment of habitat vulnerability (expert judgement used in the production of vulnerability layers); and
- Stage 5: Assessment of habitat condition (expert judgement used in the overall assessment of likely condition).

The use of expert judgement should not preclude the delivery of robust and transparent status assessments for marine biodiversity, although the need for, *inter alia*, a more robust and efficient assessment process which is centred on data, regularly repeated, and harmonised with, and responsive to, emerging requirements has been highlighted in recent work. Standards and guidelines are required alongside more-explicit representations of uncertainty to allow policy makers, scientists and stakeholders to both understand and trust in marine biodiversity status assessments. Whilst work is underway to address these points, the next cycle of UK marine reporting will (particularly with reference to status assessments of benthic habitats under the MSFD) continue to rely strongly on expert judgement, alongside limited monitoring data and information on the distribution of benthic habitats and pressures.

This report clearly demonstrates the pivotal role that expert judgement can play in benthic habitat vulnerability assessments and clarifies where different approaches and techniques (such as elements of good practice) might be effectively deployed. The five stage process for vulnerability assessment that is outlined, together with the recommendations that are made, should be used to provide the basis of an approach for assessing benthic habitats in the UK as part of meeting the 2018 MSFD requirements for reporting on the state of Europe’s seas.

6.3 Limitations

The supporting work for this report was undertaken as a defined commission that followed a closely defined scope. As a consequence, there inevitably are areas that might otherwise have been expanded upon. In particular, it should be recognised that the 13 selected case studies used to provide the central platform for many of the recommendations contained in this report relate specifically to the use of expert judgement in biodiversity assessments. As such they are effectively a subset of a wider range of literature that incorporates the elicitation and application of expert judgement.

In part, the recommendations for further work (below) reflect this limitation, and seek to expand on concepts or approaches that, whilst they may have been highlighted within the biodiversity literature, may be covered in more detail elsewhere. There is also the possibility that approaches not previously used for work on biodiversity assessment, but used successfully in other fields, may have been reported in the wider literature.

6.4 Recommendations for Further Work

6.4.1 **Bayesian Belief Networks**

Bayesian belief networks (BBNs) provide a rational technique for integrating both subjective opinion (e.g. expert knowledge) on probabilistic relations and quantitative empirical data (e.g. monitoring data, modelling results, etc.). They may provide a useful and practical means of re-examining the sensitivity information collected as part of the MB0102 project, with a view to developing a probability-based sensitivity assessment tool rather than the qualitative tool that is currently used. Accordingly, it is recommended that further work is commissioned to examine the potential role that might be played by adopting a BBN approach to linking pressures to feature sensitivities.

6.4.2 **Fuzzy Logic**

The use of fuzzy logic (in the form of fuzzy triangle numbers) has been suggested as a means of eliciting judgements that help capture uncertainty around judgements made by experts. As this report has focussed largely on studies in the field of biodiversity assessment, this suggestion is inevitably the product of a relatively small ‘sample size’ of case studies. Consequently, whilst there are good grounds for suggesting its consideration as a valuable technique, the underlying methods discussed in this report (fuzzy triangle numbers with centre-of-gravity defuzzification) may not be optimal.

A wider literature review examining the use of fuzzy logic as a means of accommodating inherent uncertainties in expert elicitation studies in general could help better define its role and use in the more specific context of habitat vulnerability assessment. It is likely that, by taking a more general overview than the relatively tight focus on the field of biodiversity, it would be possible to identify ‘best practice’ when applying fuzzy logic to the elicitation and use of expert judgement.

6.4.3 **Handling Uncertainty**

It is recommended that further work be considered to address the use of information on uncertainty, or confidence that surrounds expert judgements. In particular, there is the need to develop a process to derive a measure of overall confidence when combining data from different sources, and where the confidence in, or uncertainty around, each data source may have been defined or recorded using different methods.

Developing a better understanding of uncertainty in expert judgements and the development of a formalised approach to its characterisation and incorporation into biodiversity assessments has fallen outside of the scope for this report but could usefully be addressed through a targeted literature review.

6.5 Acknowledgements

The authors would like to thank staff at JNCC for their comments and input through the development of this report, and the following researchers who have been kind enough to offer support in the identification and provision of supporting literature for this work:

Villie FLARI
Food and Environment Research Agency (fera)
York, UK

Simon FRENCH
Risk Initiative and Statistical Consultancy Unit
Department of Statistics, University of Warwick, UK

Marissa McBRIDE
Australian Centre of Excellence for Risk Analysis
University of Melbourne, Australia

Karlene MILLER
Southern California Coastal Water Research Project
Costa Mesa, California, USA

Rabin NESLO
Universitair Medisch Centrum (UMC)
Utrecht, Netherlands

Steve WEISBERG
Southern California Coastal Water Research Project
Costa Mesa, California, USA

7 References

ABPmer (2011) Marine Protected Areas - gathering/developing and accessing the data for the planning of a network of Marine Conservation Zones - MB0102 Marine Biodiversity R&D Programme. Funded by Defra, JNCC, CCW, NW, Scottish Government, DOENI & Isle of Man Government.

AISH, A., ASHWORTH, J., BARRIO FROJÁN, B., BENJAMINS, S., BOLAM, S., BRAZIER, P., BROCKINGTON, S., CHANIOTIS, P., CLEMENTS, A., COGGAN, R., COLTMAN, N., CONNOR, D., COYLE, M., DAVIES, J., EDWARDS, H., EDWARDS, T., GREATHEAD, C., GREENSTREET, S., HALL-SPENCER, J., HAWKRIDGE, J., HISCOCK. K., HOWELL. K., HUGHES, A., IDE, R., JAMES, B., KELLY, C., KHAN-MARNIE, J., MACKIE, T., MANSON, F., MAZIK, K., MIESZKOWSKA, N., MOORE, J., O'REILLY, M., PIRIE, P., PRIOR, A., REACH, I., REES, I., ROBINSON, L., SANDERSON, B., SERVICE, M., STOKER, B., TELLIER, P. & VERLING, E. (2010) Charting Progress 2: Healthy and Biological Diverse Seas Feeder Report: Section 3.1: Marine Habitats.. In *UKMMAS (2010) Charting Progress 2 Healthy and Biological Diverse Seas Feeder Report (Eds*. M. Frost& J. Hawkridge*).* Published by Department for Environment Food and Rural Affairs on behalf of UKMMAS. p68-271.

BASHARI, H., SMITH, C. & BOSCH, O.J.H. (2009) Developing decision support tools for rangeland management by combining state and transition models and Bayesian belief networks, *Agricultural Systems*, **99**, 23–34.

BRAUN, P.A. & YANIV, I. (1992) A case study of expert judgment: Economists' probabilities versus base-rate model forecasts. *Journal of Behavioral Decision Making*, **5**(3), 217-31.

BURGMAN, M., CARR, A., GODDEN, L., GREGORY, R., MCBRIDE, M., FLANDER, L. & MAGUIRE, L. (2011a) Redefining expertise and improving ecological judgment. *Conservation Letters*, **4**, 81-7.

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**, e22998.

CEBC (2009) Guidelines for Systematic Review in Conservation and Environmental Management. Version 3.1. Centre for Evidence-Based Conservation, Bangor University, Wales, UK. Also available online at:
http://www.cebc.bangor.ac.uk/Documents/Reviewguidelinesversion3.0_FINAL.pdf.
[accessed July 2013].

CERTAIN, G. & SKARPAAS, O. (2010) *Nature Index: General Framework, Statistical Method and Data Collection for Norway*, Norwegian Institute for Nature Research (NINA) Report 542, 47pp.

CERTAIN, G., SKARPAAS, O., BJERKE, J-W., FRAMSTAD, E., LINDHOLM, M., NILSEN, J-E., NORDERHAUG, A., OUG, E., PEDERSEN, H-C., SCHARTAU, A-K., VAN DER MEEREN, G.I., ASLAKSEN, I., ENGEN, S., GARNÅSJORDET, P-A., KVALØY, P., LILLEGÅRD, M., YOCCOZ, N.G. & NYBØ, S. (2011) The Nature Index: A general framework for synthesizing knowledge on the state of biodiversity. *PLoS ONE*, **6**(4): e18930. doi:10.1371/journal.pone.0018930.
Available at: http://www.plosone.org/article/info:doi/10.1371/journal.pone.0018930 [accessed October 2012].

COHEN, J. (1960) A coefficient of agreement for nominal scales, *Educational and Psychological* Measurement, **20**, 37–46.

CONNOR, D. & ENSERINK, L. (2009) Report of the Utrecht Workshop – Regional Assessment, Ospar Commission, London. Publication 468/2009.

COOKE, R.M. & PROBST, K.N. (2006) Highlights of the Expert Judgment Policy Symposium and Technical Workshop. Available at: http://rff.org/Documents/Conference-Summary.pdf

COOKE, R.M. & GOOSSENS, L.J.H. (2000) *Nuclear science and technology. Procedures guide for structured expert judgment.* Project report EUR 18820EN 'Nuclear fission safety 1994-98', European Commission, Luxembourg, Euratom pp.1-55. ISBN 92-894-0111-7

COWLING, R.M., PRESSEY, R.L., SIMS-CASTLEY, R. LE ROUX, A., BAARD, E., BURGERS, C.J. & PALMER, G. (2003) The expert or the algorithm? - comparison of priority conservation areas in the Cape Floristic Region identified by park managers and reserve selection software, *Biological Conservation*, **112**, 147–67.

DCLG, (2009). Multi-criteria analysis: a manual. Department for Communities and Local Government. London, ISN 9781409810230, 168pp. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/7612/1132618.pdf [accessed July 2013].

EASTWOOD, P.D., MILLS, C.M., ALDRIDGE, J.N., HOUGHTON, C.A., & ROGERS, S.I., (2007). Human activities in UK offshore waters: an assessment of direct, physical pressure on the seabed. *ICES Journal of Marine Science*, **64**, 453-63.

ENQUIST, C.A.F. & GORI, D.F. (2008) Application of an expert system approach for assessing grassland status in the U.S.-Mexico borderlands: implications for conservation and management, *Natural Areas Journal*, **28**(4), 414–28.

EVANS, C. & CRAWFORD, B. (2000). Expert judgement in pharmacoeconomic studies. Guidance and future use. *Pharmacoeconomics*, **17**(6), 545-53.

FLARI, V., CHAUDREY, Q., NESLO, R. & COOKE, R. (2011) Expert judgement based multi-criteria decision model to address uncertainties in risk assessment of nanotechnology-enabled food products. *Journal of Nanoparticle Research*, **13**, 1813-31.

FLEISS, J.L. & COHEN, J. (1973) The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability, *Educational and Psychological Measurement*, **33**, 613–19.

FRENCH, S. (2011) Aggregating expert judgement, *RACSAM (Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas)*, **105**(1), 181-206

GEHRIS, R. (2008) A simulation study of the classical method of expert judgment combination: how many seeds and how many experts? ProQuest. 208 pp. Available at: http://search.proquest.com/docview/304639539/fulltextPDF [accessed February 2013]

GOOSSENS, L.H.J., COOKE, R.M., HALE, A.R. & RODIC´-WIERSMA, Lj. (2008) Fifteen years of expert judgement at TUDelft, *Safety Science*, **46**, 234–44

HALPERN, B.S., SELKOE, K.A., MICHELI, F. & KAPPEL, C.V. (2007) Evaluating and ranking the vulnerability of global marine ecosystems to anthropogenic threats, *Conservation Biology*, **21**(5), 1301–15.

JORGENSEN, M. (2004) A review of studies on expert estimation of software development effort. *The Journal of Systems and Software*, **70**, 37–60.

KANDLIKAR, M., RAMACHANDRAN, G., MAYNARD, A., MURDOCK, B. & TOSCANO, W.A. (2007) Health risk assessment for nanoparticles: A case for using expert judgment. *Journal of Nanoparticle Research*, **9**, 137–56.

KAUFMANN, A. & GUPTA, M.M. (1985) *Introduction to Fuzzy Arithmetic: Theory and Applications*. Van Nostrand Reinhold, New York.

KEENEY, R.L & VON WINTERFELDT, D. (1989) On the uses of expert judgement on complex technical problems, IEEE Transactions on Engineering Management, **36**(2), 83-6.

KLEYPAS, J.A. & EAKIN, C.M. (2007) Scientists' perceptions of threats to coral reefs: results of a survey of coral reef researchers, *Bulletin of Marine Science*, **80**(2), 419-36.

KNOL, A.B., DE HARTOG, J.J., BOOGAARD, H., SLOTTJE, P., VAN DER SLUIJS, J.P., LEBRET, E., CASSEE, F.R., WARDEKKER, J.A., AYRES, J.G., BORM, P.J., BRUNEKREEF, B., DONALDSON, K., FORASTIERE, F., HOLGATE, S.T., KREYLING, W.G., NEMERY, B., PEKKANEN, J., STONE, V., WICHMANN, H.E. & HOEK, G. (2009) Expert elicitation on ultrafine particles: likelihood of health effects and causal pathways. *Particle and Fibre Toxicology*, **6,**19.

KNOL, A.B., SLOTTJE, P., VAN DER SLUIJS, J.P. & LEBRET, E. (2010) The use of expert elicitation in environmental health impact assessment: a seven step procedure. *Environmental Health*, **9**, 19.

KRAYER VON KRAUSS, M.P., CASMAN, E.A. & SMALL, M.J. (2004) Elicitation of expert judgments of uncertainty in the risk assessment of herbicide-tolerant oilseed crops. *Risk Analysis*, **24**, 1515-27.

KYNN, M. (2008). The 'heuristics and biases' bias in expert elicitation. *Journal of the Royal Statistical Society Series A* **171**, 239–64.

LANDIS, J.R. & KOCH, G.G. (1977) The measurement of observer agreement for categorical data, *Biometrics*, **33**, 159–74.

LEE, J., STELZENMULLER, V. & ROGERS, S. (2009) *Further development of marine pressure data layers and ensuring the socio-economic data and data layers are developed for use in the planning of marine protected area networks (MB0106) - Report No. 1: Objective 1 Provision of geo-database containing standardised layers showing the distribution of specified activities, sites and resources with associated metadata and comments*. Available at: http://sciencesearch.defra.gov.uk/Document.aspx?Document=MB0106_10042_FRP.pdf [accessed July 2013].

LOW-CHOY, S., R. A. O'LEARY, & K. L. MENGERSEN. (2009). Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology*, **90**, 265–77.

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**, 906–20.

MACCOUN, R. J. (2011). How should expert judgment inform the legalization debate? In *Controlling Crime: Strategies and Payoffs (eds* P. Cook, J. Ludwig & J. McCrary),. University of Chicago Press.

MARTIN, T.G., BURGMAN, M.A., FIDLER, F., KUHNERT, P.M., LOW-CHOY, S., MCBRIDE, M. & MENGERSEN, K. (2011) Eliciting expert knowledge in conservation science, *Conservation Biology*, **26**(1), 29-38.

MASSART, D.L., SMEYERS-VERBEKE, J., CAPRON, X. & SCHLESIER, K. (2005) Visual Presentation of Data by Means of Box Plots. *LC-GC Europe*, **18**(4), 215–18.

MEYER, M., & J. BOOKER. (1991). *Eliciting and Analyzing Expert Judgment: A Practical Guide*, Academic Press, New York.

MONSERUD, R., & LEEMANS, R. (1992) Comparing global vegetation maps with the Kappa statistic. Ecological Modelling, **62**, 275–93.

MPA MONITORING ENTERPRISE (2012) *A framework for advancing the use of expert judgments of ecosystem condition*.
Available at: http://monitoringenterprise.org/pdf/Expert_Judgment_Draft_Framework.pdf [accessed October 2012].

MUXIKA, I., BORJA, A. & BALD, J. (2007) Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive, *Marine Pollution Bulletin*, **55**, 16–29.

NATURAL ENGLAND & JNCC (2011) Marine Conservation Zone Project; Conservation Objective Guidance. Available online at:
http://jncc.defra.gov.uk/PDF/MCZ%20Project%20Conservation%20Objective%20Guidance.pdf [accessed March 2013]

NET GAIN (2011) Net Gain Final Recommendations Submission to Natural England & JNCC. Available at:
http://webarchive.nationalarchives.gov.uk/20120502152708/http://www.netgainmcz.org/docs/Net_Gain_Final_Rec_v1_1.pdf, [accessed October 2012].

NUNES, A. & KIRLIK, A. (2005) An empirical study of calibration in air traffic control expert judgment. In: *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting, 2005*. Available at:
www.humanfactors.illinois.edu/Reports&PapersPDFs/humfac05/nunkir.pdf [accessed October 2012].

O'HAGAN, A., C. E. BUCK, A. DANESHKHAH, J. R. EISER, P. H. GARTHWAITE, D. J. JENKINSON, J. E. OAKLEY, & T. RAKOW. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*, Wiley, Chichester, UK.

OKEY, T.A., ALIDINA, H.M., LO, V., MONTENEGRO, A. & JESSEN, S. (2012) *Climate change impacts and vulnerabilities in Canada's Pacific marine ecosystems*. CPAWS BC and WWF-Canada, Vancouver, BC.

OKEY, T.A. & LOUCKS, L.A. eds. (2011) *Social-ecological assessment for the marine and coastal areas of the west coast of Vancouver Island*. Port Alberni, B.C., Canada: The Tsawalk Partnership, West Coast Aquatic. Available at: http://csi-global.org/content/sea-wcvi [accessed January 2013].

OLIVER, D.M., FISH, R.D., WINTER, M., HODGSON, C.J., HEATHWAITE, A.L. & CHADWICK, D.R. (2012) Valuing local knowledge as a source of expert data: farmer engagement and the design of decision support systems. *Environmental Modelling & Software*, **36**, 76-85.

OSPAR (2010) *Quality Status Report 2010*. OSPAR Commission. London. 176pp.

PARIS, P, GOMEZ, M. & ANDERSON, W. (1961) A rational analytic theory of fatigue, *The Trend in Engineering*, **13**, 9-14.

PENG, W., ZAN, M.A. & YI, T. (2011) Application of expert judgment method in the aircraft wiring risk assessment, *Procedia Engineering*, **17**, 440–45.

PRIOR, A., MILES, A.C., SPARROW, A.J. & PRICE, N. (2004) Development of a classification scheme for the marine benthic invertebrate component, Water Framework Directive. Phase I & II – transitional and coastal waters. Environment Agency (UK), R&D Interim Technical Report, E1-116, E1-132, 103 p. (plus appendix).

ROBINSON, L.A., ROGERS, S., & FRID, C.L.J. (2008) *A marine assessment and monitoring framework for application by UKMMAS and OSPAR – Assessment of pressures* (Contract No. F90-01-1075 for the Joint Nature Conservation Committee). University of Liverpool, Liverpool, and Centre for the Environment, Fisheries and Aquaculture Science (CEFAS), Lowestoft; 108pp.

ROMAN, H.A., WALKER, K.D., RICHMOND, H.M., HUBBELL, B.J. & KINNEY, P.L. (2008). Expert judgment assessment of the mortality impact of changes in ambient fine particulate matter in the U.S. *Environmental Science and Technology*, **42**, 2268–74.

SCHOLES, R.J. & BIGGS, R. (2005) A biodiversity intactness index, *Nature*, **434**, 45–9.

SIMOLA, K., GANDOSSI, L., MENGOLINI, A. & BOLADO-LAVIN, R. (2005) Application of formal expert judgement to the evaluation of structural reliability. In *Proceedings of a Workshop on the Use of Expert Judgement in Decision Making*, , European Commission Joint Research Centre Institute for Energy (JRC-IE), pp 175-81.

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**, 512–23.

STERN, P.C., FINEBERG H.V., editors (1996) *Understanding Risk: Informing Decisions in a Democratic Society*. Committee on Risk Characterisation, National Research Council. National Academy Press, Washington, DC.

TECK, S.J., HALPERN, B.S., KAPPEL, C.V., MICHELI, F., SELKOE, K.A., CRAIN, C.M., MARTONE, R., SHEARER, C., ARVAI, J., FISCHHOFF, B., MURRAY, G., NESLO, R. & COOKE, R. (2010) Using expert judgment to estimate marine ecosystem vulnerability in the California Current, *Ecological Applications*, **20**, 1402–16.

TEIXEIRA, H., BORJA, A. WEISBERG, S.B., RANASINGHE, J.A., CADIEN, D.B., DAUER, D.M., DAUVIN, J-C., DEGRAER, S., DIAZ, R.J., GRÉMARE, A., KARAKASSIS, I., LLANSÓ, R.J., LOVELL, L.L., MARQUES, J.C., MONTAGNE, D.E., OCCHIPINTI-AMBROGIM, A., ROSENBERG, R., SARDÁ, R., SCHAFFNER, L.C. & VELARDE, R.G. (2010) Assessing coastal benthic macrofauna community condition using best professional judgement – developing consensus across North America and Europe, *Marine Pollution Bulletin*, **60**, 589-600.

TEN BRINK, B.J.E. & TEKELENBURG, T. (2002) *Biodiversity: how much is left? The Natural Capital Index framework (NCI)*. RIVM report 402001014. Bilthoven.

THOMPSON, B., WEISBERG, S.B., MELWANIA, A., LOWEA, S., RANASINGHE, J.A., CADIEN, D.B., DAUERD, D.M., DIAZE, R.J., FIELDS, W., KELLOGG, M., MONTAGNE, D.E., ODE, P.R., REISH, D.J. & SLATTERY, P.N. (2012) Low levels of agreement among experts using best professional judgment to assess benthic condition in the San Francisco Estuary and Delta, *Ecological Indicators*, **12**, 167–73.

THORNE, M.C. & WILLIAMS, M.M.R. (1992) A review of expert judgment techniques with reference to nuclear safety, *Progress in Nuclear Energy*, **27**(2-3), 83-254.

TILLIN, H.M., HULL, S.C. & TYLER-WALTERS, H. (2010) *Development of a Sensitivity Matrix (pressures-MCZ/MPA features).* s.l.: Report to the Department of Environment, Food and Rural Affairs from ABPMer, Southampton and the Marine Life Information Network (MarLIN) Plymouth: Marine Biological Association of the UK. Defra Contract No. MB0102 Task 3A, Report No. 22., 2010. Report No 22 Task 3 Development of a Sensitivity Matrix (pressures-MCZ/MPA features).

UK MARINE MONITORING AND ASSESSMENT STRATEGY (2010) *Charting Progress 2 Healthy and Biological Diverse Seas Feeder report*. (*Eds*. M. Frost, & J. Hawkridge). Published by Department for Environment Food and Rural Affairs on behalf of UKMMAS. 682pp. Available at: http://chartingprogress.defra.gov.uk/healthy-and-biologically-diverse-seas-feeder-report [accessed October 2012].

U.S. ENVIRONMENTAL PROTECTION AGENCY (2009). *Expert Elicitation Task Force Draft White Paper*. US Environmental Protection Agency Washington DC. Available at: http://www.epa.gov/osa/pdfs/elicitation/Expert_Elicitation_White_Paper-January_06_2009.pdf [accessed October 2012].

VAN DER FELS-KLERX, H.J., COOKE, R.M., NAUTA, M.N., GOOSSENS, L.H. & HAVELAAR, A.H. (2005) A structured expert judgment study for a model of *Campylobacter* transmission during broiler-chicken processing. *Risk Analysis*, **25**, 109-24.

VAUGHAN, D. *et al.* (n prep.) *Assessing the structures and functions parameter of FCS for reef and sandbanks in UK offshore waters*, Joint Nature Conservation Committee.

WALTON, D. (1997) *Appeal to Expert Opinion: Arguments from Authority*. Pennsylvania State University Press, Pennsylvania.

WARD, T.J. (2011) *SOE 2011: National Marine Condition Assessment – Decision Model and Workshops*. Report prepared for the Australian Government Department of Sustainability, Environment, Water, Population and Communities on behalf of the State of the Environment 2011 Committee. Canberra: DSEWPaC, 2011.

WOOLDRIDGE, S. (2003) *Bayesian Belief Networks*, Centre for Complex System Science, CSIRO, Canberra. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.112.6230&rep=rep1&type=pdf [accessed March 2013]

YAGER, R.R. & FILEV, D. (1993) On the issue of defuzzification and selection based on a fuzzy set, *Fuzzy Sets and Systems*, **55**, 255–71.

ZACHARIAS, M.A. & GREGR, E.J. (2005) Sensitivity and vulnerability in marine environments: an approach to identifying vulnerable marine areas, *Conservation Biology*, **19**(1), 86-97.

ZADEH, L. (1965) Fuzzy Sets, *Information and Control*, **8**, 338-53.

ZICKFELD, K., LEVERMANN, A., MORGAN, M.G., KUHLBRODT, T., RAHMSTORF, S. & KEITH, D.W. (2007) Expert judgements on the response of the Atlantic meridional overturning circulation to climate change. *Climate Change*, **82**, 235-65.

Appendix 1. Search Protocol for Selection of Case Studies

The following text outlines the stages employed to identify candidates for inclusion in the review and analysis of case studies considered under Objectives 2 and 3 of the present study (see Section 2.1 of the main text).

Stage 1: Identify the main question:

In this case: "Review of case studies and recommendations [c] for the inclusion of expert judgement [b] in marine biodiversity status assessments [a]?"

[a] subject, [b] intervention, [c] outcome.

Subsequently, break the main question down to enable a more targeted search:

- What is expert judgement? Theory and methodologies behind it?
- Where has expert judgement been used to make marine status assessments?
- What other disciplines have used expert judgement?
- What are the recommendations for best and worst practice from expert judgement?

Stage 2: Construct a search strategy:

- Record all search terms used
- Assess relevance of returned results
- Save relevant searches for access during later stages of the search
- Follow up on relevant references from within the papers (paper chase)

Stage 3: Identify potential sources of information:

- ISI Web of Science
- Science Direct/SciVerse
- Scopus
- JSTOR
- BIOME
- ETHoS, www.findathesis.com
- Scirus
- PubMed
- Google and, where applicable, other internet search facilities
- Conservation Agencies, Environment Agency, Defra, Cefas, MarLIN online catalogues and discussions with contacts within these organisations
- Article reference sections
- Literature held in IECS (EndNote Libraries)

IECS has full access to the University library facilities which include:

- On-site collection;
- Electronic access to a wide range of online journal articles;
- Access to the British Library;
- Online literature search facilities (Scopus, Web of Science, Science Direct; and
- SciVerse).

Stage 4: Define search terms used:

The terms searched for included:

- Expert judgement
- Expert judgment (American spelling brought up more results)
- Expert judgement/judgment AND assessment
- Best professional judgement/judgment
- Expert elicitation
- Elicitation AND Elicit
- Value judgement/judgment
- Expert opinion
- Expert knowledge
- Marine biodiversity
- Biodiversity assessments

The search also included publications by specific authors known to specialise in the field of expert judgement.
Interchangeable words were used to search e.g. marine biodiversity assessments = species assessments, marine protected areas, conservation assessments etc.
A full list of search terms was developed based on the literature found and the necessity to expand/reduce the number of hits.

Stage 5: Assess the quality and relevance of the information:

- Irrelevant titles were rejected. Broadly relevant titles were considered but rejected if the article did not contain relevant information.
- Articles which did not directly relate to the study were rejected.
- Studies that provide the application of expert judgement to other disciplines were included.
- Studies which discussed the theory and methodologies behind expert judgement were also saved.
- Biased, emotive articles were rejected, as were studies that did not have a high degree of objectivity.
- Articles needed to have clearly stated aims, objectives and hypotheses to be included.
- Studies which were not genuinely informative were excluded.

Priority was given to peer reviewed scientific literature and Government Agency reports (written by qualified, practising scientists) but others were considered provided that they met the above criteria for objectivity, experimental design etc.

The above process was based on:

CEBC (2009) *Guidelines for Systematic Review in Conservation and Environmental Management. Version 3.1*. Centre for Evidence-Based Conservation, Bangor University, Wales, UK. Also available online at: http://www.cebc.bangor.ac.uk/Documents/Reviewguidelinesversion3.0_FINAL.pdf [accessed July 2013].

Appendix 2. Examples of Expert Judgement Used in Other (Non-Marine) Disciplines

The following bibliography provides some examples of where expert judgement has been used across a range of different disciplines. This listing should be used for indicative purposes only and is not intended to be exhaustive.

Climate Change

ALLALI, A., BOJARIU, R., DIAZ, S., ELGIZOULI, I., GRIGGS, D., HAWKINS, D., HOHMEYER, O., JALLOW, B.P., KAJFEZ-BOGATAJ, L., LEARY, N., LEE, H. & WRATT. D (Eds) (2007) *Climate Change 2007: Synthesis Report. An Assessment of the Intergovernmental Panel on Climate Change* [This underlying report, adopted section by section at IPCC Plenary XXVII (Valencia, Spain, 12-17 November 2007), represents the formally agreed statement of the IPCC concerning key findings and uncertainties contained in the Working Group contributions to the Fourth Assessment Report.]

WARDEKKER, J.A., de JONG, A., van BREE, L., TURKENBURG, W.C. & van der SLUIJS, J.P. (2012) Health risks of climate change: An assessment of uncertainties and its implications for adaptation policies, *Environmental Health*, **11**, 67.

ZICKFELD, K., LEVERMANN, A., MORGAN, M.G., KUHLBRODT, T., RAHMSTORF, S. & KEITH, D.W. (2007) Expert judgements on the response of the Atlantic meridional overturning circulation to climate change, *Climate Change*, **82**, 235-65.

Economics

BRAUN, P.A. & YANIV, I. (1992) A case study of expert judgment: Economists' probabilities versus base-rate model forecasts, *Journal of Behavioral Decision Making*, **5**(3), 217-31.

ZOPOUNIDIS, C., DOUMPOS, M. & MATSATSINIS, N.F. (1997) On the use of knowledge-based decision support systems in financial management: A survey, *Decision Support Systems*, **20**, 259-77.

Air Quality/Public Health

INDUSTRIAL ECONOMICS (2004) *An Expert Judgment Assessment of the Concentration-Response Relationship Between Pm2.5 Exposure and Mortality*. Report prepared for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711.

KANDLIKAR, M., RAMACHANDRAN, G., MAYNARD, A., MURDOCK, B. & TOSCANO, W.A. (2007) Health risk assessment for nanoparticles: A case for using expert judgment, *Journal of Nanoparticle Research*, **9**, 137–56.

KNOL, A.B., SLOTTJE, P., van der SLUIJS, J.P. & LEBRET, E. (2010) The use of expert elicitation in environmental health impact assessment: a seven step procedure, *Environmental Health*, **9**, 19.

ROMAN, H.A., WALKER, K.D., RICHMOND, H.M., HUBBELL, B.J. & KINNEY, P.L. (2008) Expert judgment assessment of the mortality impact of changes in ambient fine particulate matter in the U.S., *Environmental Science and Technology*, **42**, 2268–74.

VRANA, I., VANÍCEK, J., KOVÁR, P., BROZEK, J. & ALY, S. (2012) A group agreement-based approach for decision making in environmental issues, *Environmental Modelling & Software,* **36**, 99-110.

Food Industry

FLARI, V., CHAUDREY, Q., NESLO, R. & COOKE, R. (2011) Expert judgement based multi-criteria decision model to address uncertainties in risk assessment of nanotechnology-enabled food products, *Journal of Nanoparticle Research*, **13**, 1813-31.

KRETZSCHMAR, U. & SCHMID, O. (2011) Quality and safety aspects of organic and low-input food processing: Results of a Delphi survey from an expert consultation in 13 European countries, *NJAS - Wageningen Journal of Life Sciences*, **58**, 111–16.

OLIVER, D.M., FISH, R.D., WINTER, M., HODGSON, C.J., HEATHWAITE, A.L. & CHADWICK, D.R. (2012) Valuing local knowledge as a source of expert data: farmer engagement and the design of decision support systems, *Environmental Modelling & Software*, **36**, 76-85.

SOON, J.M., DAVIES, W.P., CHADD, S.A. & BAINES, R.N. (2012) A Delphi-based approach to developing and validating a farm food safety risk assessment tool by experts. *Expert Systems with Applications*, **39**, 8325–36.

Industrial Applications (nuclear, engineering, flood risk, aerospace)

FAILING, L., HORN, G. & HIGGINS, P. (2004) Using expert judgment and stakeholder values to evaluate adaptive management options, *Ecology and Society*, **9**(1), 13.

JRC (2005) *Workshop on the Use of Expert Judgement in Decision Making*. June 21-23, 2005 (and associated papers) Available at: http://www-cadarache.cea.fr/fr/reseau/pdf/WEJ.pdf [accessed July 2013].

KEENEY, R.L & VONWINTERFELDT, D. (1989) On the uses of expert judgement on complex technical problems, *IEEE Transactions on Engineering Management*, **36**(2), 83-6.

NUNES, A. & KIRLIK, A. (2005). An empirical study of calibration in air traffic control expert judgment. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting, 2005.* Available at: www.humanfactors.illinois.edu/Reports&PapersPDFs/humfac05/nunkir.pdf [accessed October 2012].

PENG, W., ZAN, M.A. & YI, T. (2011) Application of expert judgment method in the aircraft wiring risk assessment, *Procedia Engineering*, **17**, 440–5.

ROSQVIST, T. (2003) *On the use of expert judgement on the qualification of risk assessment. VTT Industrial Systems*. Dissertation for the Degree of Doctor of Technology , Helsinki University of Technology, ESPOO, Finland.

SIMOLA, K., GANDOSSI, L., MENGOLINI, A. & BOLADO-LAVIN, R. (2005) Application of formal expert judgement to the evaluation of structural reliability. In *Proceedings of a workshop on the Use of Expert Judgement in Decision making*, the European Commission Joint Research Centre Institute for Energy (JRC-IE), pp 175-181.

THORNE, M.C. & WILLIAMS, M.M.R. (1992) A review of expert judgment techniques with reference to nuclear safety. *Progress in Nuclear Energy*, **27**(2-3), 83-254.

Medicine/Health

ARIMONE, Y., BEGAUD, B., MIREMONT-SALAME, G., FOURRIER-REGLAT, A., MOORE, N., MOLIMARD, M. & HARAMBURU, F. (2005) Agreement of expert judgment in causality assessment of adverse drug reactions, *European Journal of Clinical Pharmacology*, **61,** 169–73.

BRUINE De BRUIN, W., FISCHHOFF, B., BRILLIANT, L. & CARUSO, D. (2006) Expert judgments of pandemic influenza risks, *Global Public Health*, **1**(2), 178-93

EVANS, C. & CRAWFORD, B. (2000) Expert judgement in pharmacoeconomic studies. Guidance and future use. *Pharmacoeconomics*, **17**(6), 545-53.

Software

JORGENSEN, M. (2004) A review of studies on expert estimation of software development effort, *The Journal of Systems and Software*, **70**, 37–60.

Others

ALIGICA, P.D. & HERRITT, R. (2009) Epistemology, social technology, and expert judgement: Olaf Helmer's contribution to futures research, *Futures*, **41**, 253–59.

MACCOUN, R. J. (2011) How should expert judgment inform the legalization debate? In *Controlling crime: Strategies and Payoffs (eds* P. Cook, J. Ludwig & J. McCrary),. University of Chicago Press.

Appendix 3. Application of Expert Judgement in Other Disciplines

Case Study 1. Expert Elicitation on Ultrafine Particles: Likelihood of Health Effects and Causal Pathways (Knol *et al.,* 2010)

Knol *et al.* (2010) use three separate public health studies that make use of expert judgement to provide case examples to support a proposed seven step framework (see Figure A3.1) for the application of expert judgement (or 'expert elicitation') within the context of integrated environmental health impact assessment.

Figure A3.1. Seven step procedure for organising a formal expert elicitation (after Knol *et al.*, 2010)

The three example studies used in their report relate to:

- The potential health effects of exposure to ultrafine particles (Knol *et al.*, 2009)
- The potential adverse effects of genetically modified herbicide-resistant crops on agricultural and cultivation practices (Krayer von Krauss *et al.*, 2004); and
- Broiler-chicken processing and potential exposure to *Campylobacter* (van der Fels-Klerx *et al.*, 2005).

These particular case studies were chosen because they differed on three main issues (the type of uncertain information to be elicited; the purpose of the elicitation or how the elicited information is intended to be used; and the resources available to the study) and, hence, used different designs.

Why Was Expert Judgement Used?

The following text is based extensively on Knol *et al.* (2010). Integrated environmental health impact assessment (IEHIA) is defined as a process that aims to support policy-making by comprehensively assessing environmental health effects, whilst taking account of underlying complexities. Within this context, professionals are confronted with questions about the overall impact of environmental stressors on public health and about the beneficial effects of policy measures intended to reduce environmental exposures. However, because of limitations and inconsistencies in the underlying scientific knowledge base, it is often difficult to address questions such as these. Developing a formalised process for employing expert elicitation is one means of moving towards a structured and transparent way to address such uncertainties. Such a process can also serve as a way of presenting limited available knowledge in order to inform policies which have to be made before conclusive scientific evidence becomes available.

Knol *et al.* (2009) describe the elicitation of expert judgement as being a useful means of gaining insight into environmental health issues about which current evidence is limited or inconclusive. It provides a temporary summary of the limited available knowledge. Although expert elicitation is most commonly used to estimate quantitative values, it can also provide insight into qualitative issues or conceptual (causal) models. Expert elicitation may help to structure a problem and can be used to focus new research on the most salient uncertain issues. As such, expert elicitation can be applied widely in environmental health research, and provide useful contributions relevant to all phases of an IEHIA. The transparency and reproducibility, and most likely also the quality of the elicited information, increases when the expert elicitation is carried out according to a systematic protocol.

Benefits and Limitations of the Application of Expert Judgement

Knol *et al.* (2009) allude to potential cost implications for expert elicitation, noting that the resources available for expert elicitation were only sufficient in two of the three case studies that they present. The total budget of the *Campylobacter*-elicitation work was relatively high (c.50k Euros) but allowed the project team to interview each of 12 experts individually. This individualised approach was followed as it was thought that the experts would be more motivated to cooperate if it was apparent that their opinions were considered important enough for personal interviews. A similar approach (individual interviews) was followed in the genetically modified (GM) crop elicitation work that is reported by Knol *et al*. (2009), although the project team involved considered that different experts might have had different understandings of the questions (a situation that would have been countered to some extent if group workshops had been held). Indeed, this was recognised as one of the main benefits of group elicitation recognised in the ultrafine particle elicitation.

Knol *et al.* (2009) conclude that formal expert elicitation can support environmental health research in various ways and is a valuable and necessary method to improve understanding and inform assessments and policies, as well as help prioritise research agendas. This conclusion is unlikely to be limited solely to the field of environmental health research.

Use of Expert Judgement in Decision Making

The work presented by Knol *et al.* (2009) does not represent a simple decision-making study. However, the work that is reported provides a useful seven step procedure for organising a formal expert elicitation (presented in detail in Annex 1 to this Appendix).

The discussion of this seven step procedure which they provide, with reference to the three key studies that were considered, provides a number of useful pointers that are likely to be applicable to, or might otherwise inform, the development of expert judgement within the context of marine biodiversity assessments.

Knol *et al.* (2009) suggest that natural scientists tend to be more sceptical towards using expert elicitation methods than scientists from other disciplines. In exact disciplines, expert elicitation is not perceived as a reliable or rigorous scientific method (compared to those used in empirical studies) and consequently the results of formal expert elicitation are often considered as being inherently less accurate. It is noted that such criticism can often be traced back to a lack of knowledge about formal expert elicitation, or a disproportionate trust in the quality and relevance of empirical data. It can also be argued that empirical data may also often contain many - often implicit - expert judgements and that by making such judgements explicit and transparent, as is done in formal expert elicitation, criticism should in fact decrease instead of increase.

In terms of decision making however, Knol *et al.* (2009) conclude that expert elicitation can be a useful means to gain insight into issues (e.g. environmental health) about which current evidence is either limited or inconclusive, providing a temporary summary of the limited available knowledge. As such it can be used as a relatively quick and inexpensive (albeit lower quality) substitute for time- or money-consuming research. The results can be used to develop a basis for action in cases where problems are too urgent or stakes are too high to postpone measures until more complete knowledge is available.

Case Study 2. Application of Formal Expert Judgement to the Evaluation of Structural Reliability (Simola *et al.,* 2005)

Simola *et al.* (2005) provide an example of one of the many studies that have applied expert judgement to areas relating to structural integrity.

The main objective of the paper by Simola *et al.* (2005) was not to solve a real structural integrity problem, but rather to obtain useful experience for developing guidelines for formal expert judgement applications in the area. Whilst in this sense the basis of the paper is conceptual, the work undertaken directly mirrors that which would be required for a real-world application and, as such, it provides a valuable reference for the application of expert judgement in this area.

The selected experts were playing the role of engineering consultants on behalf of a fictional nuclear utility operating 26 units. Several cracks, all found in a specific circumferential weld of the reheat inlet penetrations, had occurred throughout a fleet of 14 units at a second utility. The affected components are high safety-significant; a guillotine break of one of the circumferential welds could lead to serious consequences. For this reason, the second utility was forced to shut down some of its units, while the others operated at reduced power. The nuclear regulator has required a probabilistic assessment of the structural integrity of the similar component at the first utility. However for several reasons (e.g. pressure to avoid shutting down further plants during the winter months, lack of time, lack of required expertise) the only way forward as a short-time solution is identified in the use of expert judgements.

Underlying Approaches and Philosophy – Why Was Expert Judgement Used?

The following text is based extensively on Simola *et al.* (2005) who apply the use of expert judgement in a study examining aspects of structural integrity in the nuclear industry. They suggest that aspects of engineering such as this require a move towards the production of more realistic estimates for problems involving structural integrity, where probabilistic approaches are used to model uncertainties.

Within the nuclear industry there are several degradation mechanisms (e.g. those that relate to a 'loss of cooling accident') for which no validated structural reliability tools are available. Even for the better-known mechanisms, different models may produce quite different results.

Whilst formal expert judgement has become a rather well-established tool in connection to risk assessments, its application in the field of structural reliability of nuclear power plant components seems to be rather limited; the scarcity of operating experience and the quality of data both tend to limit the use of statistical approaches.

An internal case study was initiated within the Joint Research Centre, Institute for Energy (JRC-IE) to examine specific features of the application of formal expert judgement in structural reliability issues. The main aim was not to solve structural integrity problems *per se*, but rather to obtain useful experience for establishing guidelines for future formal expert judgement applications, effectively developing a methodology to help address issues of structural reliability.

Benefits and Limitations Inherent in the Application of Expert Judgement

Given such an information-poor environment the use of structured expert judgement may provide the optimum means for obtaining numerical estimates for structural reliability issues.

Simola *et al.* (2005) outline a series of requirements for selecting experts for eliciting expert judgement, citing Cooke & Goossens' Procedures Guide for Structured Expert Judgement (Cooke & Goossens, 2000):

- Reputation in the field of interest;
- Experimental experience in the field of interest;
- Number and quality of publications in the field of interest;
- Familiarity with uncertainty concepts;
- Diversity in background;
- Awards;
- Balance of views;
- Interest in the project; and
- Availability for the project.

Whilst these requirements appear reasonable, a potential limitation to the study was identified at the outset: because of the nature of the problem being addressed it was not possible to completely fulfil all of the requirements for the case study. This may be a problem that becomes manifest in other studies.

Evaluation of the hypothetical case study proposed by Simola *et al.* (2005) used a relatively small group of experts. Four technical experts were invited to take part although only one had previously had hands on experience on probabilistic fracture mechanics (the others had experience mainly in deterministic structural analyses). One of the experts made clear from the beginning that he did not consider himself as a specialist in structural integrity calculations, but rather in manufacturing and welding techniques. Only one of the experts declared that he had previous experience on assessing 'leak-before-break' scenarios. However all four experts showed a keen interest in taking part.

Furthermore it was noted that, even though the experts taking part may be familiar with most of the concepts involved, they may lack specific knowledge (e.g. of the subjective interpretation of probability) and may not be aware of potential biases that may relate to their judgements.

In terms of benefits however, the principal advantage of the approach investigated by Simola *et al.* (2005) lay in its ability to provide an estimate of likely failure rates for what was a safety-critical component; an assessment that would have been severely limited given reliance solely on the 'regular' methodologies of detailed empirical modelling.

Use of Expert Judgement in Decision Making

A meeting of all four experts was used to provide a forum in which the experts had an opportunity to discuss their analytical approaches without providing final quantitative estimates of probability of failure. At this point most of the experts had not actually performed their analyses. After this meeting, and having given the experts more time to finalise their assessments, a series of expert elicitation interviews were held on an individual basis.

Differences across the judgements made by the four experts were noted in a number of areas, for example:

- The underlying analysis approaches used by the experts;
- The use of different levels of simplification when applying relevant models;
- Differences in the definitions of key terms (e.g. 'failure') as used in the questions to elicit the expert judgement; and
- Differences in the interpretation of questions (e.g. estimating the probability of failure over an unlimited period or just within the typical lifetime of a nuclear plant).

However, all four experts were in agreement over two major areas: the assumption that fatigue was the only possible crack growth mechanism and so modelling of crack propagation could be based on the so-called Paris law (Paris *et al*., 1961); and the assumption that the growth rate of cracks would be slow and so a leak or failure during the life of the plant would only be possible in instances where the initial crack is very large.

A comparison of the seven step procedures as presented by Knol *et al.* (2009) and by Simola *et al.* (2005) are presented in more detail in Annex 2 to this Appendix.

Case Study 3. Expert Judgements on the Response of the Atlantic Meridional Overturning Circulation to Climate Change (Zickfeld *et al.,* 2007)

The study presented by Zickfeld *et al.* (2007) describes an expert judgement elicitation exercise concerning the possible effects of global climate change on the Atlantic Meridional Overturning Circulation (AMOC). The exercise sought to examine the range of opinions within the climatic research community about the physical processes that determine the current strength of the AMOC; its future evolution in a changing climate; and the consequences of potential AMOC changes.

Underlying Approaches and Philosophy – Why Was Expert Judgement Used?

A number of lines of evidence are available that can be used to infer possible changes in the AMOC in the face of ongoing climate change although none allow definitive predictions to be made. A new paradigm is required to draw these lines of evidence together; one possible strategy for achieving this is by expert consensus review. Such an approach may be complemented by the (quantitative) elicitation of individual expert judgement to provide an explicit quantification of uncertainty that would not be available in the formal literature.

Zickfeld *et al.* (2007) note that, as well as being applied to Bayesian decision analysis across a variety of business planning and other applications, expert judgement has also been widely used to assist decision making in climate and other areas of environmental policy.

Benefits and Limitations Inherent in the Application of Expert Judgement

The elicitation of quantitative individual expert judgement allows not only the synthesis of several disparate lines of evidence but also permits the quantification of uncertainty. It is pointed out by Zickfeld *et al.* (2007) that the consensus reviews alone may understate uncertainty as the diversity of opinion is not captured as effectively.

The process adopted for the elicitation of expert judgement was (when taking the preparation work into account) relatively lengthy whilst the manpower resource for the actual elicitation itself was high. Overall, the process involved the development of a 60 page written interview, which was refined and tested over the course of a two year period. Whilst there was a desire for it to be comprehensive, the content of the questionnaire was constrained by the objective of being able to undertake it during a day-long interview. Face to face interviews were carried out on an individual basis with 12 leading climate scientists. Each interview generally lasted five to seven hours and was undertaken at the expert’s home institute.

Zickfeld *et al.* (2007) considered a combination of different elicitations including:

- *ranking* (for identifying views on relative importance);
- *scoring* (for judgements on the sufficiency or ability of knowledge or models, e.g. on a scale of one to five: 1 = poor; 5 = very good); and
- *quantification* (e.g. estimation of changes in mean annual temperature associated with given changes in strength of AMOC).

Note – information on the latter (quantification) was elicited in the form of a (subjective) probability distribution. Subsequently these data could be reproduced as a series of ‘box and whisker’ type plots (see, for example, Massart *et al.*, 2005) providing a visual indication of the range of uncertainty recorded both by each expert and across all experts (for example see Figure A3.2, which shows elicited probability distributions for the present-day AMOC strength where vertical tick marks encompass the 90% confidence interval, the box spans the 50% confidence interval and the dot marks the median).

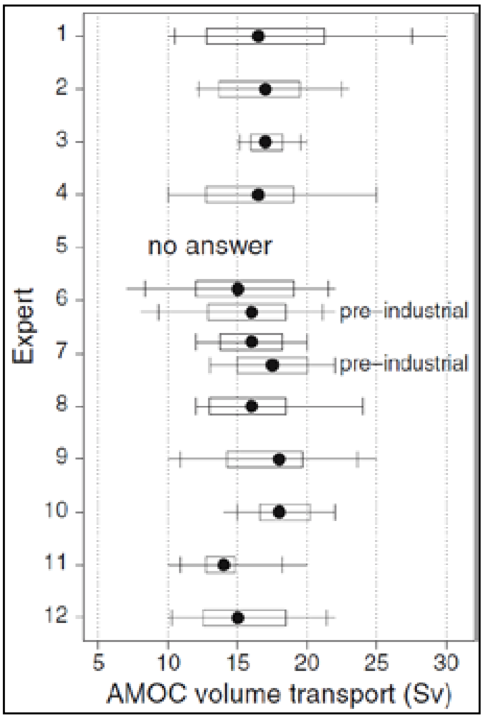


Figure A3.2. Example of graphical representation of elicited expert judgement (reproduced from Zickfeld *et al.*, 2007; see text for details)

Amongst other factors, expert judgement was elicited in order to derive:

- A ranking of physical processes in terms of their relative importance in determining the strength of the AMOC;
- An assessment of how well each of a range of potentially important physical processes is known;
- Judgement of the ability of state-of-the-art climate models to represent relevant physical processes;
- Judgement of relative importance of forcing factors or physical processes in determining AMOC response to global climate change; and
- Judgement of ability of state-of-the-art climate models to predict relevant physical processes.

Such information could not be readily derived by alternative means.

Use of Expert Judgement in Decision Making

The study by Zickfeld *et al*. (2007) was, in part, founded on using expert judgement to develop predictions on the likelihood of different future scenarios. Such predictions can be used to make decisions regarding future funding and research options. Some of the key outcomes, based on expert judgement, are discussed briefly below.

One preeminent view regarding the dominant physical processes responsible for determining the long-term mean pre-industrial strength of the AMOC was identified from the experts' responses. According to this view, heat fluxes and diapycnal mixing (i.e. mixing due to differences in seawater density) are key to determining the current state of the AMOC. Interestingly they judged that wind-driven upwelling in the Southern Ocean is relatively unimportant, in contradiction to theories proposed by some authors. Diapycnal mixing was indicated by most experts to be only poorly known, as opposed to, for example, the heat fluxes and atmospheric freshwater transport.

Almost all experts indicated that the most important forcing factors determining the response of the AMOC to increasing CO_2 concentrations are changes in the heat and freshwater fluxes in the North Atlantic.

Experts' best estimates of the weakening of the AMOC (in response to two alternative scenarios of doubling and quadrupling of the pre-industrial atmospheric CO_2 concentration) by the year 2100 range from c.2-55% and 10-90% for the doubling and quadrupling scenarios respectively. It was notable that the latter (expert judgement) estimate is much larger than the range of responses that had been simulated by state-of-the-art climate models (10-50%).

When asked to design a 15-year research program about the AMOC funded at US$ 500m (US) per year, experts tended to allocate the largest budget to long-term observations of circulation, hydrographic measurements, and coupled climate modelling. Surprisingly, experts would invest relatively little money in research on mixing processes in the ocean, although these (in particular diapycnal mixing) have been identified as key in determining the long-term mean state of the AMOC and assessed as being only poorly known.

Case Study 4. Expert Judgement Based Multi-Criteria Decision Model to Address Uncertainties in Risk Assessment of Nanotechnology-Enabled Food Products (Flari *et al.,* 2011)

Nanotechnology involves the manipulation of matter at a very small scale, generally between 1 and 100 nm, and exploits novel properties and functions that occur in matter at this scale. The advent of nanotechnologies has unleashed enormous prospects for the development of new products and applications for a wide range of industrial and consumer sectors.

Flari *et al.* (2011) describe how the known and projected applications of nanotechnology for the food sector (so far) fall into four main categories:

- processing or formulating foodstuffs to form nanostructures;
- using nano-sized, nano-encapsulated or engineered nano-additives in food;
- using nanotechnology-based materials and devices for food safety and traceability; and
- incorporating engineered nanomaterials (ENMs) in plastic polymers to develop improved, 'active', or 'intelligent' materials for food packaging (currently the largest market share of existing applications)

In consumer applications, ENMs may be present either as free particulates or in a bound, fixed or embedded form within larger objects and articles. Whilst the first three of these areas of application may not pose an immediate risk to the consumer (although they may have an adverse impact on the receiving environment after disposal) products and applications containing ENMs can give rise to direct exposure to free nanoparticles (either via inhalation, skin application, ingestion or intravenous delivery) and are therefore of particular concern.

Nanotechnology applications within the food industry have raised a number of concerns and issues, and questions have been raised over whether the current risk assessment paradigm and regulatory frameworks, which have evolved to deal with conventional materials, would be applicable and adequate for the new materials and products of nanotechnologies. Currently, risk assessment of nanotechnology enabled food products is fraught with difficulties due to the many uncertainties and knowledge gaps; Flari *et al.* (2011) suggest a means of using expert judgement to address these deficiencies.

Underlying Approaches and Philosophy – Why Was Expert Judgement Used?

Flari *et al.* (2011) suggest that, at present, uncertainties surrounding risk assessments are difficult to address due to the lack of knowledge of possible interactions of nanomaterials at the molecular and/or physiological levels, and their potential effects on human health either directly (i.e. ingestion of food items that may contain ingredients manufactured via nanotechnology processes) or indirectly (i.e. via environmental exposure). This level of uncertainty can only be addressed by expert judgement although, as the technology itself is at a very early stage of development, it can be expected that experts' opinions will vary.

The issue in question then becomes:

- how to capture experts' current knowledge and uncertainties, and
- to understand how experts use their knowledge when thinking about possible risk of nanotechnology-enabled food products.

Benefits and Limitations Inherent in the Application of Expert Judgement

Given the relative novelty of the topic area, a major and significant limitation to the effective elicitation and application of expert judgement has been identified as being the high level of incomplete knowledge for nanotechnology-enabled products. Indeed, Flari *et al.* suggest that it is widely accepted and disseminated that, currently, the risk assessment of nanotechnology-enabled food products is problematical due to the lack of appropriate data to assess potential hazard and exposure. They propose that a combined approach involving expert elicitation and the application of a multi-criteria decision model (MCDM) as a possible solution to this problem.

The authors acknowledge that the work 'involved a number of successive, lengthy steps' to collect, analyse and model information collected via elicitation of expert judgement and to validate the model that was subsequently developed.

Expert judgement was elicited remotely (via software with a customised interface)[23]. This approach enabled participants from across the global scientific community to be readily engaged. No financial reward was offered to experts for participation in the exercise; they worked *pro-bono*. Further, it was agreed that their individual inputs would remain confidential and that their anonymity would be preserved (the latter being especially easy to achieve when working remotely).

A subsequent assessment by participants at a workshop, although issued with caveats produced agreement that the approach was of significant value in aiding the assessment of safety of nanotechnology-enabled food products. The MCDM produced could serve as a screening or a first tier tool to distinguish products that could be considered as 'potentially safe' from the ones for which far more detailed risk assessment may be needed. On the other hand, workshop participants also acknowledged that the approach is novel, and a

---

[23] A number of commercial web-based 'survey' or 'questionnaire' applications are available that could be used to good effect in eliciting individual-based expert judgement. Without recommending any particular product, possible options include:

- Survey Monkey - http://www.surveymonkey.com/
- Smart Survey - http://www.smart-survey.co.uk/
- Checkbox - http://www.checkbox.com/
- Dot Survey - http://www.dotsurvey.com/

number of shortcomings of the current version of the model were identified (largely relating to the criteria that had been used).

As the model is highly dependent on the elicited ranking preferences, any weaknesses of the method followed to elicit those would, predictably, be reflected in the model's outputs and its feasibility as a decision making support tool.

Application of the precautionary principle is always an option for risk management, and it's appropriate to use within a political and/or legal framework for controlling the introduction of new nanotechnology applications. However, care needs to be given over how precautionary or non-precautionary the approach taken is (with inadequate precaution being potentially harmful, and excessive precaution halting the development of possible beneficial applications and effectively losing useful pointers towards further research needs). The application of a MCDM has been seen to be useful in capturing expert judgement on novel issues and this approach can be developed further to be used as a decision support tool that would remove the need for more subjective decisions on what level or precaution might be appropriate.

Use of Expert Judgement in Decision Making

Flari *et al.* specified ten criteria and characteristics of nanoparticles that have been identified as being of potential importance in determining their potential harmful properties and used these to characterise 26 hypothetical nanotechnology-enabled food products. Experts were invited to identify, and to rank the top five product scenarios that would be likely to trigger the least potential human health concern ('potentially safe products'). This exercise was repeated for those five product scenarios that would be likely to trigger the greatest potential human health concern ('potentially unsafe products', leaving 16 of the 26 hypothetical product scenarios unranked.

An MCDM approach was subsequently applied to derive weights characterising the importance of each of the criteria. As a consequence, the expected importance ranking of any 'new' product (i.e. one not originally included or assessed in the initial modelling phase) could be computed.

Annex 1 (to Appendix 3). Seven Step Process for Elicitation of Expert Judgement (after Knol *et al.*, 2010)

Step 1: characterisation of uncertainties

Expert judgement is often characterised as one means of dealing with uncertainty in data; other methods include modelling missing data, or performing sensitivity or scenario analyses. Uncertainties can take several forms. They may, for example, be: quantitative or qualitative; reducible or permanent; dependent on different measurement methods or on different personal values held by the experts. The identification or characterisation of the different types of uncertainty within a given study, for example through the application of an uncertainty typology, may help point to methods to deal with these uncertainties.
The uncertainty typology presented by Knol *et al.* covers six different dimensions:

1. the location of uncertainty;
2. its nature;
3. its range;
4. its level of recognized ignorance;
5. its level of methodological unreliability; and
6. its level of value diversity among analysts.

The *location* of uncertainty specifies where within an assessment uncertainty manifests itself. For example uncertainty could be: around the parameters and input data; regarding the underlying model; or within the context of the assessment itself.

The *nature* of uncertainty characterises which of the primary causes of uncertainty is in evidence, i.e. *epistemic uncertainty* (the result of incomplete knowledge) or *aleatory / ontic uncertainty* (natural uncertainty or that due to the intrinsic properties of the system). Most of the examples considered in the literature involve epistemic uncertainty, i.e. there is a lack of knowledge and expert elicitation is used to fill the gap. However, aleatory or ontic uncertainty can also be the subject of an expert elicitation, if its extent is poorly known (e.g. cases involving the natural variability of a system, such as weather conditions or activity patterns).

The *range* of uncertainty relates to whether it can be expressed in statistical terms (i.e. as a subjective probability distribution), or in terms of scenarios (i.e. as a range of plausible events, without any definitive information about the relative likelihood of each scenario). Understanding the range of uncertainty is important as it directly affects the format of information to be elicited, e.g. as subjective probability density functions or as relative likelihoods of scenarios.

Recognized ignorance deals with aspects of uncertainty for which we cannot establish any meaningful estimate. Expert elicitation can be used in such cases to give insight into what is not known and to what extent this is considered important.

Methodological unreliability reflects any weaknesses in the methodological quality of an assessment (either in whole or in part). Such unreliability may, for example, relate to the assessment's theoretical foundation, empirical basis, reproducibility or acceptance within the peer community. Expert elicitation can be used to identify areas of potential methodological uncertainty or to prioritize them for future attention.

The final dimension of uncertainty distinguished by Knol *et al.* is *value diversity among analysts*, which relates not to numerical values but to personal values and normative judgements held by scientists. Value diversity occurs when different, potentially valid choices can be made about assumptions in an assessment. Assessors making these choices may have different normative values and hence make different choices: expert opinions may well vary on issues because experts may rely on different personal norms and beliefs.

Step 2: Scope and format of the elicitation

The scope of an expert elicitation exercise is often limited by resources (both time and money). Resource limitations may constrain: the number of experts that can be approached; the degree to which they can be compensated for time and additional expenses; and the degree to which international experts are invited to contribute. As might be expected, where the information to be elicited is critical for policy making, when the outcomes are likely to be used in delicate decisions (e.g. legal proceedings), or where value diversity is high amongst the identified experts, a more elaborate expert elicitation process is warranted.

How many experts?

There is no absolute guideline on which to base the number of experts to be invited although the highlights of the Expert Judgment Policy Symposium and Technical Workshop (reported by Cooke and Probst, 2006) suggest that a panel of expert elicitation practitioners recommended that at least six experts should be included otherwise there may be questions about the robustness of the results. The feeling of the practitioners was that beyond 12 experts the benefit of including additional experts begins to diminish.

Group or individual elicitation?

Resource-wise, personal interviews inevitably consume more time for those organising the expert elicitation but, because of the losing the need to hire venues, may require less money compared against group elicitation sessions.

Whilst group interaction can give rise to ancillary benefits such as the sharing of knowledge and a better appreciation of different disciplinary viewpoints, events may suffer from the dominance of a small number of 'influential experts' and the implicit suggestion of the 'need to achieve consensus'[24]. By contrast, individual interviews may allow for more targeted questions and explanation.

Interviews or surveys?

Knol *et al.* note that information can be elicited from experts in various ways (e.g. by conducting interviews, by having questionnaires filled out, or by using specific software) although they suggest that, in general, face-to-face interviews are preferable. Face-to-face interviews (whether one-to-one or in a group format) leave more room for explanation, experts might be more motivated to join, and they may feel more responsible for providing informed judgements to an interviewer or a group than to an anonymous questionnaire. On the other hand, internet or postal questionnaires are less expensive, their content can be better standardised than the content of personal interviews, and experts may complete them at their leisure.

Step 3: Selection of experts

Types of experts

The selection of experts requires careful consideration; experts selected to take part in an elicitation can affect both the nature of the outcomes and their acceptability within the wider community. It should be noted that it is becoming increasingly recognized that non-

[24] Although the traditional Delphi method is based around consensus building, disagreement among experts may in fact highlight important information (e.g. on uncertainty or system variability) and, as a consequence, looking for consensus is not always appropriate.

professionals can also contribute valuable information and perform well in the elicitation of subjective opinion.

Balance
When there is a high degree of value diversity (see 'Typology of uncertainty', above), when there are high stakes involved (e.g. in legal proceedings), or when results need to be accepted by a wide peer-community, it is particularly important to have a well-balanced expert panel. Opposing views need to be justly represented in the panel and experts should preferably not have strong commitments to a particular outcome. The use a formal selection procedure is recommended in order to enhance such balance.

Availability of expertise
The elicitation of information from experts is inevitably reliant on the availability of expertise in the scientific community; experts cannot make up knowledge that does not yet exist. When issues are highly uncertain, controversial, unquantifiable or associated with potentially irreversible damage; or when decision stakes are very high there may simply be insufficient expertise available to derive any valid judgements.

Although it is possible to compensate to some degree through further training (for example in the production of subjective probability distributions) or the facilitation of extensive discussions between experts, this cannot fully compensate for the non-availability of expertise. Under such circumstances expert elicitation cannot be regarded as a panacea.

Step 4: Design of the elicitation protocol

Type of information to be elicited
The elicitation protocol covers both the questions to be asked to obtain expert judgement as well as the desired format for the answers. Expert judgement can provide both for quantitative and qualitative estimates, as well as the construction or evaluation of conceptual (causal) models. It is common for quantitative information to be elicited in the form of a number, its unit (e.g. grams or euros) and its uncertainty.

Performance and (internal) consistency of experts
Some experts may be better at making valid judgements than others. However, it is often difficult to check such performance of the experts, as the 'true values' of their estimates are unknown.

Wording of questions
The wording or phrasing of questions may substantially affect the responses provided. Knol *et al.* report that even slight rephrasing of the same question has been shown to lead to differences in (quantitative) responses of 4 to 15%. Consideration of potential sources of linguistic uncertainty (which can be classified into four main types: vagueness, context dependence, ambiguity, and underspecificity)[25] is therefore important.

[25] Addressing linguistic uncertainty:

Vagueness - should be addressed by providing clear definitions;

Context dependence - sufficient background information should be provided regarding the context within which the statements or questions are set;

Linguistic ambiguity - be careful with the use of homonyms (when words can have more than one meaning and it is unclear which meaning is meant); and

Underspecificity - enough detail should be provided to ensure that excessive room for interpretation is reduced.

Motivation for answers
Asking experts to record (preferably in written form) motivations for their judgements, and to identify issues that affected them, reduces the chance of heuristics and biases to remain unrecognised, and increases the proper interpretation of final results and potential outliers.

Heuristics and bias
People use various heuristics (which may, for example, relate to availability, representativeness, anchoring or adjustment)[26] when judging uncertain information – and these may introduce bias into the outcome. Motivational bias, occurring when the response of an expert is influenced by factors such as moral or professional responsibility, legal liability or peer credibility, may also affect results.

Step 5: Preparation of the elicitation session
Prior to the actual elicitation session the selected experts can be provided with a program for expert elicitation together with the protocol and the questions to be posed. In addition, any background information can be provided to the experts in the form of a briefing document which should seek to balance any potential disparate and disciplinary views (especially when a high level of value diversity exists). Such a briefing document might contain: an outline of the nature of the problem and the uncertainties related to it; key literature and/or a (qualitative or quantitative) summary of the literature; and information on both the elicitation procedure and the nature of heuristics and biases.

Step 6: Elicitation of expert judgements
Introduction of the scope and purpose of the expert elicitation
In order to familiarise experts with the subject matter a brief introduction regarding the field of interest and the purpose of the meeting should be given, allowing for discussion of the uncertainties at hand and the format for elicitation to be discussed, as well as outlining what is expected from them and how results will be used and distributed. This context-setting is important as, for example, some experts may be comfortable with making judgements in an academic context but less so if the outputs will form the basis for policy development.

Pre-elicitation training
The use of training questions is advised where quantitative estimates are to be made as most experts are unfamiliar with quantifying their degree of belief in terms of probabilities. Training questions also serve to explain the format of the elicitation (se also Step 5). Experts may also need to be made aware of potential heuristics and biases.

---

[26] Heuristics and bias:

Availability bias arises if the expert is affected by the ease of recall or the memory of recent experience;

Representativeness bias refers to inappropriate generalisation of specific knowledge, or to paying too much attention to specific details at the cost of background information; and

Anchoring and adjustment both relate to the procedure whereby an expert first selects a starting point (an anchor) as a first approximation and then adjusts this value to reflect supplementary information; results are typically biased towards the anchor.

Elicitation of judgements
The previous steps provide a framework for reducing the potential for significant divergence in judgement between experts due to the underlying elicitation process. In addition to the underlying process there are further explanations for differences in judgements between experts, including:

- different background information on which the experts base their judgement;
- different interpretation of the linguistic descriptions; and
- disagreement among experts on a more fundamental level.

Whilst the first two (especially the second) need to be avoided, the third represents the underlying value diversity and cannot be readily factored out.

Post-elicitation feedback
There are several benefits associated with the provision of post-elicitation feedback. It provides an opportunity for reflection and revision, and it stimulates discussion (as individual results can be seen in relation to the judgements of others, in turn allowing differences in interpretation and potential mistakes to be identified). Koln *et al.* note that this may have the associated risk of (un)consciously provoking experts with extreme ratings to move towards what most others reported, resulting in an unwanted regression to the mean.

Step 7: Possible aggregation and reporting
Aggregation of results
Expert judgements can be summarised or presented individually; Knol *et al*. suggest that there is no consensus about the conditions under which aggregation is warranted and, if so, in what way. The diversity of expert views carries valuable information and should be part of the open reporting of the study results. Knol *et al*. conclude that, as the fraction of experts who give a particular estimate might not be proportional to the probability of that estimate being correct, combining judgements might become problematic. It is recognised however that some form of aggregation may sometimes be necessary in order to facilitate the subsequent use and comparison of results.

Reporting judgements
The results of an elicitation should be reported alongside the procedure used. In addition, the aim of the elicitation and the anticipated use of the results need to be made clear to ensure that experts' judgements are used in the context for which they were intended.

Annex 2 to Appendix 3. Comparison of Seven Step Procedures as Presented by Knol *et al.* (2010) and by Simola *et al.* (2005)

In general terms the two alternative seven step procedures discussed in the body of this report (as presented by Knol *et al*. and by Simola *et al*.) are not too dissimilar. Differences are largely due to the emphasis on, and apparent detail within, each of the seven stages (although information provided in the paper by Simola *et al*. is very brief). The table below shows the two seven step procedures side-by-side.

It is not expected that either procedure is likely to have significant advantages over the other. However as the procedure presented by Knol *et al*. is supported by detailed text within their report, it is suggested that this procedure is considered further for this review.

<table>
<tr><th>Knol et al.</th><th>Simola et al.</th></tr>
<tr><td>Step 1 - Characterisation of uncertainties</td><td rowspan="2">Step 1 - Identification and selection of issues about which judgements of experts should be made</td></tr>
<tr><td>Step 2 - Scope and format of the elicitation</td></tr>
<tr><td>Step 3 - Selection of experts</td><td>Step 2 - Identification and selection of experts</td></tr>
<tr><td>Step 4 - Design of the elicitation protocol</td><td rowspan="2">Step 3 - Training of experts and definition of variables to be elicited</td></tr>
<tr><td>Step 5 - Preparation of the elicitation session</td></tr>
<tr><td rowspan="2">Step 6 - Elicitation of expert judgements</td><td>Step 4 - Individual work of experts</td></tr>
<tr><td>Step 5 - Elicitation</td></tr>
<tr><td rowspan="2">Step 7 - Possible aggregation and reporting</td><td>Step 6 - Analysis and aggregation of results and, in case of disagreement, attempt to resolve differences</td></tr>
<tr><td>Step 7 - Documentation of results, including expert reasoning in support of their judgement</td></tr>
</table>

Appendix 4. Initial Identification of References for use of Expert Judgement Within Biodiversity Assessments

The following listing presents the references initially identified to provide the basis for the review reported within Section 3 of this report, and indicates the scope of studies that have employed expert judgement within biodiversity assessments over recent years. The references selected by boxes indicate the 13 case studies that were subsequently selected for review.

ALLAN, J.D., YUAN, L.L., BLACK, P., STOCKTON, T., DAVIES, P.E., MAGIEROWSKI, R.H. & READ, S.M. (2012) Investigating the relationships between environmental stressors and stream condition using Bayesian belief networks, *Freshwater Biology*, **57** (Suppl. 1), 58–73.

ANDREASEN, J.K., O'NEILL, R.V., NOSS, R. & SLOSSER, N.C. (2001) Considerations for the development of a terrestrial index of ecological integrity, *Ecological Indicators*, **1**, 21–35.

AUBRY, A. & ELLIOTT, M. (2006) The use of environmental integrative indicators to assess seabed disturbance in estuaries and coasts: Application to the Humber Estuary, UK, *Marine Pollution Bulletin*, **53**, 175–85.

BAN, N. & ALDER, J. (2008) How wild is the ocean? Assessing the intensity of anthropogenic marine activities in British Columbia, Canada, *Aquatic Conservation: Marine and Freshwater Ecosystems*, **18**, 55-85.

BASHARI, H., SMITH, C. & BOSCH, O.J.H. (2009) Developing decision support tools for rangeland management by combining state and transition models and Bayesian belief networks, *Agricultural Systems*, **99**, 23–34.

BORJA, A, FRANCO, J., VALENCIA, V., BALD, J., MUXIKA, I., BELZUNCE, M.J. & SOLAUN, O. (2004) Implementation of the European water framework directive from the Basque country (northern Spain): a methodological approach, *Marine Pollution Bulletin*, **48**, 209–18.

BURGMAN, M.A., POSSINGHAM, H.P., LYNCH, A.J.J., KEITH, D.A., MCCARTHY, M.A., HOPPER, S.D., DRURY, W.L., PASSIOURA, J.A. & DEVRIES, R.J. (2001) A method for setting the size of plant conservation target areas, *Conservation Biology*,**15**(3), 603–16.

CAIN, J. (2001) Planning improvements in natural resources management - Guidelines for using Bayesian networks to support the planning and management of development programmes in the water sector and beyond. CEH, Wallingford.

CALEWAERT, J-B. (2006) *Environmental risk assessment of marine activities*. Available at: http://www.marbef.org/wiki/Environmental_risk_assessment_of_marine_activities [accessed December 2012].

CERTAIN, G., SKARPAAS, O., BJERKE, J-W., FRAMSTAD, E., LINDHOLM, M., NILSEN, J-E., NORDERHAUG, A., OUG, E., PEDERSEN, H-C., SCHARTAU, A-K., VAN DER MEEREN, G.I., ASLAKSEN, I., ENGEN, S., GARNÅSJORDET, P-A., KVALØY, P., LILLEGÅRD, M., YOCCOZ, N.G. & NYBØ, S. (2011) The Nature Index: A General Framework for Synthesizing Knowledge on the State of Biodiversity. *PLoS ONE*, **6**(4): e18930. doi:10.1371/journal.pone.0018930. Available at: http://www.plosone.org/article/info:doi/10.1371/journal.pone.0018930 [accessed October 2012].

CHAN, K.M.A., SHAW, M.R., CAMERON, D.R., UNDERWOOD, E.C. & DAILY, G.C. (2006) Conservation Planning for Ecosystem Services. *PLoS Biol.,* **4**(11): e379. doi:10.1371/journal.pbio.0040379. Available at: http://www.plosone.org/article/info:doi/10.1371/journal.pbio.0040379 [accessed December 2012].

CORK S. (2011) *A framework for assessing resilience in SoE 2011 reporting*. Report prepared for the Australian Government Department of Sustainability, Environment, Water, Population and Communities on behalf of the State of the Environment 2011 Committee. DSEWPaC, Canberra.

COWLING, R.M., PRESSEY, R.L., SIMS-CASTLEY, R. LE ROUX, A., BAARD, E., BURGERS, C.J. & PALMER, G. (2003) The expert or the algorithm? - comparison of priority conservation areas in the Cape Floristic Region identified by park managers and reserve selection software, *Biological Conservation*, **112**, 147–67.

COYLE, M.D. & WIGGINS, S.M. (2010) European Marine Site Risk Review. *Natural England Research Reports, Number 038*.

CZEMBOR, C.A., MORRIS, W.K., WINTLE, B.A. & VESK, P.A. (2011) Quantifying variance components in ecological models based on expert opinion. *Journal of Applied Ecology*, **48**(3), 736–45.

DAI, J.J., LORENZATO, S. & ROCKE, D.M. (2004) A knowledge-based model of watershed assessment for sediment, *Environmental Modelling & Software*, **19**, 423-33.

De LANGE, H.J., LAHR, J., van der POL J.J.C. WESSELS, Y. & FABER, J.H. (2009) Ecological vulnerability in wildlife: an expert judgment and multicriteria analysis tool using ecological traits to assess relative impact of pollutants, *Environmental Toxicology and Chemistry*, **28**(10), 2233-40.

ENQUIST, C.A.F. & GORI, D.F. (2008) Application of an expert system approach for assessing grassland status in the U.S.-Mexico borderlands: implications for conservation and management, *Natural Areas Journal*, **28**(4), 414–28.

FISHER, R., O'LEARY, R.A., LOW-CHOY, S., MENGERSEN, K. & CALEY, M.J. (2012) A software tool for elicitation of expert knowledge about species richness or similar counts, *Environmental Modelling & Software*, **30**, 1-14.

GIUPPONI, C. (2007) Decision support systems for implementing the European Water Framework Directive: The MULINO approach, *Environmental Modelling & Software*, **22**, 248-58.

GRAHAM-BRYCE, I. (2011) Independent review of the evidence process for selecting marine special areas of conservation, available at: http://www.defra.gov.uk/publications/files/pb13598-graham-bryce-independent-review-marine-sacs-110713.pdf [accessed January 2013].

HALPERN, B.S, LONGO, C., HARDY, D., MCLEOD, K.L., SAMHOURI, J.F., KATONA, S.K., KLEISNER, K., LESTER, S.E., O'LEARY, J., RANELLETTI, M., ROSENBERG, A.A., SCARBOROUGH, C., SELIG, E.R., BEST, B.D., BRUMBAUGH, D.R., CHAPIN, F.S., CROWDER, L.B., DALY, K.L., DONEY, S.C., ELFES, C., FOGARTY, M.J., GAINES, S.D., JACOBSEN, K.I., LEAH BUNCE KARRER, LESLIE, H.M., NEELEY, E., PAULY, D., POLASKY, S., RIS, B., ST MARTIN, K., STONE, G.S., SUMAILA, U.R. & ZELLER, D. (2012) An index to assess the health and benefits of the global ocean, *Nature*, **488**, 615-20.

HALPERN, B.S., WALBRIDGE, S., SELKOE, K.A., KAPPEL, C.V. MICHELI, K., D'AGROSA, C., BRUNO, J.F., CASEY, K.S., EBERT, C., FOX, H.E., FUJITA, R., HEINEMANN, D., LENIHAN, H.S., MADIN, E.M.P., PERRY, M.T., SELIG, E.R., SPALDING, M., STENECK, R. & WATSON, R. (2008) A global map of human impact on marine ecosystems, *Science*, **319**, 948–52.

HALPERN, B.S., SELKOE, K.A., MICHELI, F. & KAPPEL, C.V. (2007) Evaluating and ranking the vulnerability of global marine ecosystems to anthropogenic threats, *Conservation Biology*, **21**(5), 1301–15.

HIDDINK, J.G., JENNINGS, S. & KAISER, M.J. (2007) Assessing and predicting the relative ecological impacts of disturbance on habitats with different sensitivities, *Journal of Applied Ecology*, **44**, 405–13.

KLEYPAS, J.A. & EAKIN, C.M. (2007) Scientists' perceptions of threats to coral reefs: results of a survey of coral reef researchers, *Bulletin of Marine Science*, **80**(2), 419-36.

KRUEGER, T., PAGE, T., HUBACEK, K., SMITH, L. & HISCOCK, K. (2012) The role of expert opinion in environmental modelling, *Environmental Modelling and Software*, **36**, 4-18.

KUHNERT, P.M., MARTIN, T.G. AND GRIFFITHS, S.P. (2010) A guide to eliciting and using expert knowledge in Bayesian ecological models, *Ecology Letters*, **13**, 900–14.

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**, 906–20.

MACMILLAN, D.C. & MARSHALL, K. (2006) The Delphi process – an expert-based approach to ecological modelling in data-poor environments, *Animal Conservation*., **9**(1), 11-19.

MARCOT, B.G. THOMPSON, M.P., RUNGE, M.C., THOMPSON, F.R., MCNULTY, S., CLEAVES, D., TOMOSY, M., FISHER, L.A. & BLISS, A. (2012) Recent advances in applying decision science to managing national forests, *Forest Ecology and Management*, **285**, 123–32.

MPA Monitoring Enterprise (2012) *A framework for advancing the use of expert judgments of ecosystem condition*. Available at: http://monitoringenterprise.org/pdf/Expert_Judgment_Draft_Framework.pdf [accessed October 2012].

MUXIKA, I., BORJA, A. & BALD, J. (2007) Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive, *Marine Pollution Bulletin*, **55**, 16–29.

NESLO, R., MICHELI, F., KAPPEL, C.V., SELKOE, K.A., HALPERN, B.S. & COOKE, R.M. (2008) *Modeling Stakeholder Preferences with Probabilistic Inversion: Application to Prioritizing Marine Ecosystem Vulnerabilities*. In: I. Linkov, E. Ferguson & V.S. Magar (eds.) *Real Time and Deliberative Decision Making: Application to Risk Assessment for Non-chemical Stressors,* Springer, Amsterdam.

NOAA; National Oceanic and Atmospheric Administration (2004) *A monitoring framework for the National Marine Sanctuary System*. Available at: http://sanctuaries.noaa.gov/library/national/swim04.pdf [accessed December 2012].

OKEY, T.A., & LOUCKS, L.A. (2011). *Identifying and selecting indicators of social-ecological health. In: Draft social-ecological assessment for the marine and coastal areas of the West Coast of Vancouver Island*. A report published by the Tsawalk Partnership, West Coast Aquatic. Port Alberni, BC. Available at: http://www.westcoastaquatic.ca/sea/social-ecological-assessment [accessed July 2013].

OKEY, T.A., ALIDINA, H.M., LO, V., MONTENEGRO, A. & JESSEN, S. (2012) *Climate change impacts and vulnerabilities in Canada's Pacific marine ecosystems*. CPAWS BC and WWF-Canada, Vancouver, BC.

PAGE, T., HEATHWAITE, A.L., THOMPSON, L.J., POPE, L. & WILLOWS, R. (2012) Eliciting fuzzy distributions from experts for ranking conceptual risk model components, *Environmental Modelling & Software*, **36**, 19-34.

PERERA, A.H., DREW, C.A. & JOHNSON, C.J. (eds.) (2011) *Expert knowledge and its application in landscape ecology*, Springer, 307pp

PRESSEY, R.L, COWLING, L.J. & ROUGET, M. (2003) Formulating conservation targets for biodiversity pattern and process in the Cape Floristic Region, South Africa, *Biological Conservation*, **112**, 99-127

PRICE, J., SILBERNAGEL, J., MILLER, N., SWATY, R., WHITE, M. & NIXON, K. (2012) Eliciting expert knowledge to inform landscape modeling of conservation scenarios, *Ecological Modelling*, **229**, 76-87.

PRIOR, A., MILES, A.C., SPARROW, A.J. & PRICE, N. (2004) *Development of a Classification Scheme for the Marine Benthic Invertebrate Component, Water Framework Directive*, Environment Agency R & D Interim Technical Report E1-116, E1-132; Environment Agency, National Marine Service.

REGAN, T.J., MASTER, L.L. & HAMMERSON, G.A. (2004) Capturing expert knowledge for threatened species assessments: a case study using NatureServe conservation status ranks, *Acta Oecologica*, **26**, 95–107.

RONDININI, C. (2011) *A review of methodologies that could be used to formulate ecologically meaningful targets for marine habitat coverage within the UK MPA network*, JNCC Report No. 438; JNCC, Peterborough, UK.

RYU, J., LESCHINE, T.M., NAMB, J., CHANG, W.K. & DYSON, K. (2011) A resilience-based approach for comparing expert preferences across two large-scale coastal management programs, *Journal of Environmental Management*, **92**, 92-101.

SELKOE, K.A., HALPERN, B.S. & TOONEN, R.J. (2008) Evaluating anthropogenic threats to the Northwestern Hawaiian Islands, *Aquatic Conservation: Marine and Freshwater Ecosystems*, **18**, 1149-65.

SKARPAAS,O. & CERTAIN, G. (2011) The Nature Index – integrating data and expert judgment on biodiversity, paper for the special session 'Biodiversity measurement and policy: from theory to practice' at the ESEE conference, Istanbul 14-17 June 2011.

SMITH, R.J., EASTWOOD, P.D., OTA, Y. & ROGERS, S.I. (2009) Developing best practice for using Marxan to locate Marine Protected Areas in European waters, *ICES Journal of Marine Science*, **66**, 188-94.

SMITH, R.J., GOODMAN, P.S. & MATTHEWS, W.S. (2006) Systematic conservation planning: a review of perceived limitations and an illustration of the benefits, using a case study from Maputaland, South Africa, *Oryx*, **40**(4), 400-10.

SUTHERLAND, W.J. (2006) Predicting the ecological consequences of environmental change: a review of the methods, *Journal of Applied Ecology*, **43**, 599-616.

TECK, S.J., HALPERN, B.S., KAPPEL, C.V., MICHELI, F., SELKOE, K.A., CRAIN, C.M., MARTONE, R., SHEARER, C., ARVAI, J., FISCHHOFF, B., MURRAY, G., NESLO, R. & COOKE, R. (2010) Using expert judgment to estimate marine ecosystem vulnerability in the California Current, *Ecological Applications*, **20**, 1402–16.

TEIXEIRA, H., BORJA, A. WEISBERG, S.B., RANASINGHE, J.A., CADIEN, D.B., DAUER, D.M., DAUVIN, J-C., DEGRAER, S., DIAZ, R.J., GRÉMARE, A., KARAKASSIS, I., LLANSÓ, R.J., LOVELL, L.L., MARQUES, J.C., MONTAGNE, D.E., OCCHIPINTI-AMBROGIM, A., ROSENBERG, R., SARDÁ, R., SCHAFFNER, L.C. & VELARDE, R.G. (2010) Assessing coastal benthic macrofauna community condition using best professional judgement – developing consensus across North America and Europe, *Marine Pollution Bulletin*, **60**, 589-600.

THOMPSON, B., WEISBERG, S.B., MELWANIA, A., LOWEA, S., RANASINGHE, J.A., CADIEN, D.B., DAUERD, D.M., DIAZE, R.J., FIELDS, W., KELLOGG, M., MONTAGNE, D.E., ODE, P.R., REISH, D.J. & SLATTERY, P.N. (2012) Low levels of agreement among experts using best professional judgment to assess benthic condition in the San Francisco Estuary and Delta, *Ecological Indicators*, **12**, 167–73.

TILLIN, H.M., HULL, S.C. & TYLER-WALTERS, H. (2010) *Development of a sensitivity matrix (pressures-MCZ/MPA features)*. Report to the Department of Environment, Food and Rural Affairs from ABPMer, Southampton and the Marine Life Information Network (MarLIN) Plymouth: Marine Biological Association of the UK. Defra Contract No. MB0102 Task 3A, Report No. 22.

UK MMAS, Marine Monitoring and Assessment Strategy (2010). *Charting Progress 2 Healthy and Biological Diverse Seas Feeder report*. (*Eds*. M. Frost. & J. Hawkridge). Published by Department for Environment Food and Rural Affairs on behalf of UKMMAS. 682pp. Available at: http://chartingprogress.defra.gov.uk/healthy-and-biologically-diverse-seas-feeder-report [accessed October 2012].

WARD, T.J. (2011) *SOE 2011: National marine condition assessment – decision model and workshops*. Report prepared for the Australian Government Department of Sustainability, Environment, Water, Population and Communities on behalf of the State of the Environment 2011 Committee. DSEWPaC, Canberra.

WILSON, K., PRESSEY, R.L., NEWTON, A., BURGMAN, M., POSSINGHAM, H. & WESTON, C. (2005) Measuring and incorporating vulnerability into conservation planning, *Environmental Management*, **35**(5), 527–43.
WILSON, K.A. (2010) Dealing with data uncertainty in conservation planning, *Natureza & Conservação*, **8**(2), 145-50.

WORLD BANK (2004) *Score card to assess progress in achieving management effectiveness goals for Marine Protected Areas*. Available at: http://siteresources.worldbank.org/EXTBIODIVERSITY/Resources/MPA_tool.pdf [accessed October 2012].

ZACHARIAS, M.A. & GREGR, E.J. (2005) Sensitivity and vulnerability in marine environments: an approach to identifying vulnerable marine areas, *Conservation Biology*, **19**(1), 86-97.

Other Studies

Two further works (included in the listing above) were initially identified as potential case studies (Okey *et al*., 2012; and Zacharias & Gregr, 2005), although these have been excluded from the main body of this report.

Rather than developing independent expert judgement, Okey *et al*. (2012) relied on judgement previously reported by Teck *et al*. (2010). Expert-derived vulnerability values of habitat types to climate stressors from the California Current region (Teck *et al*., 2010) were used as sensitivity scores in their analysis for 18 habitat types and bottom types mapped from a compilation of sources.

Okey *et al*. (2012) note that an expert-based approach is being developed for the West Coast of Vancouver Island to estimate the impacts of climate change (temperature, acidification, and UV radiation) relative to all these stressors in a spatially explicit context (Okey and Loucks, 2011). This approach includes the assessment of the vulnerability of a set of indicators and habitat classes to these stressors by subject-expert panels. This initiative employed eight methods to collect knowledge from the various communities with a stake in the future of the social-ecological systems of the West Coast of Vancouver Island. These were envisioned as eight spokes of the wheel of community and sector engagement and knowledge collection (see Table A4.1, below).

The engagement process produced estimates of the relative degree of stress that each stressor exerts on the overall system and larger subsystems, as well as other expert-based rankings that enable the general identification and prioritisation of state and pressure indicators and management strategies.

The study reported by Zacharias & Gregr (2005) looked at a large marine area off the Canadian (British Columbian) coast, including Victoria Island and Queen Charlotte Island. It made little (if any) use of expert judgement in the context that's being considered here, and there was no substantive information on the handling of data. They used major ferry routes and shipping traffic to indicate sources of acoustic disturbance, incorporating a distance-based noise attenuation model within a GIS to create a 'likelihood map' of noise disturbance. Similar layers were created for other potential noise disturbance sources, including those associated with oil and gas extraction, and small boats with outboard engines. The authors created boundaries for vulnerable marine areas for two groups of valued ecological features, humpback whale (*Megaptera novaengliae*) and the group of offshore balaenopterid species that includes sei (*Balaenoptera physalus*) and blue (*B. musculus*) whales, by combining the joint probability of occurrence of the whale species' habitat with response surfaces for each of the four sources of stress considered (i.e. vulnerability surfaces representing the likelihood of encountering acoustic stresses).

Table A4.1. Objectives and methods of the eight spokes of the wheel of community and sector engagement and knowledge collection (after Okey & Loucks, 2011) [NCN – Nuu-chah-nulth; WCVI – West Coast Vancouver Island; WCA – West Coast Aquatic (coastal community forum, Vancouver Island)]

Aspects	Knowledge Symposium, Network, and Wiki	Online social-ecological surveys	Vision, values, issues and opportunities	Gathering existing spatial data	Spatial local knowledge interviews	Spatial sector interviews	Nuu-chah-nulth community meetings	Open house and community dialogue sessions
Objectives	Identify and document existing knowledge in the scientific, educational, First Nations, and local resource user communities. Provide a meeting for each of these different communities to express their interests Publish a proceedings of the meeting Establish knowledge network	Enable judgements and perspectives of expert knowledge holders and community members on the health of the WCVI ecosystem in general, the stressors in the system, the human values and activities, and suggested management solutions Rank relative importance of stressors	Develop coherent vision for communities and sounds Understand issues and opportunities Set priorities	Understand spatial context of social, cultural and biophysical systems Make use of existing information wherever possible	Understand NCN uses and activities Understand areas of conflict and compatibility Use local knowledge of species and habitat health and trends Understand NCN stewardship practices and principles Set priorities	Understand sector uses and activities Understand suitability, range and importance of different areas Understand areas of conflict and compatibility Understand how sectors will change in the future Set priorities	Generate public interest and support Map community issues, goals, objectives and strategies Determine acceptable uses and activities Verify vision, values, issues, opportunities and local knowledge Agree on appropriate planning units Support filling knowledge gaps as appropriate	Generate public interest and support Develop goals and objectives around certain issues Understand public uses and activities Understand areas of conflict and compatibility

Aspects	Knowledge Symposium, Network, and Wiki	Online social-ecological surveys	Vision, values, issues and opportunities	Gathering existing spatial data	Spatial local knowledge interviews	Spatial sector interviews	Nuu-chah-nulth community meetings	Open house and community dialogue sessions
Collecting data	Contributed authored contributions to the proceedings as extended abstracts. Contributed posters Facilitated conversations Initial community knowledge mapping sessions Sharing food and praying	Data were collected by designing online surveys using Survey Monkey and distributing links to the interviews via project contact lists	Semi structured interview guides Responses recorded on standard response sheets 10 interviews in each NCN community Short version included in sector spatial interviews Short version conducted at community events (fall fair, salmon fest)	Spreadsheet of geospatial data needs (developed in Google Docs to allow for collaboration) Custodians contacted Data transferred using a variety of methods	Semi structured interview guides Responses recorded using maps, response sheets, codes and audio recorders Training provided to a team of interviewers 10 interviews in each NCN community Participants selected from knowledge holders list	Semi structured interview guides Responses recorded using maps, response sheets, codes and audio recorders Interviews conducted by WCA staff	Meeting outline developed Meetings open to the public, key informant invited specifically Meetings facilitated by WCA staff and community leaders Information recorded on marine charts and flip charts	Workshops held to discuss four themes (pollution, food, habitat and species health, local use and activities) “World café” discussions facilitated by WCA staff Group discussion to summarise findings Responses recorded on marine charts and flip charts
Managing data	Data were compiled in the published proceedings and used and cited throughout the project Wiki contributions were housed on the online Barkleypedia	Data were housed on the Survey Monkey website, and summarised and downloaded from that as needed as Excel spreadsheets and in other formats. Summaries were also downloaded	Responses typed into spreadsheet	Storing geospatial data on GIS Database	Charts digitised from digital photo Spatial responses entered into GIS database; narrative responses typed into spreadsheet Audio recording used as back-up Research materials archived	Charts digitized from digital photo Spatial responses entered into GIS database; narrative responses typed into spreadsheet Audio recording used as back-up Research materials archived	Charts digitised from digital photos Flip chart notes typed into standard table format	Responses and discussion typed into spreadsheet Charts digitised from digital photos

Aspects	Knowledge Symposium, Network, and Wiki	Online social-ecological surveys	Vision, values, issues and opportunities	Gathering existing spatial data	Spatial local knowledge interviews	Spatial sector interviews	Nuu-chah-nulth community meetings	Open house and community dialogue sessions
Analysis	Each contributing author provided their own analyses Content in contributions were used in selection of indicators, etc.	The answers to each question was interpreted and analysed individually Packaged figures were sometimes downloaded	Spreadsheets saved as .txt files, responses coded Qualitative analysis software used to analyse data	Map printed Key informants asked to comment on the accuracy of data	Spreadsheets saved as .txt files, responses coded Qualitative analysis software used to analyse data	Spreadsheets saved as .txt files, responses coded Qualitative analysis software used to analyse data	Qualitative analysis software used to analyse data Spatial layers are combined to describe uses, activities, attributes and stressors within planning units	Qualitative analysis software used to analyse data Spatial layers are combined to describe uses, activities, attributes and stressors within planning units
Outputs	Published proceedings Online wiki Barkleypedia Knowledge network established	Results were summarised in this IEA	Summary reports Evaluation criteria for trade off scenarios Goals, objectives and strategies for communities and sectors	Geospatial database used to define and describe planning units and evaluate scenarios	Geospatial database used to define and describe planning units and evaluate scenarios	Geospatial database used to define and describe planning units and evaluate scenarios	Goals, objectives and strategies Geospatial used to define and describe planning units and evaluate scenarios	Geospatial database used to define and describe planning units and evaluate scenarios

Additional references

During the initial literature search process a number of additional references were identified which, whilst not relating directly to the use of expert judgement in biodiversity assessment studies, may provide useful background information on approaches and methodologies relating to expert judgement. These additional references are provided below.

ARNOLD, V., CLARK, N., COLLIER, P.A., LEECH, S.A. & SUTTON, S.G. (2006) The differential use and effect of knowledge-based system explanations in novice and expert judgment decisions, *MIS Quarterly*, **30**(1), 79-97.

CUSHEN, M., KERRY, J., MORRIS, M., CRUZ-ROMERO, M. & CUMMINS, E. (2012) Nanotechnologies in the food industry – recent developments, risks and regulation, *Trends in Food Science & Technology*, **24**, 30-46.

GIORDANOA, R. & LIERSCH, S. (2012) A fuzzy GIS-based system to integrate local and technical knowledge in soil salinity monitoring, *Environmental Modelling & Software*, **36**, 49-63.

GREGORY, R., FAILING, L., OHLSON, D. & McDANIELS, T. (2006) Some pitfalls of an overemphasis on science in environmental risk management decisions, *Journal of Risk Research*, **9**, 717-36.

JAMES, A., LOW-CHOY, S. & MENGERSEN, K. (2010) Elicitator: An expert elicitation tool for regression in ecology, *Environmental Modelling & Software*, **25**, 129-45.

JANSSEN, J.A.E.B., KROL, M.S., SCHIELEN, R.M.J., HOEKSTRA, A.Y. & de KOK J.-L. (2010) Assessment of uncertainties in expert knowledge, illustrated in fuzzy rule-based models, *Ecological Modelling*, **221**, 1245-51.

KRUEGER, T., PAGE, T., SMITH, L. & VOINOV, A. (2012) A guide to expert opinion in environmental modelling and management, *Environmental Modelling & Software*, **36**, 1-3.

LOW-CHOY, S. O'LEARY, R. & MENGERSEN, K. (2009) Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models, *Ecology*, **90**(1), 265-77.

OECD & JRC (2008) *Handbook on constructing Composite Indicators: methodology and user guide,* OECD, Paris, 162pp.

OECD (2003) *Environmental indicators: Development, measurement and use,* OECD, Paris, 37pp.

RAMANATHAN, R. (2006) Data envelopment analysis for weight derivation and aggregation in the analytic hierarchy process, *Computers & Operations Research*, **33**, 1289-307.

RINDERKNECHT, S.L., BORSUK, M.E. & REICHERT, P. (2012) Bridging uncertain and ambiguous knowledge with imprecise probabilities, *Environmental Modelling & Software*, **36**, 122-30.

SAB (2009) *Valuing the Protection of Ecological Systems and Services; a Report of the EPA Science Advisory Board*. EPA Science Advisory Board, Washington DC, USA.