NORWEGIAN SCHOOL OF ECONOMICS AND BUSINESS ADMINISTRATION
Bergen, spring 2007

Master's profile: Business Analysis and Performance Management
Supervisor: Mikael Rönnqvist

Application of Multiple Criteria Decision Analysis to compare Environmental Impact Factors in Statoil

by
Kay Meløysund

PREFACE

This master's thesis is written on the background of a collaborative work with scientists at the Energy and Environment department at Statoil Research Centre. It documents the application of Multiple Criteria Decision Analysis (MCDA) to compare Environmental Impact Factors (EIFs) from January to June 2007. The work has been headed by a project team within the integrated HSE risk management project, in which the author has been given the opportunity to participate. This thesis reflects the author's perspectives on the multiple application aspects, and cannot be interpreted as representative for the viewpoints of other project members. These perspectives are nevertheless shaped and sharpened through invaluable discussions and interactions with other key figures.

I would like to thank my colleagues in the project team, Ingunn Nilssen, Ståle Johnsen and Mathijs G. D. Smit, for the insights gained through extensive discussions and reflected feedbacks. Mathijs' patient explanations of how to assess environmental risk, Ståle's linear doctrines and clarifying to-the-point formulations and in particular Ingunn's overall devotedness to the project and to my study have been indispensable.

Furthermore, I would like to thank my supervisor Mikael Rönnqvist and my co-advisor Fred Wenstøp for sharing their vast knowledge of decision analytic theory and application with me. They have guided me through the labyrinths of MCDA and given me the opportunity to concentrate on how this mindset can be applied to Statoil in a best achievable way.

I would also like to express my gratitude to the other members of the integrated HSE risk management project for their contributions. The engaged and constructive expert panel constituted by Edgar, Sigurd, Ellinor and Marianne and the critical eyes of Espen and Jakob have kept me busy. The scientists at the Energy & Environment department at Rotvoll also deserve an acknowledgement for willingly answering all kinds of tedious questions.

Trondheim, June 19, 2007

Kay Meløysund

ABSTRACT

In this study, Multiple Criteria Decision Analysis (MCDA) is applied to the risk assessment framework of Environmental Impact Factors (EIFs) in Statoil. The objective for the application is to integrate EIFs to an indicator for overall environmental risk related to emissions and discharges from petroleum activities and operations. To reach this indicator, expert judgements of the relative importance of environmental compartments are considered to be essential. The study is a part of the integrated HSE risk management project at Statoil and is based on the principles and experiences from the MCDA trial session in 2006.

To further investigate and refine the approach, the case study of drilling technology alternatives at the Norne field is applied. The Statoil goals of zero harm to the environment and continuous improvement of environmental performance form the basis of the problem design. Five decision alternatives are identified and relevant EIF scores for these alternatives are assembled or estimated. The EIFs are tailored to act as decision criteria that reflect the needs of scientific accuracy and practical viability, and the scores are accordingly modified. The special features of the EIF for air emissions require a different approach for this factor.

For each alternative, criteria scores at the compartment level are aggregated and weighted through the use of an optimisation model, and a total performance indicator for each alternative is identified. Even though the EIF scores are calculated on the basis of generic data, the area-specific sensitivity of environmental compartments results in importance weights that are limited to a pre-defined area. The set of weights for the relevant influence area in the Norne case is elicited through two expert panel sessions.

As a response to challenges at the first session, the problem design is additionally modified. The most important adjustments are related to weight elicitation on a unit basis and the introduction of “risk scores”. Due to similarity of data in the Norne case and a temporarily exclusion of air emissions, three decision alternatives are identified as equally optimal. The results from the second session indicate that the current problem design has increased the feasibility of the MCDA approach, but that challenges like integration of air emissions and relevance of sensitivity aspects remain.

INDEX

ABBREVIATIONS AND TERMINOLOGY

ALARP	As Low As Reasonably Practical
Compartment sensitivity	The characteristics of a particular environment that influence its importance relative to other environments
Discharges	Generic term for releases to all environments except from air.
EIF	Environmental Impact Factor. Indicator for potential impacts on species as a result of emissions and discharges.
Emissions	Generic term for releases to air.
Exposure	Level of toxicants and/or no-toxic stressors in an environment to which species are exposed.
Event	A happening that can lead to an impact. Events can be both regular (probability of occurrence = 1) and acute (probability of occurrence < 1). Emissions and discharges are examples of events.
HSE	Health, Safety and Environment
Impact	Represents a measure for harm or benefit to people and environment.
MAUT	Multi-Attribute Utility Theory
MAVT	Multi-Attribute Value Theory
MCDA	Multiple Criteria Decision Analysis.
NOEC	No-Effect Concentration
msPAF	multi-substance Potentially Affected fraction of Species
PEC	Predicted Environmental Concentration
PNEC	Predicted No-Effect Concentration
Risk	Combination of the probability of an event's occurrence and the adverse impacts from this event. It comprises potential impacts deriving from both regular and acute events. The term is used interchangeably with 'potential impact'.
Species sensitivity	Species' vulnerability to exposure
SSD	Species Sensitivity Distribution
Threshold level	The limit above which unacceptable environmental effects are likely to occur
WTP	Willingness-To-Pay

1. INTRODUCTION

"If one does not know to which port one is sailing, no wind is favourable."

Seneca

1.1 Background

One of Statoil's overall goals is to cause zero harm to people and environment. In more operational terms this is expressed as a commitment to "reducing the negative impact of our activities and products on health and the environment", and to "continuously evaluate and improve our performance" (http://www.statoil.com/hse).

Decision making is a choice between alternatives. These decision alternatives often differ considerably with regards to both amount and distribution of potential impacts to people and environment (i.e. HSE risks). To be in line with the overall goal of zero harm, the following question therefore has to be addressed: Given the impacts each alternative may have to people and the environment, which alternative should be considered to represent the least risk?

To answer this, the risks first have to be quantitatively predicted. Statoil has developed, in cooperation with other oil companies and research institutions, a framework of so-called environmental impact factors (EIFs). This framework constitutes a comprehensive tool for environmental risk assessment related to potential impacts from emissions and discharges. Unfortunately, the identification of the combination of EIFs that represents the least risk is far from trivial:

- EIFs are themselves highly complex and difficult to comprehend, as they are already a product of several variables and weighting procedures.
- Although ongoing projects in Statoil are attempting to rescale EIFs to enhance commensurability, some of them will most likely continue to be measured in different units and terms.

- Some environmental compartments at risk are simply more valuable than others. As an example, even if the impact on water column were equal to the impact on sediment in numerical terms, the water column could be considered more important and the impact therefore considered worse.

And yet, comparison to other relevant environmental aspects, health and safety issues and other corporate goals remains before the preferred decision alternative can be identified.

1.1.1 HSE risk integration project

The objective for the integrated HSE risk management project at Statoil is "to enable the company's business areas to identify, select and document an optimal solution for meeting the HSE strategy and the zero harm principle in given cases." ('Integrated HSE project manual' 2005:4). Integration of EIFs to an overall environmental risk index (EIF_{total}) and further integration of H, S and E indexes to an overall HSE risk index are decided to be important deliverables for this project.

A trial session of EIF comparisons through different approaches was held in January 2006. The conclusion was that the methodology of Multiple Criteria Decision Analysis (MCDA) seems feasible for this purpose, but that more investigation and refinement still was needed (Wenstøp 2006).

1.2 Goal and objectives for this study

Overall goal:

Based on the developed EIF framework and previous trial sessions, the overall goal of this study is to further investigate how an environmentally optimal alternative can be identified through MCDA methodology.

Concrete objectives:

1) For a given case: Design and apply an approach based on MCDA to identify an EIF_{total} for each decision alternative. The approach should meet the needs for both scientific accuracy and practical viability. The main concerns are:
   - To establish decision criteria based on the EIFs.
   - To propose a multi-attribute function for aggregation of EIF-based decision criteria to an overall indicator.
   - To consider appropriate procedures for attributing weights to EIF-based decision criteria that reflect their relative importance.
2) Consider the feasibility of the chosen procedure, methodologically as well as organisationally. See if trade-offs to monetary units are achievable.
3) If problematic and unsolved areas remain: Frame challenges and suggest/discuss possible solutions.
4) Use information and judgements from the previous points to suggest:
   - how an EIF_{total} can be reached in future applications
   - if and/or how this index may be further applied in order to identify, select and document optimal HSE solutions.

The study is of applicative nature and the main focus is on aspects that have at some time been regarded as vital for bringing the project further. Hence, the scope of work has been modified as new and more precarious challenges have been revealed, and the objectives have been correspondingly adjusted. As a result, the core of the study in its present form is on choice, definition and modification of the decision criteria.

1.3 Working methods

The applicative nature of the study is also reflected in the working methods applied. Valuable insights and discussions are mainly a result of interaction within the project team, consisting of the author and key figures at the Statoil research department. Other important approaches employed are:

- Review of relevant decision analysis theory and Multiple Criteria Decision Analysis in particular

- Review of Statoil governing documents and reports related to decision making and environmental risk management
- Review of reports related to the Norne case study and the MCDA trial session in 2006
- Conversations and discussions with the integrated HSE risk management project members and other key figures within risk assessment
- Conversations and discussions with supervisor and co-supervisor on methodological issues
- Analysis of outcome and feedback from two expert panel sessions in Stavanger mid-April and mid-May 2007

1.4 Outline of thesis

In the following chapter, theory on decision analysis in general and MCDA in particular is reviewed. Challenges related to environmental decision making and how MCDA could serve as a tool for meeting them are also discussed. Chapter 3 is turning the focus to Statoil and looking into general characteristics of decision making in the organisation. The main focus is on current assess- and treatment of environmental risk. In chapter 4, the Norne case and the problem of identifying which drilling technology implies the least environmental risk is introduced. Features from the two preceding chapters are applied to make a design for how the case can be solved. Chapter 5 reports from the implementation of the design at the expert panel sessions and the results obtained there. Several modifications and remaining challenges are discussed. Chapter 6 draws conclusions from the insights gained in this study, and gives suggestions to further investigation and to how the proposed environmental index could actually be applied.

2. BACKGROUND FOR ENVIRONMENTAL DECISION MAKING

In order to provide theoretical foundation for the upcoming case application, the concept of decision analysis in general, MCDA in particular and special characteristics for environmental decision making are reviewed.

2.1 Decision analysis

Decision analysis is a technology designed to help individuals and organisations make wise inferences and decisions (von Winterfeldt and Edwards 1986). The paradigm of this technology comprises numerous schools and techniques; all worked out for helping decision makers to structure their approach to a problem in a way that the actions taken may be rational according to their fundamental objectives and values.

2.1.1 Why perform a decision analysis?

"To have a decision problem is to be in a situation that requires action, and there are several options available" (Seip and Wenstøp 2006: 23). Some decision problems are trivial, and decision makers have no troubles in identifying which course of action is preferred. Von Winterfeldt and Edwards (1986) emphasise however several possible problem dimensions that might complicate the cognitive process to such an extent that a more formal decision analysis is recommended. These dimensions could be summarised in four categories:

1) **Multiple conflicting objectives**

   The overall performance of a decision alternative is often determined by several criteria, which are again made up of underlying objectives. In many cases, an alternative could perform well on some of them but poorly on others, and trade-offs have to be made. When purchasing a new product, your objectives could very well be to achieve both high quality and low price at the same time, but most likely you have to trade one of them off for the other. Consequently, the decision maker has to make a subjective judgement of the size of this trade-off, or willingness-to-pay (WTP).

2) **Uncertainty**

Irrespective of the amount of conflicting objectives, decisions tend to be made within a context of uncertainty. Belton and Stewart (2002) differentiate between external and internal uncertainty. The former is connected to lack of knowledge about the consequences of a particular choice (information uncertainty); the latter is more related to the modelling process itself such as imprecision in model structuring and subjective judgements (model and preference uncertainty). For coping with uncertainty, the irrational approach is to ignore its presence. A more rational manner is to take actions to reduce it. Nevertheless, chances are high that some uncertainty will remain.

3) **Multiple stakeholders**

Decisions are often affecting the interests of other people and organisations in addition to the interests of the decision making institution. Even among decision makers, different objectives are likely to cause considerable difference in how the problem should be formulated, which alternatives that are available, how uncertainty should be assessed and how large trade-offs should be (von Winterfeldt and Edwards 1986). Relevant objectives could be difficult to identify if relevant stakeholders are not identified first.

4) **Far-reaching consequences**

Not only are consequences uncertain; they often vary considerably in when they will occur and for how long they will last. The time span may be years, or even generations (von Winterfeldt and Edwards 1986). Besides, consequences are often having secondary impacts, which again have impacts, and the longer you make the cause-effect chain, the longer the list of decision criteria will have to be. The geographic extent is also an important aspect that may complicate the picture – consequences on a local level could very well be different from global consequences.

The structure of a decision analysis assists the decision maker in taking these complicating dimensions into account in a rational way. There are situations where the decision maker has already decided what to do, but where decision analysis still might have a purpose. These are situations where decision makers want psychological comfort for their decision, help to communicate insights and considerations, or a formal justification and documentation in order to convince others. Decision analysis may also uncover new insights that alter a decision originally made (Keeney and Raiffa 1976).

2.1.2 How to perform a decision analysis

Belton and Stewart (2002) identify three main stages of a decision analytic process, which can be further split into several steps. The suggested procedure below has to be seen as more normative than descriptive; processes turn out to be iterative and even the more fundamental parts are likely to be altered as the work progresses (Keeney and Raiffa 1976).

A) PROBLEM STRUCTURING

Before a problem can be solved, it has to be identified and structured. This process starts with restricting the problem and ends with a consequence table.

1) **Restricting the problem**

   This requires first of all that the problem *frame* is restricted; cf. point on far-reaching consequences in chapter 2.1.1. There is no fixed answer on where to stop, but a guiding principle could be to only include impacts that have obvious value (Seip and Wenstøp 2006). By restricting there is a danger of sub-optimising, but problems will remain hopelessly intractable if they are not bounded, and dangers of sub-optimising will be even higher (Andrews and Govil 1995).

2) **Assessing relevant stakeholders**

   Next step is to decide how to assess the complexity of multiple stakeholders. A problem is not specified until it is clarified who the decision makers are and which stakeholders that are relevant (Seip and Wenstøp 2006). It is important to distinguish between the organisation itself as stakeholder and representatives for the organisation (i.e. individuals and departments) as stakeholders – the latter are not relevant. Mapping decision context and on which hierarchical level the model and outcome is to be applied is also a part of this step.

3) Map relevant objectives

Once relevant stakeholders are known, their values and objectives form the basis for the list of criteria according to which decision alternatives should be measured. As such, first the underlying objectives have to be mapped and made explicit. It is important to notice that decision analysis is by and large not concentrating on whether these values and objectives are appropriate or not. The focal point is on how ends can be reached by choice of rational means, not the choice of rational ends itself (von Winterfeldt and Edwards 1986). It is equally important to underline that the emphasis is on procedural rationality rather than substantively rationality. The distinction is made by Simon (1976) and referred to in Janssen (1992): "A decision process is substantively rational if the decision process results in selection of the best solution. A decision process is procedurally rational if the procedure to reach the best solution is optimal". A good decision is not necessarily leading to a good outcome. It is however hard to know in advance which alternatives that will turn out to be the best, and improving decision quality is therefore as close as we get (Janssen 1992).

4) Define a list of criteria

Criteria are chosen according to their capability of measuring attainment of objectives. Often we do not have any exact measures, and we have to choose an instrumental decision criterion that serves as an indicator of the real concern (Seip and Wenstøp 2006). Usually criteria are aggregations of larger amounts of so-called primary factors (Lahdelma et al. 2000).

The list of criteria should be as complete as possible, so that it covers all the important aspects of the problem. Furthermore, they should be operational, so that they can be measured and used meaningfully in the analysis. A third guideline is that criteria should be decomposable, so that aspects of the evaluation process can be simplified by breaking it down into parts. Criteria should also be non-redundant, so that double counting of impacts can be avoided; and minimal, so that the problem dimension is kept as small as possible (Keeney and Raiffa 1976).

5) **Identify viable alternatives**

The identification of viable decision alternatives is also vital. Alternatives that clearly performs worse on every criterion compared to another can be eliminated from the beginning (Seip and Wenstøp 2006). Keeney (1992:48) suggests that decision makers' values should be mapped prior to identifying alternatives, as the other way round tends to "...anchor the thought process, stifling creativity and innovation". However, considering alternatives may be a helpful tool in identifying values (Belton et al. 1997).

6) **Compute a consequence table**

Having defined all relevant criteria and having calculated a score, quantitative or qualitative on each criterion for each alternative, the results should be summarised in a matrix/consequence table, or alternatively a decision tree. This constitutes the basis for further model building.

B) MODEL BUILDING

In most problem situations that require decision analysis, criteria are expressed on different scales and in different units with a differing degree of certainty. Besides, the relative importance of criteria and criteria scores may vary substantially. The model has to compensate for both of these aspects through aggregation and weighting:

1) **Aggregate criteria**

A way to modify and aggregate criteria scores must be chosen so that a total performance of an alternative may be calculated. This is inter alia referred to as choosing an evaluation method (Janssen 1992), preference model (Belton and Stewart 2002), or decision aid method (Lahdelma et al. 2000).

2) **Weight criteria**

A way to further modify criteria scores must also be chosen so that differences in importance are taken into account. For most evaluation methods this is often referred to as eliciting criteria importance weights (Seip and Wenstøp 2006), and is equivalent to assessing trade-offs between criteria. The interpretation of weights is highly dependent on the evaluation method applied, and they should accordingly be elicited only after this choice has been made (Vincke 1992).

C) MODEL APPLICATION

When the problem frame, objectives, criteria and alternatives are given, the assessment of the consequence table is more a task of information gathering. Moving on to the second stage, most decision analytic tools require judgements of a more subjective character. This is reflected both in how criteria scores should be converted to preference variables and subsequently how these variables should be weighted (Seip and Wenstøp 2006).

After these judgements have been made, it should be possible to calculate a total performance for each alternative, and rank them accordingly. Due to both internal and external uncertainty in data and processes, it is recommended that an analysis of the sensitivity of the ranking is performed in order to provide decision makers with further insight (Janssen 1992).

As the scope of the study is to employ the MCDA methodology, further attention is given to schools within this category of decision analytic methods. An outline for why a subjective treatment is needed for the Statoil case is given in the chapters 2.3.2 and 3.3.1.

2.2 Multiple Criteria Decision Analysis

2.2.1 General overview

Multiple Criteria Decision Analysis (MCDA) methods concentrate on problems where the complexity of multiple conflicting objectives is present. Belton and Stewart (2002:2) use the MCDA expression "[…] as an umbrella term to describe a collection of formal approaches which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that matter."

All MCDA approaches aim at assisting decision makers to integrate objective measurement with value judgement, and to make subjective assessments explicit and manageable. The focus is not on removing the need for difficult judgements, but rather on making judgmental procedures and results consistent and transparent. (Belton and Stewart 2002).

The frameworks within the MCDA field may vary from quite simple approaches to more advanced models based on mathematical programming that require extensive information on criteria and preferences. They all share however the main characteristics of decision analytic approaches, such as the necessity of a matrix with scores deriving from a finite number of alternatives and criteria, and an element of importance weighting of these scores (Greening and Barnow 2004).

2.2.2 Specific MCDA schools

The main differences in MCDA approaches are expressed in the design of the evaluation method. These are based on different theoretical foundations such as optimisation, goal aspiration, outranking, or a combination of these (Linkov et al. 2004).

1) **Optimisation models**

   Optimisation models employ numerical measures when converting and weighting criteria scores into a total performance indicator for each alternative (Linkov et al. 2004). These models are also referred to as value function methods. Such methods are compensatory of nature, i.e. for an alternative, bad outcome on one criterion can be compensated for by a good outcome on another (Belton and Stewart 2002).

   Extensively applied variants of these models are Multi-Attribute Value Theory (MAVT) and Multi-Attribute Utility Theory (MAUT) (Keeney and Raiffa 1976), which will be presented in more detail in section 2.2.3 and 2.2.4. Another widely used optimisation model is the Analytic Hierarchy Process (AHP) approach, where comparison between alternatives is based on pair-wise comparisons of decision criteria. In the AHP methodology, relative preferences are expressed on a qualitative scale instead of using and modifying value or utility functions as in MAVT/MAUT (Linkov et al. 2004). The implied meaning of the weights in AHP is perceived as hard to conceptualise for decision makers (Belton and Stewart 2002)

2) **Goal aspiration models**

   These are non-compensatory models that are based on satisfying levels of achievement for each criterion (Linkov et al. 2004). In short, decision makers rank criteria and seek improvement on the criterion considered to be the most important. When the level is satisfactory, the emphasis is moved to improve the second most important criterion. Alternatively, a mathematical programming algorithm is applied to get as close as possible to all goals/satisfying levels (Belton and Stewart 2002).

3) **Outranking models**

   If one alternative performs better than another on all criteria, the first dominates the other and the evidence favouring this conclusion is indisputable. In less obvious situations where dominance on each criterion does not exist, there could still be sufficient evidence to claim that the first alternative is at least as good as the second, and thereby outranking that alternative (Belton and Stewart 2002). As such, the approach is based on pair-wise comparisons between potential actions, or overall alternatives, rather than on each criterion per se (Georgopoulou et al. 2003). The core challenge is to establish the strength of evidence, i.e. identifying sizes of difference that imply clear preference and clear indifference between the alternatives (Linkov et al. 2004). These thresholds are difficult to assess, as the mathematical functions underlying them are hard to conceptualise. The procedure does not necessarily result in a complete ranking of alternatives (Simpson 1996).

As the point of departure for the HSE risk integration project is to apply utility functions, further attention will mostly be concentrated on MAVT/MAUT models. Discussions on choice of aggregation model are found in chapter 4.3.1.

2.2.3 More on Multi-Attribute Value/Utility Theory

In the MAVT/MAUT model, diverse criteria are transformed into one dimensionless scale (Linkov et al. 2004). The difference between MAVT and MAUT is that the former only transform criteria scores into standardised scores, whereas the latter in addition explicitly allows for score modification due to uncertainty. Hence, the utility function does not only standardise, it also includes the decision makers' attitudes to risk (Janssen 1992). For linear functions there are no differences (Janssen 1992). Of simplicity reasons, only the term ‘utility function’ and MAUT will be further applied.

The most common utility functions applied in decision analysis are additive, i.e. they are the result of a mere summation of weighted partial utility functions. A partial utility function reflects the conversion of scores for one criterion to a standardised score representing its utility. The total utility function could thus be expressed as

$$U(z) = \sum_{i=1}^{m} w_i u_i(z_i) \tag{2.1}$$

where

- z_i is the score of criterion *i* (e.g. an EIF score)
- $u_i(z_i)$ is the partial utility function (e.g. the utility of this EIF score on a 0-1 scale)
- w_i is the relative weight of this score compared to all other scores (e.g. the importance of this EIF score relative to the scores of the other criteria)
- $U(z)$ is the overall utility (e.g. the total utility of all EIF scores together, if these constituted an exhaustive set of criteria)

In the MAUT model, preferences have to be consistent with a strong set of axioms. These include inter alia that more benefit is preferred to less (or that less harm is preferred to more). To use the simple decomposed models above, one also must assume that preferences do not change with time and that preferences are independent, meaning that the subjectively assessed trade-off between levels on two criteria is not affected by the level of a third criterion (Belton and Stewart 2002), (Linkov et al. 2004), (Keeney and Raiffa 1976).

Another assumption connected to the additive versions is a so-called interval scaled property - the level of utility does not necessarily have an absolute meaning, but so does the *ratio* between two utility scores. Other functions than additive ones are of course available, but they seldom improve the validity of the process (Wenstøp 2006). In fact, for operational purposes, given z_i we want to choose *U* such that the function $u_i(z_i)$ is easy to manipulate mathematically (Keeney and Raiffa 1976). This makes it easier to maintain transparency in how the model is constructed and how the outcome is derived.

An additional assumption that could be made is that there exists a linear relationship between the criteria scores and the utility functions for each criterion such as in fig. 2.1. In other words, if you double the criterion score, you always double the partial utility deriving from this score.

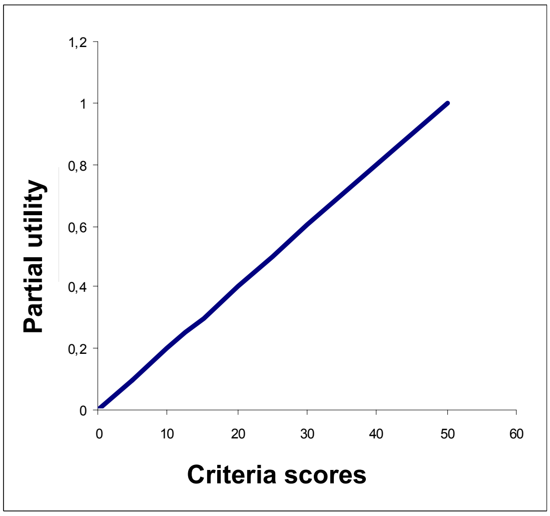


Figure 2.1 Linear partial utility function

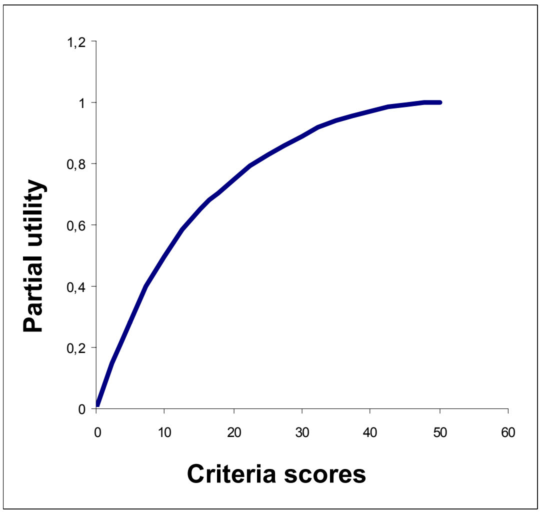


Figure 2.2 Non-linear partial utility function

An example of such a linear partial utility function is shown in fig. 2.1. This simplifies the picture, but the MAUT framework does allow for non-linear utility functions as well. As indicated by fig. 2.2, trade-offs in non-linear models could never be constant. In this case, the gain or loss in partial utility of a certain change in criteria scores will be highly dependent on the level of departure. Some studies suggest that in total, non-linear models perform inferior to linear ones (Schoemaker and Waid 1982).

If an appropriate utility is assigned to each criterion and the expected utility of each alternative is calculated, then identifying optimal strategy is the same as calculating which alternative that maximises expected utility (Keeney and Raiffa 1976).

2.2.4 Weight elicitation given MAUT

As mentioned earlier, importance weights in value/utility models can be perceived as tradeoffs: How much of criterion A is the decision maker willing to sacrifice for the benefit of criterion B (Janssen 1992). In MAUT, the criteria weights represent the criteria importance in discriminating power which is proportional to the swing from worst to best on each criterion (Choo et al. 1999). A crucial point to be made is that the importance weights represent both the intrinsic value of the criterion as well as the score range it gets in a specific situation (Belton and Stewart 2002).

Example:

Consider a decision situation where two relevant decision criteria are defined, A and B. Both of these criteria are benefits, i.e. more is better. The maximum score that is possible to obtain on these criteria is 10 on A and 20 on B, all alternatives considered. We assume furthermore that the minimum score on both criteria is 0, and that the maximum scores hence equal the swing from worst to best on each criterion. Before converting these scores to partial utilities, we assume that all partial utility functions are linear and that they are (arbitrarily) scaled to [0,1]. A score equal to a swing always achieve maximum partial utility. This is summarised in table 2.1.

Criterion i	**Swing** z_i	**Partial utility** $u_i(z_i)$
A	10	1
B	20	1

Table 2.1 Example of swings

Now, imagine that you judge the swing on criterion A to be three times as beneficial as the swing on criterion B. As a consequence, the partial utilities cannot be merely added; they have to be modified according to their relative importance. As such, the weight of criterion A has to be three times the weight of criterion B. Weights are often normalised so that all weighting factors sum up to 1. This is shown in table 2.2. If new and important criteria were added, the old weights would inevitably be altered. The relative size of the weights would nevertheless remain the same.

Criterion i	**Weighting factor** $\cancel{w}_i$	**Normalised weighting factor** w_i
A	3	0,75
B	1	0,25

Table 2.2 Example of weights

As a result, we are now able to calculate the overall performance for each alternative. Imagine two alternatives with the following scores:

alternative α) with $z_A = 8$ and $z_B = 4$, so that $u_A(z_A) = 0{,}8$ and $u_B(z_B) = 0{,}2$

alternative β) with $z_A = 5$ and $z_B = 15$, so that $u_A(z_A) = 0{,}5$ and $u_B(z_B) = 0{,}75$

Consequently, according to (2.1), the total utility will be:

alternative α) $0{,}75*0{,}8 + 0{,}25*0{,}2 = 0{,}65$

alternative β) $0{,}75*0{,}5 + 0{,}25*0{,}75 = 0{,}5625$

The conclusion is that the former is preferred to the latter.

There are numerous procedures available for how the subjective judgements of relative importance, i.e. the importance weights, can be elicited from decision makers. Von Winterfeldt and Edwards (1986) distinguish between two main set of approaches; the numerical estimation methods and the indifference methods:

1) **Numerical estimation methods**

   These methods all apply so-called "numerical ratio judgement of relative attribute importance" (Roberts and Goodwin 2002), where an attempt to quantify the degree of difference in importance between criteria is done. Often, the criteria are ranked before their relative difference is quantified. The quantification itself could for instance be done directly through point allocation, where the decision maker has to distribute a fixed number of points to all criteria involved (Shoemaker and Waid 1982). Another variant in line with the example just given is the swing weight method, where the criterion with the most important swing is chosen as a reference with a fixed number of points, and where the other criteria are given points relative to the importance of their swings (Mustajoki et al. 2004).

2) **Indifference methods**

   These methods systematically vary scores on pairs of criteria until the decision maker is indifferent between the pairs. From this, relative importance weights are implicitly calculated (von Winterfeldt and Edwards 1986).

Again, variants are abundant. Empirical studies show that elicited weights differ according to the procedure chosen (Pöyhönen and Hämäläinen 2001). Still, there is no universal answer as to which weighting procedure is preferred. No matter how sophisticated tools and methods get, the quantification of stakeholders' underlying views is still prone to be biased by human irrationality. The best way of mitigating this is to make possible procedural pitfalls explicit and to make decision makers aware of the implications of their conclusions.

2.3 Characteristics of environmental decision making

Environmental decision problems often involve most or all complicating dimensions of a decision problem cf. chapter 2.1. All decision problems are characterised by multiple objectives where value judgements between conflicting socio-political, environmental and economic aspects have to be performed (Lahdelma et al. 2000).

2.3.1 Complexities in environmental decision making

Some characteristics of environmental decision making are particularly challenging. Environmental commodities do not have a clearly defined buyer and seller – they are consequently inadequately priced, and converting them into monetary values is far from straightforward. Their value may also depend on ethical and moral principles that are not directly related to any economic use or value. At least two sources of environmental value could be addressed; one is the environment's potential to generate welfare (anthropocentric perspective), the other is the environment's intrinsic value (Janssen 1992), (Seip and Wenstøp 2006), (Linkov et al. 2004). In light of this, the often seen assumption that all impacts are negative is far from trivial.

Moreover, the information available is often incomplete, as environmental impacts occur in systems that are often insufficiently understood. Lack of information and knowledge about these systems leads to high uncertainty both when assessing impact probabilities and impact consequences (Janssen 1992).

In issues related to potential impact to the environment, the number of stakeholders is often quite high. Stakeholders are frequently divided into categories based on their perceptions of physical or economic impacts or their ability to influence the decision-making process (Greening and Barnow 2004). Stakeholders are rarely equally affected by the consequences of a decision (Janssen 1992).

Consequences are also far-reaching, both in a spatial and temporal sense. It might take generations for an impact to occur, as well as for impacts to be mitigated. Some impacts can perhaps not be mitigated at all. This poses another ethical question on the extent to which future generations can be written off, and makes comparisons between economic and environmental effects even harder. As for the spatial dimension, effects from local activity may occur on both local and global levels. The severity of impact is likely to be related to where it takes place as a result of area sensitivity and background depositions (Janssen 1992), (Smit and Karman 2006).

2.3.2 MCDA and environmental decision problems

As a consequence of these complexities, individuals will often find it difficult to make informed and thoughtful choices and value trade-offs (McDaniels et al. 1999). Still, choices have to be made. The application of MCDA methods makes sure that all relevant aspects are made explicit, including all subjective judgements. This clearly enhances the traceability and transparency of the decision making process (Lahdelma 2000), (Wenstøp and Seip 2001). The latter is crucially important in a context where the decision maker is likely to make judgements on behalf of other stakeholders.

It is furthermore argued that other decision analytic approaches such as rule-based methods and cost-benefit analysis are deficient for these purposes, as they fail in addressing the inevitable element of value judgement (Wenstøp and Seip 2001).

The need for transparency in the application process is also emphasised by Janssen (2001:108): "The fear that stakeholders will perceive MCA as a 'black box' and, therefore, reject its results, leads to the use of simple straightforward methods, such as the weighted summation, and limited interest in sensitivity analysis." This underlines the fact that designing an MCDA problem is in itself an MCDA situation, where the need for precision and accuracy has to be traded off against the need for simplicity and transparency.

MCDA has been successfully applied in a wide range of environmental decision problems (see e.g. Wenstøp and Seip 2001, Janssen 2001, Greening and Barnow 2004). In Janssen (2001), 21 applications in the Netherlands between 1992 and 2000 are reviewed. In the majority of these cases, a simple utility function variant is employed. The consequence tables included between 14 and 100 criteria and between 5 and 61 alternatives. Often, the political process that followed the MCDA application resulted in compromise alternatives. It is claimed in a conclusive remark that supporting problem definition and design appears to be a more important methodological challenge than developing more sophisticated MCDA methods.

2.3.3 Methodological requirements

The subjective element of the MCDA methods is often regarded with scepticism. It can however be argued that in environmental decision problems, an element of value judgement cannot be avoided as long as the environment is considered to be of some value in addition to a purely economic one. The concern of decision analysts should therefore not be to avoid subjectivity, but to ensure that applications of subjective models are both reliable and valid, in particular when it comes to elicitation of importance weights (Wenstøp and Seip 2001).

For an application to be *reliable* the same results should be obtained if the process were repeated. For an application to be *valid* there has to be no doubt that the decision makers really understand what is at stake when assessing trade-offs (Wenstøp 2006). For validity to be present, one important aspect is that the acting decision makers are legitimate, i.e. that they can be regarded as unbiased, responsible experts. Another important prerequisite for validity is that the scenarios used in the valuation process are as vivid, balanced and clear as possible, so that the valuators can be both rationally and emotionally involved (Wenstøp and Seip 2001).

According to Damasio (1994), reasoning is essential for making good decisions, but it is not enough. In order to be able to apply well-founded values when solving complex problems, he suggests that reasoning has to be accompanied by an acquired emotional appreciation of consequences. Consequently, affect is indispensable for rational behaviour, as rationality is not only a product of the analytical mind, but of the experiential mind as well. In certain circumstances such biological drives may however be detrimental by "…creating an overriding bias against objective facts…" (Damasio 1994:192). This can be exemplified by the fact that our sensitivity to small changes (e.g. the difference between 0 and 1 deaths) rarely is proportional to changes further away from zero (e.g. the difference between 1000 and 1001 deaths). This is an inherent bias of the experiential system (Slovic et al. 2004).

To mitigate some of these biases, Kahneman and Sugden (2005) suggest that maximising so-called *experienced utility* (utility as hedonic experience) could be a better target for decision making than maximising *decision utility* (utility as representation of preference). 'Preferences' are described as mental entities that rationally explain the choices an individual makes. Hence, preferences are revealed through observable choices and can be seen as objective measures, as it is assumed that individuals always act according to their preferences. On the other hand, 'hedonic experience' is to be interpreted as a more subjective judgement of overall happiness, i.e. the level of pleasure and pain. Maximising pleasure is therefore proposed as a better rational target for decision making that allows for emotional appreciation of the criteria involved.

However, as individuals are only boundedly rational (as Slovic (2004) pointed out), they are not necessarily making choices that will actually increase their happiness. Affective-rational measures of hedonic experience are therefore not possible to identify by observing choices, as experience and behaviour do not correspond. Consequently, expected experienced utility is difficult to estimate precisely. Two reasons for bounded rationality prevail:

1) Individuals fail to forecast to which extent they actually will adapt to a new state.
2) Individuals overstate the importance of whatever issue they are currently required to think about.

In short, the overriding bias from immediate emotional responses hampers the assessment of future "happiness".

For the scope of this study, it can be argued that these boundaries are less apparent:

- Since Statoil aims at zero impact on the environment, it is the environmental transition itself that should be avoided. The fact that affected ecosystems could be quite rapidly restored or replaced, or that some impacts even may contribute positively, is to a certain extent irrelevant. Hence, it is exactly the emotional responses to change that are relevant, not the emotional appreciation of a future state.
- Given that just potential impacts on different environmental compartments are included in the decision problem, the weights only reflect the *relative* importance of the risk across compartments. If the general importance of a potential impact is overstated, it would affect all criteria. However, a possible deviance of attention between compartments could arise, as decision makers may be prone to attribute greater importance to potential impacts in compartments they are more familiar with.

After all, expected anomalies arising from a preferential approach are perhaps not that strong. In addition, it can be argued that maximising pleasure may not be an appropriate target in environmental decision making, as objects at stake could have value beyond their capacity to create pleasurable affective states. This might mitigate some of Kahneman and Sugden's (2005) general criticism of the expected utility approach, which is chosen as the method in the present study. The important challenge of evoking the right balance of well-tempered emotions remains however (Wenstøp 2005).

3. ENVIRONMENTAL DECISION MAKING IN STATOIL

3.1 General approach for decision making in Statoil

Before turning to details on how Statoil has framed their environmental risk assessment and treatment, a normative introduction to Statoil decision making will be given through *how, when and who*.

3.1.1 How: Decision foundations

According to the Statoil Book (2007:31), decision makers shall make their choices based on:

- values and policies
- 'ambition to action'
- decision criteria and authorities
- sound business judgement

1) **Values and policies**

   "Our values, HSE and ethical requirements are at the core of all our activities." ('The Statoil Book' 2007:8). For environmental management, it is clearly stated that the fundamental guideline is to "cause zero harm to people and the environment" (http://www.statoil.com/hse). A list of Statoil values and HSE goals can be found in appendix A.

2) **'Ambition to action'**

   The 'ambition to action' is a process where long term ambitions are translated into shorter term strategic objectives. Required actions are identified within five delivery areas, with the aim of ensuring balance between financial and non-financial concerns, as well as between short term and long term focus. The five delivery areas are ('The Statoil Book' 2007):

   - People and organisation
   - Health, safety and environment
   - Operations
   - Market
   - Finance

Often these strategic objectives are conflicting in the sense that improving performance on one objective is likely to make the performance on another objective worse off.

3) **Decision criteria and authorities**

Key Performance Indicators (KPIs) are ideally measuring delivery against strategic objectives. Requirements from authorities as well as internal guidelines should influence both which indicators are chosen as KPIs as well as their specific target. For environmental issues, absolute acceptance thresholds are however rare. Targets are therefore rather based on the principle of 'continuous improvement', i.e. as long as there is a risk present, it should be reduced.

4) **Sound business judgement**

Investment decisions are based on an overall management evaluation of important factors relating to each individual investment proposal, so that sub-optimisation is avoided.

For HSE concerns, the appraisal of objectives, relevant criteria and sound business judgements require that environmental risk relevant for the decision is assessed. The assessment process is divided in three parts ('WR2266' 2007):

- Risk identification (what are the possible hazards/events?)
- Risk analysis (estimating/quantifying impacts and their likelihood)
- Risk evaluation (determining severity and significance of such impacts with respect to internal and external requirements)

In accordance with Statoil governing document WR1912 (2006), small and medium sized projects can apply a simplified HSE risk assessment process.

In the majority of cases it is possible to achieve a lower level of risk than what requirements demand. Alternative solutions and risk reducing measures will be identified and executed if costs are not excessive relative to benefits, even when the risk level is below minimum requirements. This is referred to as the As Low As Practicably Possible (ALARP) principle, and these judgements constitute the core of the "risk treatment" process ('WR1912' 2006).

In addition to the mere identification of optimal alternatives, the need for best practices and optimal HSE solutions to be documented has also become more evident ('Integrated HSE project manual' 2005).

3.1.2 When: Decision processes

Decisions can be categorised according to which phase they are related to, expressed in terms of decision gates, see fig. 3.1. All decisions from the so-called decision gate A up to decision gate 1 belong to what is called the early phase, and covers decision areas from 'country entry' through to 'feasibility'. At decision gate 1, decisions on project concretisations are made, at decision gate 2 the main concept is chosen and at decision gate 3 the entire concept design is defined. Decision gates 1-3 cover all decisions between the start of project planning to the project is finally sanctioned. Decision processes in merger and acquisition contexts are likely to follow a different structure.

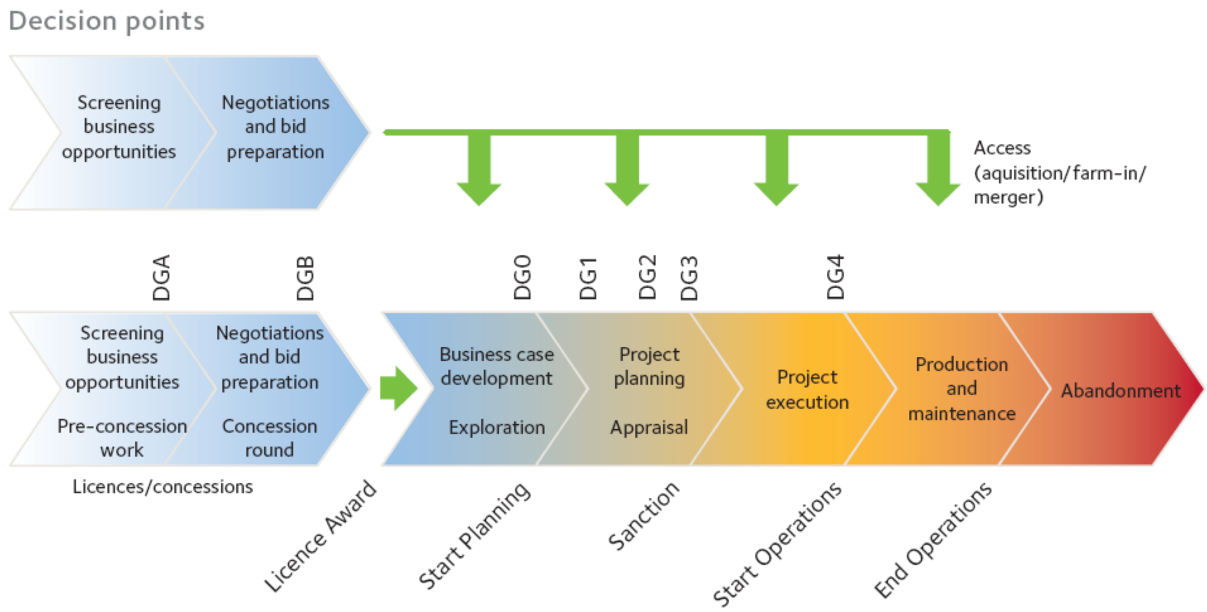


Figure 3.1 Decision points in Statoil ('The Statoil Book' 2007:31).

For HSE risk assessment and treatment in the early phase, there is currently a separate management tool under construction in Statoil named EPRA (Early Phase Risk & Opportunities Assessment). This tool is suitable for decision making in cases where information is scarce (Kinsella 2006). In this study, the focus will be on decisions and decision relevant information from gate 1 and onwards, where there is enough information available for EIF calculations.

3.1.3 Who: Decision makers

The organisation of Statoil is made up of asset-based and function based units. The former are concentrated on achieving results on directly related business activities, whereas the latter are responsible for delivering “function capabilities, best practice work processes and requirements” (‘The Statoil Book’ 2007:25). These units mainly correspond to the line role and the support role respectively, where the former is responsible for decision making according to their location on the organisational chart. When there are conflicts of interest, the line role has primacy over the support role.

In HSE issues, the Health, Environmental and Safety Technology department (HMST) attends to the supporting role of developing and applying knowledge, expertise and tools that support the overall work of reaching the HSE targets. The research department is developing these tools and also assists in applying them. The supporting role for HSE issues is always present in project organisations, embodied by representatives from HMST and the professional ladder.

The division between roles require inter alia that tools elaborated by the supporting division are transparent so that they are able to gain credibility and confidence in the line division. The procedures prescribed by a decision support tool should furthermore be compatible to the way decisions are currently made in the organisation, as organisational activity has a life of its own and may or may not be much influenced by chosen managerial instruments. (Brown and Duguid 1991). Both these considerations may conflict with the endeavour for scientific accuracy, but may on the other hand be indispensable for actual application of a new framework. This trade-off when elaborating new tools constitutes in itself a multi criteria decision making problem.

3.2 Environmental risk assessment

3.2.1 Impacts from discharge and the EIF framework

A prerequisite for HSE considerations when making decisions in projects is that the HSE risk is properly assessed. In environmental risk assessment, different assessment techniques can be applied. These techniques range from simple screening tools to very sophisticated ecotoxicological models. All tools have in common that they include a comparison of exposure and threshold levels.

A) THRESHOLD LEVELS

In general three different levels of risk assessment can be distinguished. Level 1 and 2 are based on generic data whereas level 3 is area-specific (Smit and Karman 2006):

1) The most conservative level is the so-called PEC:PNEC-level, where PEC = predicted environmental concentration and PNEC = predicted no-effect concentration. The PNEC figure represents the maximum concentration that can be present without affecting the most sensitive species. If the predicted concentration resulting from a discharge (PEC) is higher than the PNEC (PEC:PNEC ratio > 1), the tolerance level is exceeded and it is likely that adverse effects to species will occur.

2) The second level is based on probabilistic risk assessment and has the msPAF as risk endpoint, where msPAF = multi-substance potentially affected fraction of species. This approach does not only look at the most sensitive species, but also takes into account that the sensitivity to stressors among species varies. With the help of species sensitivity distribution curves (SSDs), a PNEC equivalent can be found. If the PEC figure now exceeds PNEC, it is likely that a fraction of species will be affected. The higher PEC gets, the higher this fraction will be. The PAF figure quantifies how many species will potentially be affected as a percentage of all species present in a generic ecosystem. The msPAF figure is the combination of PAF figures for all relevant stressors. The acceptance level is by international regulations defined to be exceeded when msPAF $\geq$ 5 % (e.g. van Straalen and Denneman 1989)

3) The third level is based on ecological modelling and has the effect on individuals as risk endpoint. At this level specific adverse effects on representative species are predicted. This demands considerably more data than a more generic approach, as individual dose-response relationships for each species and for each stressor have to be mapped. On the other hand, the accuracy of the environmental impact assessment is enhanced as area-specific data are applied. Effects occur when PEC $>$ NOEC, where NOEC $=$ No-Effect Concentration for single species. Acceptance levels depend on characteristics of species and area.

B) EIFs AND THRESHOLDS

Based on the three levels above, the framework of Environmental Impact Factors (EIFs) has been designed. The EIFs are indices of quantitative nature, reflecting the potential impact on species from emissions and discharges. They are widely applied as decision tools for ranking different technology solutions and for selecting measures that yield the biggest environmental gain ('Mastering challenges' 2006). Their quantitative element is related to a spatial extension where the concentration of the emitted or discharged compounds exceeds an environmental level of tolerance ('Integrated HSE project framework' 2005). In other words, it is not the amount of discharge itself that is measured, but rather the extension of the environment (expressed as amount of area/volume grid cells of environmental compartments) that is likely to be afflicted by the discharge. Consequently, an EIF value of 1 means that one unit of the compartment in question is having an unacceptable environmental level of environmental risk.

EIFs can be calculated on the basis of all three threshold levels:

- If threshold level 1 is applied, the EIF score corresponds to the number of grid cells where PEC:PNEC $>$ 1. The PNEC is calculated using generic data.
- If threshold level 2 is applied, the EIF score corresponds to the number of grid cells where msPAF $\geq$ 5%. The msPAF is calculated using generic data.
- If threshold level 3 is applied, the EIF score corresponds to the number of grid cells where PEC:NOEC $>$ 1. The NOEC is defined using area-specific data.

Another activity in the integrated HSE risk management project aims at converting EIFs to be expressed with the same threshold level (Smit and Karman 2006). The msPAF-level is chosen, implying that all EIFs should be expressing number of grid cell units with msPAF ≥ 5%. An overview of how an EIF score at the msPAF level is constructed is shown in figure 3.2.

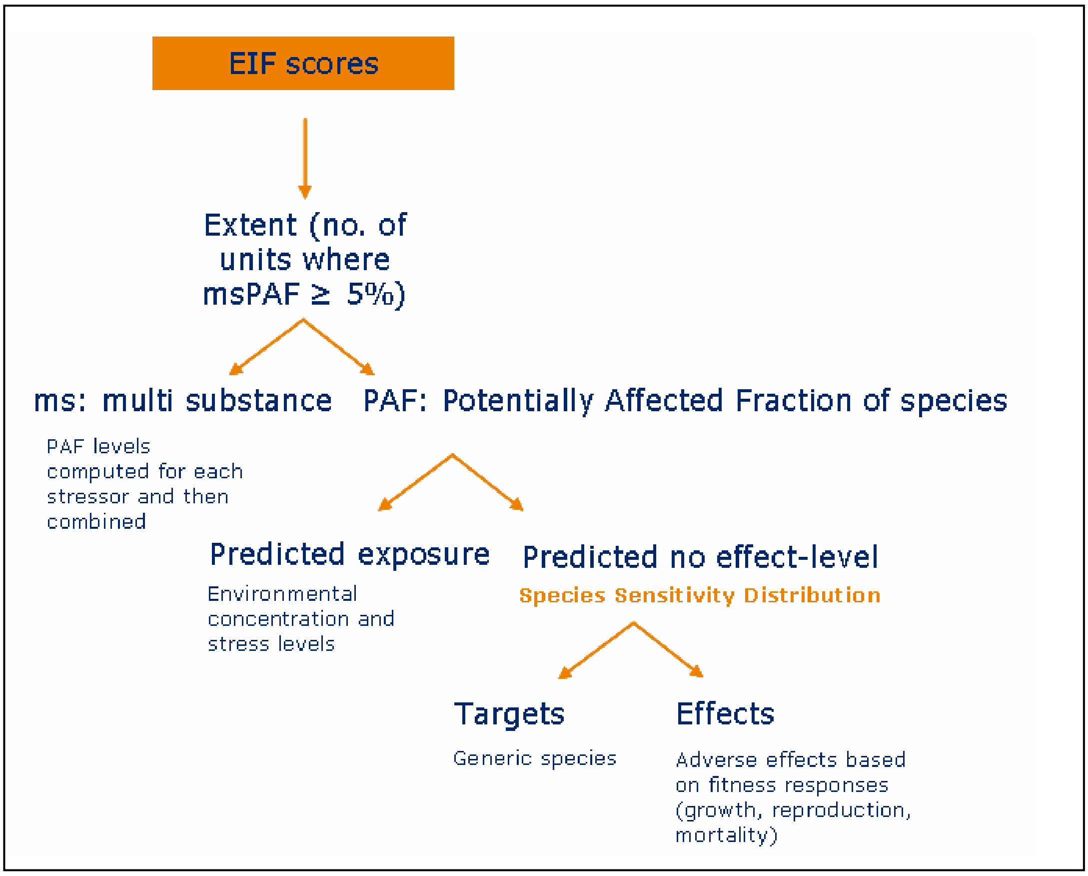


Figure 3.2 Construction of an EIF at the msPAF level

C) CURRENT FEATURES OF THE EIFs

The EIFs are referred to by source of emissions and discharges and not by compartments potentially affected. So far, four different EIFs are more or less operative, comprising discharges from produced water, drilling, air emissions and acute spills. Two more EIFs are under development, seeking to integrate discharges from land facilities. An overview over all EIFs and the compartments potentially affected can be seen in table 3.1. For more detailed info on the separate EIFs, see appendix.

As presented in table 3.1, the quantitative elements of the EIFs still differ with respect to volume/area even if all were expressed at an msPAF-level. The degree of uncertainty regarding the quality of input data for modelling is furthermore highly variable among the EIFs (Smit and Karman 2006).

Name	Source	Compartments affected	Quantitative element	Availibility
EIF PW	Produced water	Water column	Volumes with msPAF≥ 5%	available
EIF DD	Drilling discharges	Water column, sediment	Volumes/areas with msPAF≥ 5%	available
EIF air	Air emissions	Soil, fresh water	No cut-off criteria	available
EIF acute	Acute spills	Water column, sea surface, sea shore	Volumes/areas with msPAF≥ 5%	available
EIF onshore	Land facilities' discharges to sea	Water column, sediment	Volumes/areas with msPAF≥ 5%	under development
EIF soil & ground water	Land facilities' discharges on land	Soil, fresh water and ground water	Volumes with msPAF≥ 5%	under development

Table 3.1 Characteristics of available and future EIFs.

As also presented in table 3.1, the quantitative element of EIF air is different from the other EIFs. In fact, converting EIF air to an msPAF level would make the tool less practicable for selecting best environmental options regarding air emissions, and it has therefore been judged as an inadequate solution ('Minutes integration meeting' 2007).

The difference lies in the way EIF air is defined. It assesses the potential for acidification, eutrophication and production of near ground ozone as a result of emissions of nitrogen, sulphur and volatile organic compounds (VOC). The EIF score does not indicate the extent of the affected compartments directly as the other ones do, as the score is not directly related to exceeding threshold levels. Unlike other discharges, air emissions often contribute with an immaterial fraction of toxic elements in relation to the often substantial background depositions in the compartments. The emissions alone are rarely harmful, but together with background depositions they may contribute to exceed the critical loads for an ecosystem.

Consequently, the background depositions are multiplicatively integrated in the EIF air, resulting that the factor is not expressing grid cells affected from the activity under consideration. These features give the EIF stemming from air emissions a strong descriptive power, as the numerical figures are always area-specific.

Another challenge with the EIF air is that the factor does not comprise impacts from emissions of greenhouse gases. Because potential impacts from greenhouse gases are on a global level, mere quantitative estimates of emissions are, for the time being, judged as being representative for indicating importance to decision makers. All emissions of greenhouse gases can be recalculated to CO_2 equivalents.

3.2.2 Environmental risks not included in the EIF framework

In the introduction, a link between optimal EIF scores and optimal environmental performance was established. There are however several aspects of environmental risk not comprised by the EIF framework. Statoil's governing document WR2266 (2007) states that in addition to impacts from emissions, discharges and waste streams, risks related to the following areas should be identified:

- Energy consumption
- Use of land
- Utilisation of natural resources (including use of fresh water when this is a limited resource)
- Products
- Reputation (including stakeholder concern)

In addition to these areas again, Statoil's sustainability report (2006) states that consequences of global warming deriving from emissions of greenhouse gases overshadow all other environmental problems. Moreover, "preserving biodiversity is a key element in sustainable development and occupies a central place in our environmental work" ('Mastering challenges' 2006:42). It can be argued that none of these two concerns are sufficiently comprised by the assessment tools as they are applied today.

In WR2266 (2007), it is emphasised that potential impacts from emissions (except greenhouse gases), discharges and waste streams should be quantified by established management tools such as the EIF framework. If there are potential impacts from other sources, they "shall be analysed through the impact assessment process" (p. 10). No other comprehensive, quantitative risk assessment framework than the EIFs exist, although Life Cycle Analysis (LCA) has proven to be a suitable tool for assessing parts of the environmental risk related to use of resources and products (Gulbrandsøy and Solberg 2006), (Goedkoop and Spriensma 2001).

3.3 Environmental risk treatment

As normatively defined in WR2266 (2007), risk treatment at Statoil "involves identifying the range of options for risk mitigation, assessing these options, prioritizing, preparing risk treatment plants and implementing them" (p. 11). In the risk assessment phase, the unacceptable environmental risks related to specific actions are mapped. In the risk treatment phase, decision makers have to judge which alternative that makes the overall risk as low as reasonably practical.

As it is today in Statoil decision making, no formal comparison of EIFs or other environmental risks take place to facilitate this judgement. Whatever risk that is separately regarded as relevant, expressed through quantitative or qualitative measures, is brought to the project decision makers by the HSE representative for further treatment.

For the purpose of this study it is convenient to split the normative judgement in two parts:

1) Which decision alternative is environmentally optimal when taking all relevant environmental aspects into account?
2) When including other values and decision criteria, which alternative should finally be chosen?

Both trade-off situations require a profound framing of the decision problem.

3.3.1 Identifying an environmentally optimal alternative

The first challenge is to grasp the overall potential environmental impact as measured by the EIFs. It has already been claimed that a straightforward aggregation of different EIF scores has no meaning. Furthermore, it is rare that one alternative is superior to all others on every EIF so that a clearly dominant alternative can be recognised. In order to identify the overall performance of all EIFs deriving from an alternative, the relative importance of the EIF scores has to be identified first. As pointed out by Wenstøp (2006), there are in principle two approaches for how such trade-offs could be made; judgements involving natural science and socio-economic judgements (referred to as objective) and judgements by experts (referred to as subjective).

Environmental science can alleviate some of the differences in the quantitative elements of the EIF scores; cf. the project of expressing all EIFs at an msPAF level. Other aspects of the differences in the quantitative elements are harder to account for by natural science. Volume is different from area, and there are for the time being no given answers as for how equivalents can be calculated. Besides, even if the quantitative elements were totally similar, we would still have to allow for differences in relative importance between potentially affected compartments. One unit of water column at risk is not necessarily judged as severe as one unit of fresh water at risk, and this judgement is likely to differ from location to location in line with differing compartment sensitivities.

Thus, trade-offs have to be a result of a value judgement of the importance of ecosystems and extension differences, and it has to be made through the content of complex indicators, site-specific knowledge and attitude to uncertainty. It is hard to imagine that these aspects are so feasible for laymen that a reliable and valid judgement of what is optimal from an environmental point of view could be achieved through social surveys.

Hence, a subjective treatment is needed, and it has to be performed by experts that are able to comprehend the criteria and appreciate the underlying values that have been regarded as relevant. If these values correspond to the corporate values, then corporate experts should be used.

The second challenge is how the potential impact as measured by the EIFs can be compared to other relevant environmental risk. For the time being and for managerial purposes, it is assumed that EIF scores and other EIF related information are sufficient for assessing overall environmental risk. This assumption will be problematised in chapter 5.3.4 and 5.3.5.

3.3.2 Identifying an overall optimal alternative

In a real decision making situation, there are several reasons for why the environmentally best solution is not always chosen:

- Health and safety concerns also constitute important decision criteria. Operations that are optimal from an environmental point of view could at worst be fatal for one or both of these matters.
- Even if an optimal HSE alternative is identified, there are still objectives and targets in other delivery areas that have to be considered. There is for instance likely to be a trade-off between reducing risks and increasing costs.

In actual overall decision making, cost/benefit or cost/effectiveness analyses are often performed when adequate figures are available. Other important aspects are presented through separate quantitative or qualitative figures that indicate performance, e.g. interval scales and “traffic lights”. Others again are not represented by separate indicators. As a result, if some criteria were made more visible, it could enhance the likelihood that trade-offs made actually reflect the underlying values of the decision makers. A suggestion for how MCDA could have been applied at an overall basis have been elaborated, but has reportedly not yet been tried out in an actual decision making setting (Aksnes 2007).

There have previously been attempts to make bilateral trade-offs between specific environmental risk factors and costs. In connection with the zero discharge report to the Norwegian Pollution Control Authority in 2002, a willingness-to-pay figure for reducing EIF produced water by 1 was monetised to NOK 200 000 through simplified cost/benefit considerations. Thus, if EIF produced water and costs were the only decision criteria, no further judgements would be needed to identify the overall optimal alternative. The figure has not been updated and is currently not in use (Furuholt pers.comm. 2007).

A general aim for the supporting roles at Statoil is to provide comprehensive decision support for concept selection, technical solutions and execution of activities, so that the best available alternative on an overall basis is chosen. The development of an environmental risk indicator fulfils the aim by taking an important step to facilitate further treatment of environmental risk and at the same time elucidate its presence.

4. CASE PRESENTATION AND METHODOLOGY DESIGN

The case applied in this study is a modified version of the case that was applied during the trial session in 2006; choice of drilling technology at Norne. The limited time frame for this study did not allow for alternatives, as no other cases have been thoroughly developed since then. Experience from the 2006 session, the work of converting EIFs to an msPAF level and further conceptualisation have all contributed to the design of the current application process.

But, as mentioned in chapter 2, processes tend to be iterative. This process has certainly not been an exception. Several aspects have therefore been modified as the work has progressed. The initial application procedure is outlined in this chapter, while the results, amendments and further results are all presented in chapter 5.

4.1 Case: Drilling at Norne

Norne is an oil field in the Norwegian Sea, 80 km north of Heidrun and 182 km west of Sandnessjøen. The size of the field is approximately 9x3 km, and production started in 1997. The oil and gas is located in sandstone of Jurassic age (Dokka and Midttun 2006), (Knudsen et al. 2006).

The case study is related to the drilling of a production well at this field. Two problem aspects are identified at the outset of the study; one with regard to which mud system that should be adopted (oil based or water based), the other is related to treatment of drill cuttings. The choice of waste management will depend on the choice of mud system.

The relevant EIF scores have to a large extent already been assessed. Two areas are incomplete:

- Risk related to onshore activities (EIF framework not complete)
- Risk related to waste management

As a result, some potential impacts are assessed on the basis of other available data.

The focus for this application is however on further framing, aggregation and weighting, and to evaluate whether the approach as such is feasible. At this juncture it is not an objective to construct weights for real decision making. The assessed risk should nevertheless be close to realistic figures to enhance decision makers grasp of what is at stake.

4.2 Problem structuring

For this application, the main lines as prescribed in chapter 2 are followed. Essential parts of the problem structuring were already done through decisions in the HSE integration project as well as through the case study application in 2006.

4.2.1 Restricting the problem

The focus will be on integrating environmental risk related to discharges and emissions that are, or will be, captured by the EIF framework. A cost element will also be included in the consequence table to investigate the expert panel's enthusiasm of making monetary trade-offs, but it will not be included when calculating the EIF_{total}. Restrictions are essentially made within three categories: time, place and considerations, and are a result of discussions in the project team.

1) **Time**

   All potential impacts occurring within five years are included. Potential impacts after this point are disregarded. This is trivial for our case, as all potential impacts where the scores significantly differ have a short and limited time horizon.

2) **Place**

   Only potential impacts as a direct result of the drilling activity and management of drilling waste are assessed.

3) **Considerations**

   The importance of potential impacts should be judged on an environmental scientific basis, not involving political aspects. This is related to an expressed concern among researchers to keep the indicator as scientific as possible as long as possible. Political considerations are still essential for the final decision, but it is assumed that these are integrated at a later stage. This restriction is further discussed in chapter 5.3.4.

4.2.2 Relevant stakeholders

The final decision maker in this case is assumed to be a fictive project organisation guided by Statoil values. Other stakeholders that have an interest in which decision is made may have other values. Which stakeholders are relevant have to be identified so that the corresponding set of values can be mapped.

The public have an interest in keeping the impact from petroleum activities at Norne as low as possible. Through governing authorities, public thresholds and requirements are set to ensure that decisions taken do not imply unsustainable harm to the environment. It is assumed that all initial alternatives that do not comply with these absolute requirements are withdrawn from further processing.

The use of an internal expert panel will result in weights that reflect relative importance of environmental impacts from a corporate point of view. But as long as the analysis sticks to trade-offs between environmental impact factors, and the corporate underlying value of zero harm supposedly is representative for the underlying value of the public as well, there should be no larger divergence between corporate and societal weights.

4.2.3 Relevant objectives

The fundamental objective for the current decision problem is zero harm. As long as activities are run, an absolute zero-tolerance to even the minor risk is not practically feasible. A further split into more operational sub-objectives therefore implies a translation of zero harm into minimising impacts. These impacts could occur in several environmental compartments. The sum of EIFs to each compartment serve as indicators for the degree of achievement of these sub-objectives, as shown in the value tree in figure 4.1.

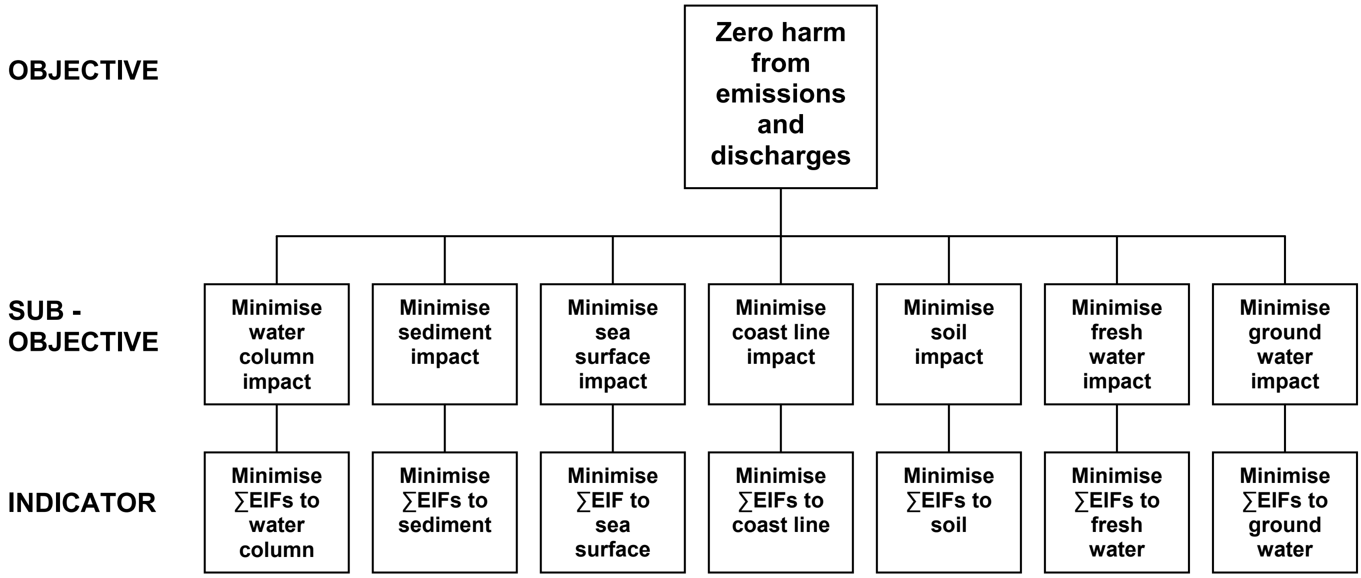


Figure 4.1 Value tree for the current decision problem.

It is important to have in mind that although this is an appropriate hierarchy for the current decision problem, zero harm from emissions and discharges is far from the only corporate value at Statoil. When applying the outcome from this decision problem on later decision stages, objectives like minimising economic costs come into play, cf. chapter 3.1.1. This constitutes a new decision problem with different ends and consequently different means, and additional trade-offs have to be made.

4.2.4 Choice of criteria

Having established the EIFs as indicators for attainment of the overall objective, they naturally constitute the basis for the criteria in our decision problem. However, due to their incomprehensiveness, the EIFs have to be framed so that they are more comparable and take into account other aspects than a mere extension. Several modifications to the original EIF scores are suggested by Smit and Karman (2006) and the project team:

1) **Same threshold levels**

   This is achieved through the msPAF conversion project referred to in chapter 3.2.1, except for EIF air. Although similar threshold levels are not indispensable for further modelling, it will highly facilitate further comparisons. This was one of the main objections to the 2006 application – the incommensurability of the EIFs made the trade-offs too challenging.

2) **Distribution to compartments**

The importance of a potential environmental impact is not dependent on what source leads to the impact, but rather on the sensitivity of the environmental compartment that is affected. If EIF scores are bundled on the compartment level, these could be the point of departure for further treatment. Following the EIFs in table 3.1 and the value tree in figure 4.1, the distribution could be done as shown in figure 4.2.

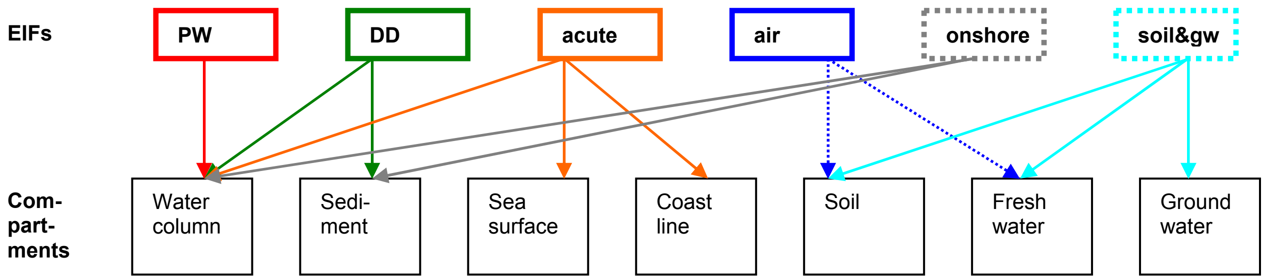


Figure 4.2 *Distribution of EIF values to potentially affected compartments. EIF air is currently not distributed, but arrows show which compartments that could be affected by air emissions. EIF onshore and EIF soil & ground water are not fully developed.*

By aggregating different EIFs, decision makers are assumed to be indifferent to the variable quality of input data in the modelling. The EIF air demands special treatment due to its dissimilar composition, and it cannot be straightforwardly aggregated with the contributions from the other sources to compartment sums. As some EIFs are not yet developed, EIF air is for the moment the only factor contributing to soil and fresh water risk in the Norne case. For the time being, it is therefore chosen to keep the EIF air scores unchanged and non-distributed, serving as an indirect indicator for potential soil and fresh water impact. This approach is however not feasible for a more generic procedure. Further integration of air emissions is discussed in chapter 5.3.2.

3) Integrating impact duration

Even though the EIF scores have been calculated over a one-year-window, the score itself is a pure extension score. However, a potential impact that is expected to last longer represents a higher environmental risk, and this have to be accounted for. Two assumptions are made:

- Risk is increasing linearly with impact duration, i.e. a potential impact with an extent of 1 lasting for 12 months represents a twelve times higher risk than a potential impact with an extent of 1 lasting for 1 month.
- Time and space are factors that are mutually substitutable, i.e. a potential impact with an extent of 1 lasting for 12 months represents the same environmental risk as a potential impact with an extent of 12 lasting for 1 month.

As such, integrating duration is simply done by multiplying the original extent value with the potential impact duration in terms of fraction of year.

4) Comparing regular and acute events

The EIF framework predicts environmental risk stemming from regular emissions and discharges as well as environmental risk related to possible acute events. In the latter cases, extents can be extremely large, but on the other hand, the probability for a hazardous event to occur is often close to zero. One assumption is made:

- Decision makers are indifferent to whether the risk stems from acute or regular events, i.e. a potential impact of 1 from a certain event represents the same environmental risk as a potential impact of 100 from an event with a probability of occurrence of 0.01.

As such, the original extent scores are simply multiplied with the probability that the event leading to this extension will occur. By doing this, scores are brought down to comparable levels and “risk scenarios” combining real decision alternatives with imaginary ones are not necessary.

The assumption might seem strong. One argument for upholding it is nevertheless that the extent scores themselves, irrespective of whether the event is regular or acute, are already complex products of different uncertain dimensions:

- Even if the hazardous event were bound to happen, the actual exposure of toxic and/or stressful components would still be highly variable.
- Even if the exposure were given, there is still uncertainty remaining about whether certain targets would be affected or not.
- Even if they were sure to be affected, the characteristics and severity of the effects would still not be given.

It is essential to have in mind that EIFs at an msPAF level do not apply site specific information; they are generic management tools using toxicity data from lab studies and the precautionary principle to describe risk difference between a set of alternatives. Moreover, with reference to the discussion of bounded rationality in chapter 2.3.3, it could be argued that by converting all values to expected values, a possible source of irrationality when confronted by large numbers is eliminated.

Hence, if only considerations from an environmentally scientific point of view are to be taken into account, there are few overriding reasons for why expected scores should not be appropriate. If including reputation risk however, the importance of a potential impact will depend largely on whether it stems from regular or acute events.

5) **Disregarding severity degree**

The EIF score at the msPAF level of 5 % is given irrespective of whether the fraction is high or low, as long as the threshold is exceeded. Smit and Karman (2006) suggest that including an msPAF degree factor could be a way of accounting for differences in the severity of the effects. The risk endpoints would however be considerably altered, making them even harder to grasp for decision makers that are used to consider the original EIF extent values. Besides, as the principle of continuous improvement is overriding, actions have to be considered as long as the threshold is exceeded, regardless of by how much. It was therefore decided by the project team not to include this factor in the criteria scores.

As a result, our set of criteria will consist of environmental risk endpoints for up to seven different compartments. These endpoints are defined as product of three factors – extent (i.e. the old EIF score with no. of units where msPAF $\geq 5\%$), probability of occurrence (1 for all regular discharges, <1 for all acute discharges) and duration (impact presence measured in fraction of years). Consequently, the score expresses the expected extent in yearly units, subsequently simplified to EIF scores. In formal terms:

$$z_i = score_i = EIF_i \cdot p_{(occurrence)} \cdot dur_{(impact)} \tag{4.1}$$

where

- i is the compartment potentially affected
- EIF_i is the extent score for compartment i
- $p_{(occurrence)}$ is the probability that the event leading to EIF_i will occur
- $dur_{(impact)}$ is the expected duration of the EIF_i measured in fraction of years

As shown in table 4.1, the criteria requirements presented in chapter 2.1.2 seem to be adequately fulfilled:

Criteria requirements	Characteristics for the current case
Complete	The compartments cover an exhaustive list of relevant areas potentially affected. All other elements than environmental aspects related to emissions and discharges are deliberately omitted. For the emissions and discharges, the best available framework with critical modifications is applied.
Operational	All major decisions require a quantitative assessment of relevant environmental risks. When the EIF framework is complete, all necessary scores will be available.
Decomposable	The criteria are put together by numerous sub-criteria that are easy accessible. The score could be decomposed to extent, probability of occurrence and duration, and the extent could be further decomposed if necessary.
Non-redundant	The compartments are not overlapping.
Minimal	The scope has been limited as much as possible.

Table 4.1 Fulfilment of criteria requirements, cf. chapter 2.1.2.

4.2.5 Choice of alternatives

Two main drilling alternatives were identified at the outset of the planning: drilling all four sections with water based fluids (WBM) or drilling the upper two with water based fluids and the lower two with oil based fluids (OBM). For simplicity, these two alternatives will subsequently be referred to as the WBM and OBM alternative respectively. For the OBM alternative, several options for waste handling exist. In 2006, waste handling was left out. For this year's application, a set of alternatives were identified by the project team. These are shown in the decision tree below (fig. 4.3). The alternatives identified do not constitute an exhaustive list of options that would be available in a real decision making situation, but all main alternatives are nevertheless included.

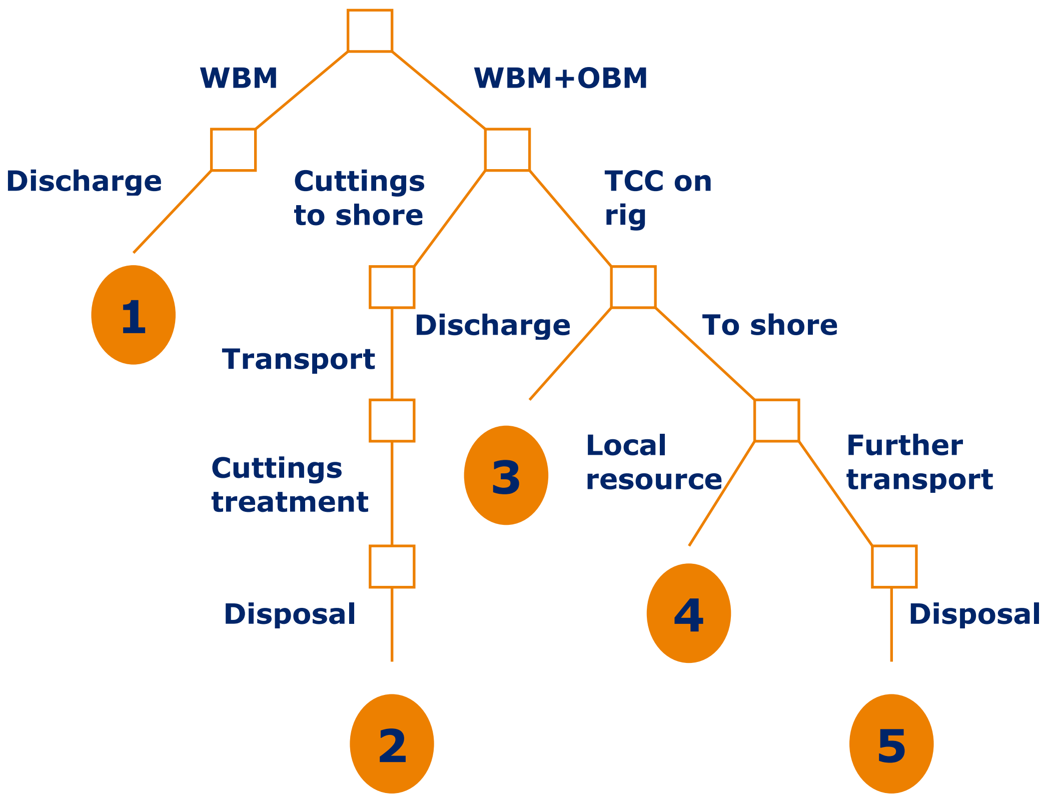


Figure 4.3	*Alternative 1:*	*Water based drilling method with offshore discharges.*
	Alternative 2:	*Oil and water based drilling method with transport of cuttings to shore for further treatment and disposal.*
	Alternative 3:	*Oil and water based drilling method with treatment of cuttings offshore and subsequent discharge of treated drill cuttings.*
	Alternative 4:	*Oil and water based drilling method with treatment of cuttings offshore and subsequent transport of treated drill cuttings to land for local reuse.*
	Alternative 5:	*Oil and water based drilling method with treatment of cuttings offshore and subsequent transport of treated drill cuttings to land for disposal.*

4.2.6 Consequence table

The consequence table in table 4.2 gives the basis for risk treatment. EIF scores not previously calculated for the case were gathered from other applications and adapted to the

Norne case (Rye et al. 2005), (Knudsen et al. 2006), (Smit and Karman 2006), (Arvesen and Pehrson 2003), (Larssen et al. 2005), (Gjerstad et al. 2005), (Rye and Ditlevsen 2005), (Paulsen et al. 2003). For a complete list of EIF calculations and details behind the figures, see appendix C.

Alternatives / Criteria	Units	1 WBM	2 OBM/cut	3 TCC/disch	4 TCC/reuse	5 TCC/disp
Water column	100*100*10m	1542,3	1541,3	1568,5	1541,3	1541,3
Sediment	100*100m	86,0	5,0	5,0	5,0	5,0
Sea surface	100*100m	8,6	8,6	8,6	8,6	8,6
Coast line	100*100m	3,0	3,0	3,0	3,0	3,0
EIF_air	no unit	0,00473	2,15323	0,00475	0,00477	0,94915
Cost	1000 NOK	62350,0	51230,0	52090,0	52330,0	52660,0

Table 4.2 Consequence table for the Norne case. The scores show expected extent in yearly units for the first five years for all environmental criteria on all alternatives. The alternatives correspond to the alternatives outlined in figure 4.3. Costs are simplified investment cost estimates.

Two of the EIF sources – EIF produced water (with potential impact to water column) and EIF acute (with potential impact to water column, sea surface and coast line) are estimated to be similar for all alternatives. As these discharges constitute the most important risks, the total compartment scores are fairly similar across the alternatives. The main differences are due to the following aspects:

- discharge of drill cuttings in alternative 1 (with a higher potential impact to sediment)
- discharge of treated drill cuttings in alternative 3 (with a higher potential impact to water column)
- variable severity of air emissions because of different cuttings treatment methods, transport needs and location of treatment.

The project team discussed whether the risk should be presented in real units, i.e. as m^2 m^3 and not as 100x100m and 100x100x10m. The former would have given the scores a stronger reference to reality, but the figures would on the other hand have been of a size that probably would make them difficult to grasp and subsequently compare. Besides, decision makers are already used to the artificial units through work with the EIFs. It was therefore a unanimous recommendation in the project team to keep the artificial units for further processing.

As shown in the table, EIF air is kept as a separate criterion and not distributed to compartments. Furthermore, an estimate of the main costs related to each alternative has been included, such as purchase of chemicals, rig time and waste handling. For details, see appendix C. The cost figure is by no means exhaustive or accurate, but is included to see if a trade-off assessment between compartment risk and costs is feasible. The higher costs for alternative 1 are mostly related to longer rig time when drilling with WBM compared to OBM.

4.3 Model building

4.3.1 Choice of aggregation method

The HSE integration project specifies that the sub-objective for integration of environmental indices is “to develop a total environmental index integrating and comparing all relevant compartment specific EIFs” (‘Integrated HSE project manual’ 2005:7). As such, the aggregation method applied is not only supposed to identify the best alternative, it is also supposed to quote an alternative’s performance in numerical terms. The application of an optimisation model is therefore already settled, as neither goal aspiration models nor outranking models will result in numerical indicators. The framework for HSE integration (2005) has furthermore specified that the indicator should be a weighted, additive summation of sub-indices, suggesting a multi-attribute utility model to be applied.

Given the context for our decision problem and the further application of the model outcome, this restriction is reasonable:

- Goal aspiration models are not adequate, as they are based on reaching satisfying levels for each criterion. In line with Statoil’s guiding principle of continuous improvement, there are no acceptance criteria in terms of EIFs. As long as the EIF values are above zero, an unacceptable risk is present and measures to reduce it should be considered.

- Outranking models apply thresholds of preference and indifference between alternatives, and do not conflict with the lack of acceptance criteria as the goal aspiration models do. The already challenging complexity of the EIFs requires however that the methodology applied do not add unnecessary intricacy to the conceptualisation process. As such, the mathematical functions of outranking models are not recommendable.

Moreover, since other considerations than environmental performance have to be incorporated at later stages, a mere identification of the optimal alternative is likely to be insufficient. The already quantified EIFs allow furthermore for more sophisticated results to be obtained. Due to the comprehensive assumptions taken for the quantitative variables, the attention and importance given to them could on the other hand be questioned. The figures are adequate for assessing relative differences, but interpreting them to represent absolute risk levels is erroneous.

Again, the need for simplicity and transparency in the EIF integration process is an argument for choosing aggregation methods that are intuitively easy to understand and easy to use. As a result of this, the Analytic Hierarchy Procedure (AHP), with its tedious questions and extensive implementation phase, is considered to be less appropriate than utility functions. The indistinct meaning of AHP importance weights is also a major objection to the method, as one of the main challenges for reaching a valid EIF_{total} is to provide for thorough understanding of what considerations the weights should represent.

4.3.2 MAUT application

Several strong axioms underlie the MAUT model. In addition, there is a need for methodological simplicity, resulting in even more assumptions. Methodological axioms and assumptions for the current application are:

- Less environmental risk is preferred to more

- Preferences related to potential compartment impacts do not change with time
- For all alternatives in question, the effect on total performance stemming from high EIF scores in one compartment can be compensated for by low EIF scores in another, i.e. a compensatory model is appropriate
- The compartment scores are preferentially independent, i.e. the importance (or weight) of a score in one compartment is independent of the scores in the other compartments.
- The total utility can be seen as a sum of partial utility contributions from each criterion, i.e. the partial utility of the score in one compartment is independent of the scores in other compartments. Consequently, an additive utility function is representative for the total performance of an alternative:

$$U(alternative) = \sum_{i=1}^{m} w_i u_i(score_i) \tag{4.2}$$

where

  - i is the compartment potentially affected
  - $score_i$ is the expected extent as defined in (4.1)
  - $u_i(score_i)$ is the partial utility of $score_i$
  - w_i is the relative weight of this score

- The EIFs increase proportionally to potential end impacts, and as it is assumed that partial utility decrease proportionally to potential end impacts, all partial utility functions $u_i(score_i)$ are linear. This assumption is a prerequisite for a future identification of a willingness-to-pay constant for reducing an EIF.

The majority of these assumptions were presented to the expert panel during the 2006 application, and were then judged as acceptable. Please note that the application of a utility function under these assumptions also implies that decision makers are risk neutral, so that the utility of the expected outcome equals the expected utility:

$$U(E[x]) = E[U(x)] \tag{4.3}$$

This is in accordance with the discussion previously in this chapter on the use of expected values.

Before converting EIF scores into partial utility scores on a 0-1 scale, the identification of reference points is required. Claiming that the best achievable EIF score should be set to 0 should not be too controversial, but identifying the worst thinkable EIF score on a general level is on the other hand not an easy task. As a consequence, the project team has suggested to use a global scale for assessing best possible partial utility (as the globally best EIF is 0, an EIF of 0 gives partial utility of 1) whereas a local scale is applied for assessing worst possible partial utility (worst EIF score among alternatives gives partial utility of 0).

This step, along with the assumption of partial linearity, allows for the re-writing of the partial utility function to

$$\frac{score_i^{\max} - score_i}{score_i^{\max}} \tag{4.4}$$

for each criterion and each alternative. Hence, the total utility function for each alternative can be expressed as:

$$U(alternative) = \sum_{i=1}^{m} w_i \frac{score_i^{\max} - score_i}{score_i^{\max}} \tag{4.5}$$

This implies the following characteristics:

- If $score_i = score_i^{max}$, the contribution to total utility for this alternative is 0, irrespective of weighting coefficient.
- If $score_i = 0$, the contribution is initially 1 and subsequently weighted by the weighting coefficient.
- If $0 < score_i < score_i^{max}$, the contribution is initially between 0 and 1 and subsequently weighted.

4.3.3 Weight elicitation

As a rule of thumb, if there are important criteria aspects that are not accounted for in the scores, they have to be taken into consideration when eliciting score weights. Besides, the score range itself has to be reflected in the weights.

Recall that a swing is the difference resulting from a move from what has been defined as the worst score to what has been defined as the best. Table 4.3 shows the swings for each criterion in the Norne case and the partial utility that follows from such a swing. For a criterion representing a harm rather than a benefit, a swing from worst to best is negative and corresponds to a move from maximum to minimum score. This applies to all criteria in our case.

Compartment	Swing	Partial utility
Water column	-1568,5	1
Sediment	-86,0	1
Sea surface	-8,6	1
Coast line	-3,0	1
EIF_air	-2,15323	1
Cost	-62350,0	1

Table 4.3 Criteria swings

The question is then – which swing do decision makers find most beneficial, which do they find second most beneficial etc. Afterwards, the question is to quantify the difference in benefit - what is the benefit of the second most important swing compared to the most important, what is the benefit of the third etc. In order to answer these questions, decision makers should consider the following dimensions:

1) **Range**

   The size of the swing is highly decisive for the answers. In this example, the maximum water column score is considerably higher than the scores in the other compartments, and all other things equal, this should be reflected in the weights.

2) **Extension**

   The units applied are not equal across the criteria. Water column is a volume unit, whereas sediment, sea surface and coast line are areas. The EIF air value is not related to any extension and costs are measured in NOK 1000. Consequently, decision makers have to consider these differences when answering.

3) **Compartment sensitivity**

   Even if range and extension were similar, the importance of the swings would, for the compartment criteria, still differ due to characteristics of the compartment in question and of the area for which the risk is representative. In the EIF air value, background depositions are however already incorporated, and area sensitivity has already been considered.

As a consequence, the weights elicited will always be case-specific (as range will be different in other cases) and area-specific (as compartment sensitivity will be different in other areas).

The procedure described above corresponds to the swing weight method of weight elicitation, which is a widely applied variant of numerical estimation methods. In the 2006 trial session, weights were elicited by pair wise comparison of criteria. Since the expert panel then was reportedly more eager to distribute weights directly rather than searching for indifference combinations, the facilitators decided to keep the main focus on numerical estimation methods. An updated version of the data support tool "Pro&Con" was made available for graphical support and easier calculations (Wenstøp 2007).

4.3.4 Sum-up: Application 2007 vs. 2006

The current design has several similarities to the 2006 trial session. The use of an expert panel and a MAUT utility function are features that are repeated for the 2007 case study. Substantial modifications have however been done to the definitions of the criteria and the alternatives. The most important modifications are listed in table 4.4.

	Case study 2006	**Case study 2007**
Level of EIF expression	Source of emissions	Compartments affected
Tolerance thresholds	Various	msPAF (except EIF_air)
Weighting method	Indifference methods	Numerical estimation methods
Uncertainty	"Certain" scores, risk scenarios	Expected scores, real scenarios
Duration of impact	Excluded	Included
Waste handling	Excluded	Included
Costs	Excluded	Partially included for testing
Weight considerations	Environmental and political	Environmental

Table 4.4 Summary of main changes from 2006 to 2007 application.

5. RESULTS, ANALYSIS AND IMPLICATIONS

The initial problem design as outlined in chapter 4 was presented for an expert panel to elicit weights. It soon became clear that the design had to be modified. An additional expert panel session was therefore set up to complete the weight elicitation. The results from the first session, challenges met there and measures chosen for solving them are explained in chapter 5.1, whereas the actual weighting results from the second session are presented and analysed in chapter 5.2. Chapter 5.3 is describing challenges that still remain and gives an outline of possible solutions for how they could be treated.

5.1 Results and analysis of first application round

5.1.1 Process

The session was held with four expert panel participants, three observers and two facilitators. An important factor when composing the panel was to ensure that different professional backgrounds and viewpoints within the domain of environmental management were present. All members had substantial background knowledge of the EIF framework. Two of the members in the expert panel also participated in the trial session in 2006, and were accordingly more or less familiar with the basic principles of the MCDA methodology and the problem frame for the Norne case.

The panel immediately started to discuss challenges presented by the current problem design, and to what extent these challenges had to be met before the weight elicitation process could continue.

5.1.2 Challenges identified

The most important remarks and challenges discussed by the expert panel at the first session are listed below:

- For the methodology to actually be applied in projects, the weights have to be of a more generic character so that the weight elicitation process does not have to be repeated for each application.

- The dissimilar definition of EIF air makes it very hard to trade it off to the other compartments.
- Several dimensions to consider simultaneously, i.e. range, extension and sensitivity, makes it close to impossible to reach weights to which one feels comfortable.
- The stressing of EIFs as extension scores with msPAF characteristics is troublesome
- The focus on positive utility instead of negative impact is confusing, as the usual approach to EIFs is to think in terms of impact rather than benefit.
- The boundaries for which aspects to include when assessing compartment sensitivity are not clear enough
- A purely verbal presentation of certain sensitivity aspects does not provide sufficient cognitive and emotional background to grasp the necessary characteristics of the areas potentially affected.
- The presence of data that are similar to all alternatives makes it hard to grasp the real differences between them.
- The absence of data, as for terrestrial impacts from other sources than air emissions, gives less confidence to the process.
- Making monetary trade-offs are difficult, in particular when the cost figures are incomplete, and they constitute trade-offs that are outside the mandate of the expert panel.

5.1.3 Meeting challenges

In order to meet these challenges, considerable efforts were made to modify the problem design after the first application round. Possible measures suggested at the meeting and other proposals were object for profound discussions within the project team. Finally, six important measures were chosen:

1) **Changing from alternative based weights to unit based weights**

   The first measure chosen was to elicit weights on the basis of generic units rather than specific alternatives. This technique was also part of the initial design discussions before the first application round, but it was discarded due to expected time limitations. The main differences are as follows:

- With the alternative-based design, importance weights depend on the score ranges. A higher swing is more important to avoid than a lower, all other things equal. The score ranges are case-specific, as swings in other cases will most likely be different. Consequently, the design requires decision makers to elicit case-specific importance weights.
- With the unit-based design, artificial alternatives are created where maximum ranges correspond to one unit for each criterion. As such, all swings are equal and have a score of 1. Only after eliciting weights on this basis, weights are applied to the scores of the real alternatives so that an overall performance may be calculated.

Changing to the latter approach has several implications:

i. The consequence table for weight elicitation is modified to contain as many artificial alternatives as there are criteria. Alternative 1 has a score of 1 on criterion 1, and 0 for all other criteria. Alternative 2 has a score of 1 on criterion 2, and 0 for all other criteria, etc. As such, the swings have a score of 1 for all criteria. Differences in ranges are consequently ruled out, and decision makers have one dimension less to grasp when eliciting weights.

ii. There is no longer a utility function of the form as presented in (4.2), as the importance weights and the actual scores stem from two different sets of alternatives. The importance weights elicited from the artificial set represent the benefit of reducing one unit of each criterion relative to reducing one unit of the other criteria. This is similar to saying that they represent the impact of increasing one unit of each criterion relative to increasing one unit of the other criteria. Once more assuming linearity, these weights can be multiplied with the corresponding scores from the set of real alternatives. The total performance for an alternative, in terms of total impact relative to the other alternatives, is found by adding up weighted scores for all criteria. The formula for this total impact is written as:

$$I(alternative) = w_{i(unit)} z_i \tag{5.1}$$

where

- z_i is the same as before, expected extent in yearly units
- $w_{i(unit)}$ is the importance weight for compartment i, elicited on a unit-basis

As long as both the EIF score and the related importance weight are positive, there will be some contribution to overall impact. This corresponds to a utility less than 1 on the former [0,1] scale, as a utility of 1 was only achieved when the EIF score or the importance weight was zero. There are on the other hand no longer upper limits for how large the impact contribution could be in numerical terms. In the former model, a criterion with maximum score of all alternatives gave zero contribution to overall utility. In the current model, contribution to impact is linearly extended with the EIF score. The performance is consequently no longer on a [0,1] scale, but the relative differences between the alternatives remain the same.

iii. No case specific knowledge is longer requested. This should in principle be trivial as importance of impact figures is supposed to be independent of the pollutive source. A drawback is however that it could be harder for decision makers to become emotionally involved. Another aspect is that compartments' relative importance could differ internally, e.g. if impacts occur in water column closer to shore, water column should be given a higher weight than if impacts occur in water column offshore. When no case specific information is given, this judgement is harder to make. Area specific knowledge is however still needed, as the relative importance of compartments is highly dependent on characteristics of the local environment.

2) Leaving out EIF air

In order to facilitate the time-restricted second trade-off session, and to ensure that importance weights actually were obtained, the project team decided to temporarily leave out impacts from air emissions. There is a strong agreement however that air emissions are important for the overall risk picture, and that a way to include them should be identified. Possible ways of integrating air emissions will be further discussed in chapter 5.3.3.

3) **Splitting weights and assuming extensional indifference**

By changing to unit-based weighting, the range dimension was eliminated. There was however a great concern in the project team that the dimensions left to consider, extension and compartment sensitivity, still would make the trade-off situation too complex to grasp for the expert panel. It was consequently suggested to split the trade-off in two:

- A first step where the EIF scores are weighted for generic compartment differences, such as extension in area/volume, into same-scaled "risk scores"
- A second step where these risk scores are further weighted for differences in sensitivity between the actual compartments.

The content of the first step requires a more profound explanation:

i. *Interpretation of risk scores*

The risk scores resulting from the first step should be interpreted in terms of a preliminary scale where similar scores represent a potential impact of similar importance. *Before* assessing actual compartment sensitivity, similar scores should thus be equally important to reduce for decision makers, irrespective of compartment origin. This is achieved by ruling out differences in generic compartment qualities (e.g. the difference between area and volume units) in the first weighting procedure.

ii. *Interpretation of a generic compartment*

The msPAF scores (applied in all EIFs but EIF air) are calculated on the basis of generic data and are consequently indicators for impact to a generic compartment. The EIF air scores depend certainly on the area where the impact is likely to occur, but an EIF score of 1 could be said to represent generic severity. Hence, if trade-offs between these generic scores are feasible, the first step could be carried out without considering area specific information.

iii. *Interpretation of generic compartment weights*

Technically, what is done here is nothing more than a redefinition of z_i from (4.1), allowing for this two-step procedure:

$$z_i = score_i = w_{i(gen.unit)}(EIF_i \cdot p_{(occurrence)} \cdot dur_{(impact)}) \tag{5.2}$$

Similar to the procedure in (5.1), the generic weight is elicited on a unit basis. The generic importance weight $w_{i(gen.unit)}$ should here be interpreted as the generic importance of decreasing one unit of "expected yearly extent" in compartment *i* relative to decreasing one unit in the other compartments. z_i represents from now on the more abstract risk scores, not EIF scores.

The content of the second step also deserves a clarification:

All modifications due to the characteristics of the actual ecosystems potentially affected belong to the second step. As such, the importance weight $w_{i(unit)}$ should be interpreted as the importance of decreasing the risk score by 1 within each area specific compartment relative to decreasing the risk score by 1 within other area specific compartments.

One possible application of this model would be to let the expert panel elicit weights for both steps. An overall advantage of this approach is that even fewer dimensions have to be considered simultaneously. A possible disadvantage is that it could be hard to distinguish generic compartment aspects from case related aspects.

Assumption for this case: Generic compartment indifference

For the purpose of this study, the project team decided to make two additional assumptions that made the entire first step redundant:

- The potential impact connected to an expected yearly EIF of 1 volume unit is in principle independent of which volume related compartment is affected. Similarly is the potential impact connected to an expected yearly EIF of 1 area unit in principle independent of which area related compartment is affected.
- The potential impact connected to an expected yearly EIF of 1 volume unit (100x100x10 m) is in principle similar to the potential impact connected to an expected yearly EIF of 1 area unit (100x100 m).

These assumptions might seem arbitrary and they could be debated. Still, given the way the risk assessment tools are designed, it could be argued that they are just as reasonable as any other judgements that could be made. The amount and quality of generic information underlying the msPAF calculations differ considerably across compartments, rendering generic compartment weights inaccurate no matter which approach is chosen. Moreover, from a bird's-eye view (reducing volume units from three to two dimensions), the projection of a volume unit and an area unit is the same.

Applied to our model, the assumptions imply that the first step weights, $w_{i(gen.unit)}$, are equal for all i. Instead of normalising all weights to sum up to one, it is convenient to give each weight a value of 1. Hence, the numerical figures representing z_i will be the same as in the first application round. This time they are however representing risk scores, and the $w_{i(unit)}$ have to be elicited according to this.

For future applications however, these assumptions could be abandoned. Then, instead of keeping the extension definitions and weight them differently, one should consider to maintain equal weights and adjust the extent of a unit until indifference is reached.

4) Improving area presentation

The presentation of the potential area of influence was expanded and improved. The area was set to be between Lofoten and Trondheim. For this region, maps of oceanic currents were prepared, along with species distribution maps. These graphics were also made available on paper. The original sensitivity list was rendered more comprehensive to include information on background levels, red list species, presence of corals, spawning ground and other important resources for all compartments. This information was at the same time supposed to give clearer guidelines as to which environmental aspects to take into account in the sensitivity dimension. An excerpt of this information can be found in the appendices D, E and F.

5) **Removing all data related to production**

   As pointed out earlier, scores from real alternatives are irrelevant when eliciting weights on a unit basis. However, the project team judged it as reasonable to allow for a subsequent modification of the weights if the panel would feel uncomfortable with the actual performance scores. As such, there was still a need for removing static data that were similar across alternatives. All data related to production, i.e. all scores for produced water discharges, and all acute scores for year 2-5 were consequently removed. The project team assumed that the EIF acute risk estimated for year 1 was related to the drilling activity.

6) **Removing costs**

   By applying a unit based approach, complete cost data are no longer needed for making monetary trade-offs to other criteria. This does however not mitigate the fact that the panel considered monetary trade-offs to be outside the mandate for their work. The project team decided therefore to disregard all cost figures.

   Leaving out costs in this round do not disqualify from monetary trade-offs with specific compartments at a later stage. By making a trade-off between e.g. water column risk and costs, willingness-to-pay factors for all other compartment risks are implicitly achieved through the already elicited set of compartment weights.

An overview of how the six measures meet the initially listed challenges is given in table 5.1. An updated consequence table is presented in table 5.2, showing that scores are generally lower than in the first round. After the exclusion of air emissions and costs, alternative 2, 4 and 5 have equal scores. The major differences are found for potential impact to sediment (discharging water based cuttings in alt. 1) and water column (discharging TCC treated drill cuttings in alt. 3). Technically, cost does still appear as a criterion, but it is a priori given a weight of zero and the scores could be disregarded.

CHALLENGES (listed in 5.1.2)	MEASURES (listed in 5.1.3)
Generic weights wanted	1) Unit weights
EIF air dissimilarity	2) Leaving out EIF air
Multiple dimensions	1) Unit weights, 3) Risk units
Scores as msPAF	3) Risk units
Utility confusion	1) Unit weights
Sensitivity aspects	4) Area presentation
Incomplete area description	4) Area presentation
Data stiffness	5) Only drilling activity
Incomplete data	1) Unit weights, 6) Removing costs
Cost reluctance	6) Removing costs

Table 5.1 Challenges met in the first application round and measures taken to meet them before the second application round.

Alternatives / Criteria	Units	1 WBM	2 OBM/cut	3 TCC/disch	4 TCC/reuse	5 TCC/disp
Water column	100*100*10m	62,2	61,3	88,4	61,3	61,3
Sediment	100*100m	86	5	5	5	5
Sea surface	100*100m	1,7	1,7	1,7	1,7	1,7
Coast line	100*100m	0,6	0,6	0,6	0,6	0,6
Soil	100*100*10m	0	0	0	0	0
Fresh water	100*100*10m	0	0	0	0	0
Ground water	100*100*10m	0	0	0	0	0
Cost	1000 NOK	62,4	51,2	52,1	52,3	52,7

Table 5.2 Revised consequence table for real alternatives to be used in the second application round.

5.2 Results and analysis of second application round

5.2.1 Process

The second session was held with the same participants as in the first session, except for one member short in the expert panel. The MCDA framework was briefly repeated and the modifications made to the model were presented. After a thorough area sensitivity description and some discussions, the weight elicitation procedure was initiated. A modified version of the data support tool "Pro&Con" was made available, with focus on impact instead of utility and where unit weights were compared to real alternatives. This tool was used to visualise trade-offs and consequences.

The expert panel was first asked to rank the criteria by answering the question "In which compartment is it most important to avoid an impact? Which would be the next compartment, etc?" After the ranking was completed, the panel was asked "How important is it to avoid an impact in the compartment ranked second related to the compartment ranked first (as a percentage)?" These percentages were normalised to sum up to 1 by Pro&Con, and the performance for each real alternative was calculated.

The last part of the session was used to a discussion on how air emissions could possibly be included in the further integration process.

5.2.2 Results

Table 5.3 shows the ranking, the weights as percentage of the highest ranked compartment and the normalised weights. Coast line was clearly regarded as the area where impacts were most important to avoid.

RANKING AND WEIGHTS	Worst	Best	Weight	Norm. weight	Rank
Water column	1	0	40	14 %	3
Sediment	1	0	20	7 %	6
Sea surface	1	0	45	16 %	2
Coast line	1	0	100	36 %	1
Soil	1	0	40	14 %	4
Fresh water	1	0	30	11 %	5
Ground water	1	0	5	2 %	7
Cost	1	0	0		8

Table 5.3 Ranking and weights for the Norne case.

In table 5.4, the normalised weights from table 5.3 are multiplied with the scores in table 5.2 and summarised to the overall performance for each alternative. The performance of the different alternatives is also presented graphically in figure 5.1.

PERFOR-MANCE	1 WBM	2 OBM/cuttings	3 OBM/discharge	4 TCC/reuse	5 TCC/disposal
Water column	24,9	24,5	35,4	24,5	24,5
Sediment	17,2	1,0	1,0	1,0	1,0
Sea surface	0,8	0,8	0,8	0,8	0,8
Coast line	0,6	0,6	0,6	0,6	0,6
Soil	0,0	0,0	0,0	0,0	0,0
Fresh water	0,0	0,0	0,0	0,0	0,0
Ground water	0,0	0,0	0,0	0,0	0,0
Cost	0,0	0,0	0,0	0,0	0,0
RISK INDEX	43,4	26,9	37,7	26,9	26,9

Table 5.4 Performance table for the Norne case

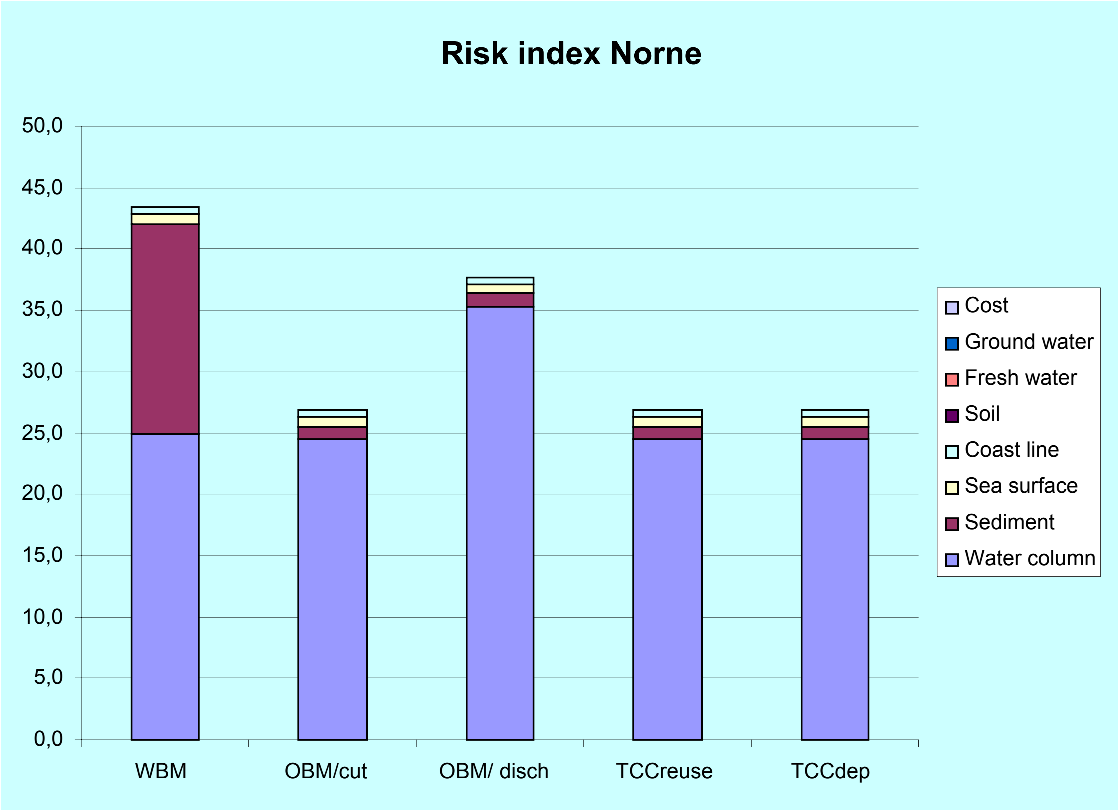


Figure 5.1 Performance chart for the Norne case

Optimal alternatives are number 2, 4 and 5. Since their risk scores are equal, their performance scores are also similar. And as these alternatives score equally or less than alternatives 1 and 3 on each criterion, the latter alternatives are dominated and could never be optimal, irrespective of weights elicited.

It is interesting to see what could have been the result if EIF air scores or costs were included:

- If the EIF air scores could be treated as the other EIFs and distributed to the compartments 'fresh water' and 'soil' with a share of 90% and 10 % respectively, alternative 4 would have been slightly better than alternative 5. The latter alternative would furthermore have been slightly better than alternative 2. As impact scores from air emissions are diminutive for this case, including them would only have altered the performance of the alternatives immaterially. However, the implicit assumption of an EIF air score of 1 representing the same risk as every other EIF scores of 1 could unquestionably be countered. If the former was differently scaled, alterations could have been considerable.

- If the costs figure presented in the consequence table were representative and included in the performance calculation, then alternative 2 would have prevailed among the three originally optimal alternatives, all other things equal. Given the weights elicited, alternative 1 has higher costs and poorer environmental performance than any other alternative, and could never be optimal. Alternative 3 would similarly always perform worse than alternative 2.

5.2.3 Sensitivity analysis

Both theory on weight elicitation and the expert panel sessions show that there is always uncertainty connected to the figures set. For decision makers, it is therefore opportune to have an additional sensitivity analysis performed. This way, it could be identified how much the weights can change before another alternative is preferred. If considerable changes can be made without altering the ranking, decision makers can feel more confident that the actual optimal alternative has been identified. Sensitivity analysis can be performed on both weights and scores:

1) **Importance weight sensitivity**

   For our case, we have already mentioned that alternatives 2, 4 and 5 are dominant and will always be optimal. If the importance of water column were increased from 40 to 62 as a percentage of coast line importance, alternative 1 (WBM) would be preferred to alternative 3 (TCC with discharge of treated drill cuttings off-shore). Altering the importance of the sediment compartment would almost only have an effect on the performance of alternative 1. At an extreme, if sediment importance were set to 0, this alternative would have been very close to the performance of the optimal alternatives. As all alternatives score equally on sea surface, coast line, soil, fresh water and ground water, altering importance weights here will have no effect on the ranking.

2) **Score sensitivity**

   Sensitivity analysis could also be performed on the scores themselves; given the weights, by how much can scores change before another decision alternative is preferred? Obviously, only minor changes among the three optimal alternatives would lead to a split in their performance. If the onshore tools had been available and if any potential impact from disposal of cuttings could have been predicted, then alternatives 1 and 3, which do not imply onshore disposal, would have improved their relative performance.

Since the scores are products of several factors, one could decompose the scores and investigate the sensitivities to changes in the factors instead, e.g. changes in probability of occurrence and duration. For this case, changes in probability of occurrence for the acute discharges will affect all alternatives similarly. In addition, there are some spill scores integrated in the scores stemming from EIF drilling discharges, If these spills were set to be occurring with a probability of 1, alternative 2 (loss of container) and alternatives 4 and 5 (hose rupture) would be slightly affected. Still, due to a very limited duration of these impacts, changes in performance scores would be immaterial. Uncertainty inherent in the extent scores could also be considered.

5.2.4 Validity and reliability

For testing the reliability of the method, the weight elicitation could have been repeated by applying other elicitation methods, e.g. pair wise comparison of weights. Again, the time frame did not allow for this. Besides, chances are high that decision makers would have remembered the figures from the primary elicitation, and consequently allocated indifference points until the weights elicited by the first method were achieved. Comparing to the results of the 2006 study is of no value, as the design of the alternatives and the comprehensiveness of the criteria have been considerably altered since then.

The validity of the process was strengthened by thorough discussions on compartment importance. Both reasoning and emotions seemed to influence the ranking and the quantitative figures, without obviously irrational arguments gaining ground. Nevertheless, it could seem that the panel members tended to attribute greater importance to compartments that were more familiar to them. This is in line with the “focusing illusion” effect outlined in chapter 2.3.3. The panel did not fully agree on the weights set, but was nevertheless able to negotiate until a final set of figures was reached. The similarity of the alternatives made it hard to use performance figures to judge whether the weights elicited were reasonable or not.

One problem reported was that some participants found it hard to grasp the meaning of the new risk scores. This was probably due to the project team's deliberate limitation of willingness to discuss methodological issues as well as a lenient application of the word "risk". In effect, these scores are no more risk scores than the old expected EIF scores, as the latter were also risk indicators. The name should therefore have been chosen with more care; EIF equivalents have been suggested. These EIF equivalents, or risk scores as they will be named through the rest of the report, are risk indicators whose reduction by one is judged to be equally important for decision makers, irrespective of generic compartment origin. If then all actual compartment sensitivities are judged to be equal, the environmentally optimal alternative is the alternative with least risk units. Hence, the only relevant consideration is compartment sensitivity and information uncertainty with respect to compartment sensitivity. Since both of these were covered by the question "in which compartment is it most important to avoid an impact", this confusion did probably not effect the outcome.

5.3 Remaining challenges

Even though substantial modifications were made after the first application round, important aspects without clear-cut answers remain. This section will discuss what the author perceives to be essential challenges and possible measures ahead of further application.

5.3.1 How should time and place be further framed?

In the second session, the expert panel still felt that the decision context related to time and place was insufficiently restricted. Hence, further specifications need to be made, in particular for season of potential impacts, compartment borders and area of validity for weights:

1) **Season of potential impact**

Compartment sensitivity varies considerably according to time of year. Some species are present only for parts of the year and effects to ecosystems are likely to be more serious in breeding/spawning seasons. The variation in sensitivity is rarely the same for all compartments. This can be mitigated in several ways:

i. *Elaborate two sets of weights*

As such, one set for spring/summer and one set for autumn/winter are elicited. If e.g. water column were highly sensitive in the former season and not in the latter, whereas sensitivity for all other compartments were unchanged, this could be accounted for. A challenge is that the EIF indices have to be split between seasons, and this is not trivial (some impacts are continuous and some temporary impacts are cross-seasonal).

ii. *Imagine worst thinkable time of year*

As such, no compartments are split. This procedure is in line with the precautionary principle. A challenge is however that some compartments could turn out disproportionately more important than others, as the sensitivity variance is not equal. This is a general problem with the precautionary principle of using maximum values – it works well when indicators are applied separately, but the actual difference in importance may be distorted when indicators are aggregated.

iii. *Assume time-average sensitivity*

This approach is similar to the approach in ii), except that average sensitivity is assumed instead of maximum sensitivity. A challenge is that the meaning of "average" can be hard to grasp.

2) **Compartment borders**

The last session rendered two challenges with the compartment definitions clear; where do the borders between compartments go (external limits) and should the compartments be further split into sub-compartments (internal limits):

- *External limits*

  The panel wanted clearer definitions of the environments belonging to a compartment. There were inter alia discussions on where the coast line ended and the soil begun. Giving a clear cut compartment definition and assuring that there are no overlap should be a trivial task.

- *Internal limits*

  The question of importance differences *within* compartments was raised, as foreseen in chapter 5.1.3. To mitigate this, relevant compartments can be split into more sensitive and less sensitive ones, e.g. water column-offshore and water column-near shore. As such, they can be weighted differently, and the accuracy of the outcome data is enhanced.

  In the project team discussions before the second round however, the cognitive advantage of keeping the number of compartments down was judged to be higher than the drawback of not being able to differentiate the weights. Splitting compartments also assumes that scores from risk assessment can be distributed according to this division, which could be complicated for certain cases. If compartments are not split, decision makers can follow the mindset from the framing of impact season – imagine the worst possible area or the area most likely to be affected.

3) **Area of validity for weights**

   It was stated during the first round that generic weights for an area were required. It was not stated how large this area should be, or whether the set of weights should be referring to the area of activity (where emissions and discharges take place) or the area of impact (where species are actually affected). The former is easier to apply, whereas the latter is more correct, as the EIFs are impact estimates and not discharge estimates.

   For most discharges, there are insignificant differences between the two approaches. For some emissions however, impacts are more regional than local. The compartments affected could then have a different location than the original source of emissions.

If an affected compartment is assumed to be located in one and only one specific area, the difference is still insignificant. Then it is a simple question of definition: emissions from activities in area A possibly affecting compartment 1 will impact areas solely comprised by area B. Consequently, decision makers should have the characteristics of area B in mind when eliciting weights for compartment 1, and area A in mind for the other compartments. If emissions from A will impact compartment 1 in both A and B, and these areas have significantly different characteristics, the question is however not as trivial.

In the case applied in this study, the Norne field is the base for most of the activities. Geographical location for impacts is said to be the area between Lofoten and Trondheim. This area comprises both offshore and onshore compartments. Following the last paragraphs, there are at least two possible sets of weights with certain characteristics and assumptions:

Alternative 1: Generic weights for activities at Norne (Haltenbanken):

- Valid for all future activities at Norne (Haltenbanken)
- Potentially affected onshore compartments due to activities have to be located geographically
- Once located, activities are assumed not to affect the same compartment in areas where sensitivity characteristics are substantially different

Alternative 2: Generic weights for potential impacts to the area Lofoten – Trondheim:

- Valid for all future potential impacts to the area Lofoten – Trondheim
- Geographic location of both offshore and onshore compartments are given
- Sensitivity for each compartment is assumed to be equal within the region

5.3.2 How could air emissions be integrated in the analysis?

The dissimilarity of EIF air and the omission of greenhouse gases in the EIF framework constitute a considerable challenge. Several alternatives for how air emissions can be integrated are assessed:

1) **EIF air / greenhouse gases as separate criteria along with all other compartments.**
   This is a return to the situation where criteria were defined through different scales, and trade-offs were judged to be hard during the first application round. Besides, decision makers have to make trade-offs between a score in compartments like fresh water and soil (which cannot be excluded from the set due to future EIF soil & ground water contributions) and a score that impact these compartments (i.e. the EIF air unit score). This would probably be close to cognitively impossible.

2) **EIF air / greenhouse gases as separate criteria along with a weighted compartment indicator.**
   This approach calls for a first round integrating compartment scores only, and a second round where EIF air and greenhouse gases are compared to this indicator. According to experiences made, the first round is feasible, but it is probably even more difficult to make trade-offs with an aggregated indicator in a second round than to separate compartments in the first.

3) **EIF air / greenhouse gases converted to risk scores and distributed to compartments.**
   Even though it has been decided that EIF air should not be converted to msPAF units, the new approach of risk scores opens up for other ways of making EIF air comparable to the other EIFs. In brief, the EIF air scores can be directly converted to compartmental risk scores if:
   - The EIF air scores can be distributed to compartments affected
   - The EIF air scores can be rescaled so that one unit here is equally important to reduce as one unit of the expected msPAF-EIFs

Afterwards, all risk scores are weighted to a total risk indicator. From a risk assessment perspective, this approach, as shown in figure 5.2, seems feasible.

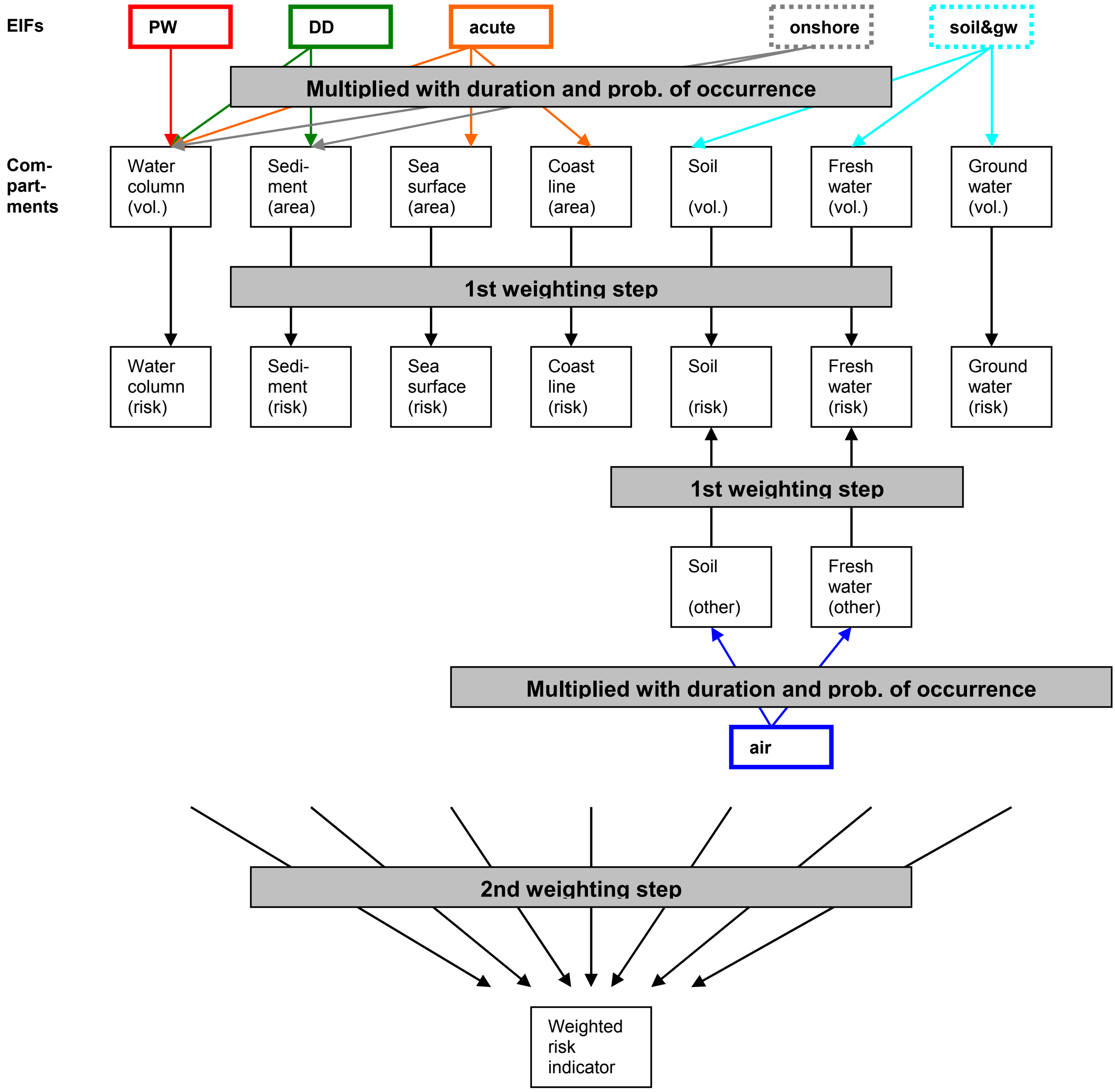


Figure 5.2 Integrating EIF air through risk scores.

The same mapping and rescaling should be feasible for greenhouse gases, but in that case, a “global compartment” has to be made. It is not given that an expert panel would be able (or should be able) to make trade-offs between local/regional compartments and global compartments.

4) **No integration of EIF air / greenhouse gases, separate political judgement by project decision makers.**
As such, there would be no integration problem. The indicator stemming from the other EIFs is presented together with scores indicating potential impacts from air emissions, and it is up to the project organisation how to apply this information. The political sensitivity of CO_2 emissions is an argument for keeping them separate and transparent.

5) **No integration of EIF air / greenhouse gases, no additional judgements made.**
For both EIF air and greenhouse gases, there are costs directly related to emissions through the CO_2-fee (currently NOK 338/ton) and NO_x-fee (currently NOK 15/kg). These costs will be a part of the total cost estimates for each alternative, irrespective of the environmental risk assessment. In 2008, offshore petroleum activities will be a part of the quota trade system for CO_2 emissions, and figures might be altered.

If the fees represent the true willingness-to-pay (WTP) for reducing emissions of these gases, it can be argued that no further impact indicators have to be included for decision making, as costs sooner or later will be considered before an overall optimal alternative is identified. This is in line with how emissions of greenhouse gases are treated today. There are at least three reasons however why this approach is still questionable, in particular for EIF air:

- Not all emissions are included by these fees. EIF air comprises other gases than NO_x and greenhouse gas emissions comprise other gases than CO_2. This could however be compensated for by calculating artificial cost equivalents for the other gases in proportion to their impact potential, and adding these equivalents to the total costs. For greenhouse gases, figures are currently available for relative impact potential.
- In investment decisions, future costs are normally discounted to present value. If air emissions are to be represented by costs, future emissions will implicitly be treated as less severe. As mentioned in chapter 2.3.1, this is

not ethically trivial. On the other hand, it can in fact be argued that cut-backs of emissions of greenhouse gases now are more valuable than cut-backs later, as these gases impact continuously from the day of emission. Furthermore, if cut-backs are to be made later instead, they have to be performed faster and more extensive for the same mitigating effects to occur. Consequently, a discounting cost regime that favours cut-backs today, all other things equal, can be defended. The CO_2 fee must however be representative for the willingness to pay for reduction today, not in the future.

- For EIF air, emission quantities are not representative for potential impact. As demonstrated by the Norne case: If one assume that all impact derives from NO_x, 1000 kg gives an EIF of 0,82 at Mongstad but only an EIF of 0,00001 at Norne. This is related to the different background depositions of stressors.

If the fees do not represent the true willingness-to-pay, the fee could be modified directly to represent true WTP, or alternatives 1)-4) could be considered. Note that no matter which of these alternatives is chosen, an impact indicator for air emissions should actually represent the importance *difference* between importance as already covered by the fees and the perceived real importance. If the impact indicator is nevertheless defined to comprise the entire importance of emissions, it can be argued that fees should be subtracted from the cost calculations in the final identification of the optimal decision alternative, so that there is no double counting of importance.

EIF air and greenhouse gases should perhaps be treated differently. The most relevant approaches are summarised in table 5.5.

GREENHOUSE GASES	EIF AIR
All emissions of other greenhouse gases than CO_2 should be converted to CO_2 equivalents according to their impact potential. **If CO_2 fee is judged as representative for WTP:** Apply CO_2 fee for all CO_2 equivalents in cost calculations for decision making purposes. Adjust discounting factor for these costs if judged as inappropriate. No further concern of emissions of greenhouse gases is necessary, except that limits and requirements are kept (alt. 5). **If CO_2 fee is judged as not representative for WTP:** Decision makers have to include an additional factor that takes the difference into account when identifying optimal decision. This could be done in several ways: • Modify the CO_2 fee directly so that it corresponds to a real WTP, and follow the procedure as if the CO_2 fee were representative (alt. 5). • Present the amount of CO_2 equivalents and let the project organisation assess the risk directly (alt. 4). • Define the unit quantity for one unit of CO_2 equivalents emissions to be equal to one unit of other risk scores. Create a "global compartment" and include this as a separate criterion. Make trade-offs between global compartment and other compartments according to their sensitivities. Consider if actual CO_2 fee should be subtracted from cost calculations for decision making purposes (alt. 3).	With reference to the discussion above, there are too many objections to NO_x fees as representative for EIF air importance for this approach to be recommended. Consequently, a separate indicator has to be calculated. There is a strong willingness to have the EIF air integrated with the other EIFs. For the time being, alternative 3 seems to be the most viable alternative.

Table 5.5 Relevant approaches for integration of air emissions

5.3.3 How could a willingness-to-pay figure for decreasing risk be identified?

A purely environmental indicator was regarded by the panel to be more solid with regards to further applications. Still, at some point, explicit or implicit trade-offs with costs are inevitable. Moreover, if parts of the environmental consequences are to be considered directly through cost figures cf. the discussions in the last chapter, these trade-offs are even needed for identifying what is environmentally optimal.

By including the monetary dimension in the trade-off process, figures for how much Statoil is willing to pay for reducing environmental risk indicators can be calculated. For smaller mitigation projects demanding a less comprehensive list of decision criteria, a set of WTP-figures may be sufficient to see if a measure should be implemented, i.e. investigating if the net environmental benefits of an action exceed action costs.

Irrespective of which decision makers that are finally supposed to make monetary trade-offs, there are several possibilities for how WTP figures can be identified within the current design of the decision problem:

1) **Cost as a criterion**

   An adequate cost unit can be defined and compared to the risk scores in the other compartments through the initial weight elicitation. This requires that the trade-off is made by the environmental expert panel.

2) **Cost compared to one compartment only**

   If a set of compartment weights is already available, it is sufficient that a cost unit is compared to a unit of one compartment only. The weights representing relative difference between compartment units will implicitly give WTP figures for all other criteria. This trade-off itself is not dependent of MCDA modelling. One approach could be to repeat the exercise some years ago when WTP for reduction of EIF produced water by one unit was set to be NOK 200 000; cf. chapter 3.3.2.

3) **Cost compared to the total environmental indicator**

   Instead of comparing cost units to compartments, it can be compared directly to the total indicator so that a WTP for reducing the indicator by one unit could be identified. However, as this indicator is of limited absolute value, the importance of the indicator is hard to grasp and the monetary trade-off has to be repeated for each case.

4) **Using the CO_2 fee**

The CO_2 fee may constitute a bridge between costs and compartment impacts if two conditions are met:

- The CO_2 fee is representative for the WTP of reducing greenhouse gases
- A trade-off between greenhouse gases and other potential impacts can be made

Consequently, we also know the WTP of reducing other potential impacts through compartment weights

In the previous section, it was claimed that if the CO_2 fee actually were representative, no further trade-offs between CO_2 and other impacts were needed. This is still true if the purpose is solely to ensure that all impacts have been included in the decision analysis, but it does not hold if the fee is to be further used as a bridge.

In order to assess the real trade-offs, the indicator for CO_2 emissions should be treated as if no other criteria reflected the importance of these emissions. As in the previous section, the trade-off between greenhouse gases and other impacts should be made through the use of risk scores and a global compartment.

5) **Using the NO_x fee**

As the relationship between NO_x emissions and environmental impact is highly variable, it is not recommended to apply the NO_x fee for establishing WTP figures.

5.3.4 What considerations should be represented by the weights?

In both the first and the second session, the presentation of compartment sensitivity for the actual area kept to purely environmental aspects. No data for human use of resources or other political dimensions were included.

However, possible sensitivity considerations when judging the importance of potential impacts can probably be presented as a continuum between purely environmental aspects and purely political aspects; see fig 5.2.

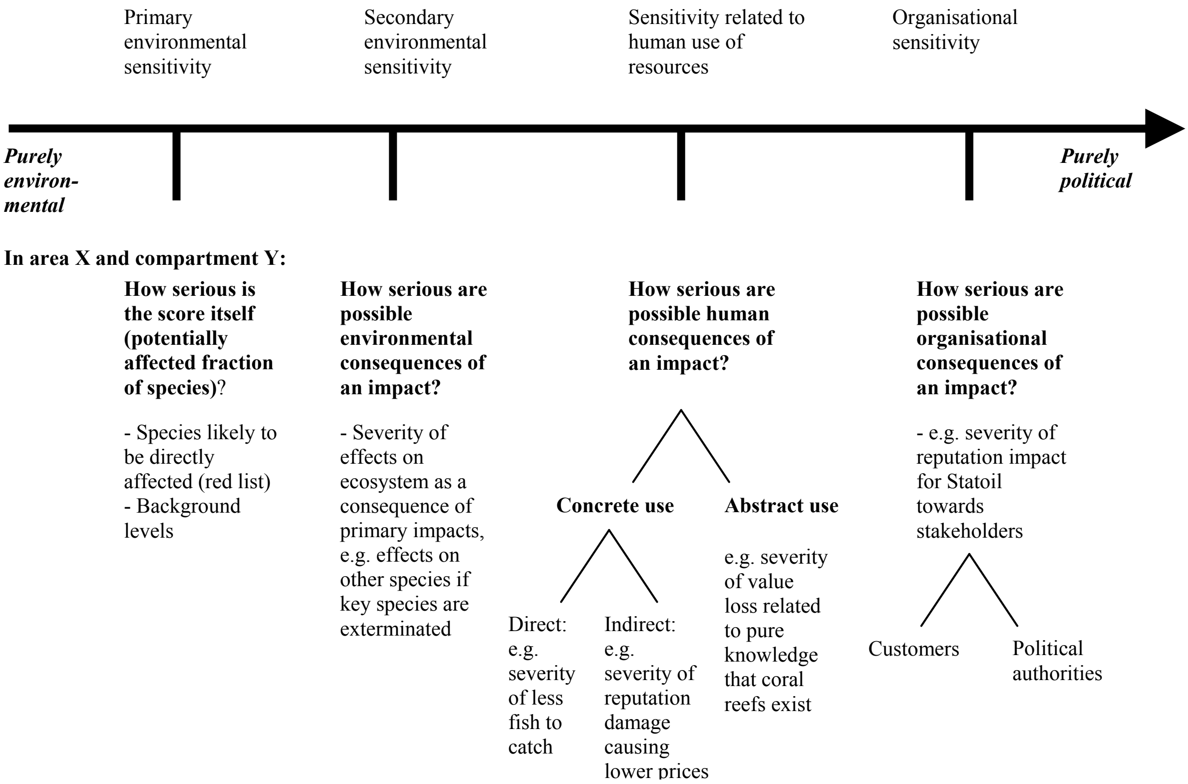


Figure 5.2 Possible sensitivity aspects that can be included when eliciting importance weights for specific compartments in a specific area.

Figure 5.2 underlines three categories of possible sensitivity aspects in addition to the sensitivity of the impact on species as measured directly by the EIF. These categories comprise secondary environmental, social and organisational sensitivity. For the two latter categories, the sensitivity could be related to direct effects of an impact as predicted by the EIF, or it could be more indirectly related through effects of a hazardous event itself. In order to identify the sensitivities, decision makers should reflect on the questions as suggested. Examples of possibly relevant features are listed below the questions.

On one hand, both the project team and the expert panel agreed that more political aspects, as reputation risk, should be taken into account at a later stage. On the other hand, sensitivities of primary and secondary environmental impacts are hard to separate, and the latter considerations were therefore unanimously included. Consequently, the remaining question is: Should considerations related to human use of resources be taken into account when eliciting weights? The expert panel participants disagreed on this issue.

There are several arguments in favour of including sensitivity related to human use of resources:

- The division between impacts on humans and impacts on other environmental elements could be seen as unnatural. Decision makers have a holistic view of a compartment's importance, and would inevitably be biased if they are asked to split this importance. Besides, the aspects describing purely environmental sensitivity are already implicitly coloured by the environment's importance for human beings. Species on the red list will probably not be regarded as equals, - even if all political dimensions were supposed to be disregarded, polar bears would probably still be considered more important than mosquitoes.
- For the time being, there are no alternative frameworks where these considerations can be accounted for. Consequences of activities to the social domain are not part of the integration project.
- Complicated frameworks may be seen as too time-consuming in projects and therefore not applied. Keeping the number of decision elements down is adapted to the way people actually work and might enhance the chances for the tool being used.

There are also several arguments in favour of the contrary; keeping aspects of human use of resources apart from the so-called primary and secondary environmental aspects and rather include them later:

- It is not necessarily easier to distinguish between political sensitivity and human use of resources than between the latter and specific environmental sensitivity. Fisheries in danger would be an aspect concerning both fishermen directly as well as other political stakeholders, and distinctions are vague. Chances that some considerations would be counted for twice (or not counted at all) are hence still present.
- Including human-related sensitivity could render judgements more vulnerable for later questioning and possible discredit. If the panel of environmental experts on the other hand kept to trade-offs made on an agreed environmental basis, the outcome would probably be less disputed.
- The more considerations that are aggregated into one indicator, the more information on how the indicator should be applied is required. Quantitative factors are often convenient for expressing concerns precisely, but they could also be somewhat exposed

to abuse if assumptions and limitations are neglected. It is therefore recommended to keep the indicators as clear cut as possible.

- If secondary effects are to be measured separately later, either through separate indicators or cost figures, the sensitivity related to them is a modifying factor for these figures, not for the EIF-related risk scores. Keeping considerations separate facilitate such an approach.

The core challenge seems to be the following: In an ideal world, weights should fully describe the importance of the scores, and the scores should fully indicate the overall risk. If not all impacts of emissions and discharges are covered by the scores, the dilemma occurs: Should the weights reflect the scores or the overall risk?

In the Norne case, the criteria are simply not complete. The risk scores contain no information at all on the range of other consequences, and applying the weights in order to adjust for these consequences could be misleading. On the other hand, if no other frameworks are available and the other consequences are estimated to be sufficiently correlated to the risk scores, some adjustment is probably better than no adjustment at all. It all depends on the extent to which the risk scores are defined to be indicating more than a potentially affected fraction of species.

5.3.5 What environmental risk should be represented by the EIF indicator?

The dilemma in the previous section requires decision makers to consider the following question:

> *Is the weighted sum of risk scores sufficient for identifying the decision alternative with least overall harm from emissions and discharges?*

Moreover, in chapter 3.2.2, it was pointed out that potential environmental impacts could derive from other sources than emissions and discharges. Neither of these areas is originally comprised by the EIF framework. From the discussions in the previous sections and chapter 3.2.2, a more comprehensive overview over possible environmental risk elements and their frameworks is given in fig. 5.3.

The upper part of the figure illustrates the complexity of risk related to emissions and discharges just reviewed. The remaining figure demonstrates that in addition to risk related to emissions and discharges, there could also be relevant risk factors related to biodiversity factors and use of limited natural resources. If a life cycle perspective is adopted, there could be risks related to activities before and after Statoil's own activities. Some environmental risk could furthermore be "hidden" in cost figures, such as emissions of greenhouse gases if these are accounted for purely by CO_2 fees. Consequently, decision makers should address the following question as well:

> *Is the weighted sum of risk scores sufficient for identifying the decision alternative with least overall harm to environment?*

It is clearly easier to give a positive answer to the first than the second question. Even though other effects than direct impact on species are not mapped, it could be argued that the risk scores are somewhat reasonable indicators of other consequences of emissions and discharges, reputation risk excluded. The inherent conservatism of the EIF calculations makes it less probable that situations where risk scores are low and other potential impacts are high will occur. On the other hand, by not making other potential impacts explicit, they run the danger of falling between chairs in decision making situations.

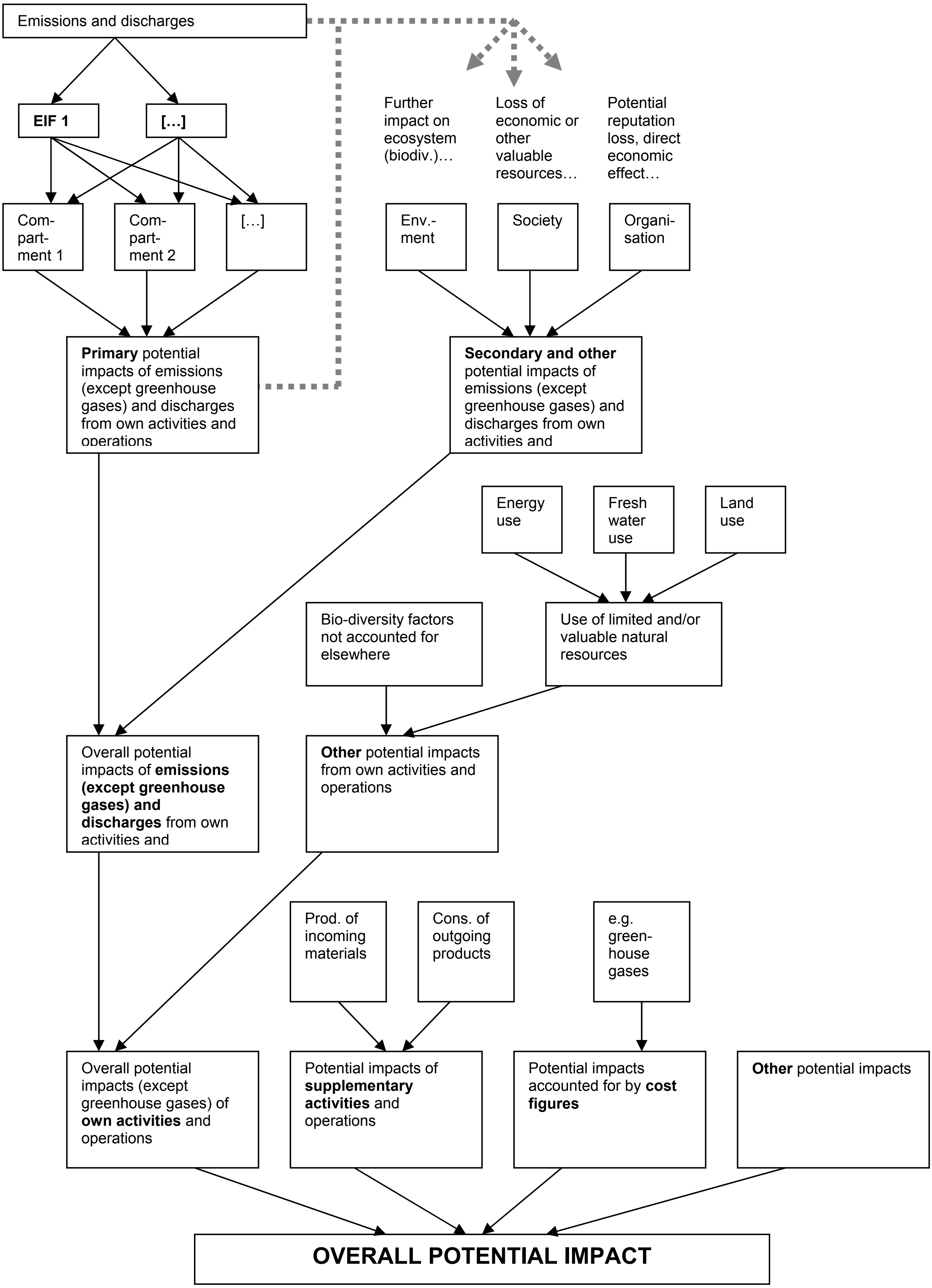


Figure 5.3 Overview of possibly relevant risk elements

6. CONCLUSIONS AND RECOMMENDATIONS

Multiple Criteria Decision Analysis does not help decision makers identify “right” and “wrong” decisions. This is a moral, value-dependent question. Neither does MCDA imply that a particular decision will be made. This is a political question. What MCDA does, is to help decision makers identify “good” and “bad” decisions based on how their expected consequences comply with a set of already given values. When this set contains values that cannot be simultaneously accomplished, judgements of their relative importance have to be made before the expected consequences can be appraised. This judgement is only rational when a combination of cognitive reasoning and emotional appreciation is present. In MCDA, the values are represented by indicating parameters that constitute ‘decision criteria’, and the judgement of their importance is reflected in the ‘importance weights’.

In environmental decision making at Statoil, the EIFs are well established parameters for environmental risk related to emissions and discharges. The underlying objectives for decision makers are to minimise potential impacts in a set of environmental compartments. As not all potential impacts can always be minimised simultaneously, the relative importance of compartments has to be elicited. This study has pursued the presumption from earlier trial sessions that the MCDA methodology and the use of an expert panel are applicable tools for this task.

For such a methodology to be reliable and valid in a multifaceted decision context, it has to make allowances for scientific accuracy as well as practical viability. Through the MCDA application on the case of drilling technology options at the Norne field, substantial modifications to the problem design have accordingly been made. These adjustments are a result of previous experiences, theoretical insights and application challenges; all implemented after comprehensive discussions in the project team. The changes are mostly related to definitions of the decision criteria and the weight elicitation procedure. The main modifications that are currently a part of the problem design are listed in table 6.1.

	Criteria definitions	Weight elicitation procedures
Modifications	• All EIFs except EIF air are distributed to potentially affected compartments • All EIFs except EIF air are expressed at the msPAF level • Probability of event occurrence and expected duration of potential impact are included by multiplication with the EIF score • All expected, yearly EIF scores are converted to "risk scores", which are for the decision maker equally important to reduce at a generic unit level • Air emissions and costs completely excluded from the criteria set	• Numerical estimation methods (swing weights) are applied instead of indifference methods • Only scientific considerations with respect to compartment sensitivity are included • Weights are elicited on unit scores rather than scores of actual alternatives, and performance is calculated by multiplying weights and actual risk score • Two-step weighting process: First to a risk score, then to a total indicator. For the time being: First step left out by assuming indifference between generic compartments.

Table 6.1 Summary of main modifications made

The overall results of the weighting sessions indicate that the modifications made have increased the feasibility of the approach. The similar indicator scores in the Norne case show however that the current design is still incomplete. Whether and when an adequate level of feasibility is reached for real-life applications depend on two aspects:

- the reasonability of the assumptions made
- a clarification of how remaining challenges should be met

For the former, all assumptions are judged to by acceptable for managerial purposes by the project team. The most delicate assumption is related to extensional indifference, but could be mitigated by adjusting unit sizes so that the first weighting step can still be omitted.

For the latter aspect, the following additional modifications seem recommendable from the author's point of view:

- Compartment limits have to be explicitly set and explained to the expert panel. The set of weights should be valid for an area of potential impact, and this area should be assessed and explicitly limited as part of the risk assessment. Decision makers should furthermore be asked to imagine the compartment to be affected at the worst time of year, assuming that differences in maximum-mean distances are negligible across compartments.

- Emissions currently assessed by EIF air should be left to the risk assessment process for distribution to compartments and further rescaling to a risk score/EIF equivalent. If trade-offs to a global compartment is feasible, greenhouse gases should be included via the risk score approach as well. If such trade-offs are judged to be hard, it is suggested to let greenhouse gases be taken into account through cost equivalents:
    - CO_2 fee for CO_2 emissions
    - CO_2 fee equivalents for other greenhouse gases
    - Adjustments of CO_2 fee if not representative for importance/willingness-to-pay
    - Adjustments to compensate for Net Present Value calculations if discounting future potential impacts is considered inappropriate

- If it is still a concern that costs should be kept apart from the weighting of environmental risk scores, the easiest accessible estimation of a WTP figure is made between reduction of a risk score in the best known compartment and costs cf. earlier assessments. If trade-offs to a global compartment is feasible however, the CO_2 fee or CO_2 fee equivalent could be applied.

- When eliciting weights, considerations of compartment sensitivity should in principle be kept within the domain of environmental science. If no alternative indicators of other effects of emissions and discharges will be made, including sensitivity considerations related to human use of resources should be considered.

For future applications, it is furthermore recommended that a more comprehensive case than choice of drilling technology at Norne is elaborated. There should inter alia be identified a set of alternatives where none of them are clearly dominant. This is assumed to better elucidate the consequences of the methodology and thus improve decision maker's judgements of its underlying feasibility.

The actual application of the indicator will be highly dependent on what decision makers find most practical and workable given the often idiosyncratic circumstances. Some general viewpoints can nevertheless be stated:

- *Only apply integrated EIFs when necessary*
  When decision alternatives have no larger divergence for what compartments are potentially affected, EIFs applied separately are sufficient for identifying the preferred solution with respect to emissions and discharges.

- *Apply the quantitative outcome with caution*
  The quantitative outcome of the MCDA approach is bounded by a set of assumptions, and must be interpreted accordingly. If the chances for such restrictions to be inadvertently lost in real applications are present, the outcome should perhaps be presented differently. One way could be to only bring forward the results verbally.

- *Consider other environmental aspects when relevant*
  The EIFs could arguably be regarded as representative for all relevant risk related to emissions and discharges. They can not be regarded as representative for environmental risk as a whole. A better link to targets related to biodiversity and possibly use of resources should be established.

- *Consider thoroughly what environmental risk to include in an HSE-indicator*
  If an EIF_{total} is the only environmental element in an integrated HSE-indicator, it is implicitly claimed that other environmental risk factors are irrelevant. Irrespective of the total amount of criteria however, a reliable and valid trade-off process between H, S and E indicators is likely to be very challenging. It is also important to underline here that the EIF_{total} indicator is expressing an alternative's performance relative to other alternatives; the numerical figure cannot straightforwardly be assumed to represent a more absolute risk level.

One apparent link for further HSE comparisons could lie in willingness-to-pay figures and the use of cost equivalents. Another possibility is to make trade-offs between one chosen criteria in H, S and E respectively. If internal sets of weights are already established *within* each area, willingness-to-pay figures in terms of units of these criteria could be elicited for all other criteria. A third option is to create categories for H, S and E and perform a two-step weighting: within a category (scientific judgements) and between categories (political judgements). Further investigation is needed to evaluate the feasibility of these approaches.

REFERENCES

Aksnes, Norvald (2007): *Making Better Decisions: Use of Multi Criteria Decision Making to rank alternative development concepts for an offshore field.* Statoil power point presentation.

Andrew, Clinton J. and Govil, Shivani (1995): *Becoming proactive about environmental risks. Regulatory reform and risk management in the US electricity sector.* Energy Policy, vol. 23:10, pp. 885-892.

Arvesen T. and Pehrson I. (2003): *Tiltakskartlegging av Nornefeltet. Strategi for nullutslipp til sjø fra Norne.* Statoil report no. M-NOR-03Z21****5813.

Belton, V., Ackermann, F., Shepherd, I. (1997): *Integrated Support from Problem Structuring through to Alternative Evaluation using COPE and VISA.* Journal of Multi-Criteria Decision Analysis, vol. 6, pp. 115-130.

Belton, Valerie and Stewart, Theodor J. (2002): *Multiple Criteria Decision Analysis. An integrated approach.* Kluwer, Boston/Dordrecht/London.

Brown, J.S. and Duguid, P. (1991): *Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation.* Organization Science, vol. 2:1, pp. 40-57.

Choo, Eng U., Schoner, Bertram and Wedley, William C. (1999): *Interpretation of criteria weights in multicriteria decision making.* Computers & Industrial Engineering, vol. 37, pp. 527-541.

Damasio, Antonio (1994): *Descartes' error. Emotion, Reason and the Human Brain.* Penguin Books Ltd, London, UK.

Fakta – norsk petroleumsverksemd (2006). Ed.: Dokka, Ane and Midttun, Øyvind. Ministry of Petroleum and Energy, Norwegian Petroleum Directorate.

Furuholt, Edgar (2007): *Personal communication 02/09/2007.*

Georgopoulou, E., Sarafidis, Y., Mirasgedis, S., Zaimi, S. and Lalas, D.P. (2003): *A multiple criteria decision-aid approach in defining national priorities for greenhouse gases emissions reduction in the energy sector.* European Journal of Operational Research, no. 146, pp. 199-215

Gjerstad, K.I., Knudsen, S., Solberg, S., de Wit, H., Høgåsen, T., Larssen, T., Aarrestad, P.A. and Stabbetorp, O.E. (2005): *Konsekvenser ved utslipp til luft fra et Energiverk ved Statoil Mongstad.* NILU report no. OR27/2005, Kjeller, Norway.

Goedkoop, Mark and Spriensma, Renilde (2001): *The Eco-indicator 99. A damage oriented method for Life Cycle Impact Assessment. Methodology Report.* www.pre.nl

Greening, L.A. and Bernow S. (2004): *Design of coordinated energy and environmental policies: use of multi-criteria decision-making.* Energy Policy, vol. 32, pp. 721-735.

Gulbrandsøy, Karen Sofie and Solberg, Marie-Louise Dale (2006): *Environmental Impact Factors and Key Performance Indicators as Management Tools in Statoil.* Master's thesis, NTNU, Norway

Janssen, Ron (1992): *Multiobjective Decision Support for Environmental Management.* Kluwer, Dordrecht/Boston/London.

Janssen, Ron (2001): *On the Use of Multi-Criteria Analysis in Environmental Impact Assessment in The Netherlands.* Journal of Multi-Criteria Decision Analysis, vol. 10, pp. 101-109.

Kahneman, D. and Sugden, R. (2005): *Experienced Utility as a Standard of Policy Evaluation.* Environmental & Resource Economics, vol. 32, pp.161-181.

Keeney, Ralph L. (1992): *Value-Focused Thinking. A path to Creative Decision-Making.* Harvard University Press, Cambridge, Massachusetts, US.

Keeney, Ralph L. and Raiffa, Howard (1976): *Decisions with Multiple Objectives*. Cambridge University Press, Cambridge, UK.

Kinsella, K. (2006): *Early Phase Risk & Opportunities Assessment (EPRA) Tool. Phase I – Tool Specification Report (Draft)*. Environmental Resources Management's London office.

Knudsen, S., McInnes, H., Larssen, T. and Høgåsen, T. (2006): *EIF-air. Drilling, production and transport Norne*. NILU report no. OR ../2006, Kjeller, Norway.

Lahdelma, R., Salminen, P. and Hokkanen, J. (2000): *Using Multicriteria Methods in Environmental Planning and Management*. Environmental Management, vol. 26:6, pp. 595-605.

Larssen, T., Knudsen, S., Bruteig, I., Aarrestad, P.A., Engen, S., Kinn, S., Johnsen, S. (2005): *EIF-Air. Environmental Impact Factor for assessment of emissions to air. Summary report*. NIVA report no. SNO 5098-2005, Oslo, Norway.

Linkov, I., Varghese, A., Jamil, S., Seager, T.P., Kiker, G. and Bridges, T. (2004): *Multi-Criteria Decision Analysis: A Framework for Structuring Remedial Decisions at Contaminated Sites*. Comparative Risk Assessment and Environmental Decision Making, eds. Linkov I. and Ramadan, A., pp. 15-54, Kluwer, Dordrecht, The Netherlands.

Mustajoki, J., Hämäläinen, Raimo P., Marttunen, M. (2004): *Participatory multicriteria decision analysis with Web-HIPRE: a case of lake regulation policy*. Environmental Modelling & Software, vol. 19, pp. 537-547.

Paulsen, J.E., Omland, T.H., Igeltjørn, H. Aas, N. and Solvang, S.Aa. (2003): *SPE 85296: Drill Cuttings Disposal, Balancing Zero Discharge and Use of Best Available Technique*. Society of Petroleum Engineers, Richardson, US.

Pöyhönen, M. and Hämäläinen, R.P. (2001): *On the convergence of multiattribute weighting methods*. European Journal of Operational Research, vol. 129:3, pp. 569-585.

Roberts, Ron and Goodwin, Paul (2002): *Weight Approximations in Multi-attribute Decision Models*. Journal of Multi-Criteria Decision Analysis, vol.11, pp.291-303.

Rye, H., Ditlevsen, M.K. (2005): *Discharges of drill cuttings and mud at Kristin field.* SINTEF report no. STF80MK F05084, Trondheim, Norway.

Rye, H., Øistein, J. and Ditlevsen, M.K. (2005): *EIF calculations for regular and accidental releases at Norne*. SINTEF report no. STF80MK F05365, Trondheim, Norway.

Schoemaker, Paul J.H. and Waid, C. Carter (1982): *An Experimental Comparison of Different Approaches to Determining Weights in Additive Utility Models.* Management Science, vol. 28:2, pp. 182-196.

Seip, Knut Lehre and Wenstøp, Fred (2006): *A Primer on Environmental Decision-Making. An Integrative Quantitative Approach.* Springer, Dordrecht, The Netherlands.

Simon, H.A. (1976): *Administrative Behaviour: A study of Decision-making Processes in Administrative Organization.* Harper and Row, New York, US.

Simpson, L. (1996): *Do Decision Makers Know What They Prefer?: MAVT and ELECTRE II.* Journal of the Operational Research Society, vol. 47, pp. 919-929.

Slovic, P., Finucane, M.L., Peters, E., and MacGregor, D.G. (2004): *Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality*. Risk Analysis, vol. 24:2, pp. 311-322.

Smit, M.G.D., Karman, C.C. (2006): *Comparison of EIFs based on dose-response modelling*. TNO-report 2006-DH-R0191, Apeldorn, The Netherlands.

Statoil document (2007): *The Statoil Book*.

Statoil document (2006): *Mastering challenges – Statoil and sustainable development.*

Statoil document (2007): *Governing document WR 2266 – HSE Risk management.*

Statoil document (2006): *Governing document WR 1912 – HSE managements in projects.*

Statoil document (2005): *Project Manual: Integrated HSE risk management.*

Statoil document (2005): *Framework for integration of HSE risk assessment: cause-effect approach (draft).*

Statoil document (2007): *Minutes, EIF integration meeting 02/19/2007*

Statoil homepage, <http://www.statoil.com/hse>, last checked 06/13/2007.

Vincke, P. (1992): *Multicriteria decision-aid.* Wiley, Chichester, UK.

Van Straalen, N.M. and Denneman, C.A.J. (1989): Ecotoxicological evaluation of soil quality *criteria*. Ecotoxicology and Environmental Safety, vol. 18, pp. 241-251.

Von Winterfeldt, Detlof and Edwards, Ward (1986): *Decision Analysis and Behavioral Research*. Cambridge University Press, New York, US.

Wenstøp, Fred (2006): *Application of Multi Criteria Decision Analysis to Compare Environmental Impact Factors*.

Wenstøp, Fred (2007): *Pro&Con software*.

Wenstøp, Fred (2005): *Rejoinder: Some Viewpoints on Emotion, Consequentialism and Multicriteria Decision Making*. Journal of Multi-Criteria Analysis, vol. 13, pp. 191-198.

Wenstøp, Fred and Seip, Knut (2001): *Legitimacy and Quality of Multi-Criteria Environmental Policy Analysis: A Meta Analysis of Five MCE Studies in Norway*. Journal of Multi-Criteria Decision Analysis, vol.10, pp. 53-64.

APPENDIX

A. Statoil values

Abstract from 'The Statoil Book' (2007):

We in Statoil

Imaginative

- Be curious and stimulate new ideas and creativity
- Chase business opportunities
- Have the courage to challenge accepted truths and enter unfamiliar territory
- Understand and manage risk

Hands-on

- Set ambitious targets and deliver on promises
- Act decisively and be loyal to decisions
- Work in teams, break down barriers and make constructive demands on each other
- Show endurance, follow through and pay attention to important details

Professional

- Always seek the most appropriate solutions and share experience
- Continuously develop sound expertise, demonstrate commercial awareness and customer orientation
- Strive for simplification and clarity, fight activities which do not add value
- Promote diversity

Truthful

- Communicate in an open and precise way
- Listen actively, give and accept constructive feedback
- Bring up ethical issues and challenges immediately

Caring

- Cause zero harm to people and the environment
- Contribute to sustainable development and accept social responsibility
- Respect the individual and continuously improve the working environment

Health, Safety, Security and the Environment

Our goal is zero harm.

All accidents can be prevented.

A high standard for health, safety, security, and the environment – HSE – has a value for itself.

- We integrate HSE in all business activities to create safe and healthy workplaces
- We conduct our business in accordance with our ethical principles
- We select suppliers based on commitment and performance
- We engage with stakeholders and communicate our ambitions and performance
- We are committed to reducing the negative impact of our activities and products on health and the environment
- We evaluate and improve our performance continuously

You and I have a common responsibility to care for each other and the environment.

B. Presentation of the different EIFs

Abstract from Smit and Karman (2006):

2.2 Evaluation of EIFs

2.2.1 EIF produced water

The EIF for produced water discharges (EIF_PW) is an indicator for the volume of water around a discharge point where adverse effects to more than 5% of the species might occur as a result of the discharge (Smit *et al.*, 2003, Johnsen *et al.*, 2000). It only assesses risk related to exposure of toxicants (natural components and added chemicals) after discharging produced water to the water column. Figure 3 in Appendix A shows the cause-effect chain for the EIF_PW.

The risk element in the EIF_PW is a result of a comparison between a predicted environmental concentration (PEC) and a predicted no-effect concentration (PNEC). Species sensitivity distributions (SSDs) are used to assess the predicted affected fraction of species (PAF) for each component group. The PAF can be considered as a risk probability, indicating the likelihood that species are affected by a specific stressor. The risk probabilities from different stressors can be combined assuming independent action. The combination of risk probabilities (PAF levels) is referred to as the multi substance PAF level (msPAF). Each water volume where the msPAF exceeds the value of 5% contributes to the value of the EIF_PW. Table 2 provides an overview of the risk elements incorporated in the EIF_PW.

Table 2 Risk elements of the EIF for produced water

Risk element	In the EIF produced water
Probability	Equals 1, emission and exposure takes place
Target	Marine species (generic)
Effect type	Adverse (toxic) effects (not specified)
Severity	Percentage of species being adversely affected (msPAF)
Extent	Volume (number of 100m x 100m x 10m grid cells)
Duration	Continuous exposure expressed as maximum value that might occur during each year of discharge, representing a worst case.

2.2.2 EIF drilling discharges

The EIF for drilling discharges (EIF_DD) consists of two different parts. One EIF for the water column (similar to the EIF for produced water) and one EIF value for the sediment compartment. The EIF for the sediment compartment is an indicator for the sediment area around a platform where adverse effects to more than 5% of the species might occur as a result of the discharge of drilling mud and cuttings (Smit *et al.*, 2006a). It assesses risk from exposure to toxicants but also from exposure to non-toxic stressors like burial, change in grain size and oxygen depletion. Figure 4 in Appendix A shows the cause-effect chain for the sediment part of the EIF_DD.

The risk element in the EIF_DD is a result of a comparison between a predicted environmental concentration (for toxic stress) or a predicted stress level (for non toxic stress) and a predicted no-effect concentration (for toxic stress) and a predicted no-effect level (for non-toxic stress). Species sensitivity distributions (SSDs) are applied to assess the predicted affected fraction of species (PAF) for each stressor. The PAF levels for each stressor are combined into a multi substance PAF level (msPAF). Each sediment area where the msPAF exceeds the value of 5% contributes to the value of the EIF_DD.

Table 3 Risk elements of the sediment part of the EIF for drilling discharges. The water column part is equal to the EIF for produced water (except for duration)

Risk element	In the EIF drilling discharges (sediment)
Probability	Equals 1, emission and exposure takes place
Target	Marine species (generic)
Effect type	Adverse (toxic) effects (not specified)
Severity	Percentage of species being adversely affected (msPAF)
Extent	Area (number of 100m x 100m grid cells)
Duration	Sediment: intermitted releases, but long term exposure, expressed as maximum value that might occur after a discharge. Water: Discharge & exposure are time limited (acute; hours/days/weeks per year).

2.2.3 EIF air emissions

The EIF for air emissions (EIF_air) is an indicator for the potential impact on habitats and vegetation of deposition from gaseous emissions from the activity, weighted by potential impact caused by deposition from emissions from other sources (Larssen *et al.*, 2005). It is based upon the critical loads concept, in which the environmental impact of air emission is expressed as the ratio of exposure (atmospheric deposition) and ecosystem sensitivity (critical load). As such, the approach is comparable to the PEC:PNEC approach which is the basis for the EIF produced water and EIF drilling discharges. Figure 5 in Appendix A shows the cause-effect chains for the EIF_air.

Emissions currently taken into account in the EIF Air calculation procedures are:
– Deposition of nitrogen (and sulphur where applicable)
– Acidification of nitrogen deposition and soils
– Impacts of nitrogen deposition on vegetation (eutrophication)
– Production/destruction of ozone as a result of NOx and VOC emissions

Dispersion of the emission of nitrogen (compounds) from the source is simulated with the INPUFF trajectory model to estimate the accumulated annual nitrogen deposition over a geographic grid. The concentrations of ozone are calculated with a dedicated photochemical computer model (EPISODE), expressed as long term average concentration or accumulated hourly values.

In the ICP Mapping Manual 2004 (source: www.icpmapping.org) the critical load is defined as '*a quantitative estimate of an exposure to one or more pollutants below which significant harmful effects on specified sensitive elements of the environment do not occur according to present knowledge.'* The definition was even further specified in the manual for the critical loads for acidification and eutrophication and the critical level for ozone respectively:
– Acidification
'the highest deposition of acidifying compounds that will not cause chemical

changes leading to long-term harmful effects on ecosystem structure and function'

– Eutrophication
'the highest deposition of nitrogen as NHx and/or NOy below which harmful effects in ecosystem structure and function do not occur according to present knowledge'
– Ozone
'the atmospheric concentration of ozone in the atmosphere above which adverse effects on receptors (…) may occur according to present knowledge'

The calculation of the absolute figure of critical loads and levels is extensively described in the ICP Mapping Manual. The basic understanding is that critical loads can either be empirically derived or derived using mass balance modelling:

– *Empirically derived critical loads* for eutrophication are based on ecosystem classification, where each EUNIS ecosystem type is assigned a critical load based on literature data or expert judgement. *Empirically derived critical loads* for acidification assign an acidity critical load to soils based on soil mineralogy and/or chemistry.
– Steady-state (SMB: Simple Mass Balance) or dynamic models can be used with varying degree of complexity. The SMB is based on all possible nitrogen fluxes in order to calculate the nitrogen mass balance, arriving at a critical load by defining the absolute value of a critical flux (mostly N-leaching).

Using the methods indicated above on a spatial grid (including geographic information on ecosystem composition and soil properties) yields a spatial map of critical loads.

The risk element in the EIF_air is a risk characterisation ratio (RCR) between a predicted load related to the activity and a critical load estimated from sensitivity data for specific environments. For each unit of area (grid cell) this RCR is weighted by the ratio of the (background) deposition from other activities and the same critical load. The value of each grid cell is then multiplied by its surface and the resulting value is added up for all grid cells to arrive at the value of the EIF (Larssen *et al.*, 2005).

Table 4 Risk elements of the EIF for air emissions (VOC that leads to tropospheric ozone is excluded, as it was not taken into account in the case study calculations).

Risk element	In the EIF air emissions
Probability	Equals 1, emission and exposure takes place
Target	Surface water (S+N), forest soil (S+N) and terrestrial vegetation (N)
Effect type	Eutrophication and acidification effects (not specified)
Severity	Exceeding of a threshold (critical load) yes or no
Extent	Area
Duration	Duration of exposure is not specified but expressed as yearly deposition

2.2.4 EIF acute

The EIF_acute combines risk probabilities for oil contamination on three different compartments; water column, sea surface and the coastline. The risk can be expressed on three different levels (level I to level III) (Nilssen *et al.*, 2005, Østby *et al.*, 2003). The definition of EIF_acute level I is: the total area where a generic threshold for oil contamination is exceeded with a likelihood of 5%. In level II and III the risk probability is not related to generic species but to specified representative species or resources. The definition of the EIF_acute level II is: the total area where the combined probability of specific resources being present and exposure exceeding the threshold for these resources exceeds 5%. In the EIF_acute level III the effect part of the risk is even more specified. EIF_acute level III is defined as: the total likelihood and extent of mortality (reduction of resources) on a specified population or habitat as a result of the exposure. Figure 6 in Appendix A shows the cause-effect chains for the different levels of the EIF_acute.

The risk elements in the different levels of the EIF_acute are different. At level I risk is expressed as a risk characterisation ratio (RCR) between a predicted concentration (PEC) and a generic threshold (PNEC). At level II risk is expressed as a RCR for a specific species. At level III not only the likelihood of effects occurring are assessed but also the also the extent of the effects is quantified.

Due to variations in weather conditions and current directions the exact trajectory of oil spills cannot be predicted on forehand. Therefore the exposure resulting from a spill scenario is assessed by multiple runs. This results in a likelihood distribution of exposures per grid cell. This probability of occurrence is also a part of the different EIF_acute levels.

Table 5 Risk elements of the EIF acute.

Risk element	In the EIF acute
Probability	Between 0 and 1 (probability of occurrence and probability of exposure)
Target	Level I: Marine biota / birds and mammals / vulnerable resources Level II: representative species or specific vulnerable resources
Effect type	Level I: adverse effects Level II: Specified effects (reduction of species or resources)
Severity	Level I and II: exceeding of a threshold (yes or no) Level III: percentage of effect
Extent	Area (km^2)
Duration	Not specified, depending on the residence time of the oil. Estimated respectively weeks/months for sea surface and water column and months/years for coastline.

C. EIF and cost calculations

1) WBM and discharge offshore

	Occurence	Extent	Prob.	Dur.	Total	∑5Y	Values derived from
EIF_DD-water	only first year	48	1	0,0192	0,9	0,9	SINTEF-rapport (2005)
EIF_DD-water (spill)	only first year	0	0	0	0,0	0,0	
EIF_DD-sea surface (spill)	only first year	0	0	0	0,0	0,0	
EIF_DD-sed	decreasing over lifetime	57	1	1	57,0	86,0	SINTEF-rapport (2005)
EIF_DD-sed (spill)	only first year	0	0	0	0,0	0,0	
EIF_air (transport)	only first year	0	1	1	0,00000	0,00000	
EIF_air (drilling)	only first year	0,00073	1	1	0,00073	0,00073	NILU-rapport (2006)
EIF_air (cuttings treatment)	only first year	0	1	1	0,00000	0,00000	
EIF_air (production)	entire lifetime	0,0008	1	1	0,00080	0,00400	NILU-rapport (2006)
EIF_acute-water column	entire lifetime	4658600	4,80E-05	0,274	61,3	306,3	TNO-rapport (2006)
EIF_acute-water surface	entire lifetime	262700	4,80E-05	0,137	1,7	8,6	TNO-rapport (2006)
EIF_acute-coast line	entire lifetime	12400	4,80E-05	1	0,6	3,0	TNO-rapport (2006)
EIF_PW	entire lifetime	247	1	1	247,0	1235,0	Arvesen og Pehrson (2003)

Additional calculations:

- EIF DD to sediment figures assumed to decline over time according to graphs in Rye et al. (2005)

2) OBM and transport of cuttings to shore

	Occurence	Extent	Prob.	Dur.	Total	∑5Y	Values derived from
EIF_DD-water	only first year	0	1	1	0,0	0,0	
EIF_DD-water (spill)	only first year	2,4	5,00E-03	0,0003	0,0	0,0	Tap av containter, SINTEF-rapport (2005)
EIF_DD-sea surface (spill)	only first year	0	0	0	0,0	0,0	
EIF_DD-sed	decreasing over lifetime	1	1	1	1,0	5,0	SINTEF-rapport (2005)
EIF_DD-sed (spill)	only first year	0,7	5,00E-03	1	0,0	0,0	Tap av container, SINTEF-rapport (2005)
EIF_air (transport)	only first year	0,9444	1	1	0,94440	0,94440	Justert ut fra NIVA-rapport Mongstad (2005)
EIF_air (drilling)	only first year	0,00073	1	1	0,00073	0,00073	NILU-rapport (2006)
EIF_air (cuttings treatment)	only first year	1,2041	1	1	1,20410	1,20410	Justert ut fra NIVA-rapport Mongstad (2005)
EIF_air (production)	entire lifetime	0,0008	1	1	0,00080	0,00400	NILU-rapport (2006)
EIF_acute-water column	entire lifetime	4658600	4,80E-05	0,274	61,3	306,3	TNO-rapport (2006)
EIF_acute-water surface	entire lifetime	262700	4,80E-05	0,137	1,7	8,6	TNO-rapport (2006)
EIF_acute-coast line	entire lifetime	12400	4,80E-05	1	0,6	3,0	TNO-rapport (2006)
EIF_PW	entire lifetime	247	1	1	247,0	1235,0	Arvesen og Pehrson (2003)

Additional calculations:

- EIF DD to sediment figures assumed to decline over time according to graphs in Rye et al. (2005)
- EIF air (transport): According to Larssen et al. (2005) and Gjerstad et al. (2005): In Mongstad, 1536 tons NO_x => EIF = 1253 (other emissions disregarded). Emissions from transport Norne – Mongstad estimated to be 3473 kg NO_x from data in Knudsen et al. (2006). This is scaled down linearly and divided by 3 due to less vulnerable area.
- EIF air (treatment): Same procedure as for transport, only that no division by 3 is performed (as all emissions occur at Mongstad). NO_x emissions from treatment estimated to be 1476 kg in Paulsen et al. (2003).

3) OBM/TCC treatment offshore and discharge of treated drill cuttings

	Occurence	Extent	Prob.	Dur.	Total	∑5Y	Values derived from
EIF_DD-water	only first year	1809	1	0,015	27,1	27,1	SINTEF-rapport Kristin (2005)
EIF_DD-water (spill)	only first year	0	0	0	0,0	0,0	
EIF_DD-sea surface (spill)	only first year	0	0	0	0,0	0,0	
EIF_DD-sed	decreasing over lifetime	1	1	1	1,0	5,0	SINTEF-rapport (2005)
EIF_DD-sed (spill)	only first year	0	0	0	0,0	0,0	
EIF_air (transport)	only first year	0	1	1	0,00000	0,00000	
EIF_air (drilling)	only first year	0,00073	1	1	0,00073	0,00073	NILU-rapport (2006)
EIF_air (cuttings treatment)	only first year	0,00002	1	1	0,00002	0,00002	Justert ut fra NILU-rapport (2006)
EIF_air (production)	entire lifetime	0,0008	1	1	0,00080	0,00400	NILU-rapport (2006)
EIF_acute-water column	entire lifetime	4658600	4,80E-05	0,274	61,3	306,3	TNO-rapport (2006)
EIF_acute-water surface	entire lifetime	262700	4,80E-05	0,137	1,7	8,6	TNO-rapport (2006)
EIF_acute-coast line	entire lifetime	12400	4,80E-05	1	0,6	3,0	TNO-rapport (2006)
EIF_PW	entire lifetime	247	1	1	247,0	1235,0	Arvesen og Pehrson (2003)

Additional calculations:

- EIF DD to water column figures for discharge of treated drill cuttings at Kristin from Rye and Ditlevsen (2005)
- EIF DD to sediment figures assumed to decline over time according to graphs in Rye et al. (2005)
- EIF air (treatment): Adjusted according to data from Knudsen et al. (2006): NO_x emissions (treatment) ≈ 3% NO_x emissions (drilling) => EIF_air (treatment) ≈ 3% EIF_air (drilling) [SO_2 emissions close to proportional].

4) OBM/TCC treatment offshore and transport of treated drill cuttings to shore for reuse

	Occurence	Extent	Prob.	Dur.	Total	∑5Y	Values derived from
EIF_DD-water	only first year	0	1	1	0,0	0,0	
EIF_DD-water (spill)	only first year	0	0	0	0,0	0,0	
EIF_DD-sea surface (spill)	only first year	64	0,025	0,0016	0,0	0,0	Slangebrudd, SINTEF-rapport Kristin (2005)
EIF_DD-sed	decreasing over lifetime	1	1	1	1,0	5,0	SINTEF-rapport (2005)
EIF_DD-sed (spill)	only first year	0	0	0	0,0	0,0	
EIF_air (transport)	only first year	1,5E-05	1	1	0,00002	0,00002	NILU-rapport (2006)
EIF_air (drilling)	only first year	0,00073	1	1	0,00073	0,00073	NILU-rapport (2006)
EIF_air (cuttings treatment)	only first year	0,00002	1	1	0,00002	0,00002	Justert ut fra NILU-rapport (2006)
EIF_air (production)	entire lifetime	0,0008	1	1	0,00080	0,00400	NILU-rapport (2006)
EIF_acute-water column	entire lifetime	4658600	4,80E-05	0,274	61,3	306,3	TNO-rapport (2006)
EIF_acute-water surface	entire lifetime	262700	4,80E-05	0,137	1,7	8,6	TNO-rapport (2006)
EIF_acute-coast line	entire lifetime	12400	4,80E-05	1	0,6	3,0	TNO-rapport (2006)
EIF_PW	entire lifetime	247	1	1	247,0	1235,0	Arvesen og Pehrson (2003)

Additional calculations:

- EIF DD to water column figures for accidental discharge from hose during offloading TCC material at Kristin (Rye and Ditlevsen 2005)
- EIF DD to sediment figures assumed to decline over time according to graphs in Rye et al. (2005)
- EIF air (treatment): Adjusted according to data from Knudsen et al. (2006): NO_x emissions (treatment) ≈ 3% NO_x emissions (drilling) => EIF_air (treatment) ≈ 3% EIF_air (drilling) [SO_2 emissions close to proportional].

5) OBM/TCC treatment offshore and transport of treated drill cuttings to shore for disposal

	Occurence	Extent	Prob.	Dur.	Total	∑5Y	Values derived from
EIF_DD-water	only first year	0	1	1	0,0	0,0	
EIF_DD-water (spill)	only first year	0	0	0	0,0	0,0	
EIF_DD-sea surface (spill)	only first year	64	0,025	0,0016	0,0	0,0	Slangebrudd, SINTEF-rapport, Kristin (2005)
EIF_DD-sed	decreasing over lifetime	1	1	1	1,0	5,0	SINTEF-rapport (2005)
EIF_DD-sed (spill)	only first year	0	0	0	0,0	0,0	
EIF_air (transport)	only first year	0,9444	1	1	0,94440	0,94440	Justert ut fra NIVA-rapport Mongstad (2005)
EIF_air (drilling)	only first year	0,00073	1	1	0,00073	0,00073	NILU-rapport (2006)
EIF_air (cuttings treatment)	only first year	0,00002	1	1	0,00002	0,00002	Justert ut fra NILU-rapport (2006)
EIF_air (production)	entire lifetime	0,0008	1	1	0,00080	0,00400	NILU-rapport (2006)
EIF_acute-water column	entire lifetime	4658600	4,80E-05	0,274	61,3	306,3	TNO-rapport (2006)
EIF_acute-water surface	entire lifetime	262700	4,80E-05	0,137	1,7	8,6	TNO-rapport (2006)
EIF_acute-coast line	entire lifetime	12400	4,80E-05	1	0,6	3,0	TNO-rapport (2006)
EIF_PW	entire lifetime	247	1	1	247,0	1235,0	Arvesen og Pehrson (2003)

Additional calculations:

- EIF DD to water column figures for accidental discharge from hose during offloading TCC material at Kristin (Rye and Ditlevsen 2005)
- EIF DD to sediment figures assumed to decline over time according to graphs in Rye et al. (2005)
- EIF air (treatment): Adjusted according to data from Knudsen et al. (2006): NO_x emissions (treatment) ≈ 3% NO_x emissions (drilling) => EIF_air (treatment) ≈ 3% EIF_air (drilling) [SO_2 emissions close to proportional].

COST FIGURES, ALL ALTERNATIVES

Figures in NOK 1000, not adjusted for time differences. No alternative costs included.

	WBM	OBM/cuttings	TCC/discharge	TCC/reuse	TCC/disposal
Drilling fluid	4750	5330	5330	5330	5330
Rig costs	57600	40000	40000	40000	40000
Waste management	0	5900	6760	7000	7330
SUM	62350	51230	52090	52330	52660

Assumptions drilling fluid:
Volume section 1+2 = 833 m3, volume section 3+4 = 116m3
Consumption of drilling fluid: WBM: 5 m3 per section m3, OBM: 2 m3 per section m3
Unit cost drilling fluid: WBM: NOK 1000 per m3, OBM: NOK 5000 per m3

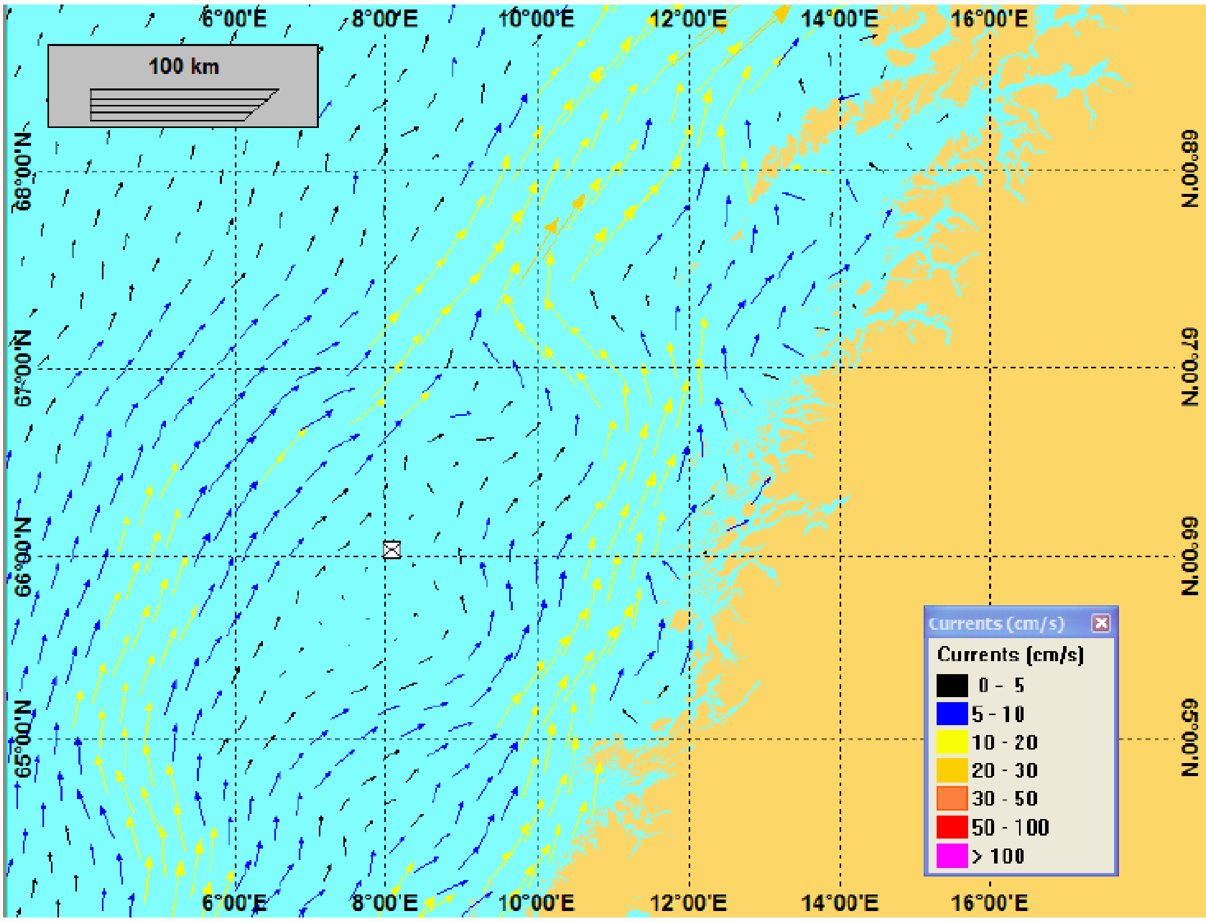
OBM method: upper two sections drilled with WBM

Assumptions rig costs:
$250 000 per day
$1 = NOK 6,4
rig time OBM = 25 days, rig time WBM = 36 days

Assumptions waste management:
Taken from Paulsen et al. (2003)
Alt. 4 slightly adjusted due to lower transport costs.

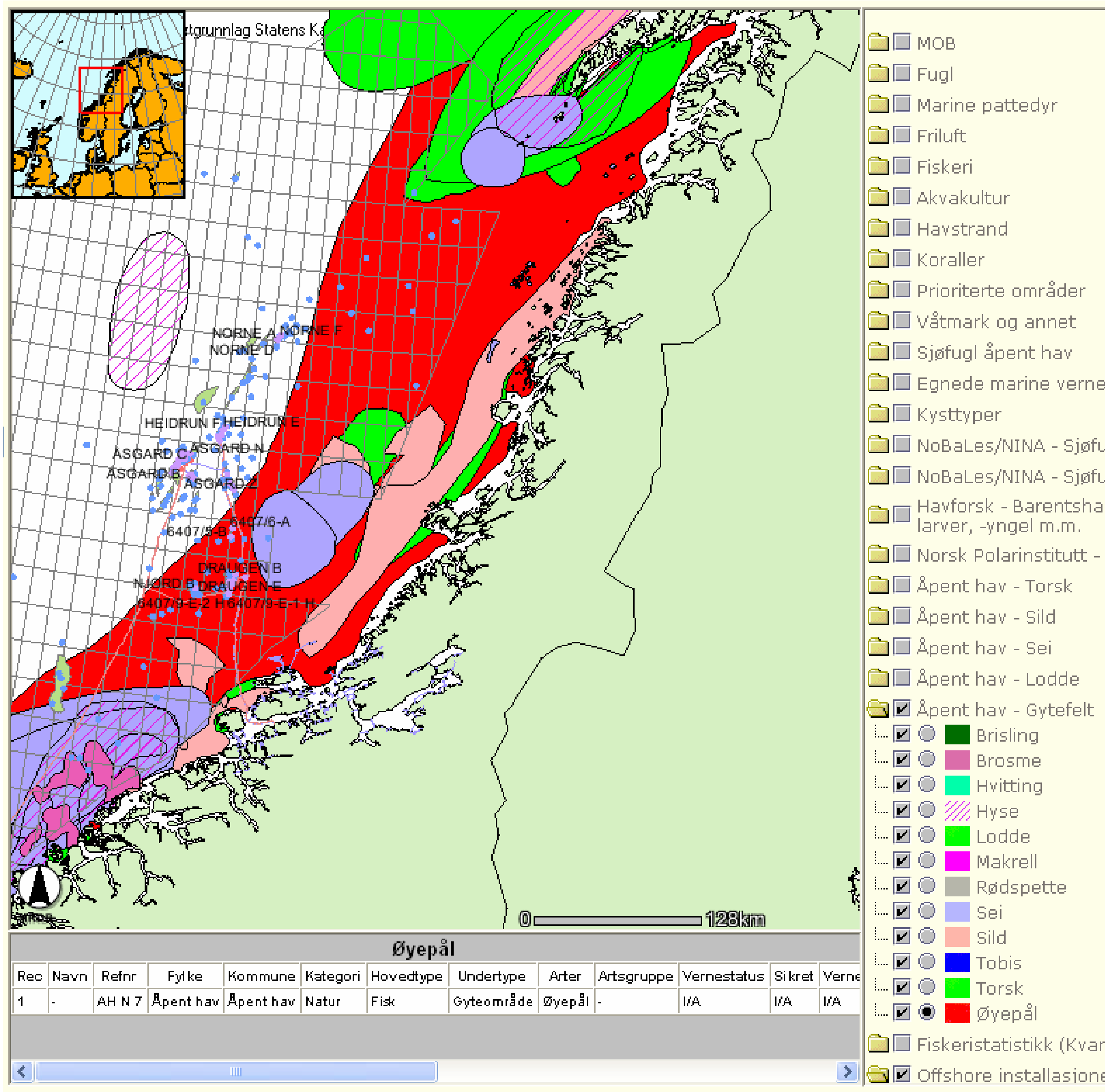
D. Current maps

Abstract from Rye et al. (2005). Example of current maps presented at the second session.



E. Distribution of species

Abstract from Marin Ressurs DataBase (www.mrdb.no). Example of maps of species' distribution presented at the second session.



F. Sensitivity list – Lofoten to Trondheim

List of sensitivity for different compartments elaborated by the project team. Note: This is not meant to be an exhaustive list, but possibly important features have been included.

		If yes: name and presence	**Other important issues**
Sediment/ seafloor	Background levels	Monitoring: THC, heavy metals	
	Species on the red list	Probable not	
	Corals	Occurrence in the whole area	
	Spawning ground	No	
	Other important recourses	Pockmarks	Mapping needed – present in the area
Water column	Background levels	?	
	Species on the red list	Cod	
	Spawning ground	No	Haddock further West - Sait, cod and Norway pout further East
	Spawning products present	Sait, Herring, Haddock - Egg and larvae	
	Other important recourses	Important bank area Plankton	Important for fish
Sea surface	Species on the red list	?	In general, little info available for open waters
	Presence of seabirds	Pelagic diving seabirds	In general, little info available for open waters
	Presence of moulting seabirds	Probable not	In general, little info available for open waters
	Presence of sea mammals	Not permanent	
	Other important recourses		
Coast line	Species on the red list	?	
	Presence of nesting species	Seabirds, sea mammals	
	Presence of vulnerable habitats	Sea meadow	
	Presence of sea mammals	Common seal, grey seal, otter	
	Other important recourses		

Soil	Background levels Species on the red list Breeding places Presence of vulnerable habitats Other important recourses	200-600 mg $N/m^2/yr$ Several lichens ? Raised bogs Boreal rain forests Nutrient poor forests Oligotrophic waters Nutrient poor alpine vegetation Wet coastal heathland	
Freshwater	Background levels present Species on the red list Presence of vulnerable habitats Other important recourses	? ? ? Drinking resources	
Ground water	Is the aquifer fresh water, and thereby a potential resource?	Yes	