Chapter 5
Handling Uncertainty and Sensitivity of ERA Acute Towards Input Parameters

Abstract Uncertainty evaluation and sensitivity testing of the functions and parameters used in ERA Acute serve two functions. ERA Acute is a deterministic model which is sensitive to the range of values used for the parameters. Parameters have inherent uncertainties as to what their true values are, and functions may have varying strength of knowledge. The individual functions were tested with respect to their sensitivity towards variation of the parameter values using both deterministic and stochastic testing. Based on the testing, an uncertainty scoring system was used to identify and prioritize the most important parameters for reducing uncertainty. Recommendations for handling the uncertainty and securing comparability in spite of uncertainty were set up as a conclusion of the studies.

5.1 Sensitivity Testing and Uncertainty Handling

Every model has some inherent uncertainty. A model is a simplified mathematical description which in a quote often accredited to Albert Einstein should be "as simple as possible, but no simpler". Both the simplifications and the detailing of sub-models and equations carry with them uncertainties.

ERA Acute is a *deterministic model* where the structure of the functions and their calculation sequences reflect how we believe that oil spills may harm the VECs in the different compartments. The output of a deterministic model is completely determined by the input parameters and structure of the model. A *stochastic model* on the other hand, has inherent randomness in the model structure and will not produce the same result, even given the same parameter value (Helton et al. 2006; Marino et al. 2008).

The *functions* are mathematical descriptions of how we understand that the impact and restoration will occur, and therefore also subject to *uncertainty* about the model framework and its scientific soundness (see e.g. Gaber et al. 2009). If our assumptions

of the mechanisms of action are uncertain, this reflects on uncertainty. As model complexity increases, the uncertainty tied to the model framework is reduced, but a more complex model uses more parameters and the data uncertainty increases (Gaber et al. 2009). For ERA Acute there are uncertainties of the model that belong to both the structure (model uncertainty) and the numerical parameters used (epistemic uncertainties).

On the condition that the functions and calculation sequences in the model are correct, if the input is not changed, the model output stays the same. In this case, uncertainty in the model output is solely affected by variation in the input parameter. This is called *Epistemic*, or reducible uncertainty (Helton et al. 2006), related to lack of knowledge of the true value of a constant parameter (Marino et al. 2008).

For the individual *parameters* used in a model it is important to distinguish clearly between:

Variability: How spread out or clustered a data set is, e.g. the (natural) variation in the measured values found in nature and
Uncertainty—The lack of certainty or knowledge about what the value of the parameter/data truly is. Such data uncertainty is specific to the individual parameter. As mentioned above, a more complex model uses more parameters and data uncertainty therefore increases (Gaber et al. 2009). ERA Acute uses many parameters.
Sensitivity Analysis (SA) tells us how the model's response can be apportioned to changes in model inputs. It is algorithm specific. For models with a high number of parameters, sensitivity analyses are useful to rank the relative importance of the factors and processes involved (Saltelli 2004).

ERA Acute is a new method and testing the sensitivity of the model to variation in the input parameters is an important part of uncertainty handling, with the goal of ensuring that ERA Acute does not under-estimate environmental risk. All data sets and parameter values have inherent uncertainties and a model consisting of a series of calculations will need some method of handling uncertainty. In the process of developing ERA Acute, the following activities were carried out:

1. Sensitivity testing of the risk functions to the variation in input parameters
2. A pilot study to score the parameters and propose feasible uncertainty handling

The functions of ERA Acute are built so that individual parameters representing biological or environmental characteristics can be improved as knowledge increases, thereby reducing uncertainty by a continuous improvement process. Sensitivity testing provides knowledge of which of the parameters that contribute most to the final endpoint values, and therefore the testing provides information on which parameters that would be most important to improve by further research if they have high uncertainty.

ERA Acute covers four compartments and uses a large number of functions. The input parameters (values and datasets) that are used are based on knowledge from few and highly diverse incidents. Validating the results of the method and applying the results with an acceptable level of (un)certainty is therefore challenging (see

Chap. 4 "Testing and Validating Against Historic Spills"). In such a case of applying a complex model and multitude of uncertain parameters, it is important to realize and accept that we do not know the "true risk" as a number as such but need to ensure that the model does not underestimate the risk and can be used to compare risks.

5.2 Methods Used in Sensitivity Testing

To test the sensitivity of the calculations towards numerical variation in the input parameters both deterministic and stochastic tests were carried out.

5.2.1 Deterministic Testing

In the deterministic testing, the impact calculations, lag- and restoration calculations within the model and its sub-models were tested by breaking them into the individual functions. By holding the other parameters constant, the input parameters were varied one by one and the resulting endpoints calculated. The results are available as graphs. The reader is encouraged to read the full reports with method description and results in: Bjørgesæter and Damsgaard-Jensen (2018) and Stephansen and Bjørgesæter (2017).

These simpler deterministic tests holding one parameter fixed at a time (One-At A-Time tests, OAT) are useful to study the direct output of varying single parameters, and thus get better acquainted with the results of the individual calculations. However, these deterministic tests are unsuitable for handling the many dimensions of variation of the input parameters, for which the global stochastic sensitivity methods are used (Marino et al. 2008).

The range in parameter values found in the literature studies during methodology development was used to define the range between the minimum and maximum values but these ranges were not used to limit the sensitivity analyses performed in the next step (see the references in the methodology development, Chap. 3 and references to Tables A.1, A.2 and A.3 in Supplementary Information 1).

A deterministic approach requires few simulations and is therefore valuable for examining models that may become costly in terms of computer time (e.g. testing oil drifts statistics used in the models).

The disadvantages of the deterministic approach include; only a few discrete outcomes are considered, it gives equal weight to each outcome, and possible interdependence between inputs are difficult to identify and quantify. Assessing the likelihood of different outcomes is therefore not possible with deterministic testing, and it is difficult to identify and rank the input parameter in terms of importance on the model output.

5.2.2 Stochastic Testing

The sub-models within ERA Acute are deterministic. To perform stochastic sensitivity testing, these models were made stochastic by using repeated random sampling (Monte Carlo (MC)) methods (Marino et al. 2008): Instead of changing the values one by one as in deterministic testing, they are assigned to a (a priori assumed) probability distribution. Configurations of model input values are then drawn randomly from the probability distribution, and the resulting set of model outputs can be seen as a random sample of the distribution of the output of interest (Helton et al. 2006). Note that stochastic analyses are sensitive to the choice of probability distribution used (e.g. Marino et al. 2008).

The result is a matrix with *n* values for each input parameter with corresponding values for the model output (model predictions, results or endpoint) (Fig. 5.1). This matrix is the input to the uncertainty and sensitivity analysis, which is performed directly on the matrix. The sensitivity analyses were carried out using the Sampling and Sensitivity Analysis Tool for Computational Modelling (SaSat) (Hoare et al. 2008a, b). For the sensitivity analysis, Pearson and Spearman correlation coefficient, Partial Rank Correlation Coefficient (PRCC) analysis and Factor Prioritization by Reduction of Variance (FPRV) were carried out (see e.g. Saltelli et al. 2000; Marino et al. 2008). Combined, these methods can rank and quantify the most important
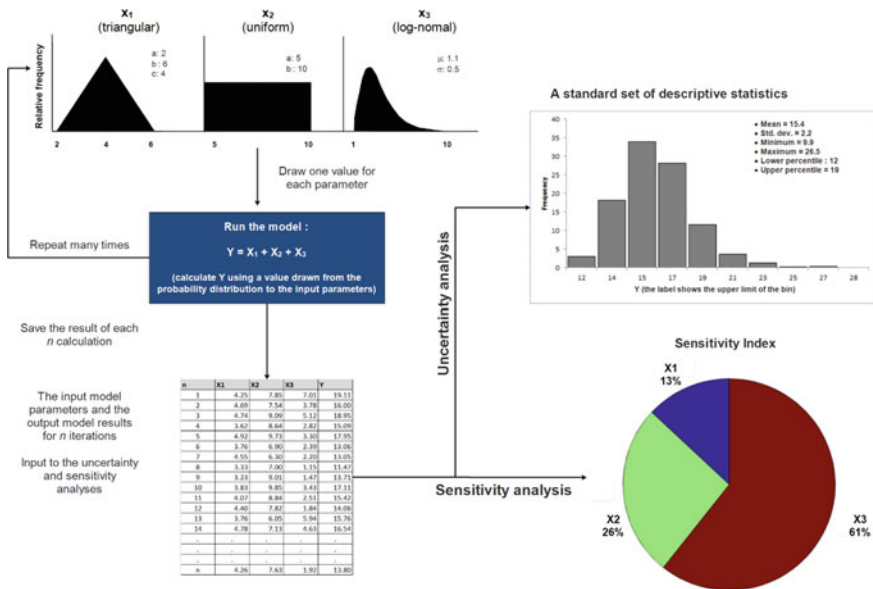

Fig. 5.1 Illustration of stochastic uncertainty and sensitivity analyses. The ERA Acute model calculations are performed in the blue box. The uncertainty analyses are performed in Excel and sensitivity analyses are performed with the MATLAB toolbox sampling and sensitivity analysis tool for computational modelling (SaSAT)

input parameters to ERA Acute. Where calculations were carried out in succession, combined formulas were used. PRCC allows independent effects of each parameter to be determined, even when the parameters are correlated. The goal is to determine which factor, once fixed to its true value by additional research, on average leads to the greatest reduction in the variance of an output. The interpretation of PRCCs assumes a monotonic relationship (relationship or function which preserves a given trend) between parameters (Marino et al. 2008). This is the case for all the (sub-) models used in ERA Acute. The rank-transformation is done to reduce the effect of non-linear data, and PRCC is a robust sensitivity measure for nonlinear, monotonic relationships (Marino et al. 2008).

The result is a sensitivity index for each input parameter to the formula, which is the *fraction of the variation in the output value that can be ascribed to the different parameters. Note that this is given the uncertainty defined by the range of natural variation (results based on literature search) and the weight of each value given by the distribution (uniform—equal weight).* If a different distribution for the initial random drawing of values had been used, the result would have been different. However, given the nature of the parameters, a uniform distribution was assumed.

The use of these statistical methods in the ERA Acute sensitivity testing is described in further detail in the project reports by Bjørgesæter and Damsgaard-Jensen (2018) and Stephansen and Bjørgesæter (2017).

Impact and restoration functions were tested for each compartment and for each relevant VEC-group within the compartment having different parameter values and/or functions.

The results from the Spearman correlation coefficient analysis are presented in the test reports by Bjørgesæter and Damsgaard-Jensen (2018) for surface, water column and shoreline compartments and Stephansen and Bjørgesæter (2017) for the seafloor compartment.

5.2.3   Example from Surface Compartment

For the deterministic testing, all parameter values used for the wildlife groups are available in the test report (Bjørgesæter and Damsgaard-Jensen 2018), as well as figures showing the results for each of the tested parameters. As part of the testing it was determined that the equation which includes the exposure time (N-let2) (Sect. 3.3.1) performs best according to the impact estimated from various field estimates.

The individual factors comprising p_{let} for the surface; $p_{beh} \times p_{phy}$ were set up with values for high, medium and low estimates of the values for each of the 13 wildlife groups. The assumption behind choice of probability distribution for the stochastic drawing of values plays an important role as described in Sect. 5.2.2.

P-values and ranking according to importance from the Spearman correlation coefficient analysis for the surface compartment are presented for the parameters used in the initial impact calculation in Fig. 5.2. If the p-value (probability of type 1

Parameter	N$_{let-1}$			N$_{let-2}$		
	Spearman Corr. Coeff.	p-values	Importance rank	Spearman Corr. Coeff.	p-values	Importance rank
Pbeh	0.19	0.00	Cov	0.13	0.00	Texp
Pphy	0.39	0.00	pphy	0.28	0.00	Cov
Cov	0.84	0.00	pbeh	0.60	0.00	pphy
Texp	-	-	-	0.68	0.00	pbeh

Sensitivity index for Nlet-1
pbeh, 7%
pphy, 20%
Cov, 73%

Sensitivity index for Nlet-2
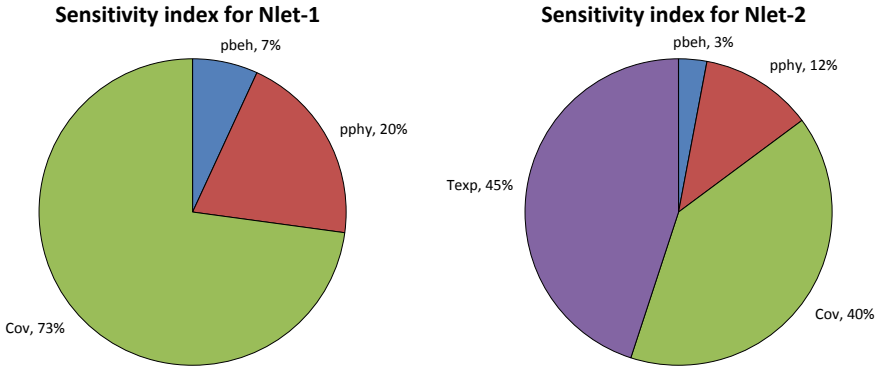pbeh, 3%
pphy, 12%
Texp, 45%
Cov, 40%

Fig. 5.2 Example of sensitivity testing: result for the sensitivity analysis of Nlet-1 and Nlet-2 for the following random variables: p$_{beh}$, p$_{hy}$, coverage (Cov) and exposure time (Texp). The relative abundance is held constant at 1

error) is below 0.05 this means that the result is statistically significant within a 95% confidence limit. The p-values were in this case ≪0.001. The pie diagram shows the sensitivity index from the FPRV.

The population loss is more sensitive to the variation in the oil drift impact parameters than to the variation in the two model parameters. This is also the case for the impact calculations in the other compartments. In the example from the surface compartment, the oil coverage in the grid cell is ranked as the most important variable for Nlet-1 (equation without exposure time) and as much as 73% of the total variance observed in Nlet-1 in Fig. 5.2 can be attributed to this parameter. Therefore, although all parameters are initially equally important, coverage and Texp (the latter for Nlet-2) are the most important parameters. Both values have inherent uncertainty from input parameters' and model soundness uncertainty of the oil drift simulation model. Logically, since coverage and exposure time represent the spreading and degradation of oil and the results are reported for cells at different distances from the spill and for a multitude of simulations representing very different weather conditions, it is natural that there is a large variation in the values. An inherent property of these tests is that if a tested parameter has a high variability it also becomes more important.

5.3 Uncertainty Scoring of the Parameters

A feasible way of handling uncertainty in ERA acute is necessary, using the results from the project development and testing whilst recognizing that many parameter values may need further refinement in the first years of use. A self-evaluation scoring system developed by DNV GL (Kruuse-Meyer 2015) was used to score the parameters and provide recommendations on the use of specific parameters. The scoring was based primarily on the results of sensitivity testing carried out using statistical and deterministic methods (Sect. 5.2) (Bjørgesæter and Damsgaard-Jensen 2018; Stephansen and Bjørgesæter 2017).

For each parameter the sensitivity-deciding elements were considered and assessed within the limits of the knowledge gained in previous project work according to:

Strength of knowledge (function where it is used): How strong is our confidence in that the risk function in which the parameter is used is a valid mathematical representation of the mechanism of impact/restoration?

Belief that the value may deviate from the average assumption: Natural variation of parameter. Do we believe that the values have a high natural tendency to vary from the base case (mean)? E.g. if a (standard deviation) (SD) is quantifiable, this can be used to assess this point.

Sensitivity of function to parameter (sensitivity index): How sensitive is the model/function to variation in the parameter?

Comments/recommendations on handling to ensure risk is not under-estimated: Recommended actions for ERA Acute use, data gathering etc.

5.3.1 Surface Compartment

The results of the scoring process based on the results of the deterministic and then stochastic testing and evaluation of surface compartment parameters are given in Table 5.1 for the impact parameters and Table 5.2 for the lag- and restoration time parameters.

Table 5.1 Summary of assessments or calculations used as basis for classification in the sea surface. Impact and impact time parameters

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)
p_{beh}	**A** Moderate/weak. Due to limited data and large natural variation it is difficult to assign a specific p_{beh} value. The assumption that behavioural factors will affect p_{exp} is strong
	B Moderate
	C Moderate
	Comments/recommendations: A higher value is conservative. Each VEC have three estimates (low, intermediate, high), using high is most conservative. Alternative, use all to obtain larger credible interval
Cov	**A** Moderate/weak. The parameter depends on other parameters evaluated as Moderate/weak. The assumption that that exposed area will affect p_{exp} strong
	B High
	C Moderate
	Comments/recommendations: A higher value is conservative. Coverage is calculated by the oil drift model. Use Best Practice for oil drift simulation set-up to ensure comparable and reliable predictions of the statistic
Texp	**A** Moderate/weak. The parameter depends on other parameters evaluated as Moderate/weak. Based on stochastic result (i.e. estimated over the whole simulation period). The assumption that the exposure time will affect p_{exp} is strong
	B High
	C High
	Comments/recommendations: A higher value is conservative. Exposure time is calculated by the oil drift model. Use Best Practice input data and setup for the ODS to ensure comparable and reliable predictions
p_{phy}	**A** Moderate/weak. Due to lack of experimental data, it is difficult to assign a specific p_{phy} values. The assumption that the physiological factors will affect p_{let} is strong
	B Low/Moderate/High, depending on VEC Low for seabirds and moderate to high for marine and aquatic mammals and sea turtles
	C Moderate
	Comments/recommendations: A higher value is conservative. Each VEC have three estimates (low, intermediate, high), using high is most conservative. Alternative, use all to obtain larger credible interval

Table 5.1   (continued)

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)
Th	**A** Moderate/weak Due to lack of experimental data, it is difficult to assign specific threshold levels for lethal oil film thickness
	B Moderate
	C High
	Comments/recommendations: A threshold value, lower value is conservative. Oil thickness is calculated by the oil drift model. Use Best Practice for ODS to ensure comparable and reliable predictions. Based on present knowledge, reducing Th from 10 to 2 μm, increases the impact with a factor of approximately 2.0–2.5, depending on the distribution of the VEC and the distance to the release point
N per cell	**A** Moderate/weak Depends on the quality of the data received from the data provider. The quality of the data for the NCS is considered high
	B High
	C Moderate/high
	Comments/recommendations: Use the best available data to reduce uncertainty. Use the same data for comparable studies. The definition of a "population" is important

5.3.2   Water Column Compartment

Results of the scoring and evaluation of water column parameters are given in Table 5.3 for the impact parameters and Table 5.4 for recovery parameters.

5.3.3   Shoreline Compartment

Results of the scoring and evaluation of shoreline parameters are given in Table 5.5 for impact parameters and Table 5.6 for recovery parameters.

5.3.4   Seafloor Compartment

Results of the scoring and evaluation of seafloor parameters are given in Table 5.7 for impact parameters and Table 5.8 for recovery parameters.

Table 5.2 Summary of assessments or calculations used as basis for classification in the sea surface. Lag time and restoration time parameters

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)
Nhab	**A** Moderate/weak The function includes various not well-defined or understood subtle effect other than acute mortality
	B High
	C Moderate
	Comments/recommendations: Using the function will increase the total recovery time, typically with 5–30% of the shoreline lag-times but depending on the importance of the affected shoreline habitats
SF	**A** Moderate/weak
	B High
	C Moderate
	Comments/recommendations: Use conservative value
t_lag (shoreline)	**A** Moderate/weak. Due to lack of experience data, it is challenging to assign specific lag time periods for different types of shoreline habitats
	B High
	C High/Moderate
	Comments/recommendations: Higher values are more conservative. Standard values for SF for different VECS and/or area are not derived. May use the same data as for calculating acute mortality (filtered for shoreline cells)
R	**A** Moderate/weak. The logistic discrete population model is a simplification of real-world population dynamics. Common R values are used for different species and populations as a standard (see b)
	B Moderate/high
	C High
	Comments/recommendations: Lower values are more conservative. The R values are conservative compared to the damage keys used in MIRA (using standard values for b, K and TLR). Field validation studies indicates that the model performs reasonably well, for population not inhibited by unknown extrinsic factors (using standard R, b, K and TLR values)
b	**A** Moderate/weak. The parameter determines the strength of intraspecific competition; a simplification of real-world population dynamics
	B High
	C High

Table 5.2 (continued)

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)
	Comments/recommendations: Lower values are more conservative. Used to reflect population growth in population inhibited by unknown extrinsic factors or the general status of the population ("poor", "intermediate", "good"). Use low b values to further increase the conservatism of the population model predictions
K	**A** Moderate/weak
	B High. Large fluctuations of population size above and below carrying capacity is common in nature
	C High
	Comments/recommendations: The carrying capacity of the environment (K) is the maximum population size that the environment can sustain. It is set equal to the population size before the oil spill release (100%) and is used as a reference point for when the population is considered recovered
TRL	**A** Moderate/weak Cut off to avoid $t_{res} = \infty$ in a logistical growth model
	B High
	C High/moderate for t_{res}, Moderate/low for RDF (effect varies with percentage population loss)
	Comments/recommendations: Higher values are more conservative. Can be chosen differently for higher level of conservatism. Using values above 95% may lead to unrealistic long Restoration times

5.4 Recommended Uncertainty Handling at This Point in Model Development

Ideally, it should be one of the goals to arrive at a quantified estimate of the degree of accuracy of the endpoints of impact and restoration modelling. However, to arrive at this, more and continuous improvement is needed. Instead, general recommendations are given for ensuring comparability and reducing variability:

- Use the *conservative* values included in the method reports and current guideline
- Use *quality* data sources from acclaimed institutions
- Seek *improved* data for the factors to which the model is most sensitive to where possible
- Use *standardised* data sets and input parameters for analyses that are to be compared.

Table 5.3 Summary of assessments or calculations used as basis for classification in the water column. Impact parameters

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)
Plet, THC Extracted from SSD-curve	**A** Strong
	B Moderate. SSD-curve based on LC50 for 24 species
	C High
	Comments/recommendations: Estimated from THC and a log-normal SSD curve with standard deviation of 0.32. A lower standard deviation is conservative (shift the SSD curve to higher THC values)
THC	**A** Moderate/weak. Vertical maxima, THC includes numerous components with varying toxicity
	B High
	C High
	Comments/recommendations: THC is calculated by the oil drift model. Use Best Practice for ODS set-up to ensure comparable results. Use a concentration grid (with many layers) that cover the same water column where the fish egg/larva are distributed
Frackilled	**A** Strong. Estimated in OSCAR during the ODS
	B Moderate
	C High/Moderate/Low (depending on setting)
	Comments/recommendations: Estimated by OSCAR during the ODS (potential acute mortality in a cell). Standard deviation (SD) of the SSD and the species sensitivity may be adjusted before one run the ODS. The species sensitivity is a safety factor. The OSCAR database LC50 values will be divided by this factor, accounting for more (factor >1) or less (factor <1) sensitive fish larva/egg
N per cell	**A** Strong. Depends on the quality of the data received from the data provider. Compared to e.g. birds the distribution is to a large degree dependent on sea currents
	B Moderate
	C Moderate/high
	Comments/recommendations: Use the best available data to reduce uncertainty and increase the quality of the predictions. Use the same data for comparable studies

Table 5.4 Summary of assessments or calculations used as basis for classification in the water column. Recovery parameters

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)	
CritDens (%)	**A** Moderate/weak	
	B High	
	C High (threshold level between two methods with different conservatism)	
	Comments/recommendations Higher values are more conservative. Expresses the threshold for when a direct relationship is modelled between larval mortality and recruitment reduction	
CritOilMort (%)	**A** Moderate/weak	
	B High	
	C High (threshold level between two methods with different conservatism)	
	Comments/recommendations Lower values are more conservative. Expresses the threshold mortality of eggs and larvae for which a proportionate relationship is calculated between killed larvae and reduced recruitment	
Annual natural mortality of immatures (%)	**A** Moderate/weak	
	B Moderate/high	
	C Not tested	
	Comments/recommendations	
Annual natural mortality of matures (%)	**A** Moderate/weak	
	B Moderate/high	
	C Not tested	
	Comments/recommendations	
Age at recruitment (year)	**A** Moderate/weak	
	B Low/moderate	
	C Not tested	
	Comments/recommendations	
Age at first spawning (year)	**A** Moderate/weak	
	B Low	
	C Not tested	
	Comments/recommendations	

(continued)

Table 5.4   (continued)

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)
Maximum age (year)	**A** Moderate/weak
	B Low
	C Not tested
	Comments/recommendations

Table 5.5   Summary of assessments or calculations used as basis for classification in the shoreline. Impact parameters

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)
Tidal range (m)	**A** Moderate/low
	B Moderate/low (Coastal tidal ranges vary considerably depending on the volume of water adjacent to the coast, and the geography of the basin. Tidal range also varies depending on the locations of the moon and sun)
	C Low
	Comments/recommendations: Lower values are more conservative. The parameter is cell specific and is used to estimate oil thickness
Slope (°)	**A** Moderate/low
	B High/moderate (the slope of the beach may vary considerable with a shoreline habitat type)
	C High
	Comments/recommendations: Higher values are more conservative. The parameter is ESI specific and is used to estimate oil thickness
OHC	**A** Moderate/low
	B High/moderate (the distribution of oil along the shoreline will also depend on factors such as current, wind, geography, that are difficult to accurately estimate outside the oil drift model)

Table 5.5 (continued)

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)
	C Moderate/high
	Comments/recommendations: Higher values are more conservative. The parameter is ESI specific and is used to distribute the stranded oil mass along the shoreline in a cell. Higher value means that more of the stranded mass is allocated to the shoreline habitat
Patchiness factor	**A** Moderate/low. Due to lack of experience data, it is challenging to assign a specific patchiness factor
	B High. Patchiness of oil may range from 1 to 100%
	C High
	Comments/recommendations: Lower values are more conservative. Fixed look-up values
Th	**A** Moderate/low. It is difficult to assign a specific threshold level for lethal oil film thickness for invertebrates and vegetation
	B Moderate
	C High (threshold value)
	Comments/recommendations: Higher values are more conservative. Threshold level for impact, 0.1 mm for invertebrates and 1.0 mm for wetland vegetation
Stranded mass (ton)	**A** Moderate/low. Basis for calculating film thickness
	B High
	C High/moderate (proportional)
	Comments/recommendations: Higher values are more conservative Stranded mass is calculated by the oil drift model. Use Best Practice for ODS to ensure comparable and reliable predictions
Shoreline length (km)	**A** Strong. Depends on the quality of the data received from the data provider
	B Low/moderate
	C High (proportional)
	Comments/recommendations: Use the best available data to reduce uncertainty and increase the quality of the predictions. Use the same data for comparable studies
Shoreline rankings	**A** Strong. Depends on the quality of the data received from the data provider
	B Moderate
	C High for recovery (lag-time and restitution)
	Comments/recommendations: ESI rankings; 1 least sensitive, 10 most sensitive

Table 5.6 Summary of assessments or calculations used as basis for classification in the shoreline. Lag-time and recovery parameters

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)	
Lag-time	**A** Moderate/low. Due to lack of experience data, it is challenging to assign specific lag-time periods for shorelines	
	B High/moderate. Variable and to a large degree depending on weather conditions	
	C High	
	Comments/recommendations: Fixed look-up values	
Recovery	**A** Moderate/low. Due to lack of experience data, it is challenging to assign specific restitution time periods for shorelines	
	B High Variable depending on invertebrate and flora communities	
	C High	
	Comments/recommendations: Fixed look-up values	

Table 5.7 Summary of assessments or calculations used as basis for classification in the seafloor. Impact parameters

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (Natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)	
Mixing depth	**A** Strong/moderate. Knowledge of what constitutes the bioturbation depth is relatively strong	
	B High uncertainty	**C** 40.0% high
	Comments/recommendations: A lower value is conservative, lower values are default for all substrates based on size of typical burrowing fauna in substrate. High natural variation: Either look for local real values or use conservative value	
Dry density	**A** Strong	**B** Low	**C** 0.5% low
	Comments/recommendations: Schultz and Zabel (2006) give general values. Low sensitivity, use defaults		
Water Content	**A** Strong	**B** Low/moderate	**C** 2.7% low
	Comments/recommendations: Use lower values as conservative		
Total org. Carbon	**A** Strong (EqP accepted methodology)	
	B High	**C** 54.9% high
	Comments/recommendations: Use conservative (lower) values. Lower values lead to higher toxicity and shorter restoration times (Higher TOC sequesters THC in sed.)	

Table 5.7 (continued)

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (Natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)	
KOW	**A** Strong (EqP accepted methodology)	
	B Moderate	**C** 1.8% low
	Comments/recommendations: Value calculated based on typical components with affinity to organic carbon in sediment. Use as implemented, can be changed, but has low impact on result	
Plet (SSD-curve used)	**A** Strong	
	B High to low depending on species sensitivity	
	C High	
	Comments/recommendations: Conservativity implemented by the curve being conservatively extrapolated from the LD5-value from a large and quality-assessed set of data (Nilsen et al. 2006). SSD curves are accepted methodology and inherent safety factor used and can be increased for sensitive species	
THCsed (used as input from OSCAR)	**A** Strong knowledge of place in ERA Acute function	
	B Is calculated by the OD model. SD is low within calculations in same model, may vary a lot between inputs from different models	
	C High (proportional)	
	Comments/recommendations: THCsed calculations in OSCAR do currently not take into consideration the grain size or TOC-content of the substrate (these factors are used by ERA Acute to modify the exposure in the initial calculations. No conservativity is included, but the other factors are chosen conservatively. The calculations in sediment in OSCAR are undergoing improvements, e.g. by possible inclusion of marine snow	
THC (WC)	**A** Strong knowledge of place in SSD-curve	
	B High uncertainty and the THC concentration is a time-averaged concentration	
	C High (proportional)	
	Comments/recommendations: The concentration is calculated as a time-averaged THC-value. This is a weakness in the approach. Use of dynamic time-steps output options (e.g. proposed in the ERA Acute Dynamic Risk Assessment incl. MIZ-proposal) could improve this. Conservativity is applied as we currently do not have available from OSCAR the THC-conc. in the lower WC, and therefore use the upper layers as for compartment WC. This is conservative	
N	**A** High strength of knowledge	
	B Moderate	**C** High (proportional)
	Comments/recommendations: Use quality data on presence or habitat area/fractions. Sampling of benthic species may lead to uncertainties, use data that are based on accepted sampling methods by accredited data sources	

Table 5.8 Summary of assessments or calculations used as basis for classification in the seafloor. Lag-time and restoration parameters

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (Natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)
$C_{threshold,sed}$	**A** Moderate strength of knowledge of function
	B High
	C High
	Comments/recommendations: Concentration of THC at which effects on faunal communities in sediment cannot be detected in monitoring studies (Renaud et al. 2008). Species may be more sensitive or less
$C_{benchmark-max,sed}$	**A** Moderate strength of knowledge of function
	B High
	C High
	Comments/recommendations: Value representing the maximum value at equilibrium. Based on data from the MOD data base (North Sea)
20 years def value	**A** Moderate strength of knowledge of function
	B High
	C High
	Comments/recommendations: Based on MOD data from North Sea, mostly sandy bottom, few sites have data on restoration times after use of oil-based drilling muds
SF	**A** Moderate strength of knowledge of function
	B High
	C High (proportional)
	Comments/recommendations: Theoretical calculation of the leaching of THC from organic carbon, simplified approach based on physical-chemical properties of THC bound to organic carbon in sediments (resuspension and redistribution may vary between substrates and is not included). The SF was introduced to the function to modify the calculated restoration time
t_{lag} (hard)	**A** Fixed value
	B High
	C High
	Comments/recommendations: Very little research available after oil spills affecting deep sea corals. Comparable incident DHOS not yet restored

Table 5.8 (continued)

Main parameter	A: Strength of knowledge (*function* where it is used) B: Belief that the value may deviate from the average assumption (Natural variation of parameter) C: Sensitivity of function to parameter (sensitivity index)		
t_{res} (Hard)	**A** Fixed value		
	B High		
	C High		
	Comments/recommendations: Very little research available after oil spills affecting deep sea corals. Comparable incident DHOS not yet restored		

Within a region, e.g. a country for which assessments should be used for applications to the authorities, this means that the industry should work together to test new values, gain common knowledge and understanding of the sensitivities as well as use common data sets. Calibration of the parameter values should be carried out after testing and documentation of the effects, and results discussed between scientists from both industry, consultancies, authorities and research institutions. The goal is *continuous, but structured and synchronised improvement.* A summary of the recommendations for each of the most important parameters is given for each compartment in Tables 5.9 and 5.10.

Table 5.9 Prioritised parameters with a potential for improvement or parameters that have a high impact on the result, with recommended action for uncertainty handling (surface compartment parameters)

Parameter	Recommendation for improvement
Cov	Use oil drift model that uses a state-of-the art calculation of oil coverage above the threshold on the surface with best practice settings
Texp	Use oil drift model that uses a state-of-the art calculation of the time with oil above the threshold level on the surface, with best practice settings. Setting a minimum exposure time could be beneficial to not underestimate impact
R	Net *fundamental* growth rate is based on demographic data (age at first and last reproduction, annual birth rate, pre-reproductive and adult survival probability) and literature review of different species and categorised into seven major groups. Updating knowledge and adding more data would increase certainty of the R values
TLR	Current restoration function is asymptotic, the threshold level for when the population is recovered is highly sensitive
b	The *realised* growth rate can be inherently different for different populations (or colonies or groups) of the same species when recovery is inhibited by known or unknown extrinsic factors (high predation, hunting, food shortage, disease etc.). Updating the knowledge and adjusting the factor (b) for these "populations" would improve certainty. A practical solution for standard environmental risk analyses is to apply three values for the b factor as a measure of the "general health" of the population/colony ("good", "medium" and "poor"). The same effect may be obtained by adjusting the net fundamental growth rate R

Table 5.10 Prioritised parameters with a potential for improvement or parameters that have a high impact on the result, with recommended action for uncertainty handling (shoreline, water column and seafloor compartment parameters)

Parameter	Recommendation for improvement
Shoreline	
Mass	High importance but proportional. Use oil drift model that uses a state-of-the art calculation of beached mass, with best practice settings
Patchiness factor	The value is a fixed value based on research. Lack of data available, could be improved with more research. Value in 2020 is 0.30 based on calibration
Slope	ESI-specific. Use best practice ESI dataset
Lag-time/Recovery time	Fixed values that could be improved with more research
Water column	
CM	Use a best practice recommendation for setting the Critical Mortality value for when the gate model is used
Seafloor	
TOC	Total organic content in the soft substrate determines the partitioning between oil adhered to the substrate and oil that is bioavailable in interstitial or gut water, and thereby the exposure and lethality. The value may vary a lot regionally depending on the background concentration of organic matter and substrate type. Monitoring studies could include this parameter for regionally/nationally improved quality of the substrate data
BDepth	Mixing depth scales the result proportionally and varies with the type of burrowing fauna. The variation in results from different studies is high. Monitoring studies could include this parameter for regionally/nationally improved quality of the substrate data
WC oil concentration	Exposure through water column determines much of the impact for all feeding modes that have exposure though water column. Best result if using oil drift modelling that provides a separate water column concentration from the bottom layer
THCsed	Start-value of oil concentration in the soft substrates. Use an oil drift model that provides a state-of-the-art calculation of oil in the sediment corrected for the substrate type (TOC-content)

References—Sensitivity Testing

Bjørgesæter A, Damsgaard-Jensen J (2018) WP2a. Sensitivity and uncertainty testing. ERA Acute Project report, ERA Acute 2A-2

Gaber N, Foley G, Pascual P, Stiber N, Sunderland E, Cope B, Saleem Z (2009) Guidance on the development, evaluation, and application of environmental models, Report, Council for Regulatory Environmental Modeling. Report to U.S. Environmental Protection Agency (EPA)

Helton JC, Johnson JD, Sallaberry CJ, Storlie CB (2006) Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliab Eng Syst Saf 91:1175–1209

Hoare A, Regan DG, Wilson DP (2008a). Sampling and sensitivity analysis tool for computational modelling: user guide. National Centre in HIV Epidemiology and Clinical Research, University of New South Wales, Sydney, Australia

Hoare A, Regan DG, Wilson DP (2008b) Sampling and sensitivity analyses tools (SaSAT) for computational modelling. Theor Biol Med Model 5:1

Kruuse-Meyer R (2015) Uncertainty estimates in environmental risk analysis—a guidance to uncertainties and sensitivities in Environmental Risk Analyses. DNV GL Internal report 2015–1204, Rev. 00

Marino S, Hogue IB, Ray CJ, Kirchner DE (2008) A methodology for performing global uncertainty and sensitivity analysis in systems biology. J Theor Biol 254(1):178–196

Nilsen H, Johnsen HG, Nordtug T, Johansen Ø (2006) Threshold values and exposure to risk functions for oil components in the water column to be used for risk assessment of acute discharges (EIF Acute). Statoil Report. 18 p. https://norskoljeoggass.no/globalassets/dokumenter/miljo/era-acute/report-2-era-acute-threshold-values-and-exposure-2006.pdf

Renaud PE, Jensen T, Wassbotten I, Mannvik HP, Botnen H (2008) Offshore sediment monitoring on the Norwegian shelf. A regional approach 1996–2006. Akvaplan-niva report no. 3487–003. 95 p

Saltelli A (2004) Sensitivity analysis in practice: a guide to assessing scientific models. Wiley, Hoboken, NJ

Saltelli A, Chan K, Scott EM (2000) Wiley series in probability and statistics. Wiley, Chichester, New York. Sensitivity analysis

Schultz HD, Zabel M (eds) (2006) Marine geochemistry, 2 edn. Springer (Textbook)

Stephansen C, Bjørgesæter A (2017) WP2a—seafloor compartment sensitivity testing and Norwegian Sea Test Case data. ERA Acute Project report ERA Acute 2A-3