SENSE
Sensitive Environments Decision Support

Akvaplan-niva

Report

WP 5 ERA Acute Methodology Uncertainty Feasibility Study

Document Number 60043.05

For Norsk Olje og Gass

Akvaplan-niva AS

Report Title:

WP 5 ERA Acute Methodology Uncertainty Feasibility Study

Author(s): Cathrine Stephansen Anders Bjørgesæter (Acona AS) Odd Willy Brude (DNV GL)	***Akvaplan-niva Report No.:*** 60043.05 ***Date: 09.05.2019*** ***No. of pages: 42*** ***Distribution:***
Client: Norsk Olje og Gass	***Client reference: Egil Dragsund***

Summary:

Based on sensitivity testing from phase 4, or qualitative assessment, the ERA Acute calculation parameters have been uncertainty scored and recommendations for handling of uncertainty are made.

Project Manager: ***Cathrine Stephansen***	***Quality Control:*** ***Geir Morten Skeie***

Table of Contents

1 Abbreviations and definitions

Abbreviation	Description
A1, A2, A3	Levels of increasingly detailed impact calculations in ERA Acute
ALARP	As Low As Reasonably Practicable
APN	Akvaplan-niva
CBR	Critical Body Residue
DSHA	Defined Situation of Hazard and Accident
ERA	Environmental Risk Assessment
ERA Acute	Environmental Risk Assessment model for acute oil spills.
ERACA	Environmental Risk Assessment and Contingency Analysis
ESI	Environmental Sensitivity Index
FPRV	Factor Prioritization by Reduction of Variance
GIS	Geographic information stem
GLR/GOR	Gas Liquid Ratio /Gas-Oil Ratio)
HPHT	High Pressure High Temperature
IMR	Institute of Marine Research (Norway)
LC	Lethal Concentration
MEMW	Marine Environmental Modelling Workbench (SINTEF-model)
MIRA	Environmental Risk Assessment method, current standard on the NCS
NCS	Norwegian Continental Shelf
NEA	North-East Arctic
NINA	Norwegian Institute for Nature Research (www.nina.no)
NORA10	Hind-cast archive of wind data
NOEC	No Observable Effect Concentration
NS	North Sea (data sets)
NSS	Norwegian Spring-spawning (herring)
ODS	Oil Drift Simulations
OOBN	Object-oriented Bayesian Networks
OSCAR	Oil Spill Contingency And Response Model (SINTEF-modell for olje-driftssimuleringer)
P95, P99, P100	Percentile values (95-, 99- and 100-percentile values)
PAH	Polycyclic aromatic hydrocarbons
PL	Production Licence
QSAR	Quantitative Structure- Activity Relationship
RDF	Resource Damage Factor
SA	Sensitivity Analysis
SD	Standard Deviation
SE	Standard Error of the Mean
SEAPOP	Norwegian Institute for Nature Research Program for monitoring of seabirds (www.seapop.no)
SF	Sensitivity Factor
SSD	Species Sensitivity Distribution
SVIM-archive	Hindcast-archive from numeric ocean models, containing current data with 4 km resolution (among other data)
SYMBIOSES	SYsteM for BIOlogy-based asSESsments
THC	Total Hydrocarbon Content

TOC	Total Organic Carbon
TLR	Threshold Limit for Recovery
UA	Uncertainty analysis
VEC	Valued Ecosystem Component

2 Summary

This feasibility study for handling of uncertainty in ERA Acute builds on the experience from case studies, comparisons and uncertainty studies, carried out in former phases of the ERA Acute development project. The ultimate ideal goal would be to be able to estimate quantitatively, the overall uncertainty of the model endpoints in the form of impacts, restoration times and resource damage factor values. However, although such studies have been a priority throughout development, the model is still “young”, and there is currently not enough data to quantify uncertainty fully. It is, however important to make sure that using ERA Acute does not under-estimate risk when using it in risk management situations.

Gaining knowledge about the sensitivity of model functions, and thereby scoring the input parameters according to importance for model results, has therefore been an important part of this study. A scoring system developed by DNV GL for MIRA has been used and the ERA Acute-factors were placed in this system. At this point, recommendations are given for ensuring comparability and quality, as well as identifying parameters for which more knowledge would improve accuracy. Using uncertainty factors is currently not recommended or deemed feasible.

3 Scope of Work – Uncertainty in Phase 5

3.1 WP5 ERA Acute method uncertainty study description

The uncertainty of the ERA Acute methodology has been addressed through the recently completed ERA Acute DEMO2000 JIP, denoted Phase 4 of the ERA Acute project (Sea JIP summary report, Stephansen et al., 2018). The work on uncertainty in Phase 5 builds further on the experience gained in studies on sensitivity and validity testing performed in Phase 4 (Brude & Bjørgesæter, 2016; Bjørgesæter & Damsgaard-Jensen, 2018; Bjørgesæter et al. 2018; Brude & Rudberg, 2018; Stephansen & Bjørgesæter, 2017 and Stephansen, 2017).

The aim of the phase 5 study is to assess the degree of uncertainty of the ERA Acute calculations and input parameters used to calculate impact and restitution time, with a goal of devising a way forward to handle uncertainty in risk assessments for which experimental data only exist for a few parameters. Thus, the purpose is to evaluate the robustness of the methodology, related to use of the precautionary principle and to propose a way of handling uncertainty that ensures that risk is not under-estimated.

The proposed work was divided into two parts, this report describes the first part of the process, which is the scope of work.

3.1.1 Part 1 Feasibility study (this report)

There are many "layers" of uncertainty in the input data to the ERA Acute model. The goal of this study is to determine at which level uncertainty handling is possible and to evaluate (if relevant) how to find reliable and significant uncertainty levels for the key parameters used in ERA Acute, to a practical level within the scope of the project. Parameters should be prioritized for uncertainty determination, based on of their importance to the output (ref results from ERA Acute DEMO2000 JIP). Knowledge gaps and a way to handle unknown (unquantified) uncertainty will be identified to the extent possible within scope of this feasibility study. Recommendations for handling uncertainty in the ERA Acute model will be proposed based on the findings, knowledge gaps and discussions.

Full sensitivity testing of impact and restoration functions were carried out most extensively in previous phases for surface and sea floor compartments.

3.1.2 Part 2: Further uncertainty analysis and uncertainty handling method (future work)

The results of the feasibility study can be used to determine the scope for an uncertainty analysis of greater depth if deemed necessary. It is important to determine the scope of any full study at a level of detail that is obtainable and practical, with the goal to achieve output that provides a reliable risk assessment and does not underestimate the environmental risk. A further study as part 2 could involve determining actual quantitative uncertainties and constructing the uncertainty handling method including to evaluate use of potential safety factors etc. through literature search, statistical studies and further sensitivity testing.

3.2 Goal of the Phase 5 Uncertainty Study

The starting point of the Phase 5 Uncertainty Handling *feasibility* study builds on the project parts carried out in Phase 4, where the 4 compartment functions were studied with respect to *sensitivity* to the input parameters. The testing was carried out using deterministic and statistical tests and conclusions were used for calibrating the model as far as possible within the scope of Phase 4. (Bjørgesæter & Damsgaard-Jensen, 2018; Stephansen & Bjørgesæter, 2017)

Uncertainty is a topic that can be studied at many levels. For every new data set or parameter that is used in the analysis, new uncertainties may arise in the results.

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 – as lack of certainty or knowledge about what the value of the parameter/data truly is. Uncertainty is parameter-specific.

Sensitivity Analysis is how the model's response can be apportioned to changes in model inputs. It is algorithm-specific.

Testing and validation has been carried out for all four compartments and contains both deterministic and stochastic testing of the sensitivity of the calculation functions to variation in parameter values, and some statistical tests of sensitivity and uncertainty. The individual testing reports are:

- WP2 a Sensitivity testing: Sea surface, water column and shoreline report (Bjørgesæter & Damsgaard-Jensen, 2018)
- WP2 a Sensitivity testing: Sea floor sensitivity testing report (Stephansen & Bjørgesæter, 2017)
- WP2 b Field validation: All compartments validation report (Bjørgesæter et al. 2018)
- WP2 c Comparison with MIRA report (Brude & Rudberg, 2018)
- WP2 c Norwegian Sea test case for Sea floor report (Stephansen, 2017)

These reports together form the experience gathered so far on the use of the ERA Acute model. The testing has resulted in increased knowledge of which parameters have the largest effect on the results of the impact and recovery calculations. These parameters need to receive the highest focus in future work and will be the values for which finding local/regional values will reduce the uncertainty of results.

Given the amount of data that are used in the ERA Acute method, and the lack of certainty of the parameter values and data sets given as input; uncertainty has to be handled although we do not know what the actual degree of uncertainty is for each input. Data sets are here included in the term "input" for simplicity, ERA Acute uses input that are both data sets in .csv or netCDF.-formats, oil drift data in .txt or netCDF (sediment)-format or VEC-parameters in .csv text files. Data sets that have a degree of uncertainty range from environmental data (e.g. substrate properties of sea floor) to sea bird abundance distributions. Parameter values regarding e.g. sea bird sensitivity are based on the fortunately limited number of actual oil spills, and the limited comparability between the spill incidents. Although extensive research is carried out following an oil spill, there are natural limitations to the statistical strength of the results. Some results are modeled estimates of losses, based on limited and uncertain counts and measurements. For example, oiled dead seabirds at sea will sink. In most instances the "control" is limited as the baseline pre-spill status may be less known. Even in the case of parameter values based on laboratory experiments, such as the *plet* values based on SSD curves constructed from experimental toxicity studies, the uncertainty is an issue.

Uncertainty in toxicological risk assessments of single-substance chemicals, where the algorithms are mathematically as simple as a PEC/PNEC approach, are sometimes handled by applying *safety factors* to ensure that adequate conservatism is applied, to account for lack of certainty in the toxicity tests behind the PNEC values, as well as for *intra and inter-species* variation in sensitivity. Safety factors have been evaluated as a solution for ERA Acute, in this pilot study. Another common approach is to represent uncertainty in input data or variables with a probability distributions or by using estimates of a minimum, intermediate and maximum value. The distribution or values may be based on historical data, simulations or subjective evaluations. The uncertainty in the input data or variables is reflected in the output resulting in a larger range of possible output values making it more likely that the "true output value" is located within the results. The distribution of output values is valuable when interpreting the results of the analysis, but some data reduction is needed to reveal the distribution in a compact form. Common statistic to use is *mean* with *standard deviation* and *percentiles*.

ERA Acute is even more complex with respect to the diversity of the compartments and the high number of functions used. The input parameters (values and datasets) that are used are as mentioned, based on knowledge from non-repetitive and diverse incidents. We therefore have a challenging situation with respect to validating the results of the method and applying the results with an acceptable level of uncertainty. In such a case of applying a complex model and multitude of uncertain parameters, it is important to realise and accept the following:

- ***We do not know the "true risk" as a number as such***

Given that an oil spill occurs, of a certain rate, duration, location and time of the year, it is not possible to know the true outcome of impacts and recovery of natural resources. As mentioned, the research following real spill incidents are based on counts of animal fatalities, measurements, modelling etc. Results are often presented in ranges from minimum to maximum estimates of e.g. losses, oil amounts, etc.

The goal of this work package is therefore to examine the feasibility of assuring that ERA Acute gives results:

that ensure that risks are not under-estimated, whilst still *differentiating the risk results enough for them to be applicable in decision-making*.

This means that it is not a viable option to apply safety factors upon safety factors uncritically. Similarly, including very large uncertainty in input parameters will result in very large uncertainty in the results. The final chapter of the validation report from Phase 4 described a proposed way forward in the form of a feasibility study (project) to determine how uncertainty of the parameter values should be handled in the model, and to further identify knowledge gaps. While acknowledging that the study of uncertainty in all details of model parameters would be an almost indefinite task, a practical and scientifically robust way forward needs to be found. This is the purpose of this study.

4 Uncertainty issues in ERA Acute

In this section we go through the main groups of input data and issues and describe them according to handling uncertainty at a practical/pragmatic level suitable for risk assessment purposes.

Definitions used:

A *parameter* is a quantity that has a true value. The value may have a level of uncertainty which may or may not be quantifiable to a certain extent. Input parameters for functions that should have different values e.g. for different species or species groups, substrates etc. are typical parameter values given in the compartment lookup-tables.

Variables can have different values due to variability. This can be spatial or temporal variation, e.g. for the driver data for oil drift simulations (wind, currents, ice) or species data sets that have temporal variations.

Model output and function uncertainty, due to lack of knowledge of the magnitude of the model error. All mechanistic modelling is a simplification of reality, meaning that the actual outcome of an event can never be predicted perfectly. The model output also relies on the training, knowledge and experience of the expert setting up the model.

4.1 Reducing model input uncertainty – external data sources

By "external" data, we mean data that are obtained through use of external models, which are not specific to ERA Acute.

4.1.1 Input data to oil drift modelling

Uncertainties in results may also arise from uncertainties in input data relating to the scenario, such as:

- rate-duration matrix, probability distribution between rates and durations
- speed and direction of wind and sea current
- spill frequencies,
- choice of reference oil type etc.,

These are important factors that have a large effect on the outcome, based on experience. *However, these factors are not included in the present study as they are not ERA Acute specific.* DNV GL (Kruuse-Meyer 2015) gave an account of these issues relating to their use in oil drift simulations for MIRA and DNV GL's evaluation is also valid for ERA Acute.

Uncertainty in the inputs named above should be handled by choosing conservatively and taking caution when comparing cases. Uncertainty in the rate-duration matrix is handled using a probability distribution and uncertainty or variation in the wind and current is handled by performing many simulations with different start dates. To ensure adequate conservativity, it is recommended e.g. to choose a representative oil type that both fits the expected fluid properties and is conservative with respect to emulsion formation and weathering properties.

Currently, the rate-duration matrixes and their probability distributions vary greatly between the blowout and kill studies that are carried out by different analysis companies (which are supplied as input, this is also discussed briefly in the report by DNV GL (Kruuse-Meyer 2015). A best practice process should be used. There is a guideline available for the NCS describing how to carry out blowout and kill analyses (rate-duration matrices) (NOROG 2017) etc.

4.1.2 Input data based on modelling

4.1.2.1 Oil drift simulations

Oil drift simulations (ODS) are carried out using a complex hierarchy of calculations, based on variation of input data of wind, currents, ice etc which are *variables* in the model input. These driver data are hind-cast archives of measured data from meteorological institutes, highly variable in nature. The complex functions determining all the physical and chemical changes that determine the fate and distribution of the oil in the four compartments are based on research, and each have an array of potential uncertainties.

It is outside the scope of the project to assess all the uncertainties involved in the actual modelling, so the early-made recommendation for use in ERA Acute was based on the need for a practical approach to handling of uncertainty due to ODS results as input to ERA Acute:

To reduce unquantified uncertainty

- Use well-researched and verified ODS models
- Use the same ODS models for risk assessments that are to be compared
- Use the same driver data in the ODS for risk assessments that are to be compared (same data source, temporal/spatial adaptation/resolution and same time series)

To include uncertainty in the predicted results due to variation in the environmental input data

- Use several simulations drawing individual historic start-dates to provide a range of outcomes on which to calculate statistical results, minimum, maximum, mean, average etc.

As a practical example of uncertainty reduction, uncertainty due to different users obtaining different results from the ODS used on the NCS, is handled in a working group that gives advice regarding the Best Practice for the use of the OSCAR model in risk assessments. This group has studied the effects of various user-determined OSCAR model settings to provide a Best Practice for use in ERA Acute, specifically for the input values used in ERA Acute; Oil film thickness, coverage, oil concentration in the water column, oil concentration in sediment and oil amounts on shore. Some input values have been used in MIRA also, and have been tested earlier.

While all the uncertainties involved in user-defined settings for oil drift simulations cannot be studied in the current project, the following user-defined parameters have been tested by the best practice group following the release of OSCAR version 10.0.1.

Refinement

The version included some necessary updates for correct calculation of coverage and exposure time, and the effects of using different pre-processing thresholds and refinement were studied. As testing the refinement factor was not included in the previous Phase 4 sensitivity testing, updated recommendations are given for ERA Acute purposes with the goal of reducing uncertainty of the calculations and ensuring adequate conservativity (not underestimate risk).

Safety factors added to the input oil parameters to account for uncertainty are currently *not* used and are not recommended applied to these types of data in ERA Acute. Uncertainty in the rate, duration and release depth (topside versus subsea) is managed using probability distributions and variation in the wind and current is managed by performing many simulations with different start dates. The range of outcomes represent the high variability in weather situations etc., and results are available at a single-simulation level, so that ranges of outcomes can be provided: Maximum impacts, means, averages or percentile values are relevant. Using a high number of start-dates improves the statistical strength of the analyses.

Uncertainty in parameters are handled using Best Practice recommendations. A summary of these relevant for ERA Acute is given below.

Common for all compartments:

- Model grid 3x3 km

Driver data 4x4 km **Model parameters**:

1) The refinement parameter is set to 3 (Best Practice)
2) Internal time step is set to 15 minutes (Best Practice)
3) Lower concentration limit (ppb) is set to 1 (New standard Value)
4) Trivial mass limit is set to 1E-08 (New parameter/New standard Value)
5) Use distance to nearest neighbour is turned Off (New recommendation from SINTEF)

Stochastic Simulation (Create Output):

1) Enable deprecated UTM projection calculations is set On (New feature)

ERA Acute and MIRA setups for ODS in OSCAR lead to very different results in the amounts of oil on the surface (used in MIRA), calculation of film thickness and the coverage above the threshold film thickness, see the case studies for Busta (Stephansen, 2019) and Best Practice document https://norskoljeoggass.no/miljo/handboker-og-veiledninger/beste-praksis---oljedriftsmodellering/ . It is therefore vital for the results that the correct setups are used.

4.1.2.1.1 Surface compartment

Pre-processing thresholds for oil film thickness are used to obtain coverage and exposure time for oil in each grid cell that may be lethal for seabirds and marine mammals. The recommended threshold value for seabirds is 2 µm and the recommended threshold value for marine mammals is 10 µm.

A practical approach to reduce the number of stochastic oil drift run is to use 2 µm as a lethal oil film thickness also for marine mammals. This is believed to be a conservative threshold thickness for marine mammals based on previously work in ERA Acute (Bjørgesæter et al, 2018) and compared to similar models as the ERA Acute (Brude and Rudberg, 2017).

The importance of threshold the threshold thickness (T) was investigated in Phase 4 (Bjørgesæter & Damsgaard Jensen, 2018). The threshold thicknesses investigated were 0, 2, 4, 6, 8 and 10 µm. The endpoints investigated were:

(1) Number of grid cells with oil film thicker than T
(2) Sea surface area with oil film thicker than T
(3) Exposure time of oil film thicker than T
(4) Impact (population loss)

The effect of lowering the threshold level from 10 to 2 micrometres had a significant effect on all endpoints at a 5% significance level (ANOVA with a Tukey's range test). The mean impact estimated with a threshold thickness of 2 µm was on average 2.3 times higher than the mean impact estimated with a threshold thickness of 10 µm, ranging from 1.9 (Atlantic puffin, coastal dataset and grey seal) to 2.9 (black-legged kittiwake). In this test, the VECs distributed along the coast (i.e. typical coastal seals such as the grey- and harbour seal) was less sensitive to lowering the threshold thickness than the VECs distributed on the open sea.

The refinement parameter is set to 3. This yields a resolution of 3 km/3 or 1 × 1 km and thus a resolution of approximately 11% for coverage in the model grid cells.

4.1.2.1.2 Water column

For using the QSAR-based calculation of fraction killed during the simulations as *plet*, exposure calculation is turned On during stochastic simulations in OSCAR. Standard values to be used are:

a. Standard deviation: 32

b. Species sensitivity: 1

The species sensitivity can be adjusted to cater for individual sensitivities, otherwise the calculation of fraction killed follows the same SSD-curve as described in Nilsen et al. 2006.

4.1.2.1.3 Shoreline compartment

Preliminary recommendations are to use the same settings as are recommended under sea surface. The resolution of the GIS layer coastline definition used to make the habitat grid coastal definition is important for the shoreline results and should be selected carefully. Further testing is needed to provide full recommendations. Future possible inclusion of back-washing and oil per ESI-type in oil drift modelling will improve the results.

4.1.2.1.4 Sea floor compartments soft substrates and lower water column

To include soft sediment substrates in the sea floor compartment, single simulations are currently used due to limitations in the stochastic model in OSCAR. There are usually fewer simulations carried out, as each simulation takes longer time to run, and this will increase uncertainty. Future plans for using batches of single simulations run in "ensemble" runs of OSCAR will include sediment in oil drift simulations. For the best possible accuracy for the water column sub-compartment of sea floor, concentrations of THC should be reported for the lower water column. Future work should consider using fraction killed if this could be possible to limit to the lower water column. The fraction killed calculated by OSCAR results in a higher lethality than using the time- averaged THC as a concentration, and this is therefore recommended to ensure that risk is not under-estimated.

4.1.2.2 Resource data sets based on modelling

Distribution data such as seabird abundance distributions are based on modelling using observation data such as colony counts, logger data etc. Currently, the data used on the NCS represent the predicted abundance (population fraction) per month in each grid cell. This number has an inherent associated uncertainty. However, it would be outside the scope of this study to quantify the uncertainty of all available data sets, the interested reader is referred to documentation of the relevant data set. Resource data are included in the uncertainty handling scoring system proposed by DNV GL for the general input data. For practical uncertainty-handling within ERA Acute use, the following recommendations apply.

To reduce unquantified uncertainty

- Use well-researched and accepted data sets
- Use the same data sets for risk assessments that are to be compared

To reduce uncertainty in the predicted results based on natural variation in the resource input data

- Use several data sets representing not average monthly or seasonally predicted abundances, but daily distributions. This is currently not implemented in ERA Acute but is proposed as a development. Daily distribution data sets are available for several alcid species as well as for larvae of some fish resources. The results will provide a range of outcomes on which to calculate statistical results, minimum, maximum, mean, average etc. However, these data also have inherent uncertainties, as they are modelled data.

4.1.3 Hind-cast data representing historic temporal and spatial variation

Driver data from hind-cast archives are input to the ODS. The single simulations described above therefore represent the historic weather/ocean current situations with a temporal and spatial resolution that is adequate.

To reduce unquantified uncertainty in the results arising from variation

- Use well-researched and accepted data sets
- Use the same data sets for risk assessments that are to be compared
- Choose a high enough resolution (temporal/spatial) optimised for practical purposes.

These types of data are used directly in the ODS, and thereby provide a range of results as described above, even though the values in themselves may be uncertain. Use of safety factors is not considered necessary for this type of data.

4.1.4 Fixed parameter values in gridded data sets (one parameter value per grid cell)

Values in grid cells such as the length of different shoreline habitats, may be based on modelled data and calibrated based on research and observations, but presented as a single value or parameter in a cell (for a month if relevant) and only one data set is used (not a multitude as above). Data sets may have uncertain values, resulting in a corresponding uncertainty in the resulting calculations.

To reduce unquantified uncertainty in the results arising from uncertainty in the cell values (lack of knowledge and spatial inaccuracy)

- Use well-researched and accepted data sets
- Use the same data sets for risk assessments that are to be compared
- Choose a high enough spatial resolution optimised for practical purposes.

For this type of data, uncertainty could be handled by applying uncertainty factors representing the degree of uncertainty.

Example: ESI data set for shoreline.

4.2 Reducing model input uncertainty – ERA Acute-specific inputs

4.2.1 Parameter values used as numerical factors in ERA Acute calculations

ERA Acute specific parameters are parameters defined in the lookup-tables and are used in the impact, lag, restitution, recovery and resource damage factor calculations. Their values are based on research in literature and expert judgment with varying quality and reliable data available. Typical values are e.g. the *plet* values of surface organisms (*pbeh* and *pphy*) and parameters relating to lag- and restoration time parameters for sea floor hard bottom, slope factors for shoreline, and fish model input data. These data have inherent uncertainty and the model's sensitivity to variation in these values was tested in Phase 4 (Bjørgesæter and Damsgaard-Jensen, 2018, Stephansen & Bjørgesæter 2017 and Stephansen, 2017).

For the surface compartment, uncertainty in the values of physiological sensitivity and **behaviour**-related probability of being exposed (pphy and pbeh) is currently handled by applying three values for each species, a high, low and intermediate (best guess) value, which provide the user with the possibility to present these as separate risk estimates. This is currently not done in other compartments and doing so would mean major changes to the functions and program algorithms.

4.3 Model framework uncertainty and sub-model niche uncertainty

Uncertainties in functions representing impact and recovery relationships arise because mechanisms of actions that are less well known may give uncertainty in the actual functions and quantified relationships between factors. However, this would be viewed more as a functional error in the model and is not included in this study. New research that leads to major changes in the knowledge would lead to the need for major revisions in the methodology.

During development of the total framework of ERA Acute sub-models have been used that may have been originally developed for a different use, other temporal or spatial scales and therefore there may be uncertainty related to this new application. This is called model niche uncertainty.

4.4 Choice of model and analysis resolution

Grid cell size for the standard NCS analysis grid is currently 10 x 10 km. A higher resolution could generally be assumed to be more accurate, but the user has to be careful of resolution artifacts and false resolution. Sound, transparent spatial resolution should match the data sets correctly.

A project is being established to include dynamic modelling with daily temporal resolution in ERA Acute, which will allow better temporal matching between the current/wind data and the VEC-data. For the currently implemented impact calculations in the water column compartment, large differences are seen between the results using time-averaged THC-concentrations vs. dynamic modelling using accumulated mortality throughout the simulation. Time-matching and challenges with the matching even of dynamic data will e.g. be discussed in the ERA Acute dynamic risk assessment project. A myriad of ideal, practical and mathematical implications and considerations are relevant. Practical limitations are e.g. modelling time versus actual benefit of higher resolution. Choice of resolution should be discussed before data sets are chosen and adapted to ERA Acute purposes.

High spatial resolution may be recommended for small spills or spills close to the coast. However, increases in spatial or timestep resolution should not override the temporal resolution represented by the number of simulations necessary to represent the weather diversity. Differences in weather conditions during the analysis period has a high impact on oil trajectories and thereby also on risk.

4.5 Other approaches to risk assessment with uncertainty

For risk assessments in the case of uncertainty, Object-oriented Bayesian networks (OOBN) are sometimes used. Bayesian networks and decision graphs are evolving, having been primarily put to use in the field of decision support systems (Jensen and Nielsen, 2007). OOBNs consist of a structure of connected nodes (sub-structures or classes) that can be seen as a function that given a certain input provides a probability distribution over a set of variables (Jensen and Nielsen, 2007). Such networks can be built for risk relationships, however the method can be faulty if we don't have full overview of the mechanisms and probabilities, and the probabilities will also be uncertain. Many OOBN-studies do not take this very well into consideration, focusing more on the network building.

For the purpose of managing risk in ERA Acute, the exercise of setting up an OOBN for ERA Acute risk functions could have its use in determining uncertainty within the whole system of functions, but given that much of the uncertainties are not quantified, this would be outside the scope of the present feasibility study.

The mechanisms of the individual functions and how the variation of the parameters influence functions were studied in Phase 4 sensitivity studies, using a deterministic method looking at the range of outcomes from the functions in ERA Acute, and stochastic testing for ranking the parameters with respect to importance.

If it becomes necessary to quantify uncertainty more than to ensure conservativity, an OOBN can be set up and one could attempt to quantify the individual uncertainties. However, given the complexity of ERA Acute with respect to number of functions, the task is formidable, and probably would not give results that can be used practically without extensive research.

4.6 Sensitivity of the model functions to variation

The factors that are used in the risk functions that are specific to ERA Acute were tested with respect to importance for model function sensitivity in Phase 4, WP2. The testing was extensive for surface (Bjørgesæter and Damsgaard Jensen, 2018) and sea floor (Stephansen and Bjørgesæter, 2018) compartments. The results are used here to assess qualitatively using the DNV GL scoring system to identify particularly important parameters for more research to reduce uncertainty (for example by including scientific research institutions in discussions) or for directly handling the uncertainty by using more conservative values within the relevant range.

The sensitivity of the sea surface, water column and shoreline input factors were studied in Bjørgesæter and Damsgaard-Jensen (2018) and for seafloor in Stephansen and Bjørgesæter (2018).

For water column, some testing was carried out for the gate model in Bjørgesæter and Damsgaard Jensen (2018) whereas the differences between the two water column impact calculations have been studied in a comparison with SYMBIOSES (DNV GL, 2018), and case studies. However, the impact functions have not been sensitivity-tested, as the SSD-curves used are well documented (Nilsen et al, 2008). The water column fish stock restoration model ("gate" model) testing has been used to score the parameter values for water column.

4.7 Amplifications of uncertainty

Summarizing values over grid cells will amplify over- or underestimation, it is therefore important to have input data that are as accurate as possible. For some uncertainties, however, such as the uncertain prediction of risk that tied to variations in oil trajectory modelling based on the variation of weather conditions is mitigated by using a high number of simulations representing different historic start-dates.

Results from single cells, and single simulations are averaged over simulations, averaged for scenarios using the results of all the simulations, and statistics should be provided with maximum, minimum and/or percentile values to indicate the range of outcomes predicted. Using 95-percentile values instead of the worst-case maximum value is often used as a "moderate worst-case approach" when the goal is not to under-estimate risk.

5 Scoring system based on phase 4 uncertainty and sensitivity testing

5.1 Approach

Model functions in ERA Acute describe the current understanding we have about the mechanisms of impact and recovery as well as the quantified relationships. In their internal guideline, which has been made available to the project, DNV GL call these *assumptions* (Kruuse-Meyer 2015). Each assumption is characterised according to the degree of certainty about them. Based on the similarity between the MIRA ERA method and ERA Acute on a number of areas (both are ERA methods) a similar approach is proposed for ERA Acute, with some alterations described herein. The approach is then also extended to include the factors that are specific for ERA Acute.

To investigate the uncertainty in the model output that is generated from uncertainty in parameter inputs, an uncertainty analysis (UA) is carried out. The purpose of UA is to quantify the degree of confidence in the existing experimental data and parameter estimates, and the importance of the parameter in the functions within the model. To assess how variations in model outputs can be ascribed to, qualitatively or quantitatively, to different input parameters, a sensitivity analysis (SA) is carried out, and is the step that follows UA (Marino et al. 2008). This was carried out in Phase 4 (Bjørgesæter and Damsgaard-Jensen (2018) for surface, shoreline and water column compartment, and Stephansen and Bjørgesæter (2018) for sea floor soft and hard substrates.

5.2 Strength of knowledge

Uncertainty related to strength of knowledge – how certain are we that the assumptions are correct?

DNV GL use the following qualitative criteria for scoring the function or parameter as strong or moderate/weak. Conditions for scoring as *Strong* are:

- The assumption is seen as very reasonable.
- There are sufficient amounts of reliable data available.
- There is broad consensus among experts regarding the assumption.
- The phenomena involved are well understood; the model used is known to give good predictions.

If one or more criteria are not fulfilled, the strength of knowledge is moderate/weak. Although ERA Acute bases its functions on peer-reviewed literature, the facts are that for many of the parameters there is limited actual data available and the strength of knowledge may be characterised as moderate or weak. Due to the fortunate few incidents, the model cannot yet be known to provide good predictions of the impact should a real incident occur, this is an inherent property of oil spill risk assessments and should not be used as an argument to not use the results. Although it is important to realise that if the strength of knowledge is low and the model is a weak representation of reality, determining the sensitivity of the model to the parameters is an academic exercise. However – this cannot stop us from seeking to improve the basis for decisions in an economically important industry. The key is, as mentioned before, to apply an adequate level of conservatism in the case of uncertainty, with the goal of ensuring that the model does not under-estimate risk.

5.3 Variation of the parameter – deviation from base case

In Kruuse-Meyer, (2015) DNV GL call this “deviation from assumption”, characterized as an assessment of the deviation from a base case value. A numerical assessment of the deviation is obtained by altering the assumption, i.e. assessing how much the variable deviates from the “base case” is done in DNV GL’s method by classifying them in three categories, depending on how likely the parameter value is to deviate from the base case, meaning it has a high natural variation in the parameter. An example of such values is the TOC-value in the sea floor soft substrates which has a recommended average (base-case) value for use in ERA Acute for each substrate, but which has high natural variation between locations):

- High: The parameter is very likely to deviate from the assumed “base case”, large deviations
- Moderate: The parameter is likely to deviate from the assumed “base case”, moderate deviations
- Low: The parameter is unlikely to deviate from the assumed “base case”, small deviations

Deviations should be quantified as far as possible, by researching the data well before use. As basis for the Monte Carlo simulations that were carried out in Phase 4, a range of values was found for many of the data.

Options: It is proposed to use a derivation of the *variance, the standard error of the mean instead of the minimum and maximum values* indicating range.

Statistical *variance* gives a measure of how the data distributes itself about the mean or expected value. Unlike *range* that only looks at the maximum and minimum values, the variance looks at all the data points and then determines their distribution.

Standard deviation is the square root of the variance and is a measure of variability commonly used to measure confidence in statistical conclusions about a value. A low standard deviation indicates that the data points tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values. A useful property of the standard deviation is that, unlike the variance, it is expressed in the same units as the data (Wikipedia).

Standard error of the mean is a derivation of the Standard deviation. When a sample mean is calculated, the goal is to assess not only the mean of this particular sample, but the mean for individual values of this type, i.e. from the population from which the sample comes, to be able to generalize from the results of the sample. (The link https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1255808/ gives an easy-to-read distinction between the two). The sample mean will vary from sample set to sample set, in some cases extremely. The standard error of the sample mean depends on both the standard deviation and the sample size, by the simple relation SE = SD/√(sample size). *To find out how widely scattered the measurements in a set of data are, we use the standard deviation. If we want to indicate the uncertainty around the estimate of the mean measurement, we use the standard error of the mean.* The standard error is most useful as a means of calculating a confidence interval. For a large sample, a 95% confidence interval is obtained as the values 1.96×SE either side of the mean.

We therefore propose to use the *Standard error of the mean* to indicate the uncertainty of the factor value estimates, and the *standard deviation* as measurement of the variation between the samples. The uncertainty relating to the quality of data and sampling techniques is beyond the scope of this study, but low-quality data should always be excluded from data sets.

Example (see Table 1): Mixing depth (BDepth) is a factor which is important in determining the concentration of oil in soft substrates, as oil reported as mass/area in OSCAR needs to be converted to mass oil /mass sediment. The table was presented as Table 4 in Stephansen & Bjørgesæter (2018) and shows data from Teal et al, 2008. It is an example of how variable parameters in nature can be. The results available were 7 different regional studies presented with the parameters mean mixing depth, standard deviation, number of samples in the study, as well as maximum and minimum mixing depths. Teal et al 2009 used these regional studies to derive a global mean mixing depth of 5.75 cm and the same statistical parameters were calculated. We know that sampling of the bioturbation (mixing depth) can lead to errors in the measurements, and thus, the actual values may well be uncertain due to errors in the sampling technique, but we also know that the bioturbation depth, being dependent on many factors, one of them being presence of burrowing fauna of various sizes, it would be only natural to expect high variation in this value locally and regionally. In the FPRV (factor Prioritisation by Reduction of Variance) test (see below) BDepth accounted for 40 % of the variation of the end result of the concentration calculation (mathematically, *partly* because it varied).

Being both a parameter with high variability (natural) and of high importance with respect to calculation sensitivity, mixing depth is a good example of a very variable factor for which the extra effort to find local, but conservative values of high quality could be worth the effort to reduce uncertainty. This would be a recommendation for *uncertainty handling* for this parameter, as only using the standard error of the mean for a *global* value would mean that one has low confidence in the global mean value. However, in the currently provided input files it is the global mean that is used as a default value that could be changed by the user. Note that the SD for the global mean cited by Teal et al. (2009) is almost the same as the mean, indicating that the data vary greatly. The same is true for the 135 samples making up the North Sea mean and SD. Teal et al. also point out the inherent uncertainty related to sampling methods and regions with small samples.

Table 1. Example: Mixing depths of sediments in different regions and seas (Teal et al. 2008). The mixing depths found in the studies vary between 32 cm (maximum value in one of the Gulf of Maine samples) and 2 mm (minimum in one of the Baltic Sea samples). Note the very different mean values and the high variation in the SDs of the sample, note also that sample sizes vary in the regions for while data were compiled.

Region	Sea	Mixing depth (mean) (BDepth) (m)	Standard dev. (SD)	No. of samples (n)	Standard error of the mean (SE)	Mixing depth (high value) (m)	Mixing depth (low value) (m)
North Atlantic temperate	Gulf of Maine	0.24	0.0822	5	0.037	0.3222	0.1578
North Atlantic temperate	Baltic Sea	0.009	0.007	40	0.0011	0.016	0.002
North Atlantic temperate	North Sea	0.027	0.023	135	0.0020	0.05	0.004
Temperate South America	Temperate South America	0.064	0.027	10	0.0085	0.091	0.037
Temperate	Temperate	0.008	0.018	5	0.0081	0.026	-0.01
Polar	Polar	0.023	0.003	6	0.0012	0.026	0.02
Southern Ocean	Southern Ocean	0.028	0.013	12	0.0038	0.041	0.015
	Global	**0.0575**	**0.0567**	**791**	**0.0020**	**0.1142**	**0.0008**

For some parameters the variation was studied in Phase 4.

5.4 Sensitivity

The sensitivity of the model to variation in parameters is classified in three categories. DNV GL (Kruuse-Meyer, 2015) describe the sensitivity in terms of whether a deviation from the base case can result in an altered *conclusion*. In MIRA this could e.g. be a change in impact category, as the impact-function is non-continuous.

- High: Small changes in the parameter from the base case value may result in altered conclusion
- Medium: Relatively large changes in the parameter from the base case value may result in altered conclusion
- Low: Large changes in the parameter from the base case value are needed to alter the conclusion.

The term “conclusion” is less clearly defined for ERA Acute than for category-based ERAs, unless the resulting values are classified, as is the case when using risk matrices where the user has to define limits for each of the categories.

In ERA Acute terms, using a continuous function for most impact calculations (not impact classes) and some of the relationships are not linear either, although they are deterministic in nature. A deterministic model is one where the output of the model is completely determined by the input parameters and structure of the model. If the input is not changed, the model output stays the same, uncertainty in the model output (on the condition that the assumed relationship is correct) is solely affected by variation in the input parameter (*Epistemic*, subjective of reducible uncertainty) (Marino et al 2008).

Quantitative evaluation criteria are proposed to be set up for ERA Acute using the results of the sensitivity analyses carried out in Phase 5. The term “conclusion” is as mentioned difficult to use in the same way as above, and instead we propose to use the stochastic simulation results from the repeated random sampling Monte Carlo-simulations (MC) (uncertainty analysis (UA)) and subsequent sensitivity analysis (SA), described in Bjørgesæter and Damsgaard-Jensen (2018) for surface, shoreline and water column compartment, and Stephansen and Bjørgesæter (2018) for sea floor. Although a strong method for UA, the MC-simulations depend on an assumption of the probability distribution within the variance of the parameter values, an assumption must be made based on scientific knowledge. This brings some uncertainty into the uncertainty analysis.

Sensitivity analysis (SA) is a method for quantifying uncertainty in any type of complex model, such as e.g. ERA Acute. The objective of SA is to identify critical inputs of a model and quantifying how input uncertainty impacts model outcome(s). The sensitivity analysis in Phase 4 was performed by carrying out a Pearson and Spearman correlation coefficient, Partial Rank Correlation Coefficient (PRCC) analysis and Factor Prioritization by Reduction of Variance (FPRV). Combined, the methods were used to rank and quantify the most important input parameters for the quantitative result.

PRCC was used in phase 4 (Bjørgesæter & Damsgaard-Jensen 2018) to determine the statistical relationships between each input parameter and each output result while keeping all the other input parameters constant at their expected value. This allows independent effects of each parameter to be determined, even when the parameters are correlated. The interpretation of PRCCs assumes a monotonic relationship (relationship or function which preserves a given trend) between parameters (which are the case for all the 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 relationships (Marino et al 2008, Bjørgesæter & Damsgaard, 2018).

FPRV is carried out 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 results from the Monte Carlo simulations were used in a FPRV sensitivity analysis which in a second step, to find out which variable parameter that has the largest influence on the resulting endpoint of the formula. Where calculations were carried out in succession, combined formulas were used. 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 simpler deterministic tests holding one parameter fixed at a time, (One-At A-Time tests, OAT) that were carried out also in the sensitivity testing was useful to study the direct output of varying single parameters, and thus get better acquainted with the results of the individual calculations, but the method is unsuitable for handling the many dimensions of variation of the input parameters, for which the global sensitivity methods are used. The OATs were not used to limit the SAs.

5.5 DNV GL evaluation matrix system for risk assessments

In Kruuse-Meyer (2015) the different input parameters that are relevant for both the oil drift simulations, resource data sets and the actual MIRA calculations are scored based on the properties of belief in deviation from assumption (low, moderate/high), sensitivity with respect to assumptions (Low, Moderate/High) and strength of knowledge (Strong, moderate/weak). The factors are classified for the three properties and classified in 6 categories called "settings".

- Setting I indicates assumptions founded in high strength of knowledge (strong), low sensitivity with respect to the assumption and belief in low deviation from the assumption made.
- At the other end of the scale setting VI indicates moderate/weak strength of knowledge, moderate/high sensitivity and moderate/high belief in deviation from assumptions.

Table 2 DNV GL Classification system for evaluation of assumptions used in risk assessments (From Kruuse-Meyer, 2015).

Table 4-1 *Settings for evaluation of assumptions in risk assessments /2/.*

Belief in deviation from assumption	Sensitivity with respect to assumptions	Strength of knowledge	
		Strong	Moderate/Weak
Low	Low	Setting I	Setting II
	Moderate/high	Setting III	Setting IV
Moderate/High	Low	Setting III	Setting IV
	Moderate/high	Setting V	Setting VI

DNV GL have in this system classified the "larger" input data uncertainties according to this system:

- For input data related to the case and spill incident statistics (usually given by client) (Table 3)
- For resource data and their adaptation (Table 4)
- For settings in oil drift simulations (OSCAR) (Table 5)

The assessments in the tables were made using results from MIRA-analyses, however ERA Acute uses the same input data and the issues are the same for most of the parameters. The tables are therefore included unchanged (inserted with permission) with the following note to the reader, regarding handling of these uncertainties:

- As mentioned, a best-practice standardization process is recommended and will be undertaken for ERA Acute to make input data on rate-duration matrices and probability distributions more streamlined, using the same method.
- Common resource data for the Norwegian Shelf (NS) are being developed as part of the ERA Acute Phase 5 project. These are developed as the best available data from scientific expertise and adapted for the purpose under consultation with the same expertise. The data will be open to all and recommended for use in ERA Acute analyses on the NCS.
- Oil drift simulations. Since the DNV GL report in 2015, the OSCAR best practice group consisting of Akvaplan-niva, Acona and DNV GL have sensitivity-tested various model versions and driver data and have recommended a best practice to ensure that the model is used in the same way making results more comparable. Recommendations are given on model and scenario parameter settings, driver data, number of simulations etc.

These best-practice and common decisions may not eliminate uncertainty per se, but will contribute to better comparison between analyses by reducing the part of variability that is due to differences in the use of underlying input models or data. Although it doesn't make ERA Acute calculations able to calculate absolutely true risk predictions, it is sufficient for managing risk by identifying risk reductions etc.

In this study, we focus the efforts of this study on the classification of the ERA Acute-specific parameters related to the compartment impact and recovery functions (Table 9) in the same manner that was done by DNV GL for the input parameters that are common to MIRA and ERA Acute. For evaluation of the input parameters/data that are common to MIRA and ERA Acute, the DNV GL scorings can be used directly.

The system is used to identify the factors for which further research would reduce uncertainty. In the present work, we identify these in the comments and propose where it is particularly important to ensure that the values used are conservative to meet the goal of not underestimating risk.

Table 3. DNV GLs evaluation of input related to the case and incident statistics to environmental risk assessments. (Kruuse-meyer, 2015).

Table 4-2 *Evaluation of variables/parameters in the environmental risk analysis.*

Def par/var	Parameter/variable	Strength of knowledge	Belief in deviation from assumption	Sensitivity with respect to assumption	Setting (I-VI)
	Input from customer				
Parameter	Rates/durations – numbers and size	**Moderate/weak** Not all four conditions are fulfilled. The method for estimating rates is well documented and acknowledged in the industry, however often based on assumptions made regarding a rather unknown reservoir.	**Moderate** The values might deviate from its base case values if new well specific information is obtained.	**Moderate** Relatively large changes are needed to alter the end risk results. Change of durations seems to have greater influence on the results than rates.	IV
Parameter	Longest duration (time to drill relief well)	**Strong** The assumption is conservative to take into account the possible time to drill a relief well if a blowout occurs.	**Low** The longest duration is a conservative number and it is unlikely that it will vary much.	**Low** Large differences are necessary to change the end risk results.	I
Parameter	Spill frequency (generic vs. well specific frequency)	**Moderate/weak** The frequency does not fulfill all conditions to become strong.	**Moderate** Can deviate some if a new event is included or a more specific WRA is performed.	**High** Small changes may result in an altered environmental risk.	VI
Parameter	Oil type	**Moderate/weak** The oil type is rarely well known before an actual weathering study has been performed. Often reference oil is used to describe the oil in the reservoir, however this oil type is chosen based on limited available information. By having more information one will have a more solid base for choosing the most accurate oil type	**Moderate**. The deviation depends on the available information regarding the expected oil type. Given a change in some of the parameters that characterize the oil, a different oil type might be more applicable.	**Low** Oil characteristics affect the weathering of oil in the environment which will affect the end results. Minor differences in oil characteristics will give limited impact. However, are the oil characteristics very different from what assumed, the differences may be profound.	IV
Parameter	GOR	**Moderate/weak** The GOR is based on the available data; however these data might not be well understood. If several reservoirs are included in the analysis, this may not be reflected in the modelling.	**Low** New or adjusted information about the well may alter the calculated value.	**Low** Large changes required in order to alter the end risk results.	II
Parameter	Well location	**Strong** The drilling location is chosen as the modelling location. Small changes in the well location will not be considered a significant change.	**Low** The well location may vary some if the input data received was not accurate enough or the customer decides to move the location due to more updated information after the analysis was performed.	**Low/Moderate** If the location is offshore the answer would be low, however if the location is close to shore it is better to assume moderate (or maybe high). For a generic analysis it is reasonable to assume the low category.	I

Table 4. DNV GLs evaluation of input related to resource data set (and adaptations) to environmental risk assessments. (Kruuse-meyer, 2015).

Def par/var	Parameter/variable	Strength of knowledge	Belief in deviation from assumption	Sensitivity with respect to assumption	Setting (I-VI)
	Resource data				
Variable	Adjustment of datasets – adjusted to 100 % in one or more seasons, or not adjusted.	**Moderate/weak** Depends on the input received from the research institutes.	**Moderate/high** Dependent on the definition of a "population".	**Moderate/high** Directly linked to the calculated population loss.	**VI**
Variable	Adjustments of regions	**Moderate/weak** The definition does not necessarily take into account the biological principles and differences in various species. Logger data shows that species from different "regions" inhabit other "regions" in other periods of the year.	**Moderate/high** A population definition based on fixed regions is a simplification of the real world.	**Moderate/high** Linked to the adjustment of datasets and definition of population. What is a population and where is the "population" located at different periods of the year? To what degree does "populations" from different regions interact?	**VI**
Variable	Use of buffer zones	**Moderate/weak** Uncertainties in the wandering ranges.	**Low** The presence of the actual nesting sites is well known, as are the "normal" wandering ranges of different species.	**Low** The nesting colony is the area with highest density, and impact is most profound is this area is affected regardless of the use of buffer zones.	**II**
Variable	Coastal seabirds vs. pelagic	**Strong** Knowledge on the breeding population of seabirds' presence in the nesting period is good. Open sea dataset reflect a higher density in the coastal areas in this period.	**Moderate/high** Exact knowledge on the seabirds' presence is probably never possible to obtain.	**Low** Dataset give similar results in the nesting period (however more concentrated population in the coastal dataset gives higher risk for the nesting population).	**III**
Parameter	Vulnerability indexes	**Strong** Based on species specific evaluations with regards to vulnerability to oil pollution, however could be differentiated in terms of recovery time as well.	**Moderate** Currently, it is not common to evaluate on a regular basis, however it should be done in the future since it affects the results directly.	**Moderate/high** Higher index results in a higher loss.	**V**
Parameter	Years of data – fish eggs and larvae	**Moderate/weak** Strength of knowledge is generally good, but not in use as per today. Data exists; we could use the data to run dynamic modelling.	**Moderate/high** High deviations from present approach (statistical oil drift modelling not matched in time with the years of data).	**Moderate/high** The larvae drift data is directly linked to calculated loss. Great variation from year to year.	**VI**
Parameter	Lower effect limit	**Moderate/weak** Based on limited data, studies not specific for different types of HC, content of toxic compounds etc., only the THC concentration.	**Moderate/high** The effect category is likely species/areas/oil type specific.	**Moderate/high** Effect categories are critical for the calculated loss.	**VI**

Table 5. DNV GLs evaluation of input related to settings and input to the oil drift model used prior to environmental risk assessments. (Kruuse-Meyer, 2015).

Def par/var	Parameter/variable	Strength of knowledge	Belief in deviation from assumption	Sensitivity with respect to assumption	Setting (I-VI)
	OSCAR oil drift model				
Parameter	Number of rates/durations included in the modelling	**Moderate/weak** The strength of the data depends on the knowledge provided before modelling. For the durations DNV GL utilizes five durations instead of three (which is more than what Add Energy recommends in the blowout and kill studies). This increases the refinement of the analysis.	**Low** The number of rates and durations should be sufficient to reflect all possible outcomes of a spill. Similar rates may be merged to reduce the number without altering the results. The number of durations can however deviate more from its base case values depending on the supplier of blow and kill study, and the company performing the analysis.	**Moderate** For rate it is expected that large deviations are required to alter the end risk results while smaller changes are required with regards to duration, however, in total it is expected that the outcome is moderate. Relatively large changes may alter the end risk results. It is not tested in this study, but this should strongly be considered.	**IV**
Parameter	Number of simulations	**Moderate/weak** The number of simulations does not fulfill all conditions to become strong.	**Low** Tests have been performed to find a sufficient number of simulations, so a change should not change the results much. The results confirm this.	**Low** Earlier studies have shown that this parameter does not affect the result significantly. This has been confirmed by tests in this study.	**II**
Parameter	Tracking time	**Moderate/weak** All four conditions are not fulfilled. Testing has proved that the impact is dependent on the spill duration. Longer spill duration could require longer tracking time, while using too long tracking time for shorter spill durations could alter give unrealistic results.	**Moderate** It is likely that the end results could change if the tracking time is changed from the chosen 15 days due to the difference in weathering time.	**Low** It is expected that large changes must be considered if the end risk results should be altered considerably.	**IV**
Parameter	Number of particles	**Moderate/weak** The number of particles for standard analyses is 2500, however for larger rate and longer durations, a higher number is required .There should be sufficient amount of particles to have at least one particle released per time step to avoid merging of already released particles. Could be evaluated on a case to case basis.	**Moderate** The number of particles may have moderate deviations from the standard value of 2500 particles for a release, independent of rate and duration.	**Low** It is expected that large changes must be considered if the end risk results should be altered considerably.	**IV**
Variable	Ocean current data	**Strong** It is believed that these data are the best available and there is a sufficient amount of data available. The models predicting the wind and current are well understood and give reliable predictions.	**Low** It is believed that the current data in the current resolution will not vary much from its base case value.	**Moderate** The data used has an impact on the results; i.e. the number of years included in the statistical material.	**III**

6 Applying the Scoring System to ERA Acute

In the following section, the parameters that were tested in Phase 4 with respect to model sensitivity have been entered into the scoring system. Instead of the "Settings" classification used by DNV GL recommendations are given on the specific parameters. Based on the results of sensitivity testing of each of the included parameters used in the risk functions (Bjørgesæter and Damsgaard-Jensen, 2018; Stephansen & Bjørgesæter 2018) the scoring system is used for the parameters that are used in four compartment risk functions.

For each parameter the following sensitivity-deciding elements are considered and assessed within the scope of the feasibility study, which is based on eth knowledge gained in previous literature studies in Phase 3 and sensitivity testing/validation phase of Phase 4.

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.

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

Comments/recommendations on handling to ensure risk is not under-estimated: Used instead of assigning a setting, recommend an action for ERA Acute use, data gathering etc.

6.1 Compartment classification tables

6.1.1 Surface compartment

Results of the scoring and evaluation of surface parameters are given in Table 6.

Table 6. Summary of assessments or calculations used as basis for classification in the sea surface.

Function	Main parameter	Strength of knowledge (*function* where it is used)	Belief that the value may deviate from the average assumption (Natural variation of parameter)	Sensitivity of function to parameter (sensitivity index)	Comments/recommendations on handling to ensure risk is not under-estimated
Impact & Impact time	p_{beh}	Moderate/weak. Due to limited data and large natural variation it is difficult to assign a specific *pbeh* value. The assumption that behavioural factors will affect p_{exp} is strong.	Moderate	Moderate	A higher value is conservative. Each VEC have three estimates (low, intermediate, high), this using high is most conservative. Alternative, use all to obtain larger credible interval.
	Cov	Moderate/weak. The parameter depends on other parameters evaluated as Moderate/weak. The assumption that that exposed area will affect p_{exp} strong.	High	Moderate	A higher value is conservative. Coverage is calculated by the oil drift model. Use Best Practice for ODS set-up to ensure comparable and reliable predictions of the statistic.
	Texp	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.	High	High	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}	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	Low/Moderate/High, depending on VEC. Low for seabirds and moderate to high for marina and aquatic mammals and sea turtles	Moderate	A higher value is conservative. Each VEC have three estimates (low, intermediate, high), thus using high is most conservative. Alternative, use all to obtain larger credible interval.
	Th	Moderate/weak Due to lack of experimental data, it is difficult to assign specific threshold levels for lethal oil film thickness	Moderate	High	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 µm 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	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.	High	Moderate/high	Use the best available data to reduce uncertainty. Use the same data for comparable studies. The definition of a "population" is important.
Lag-time	*Nhab*	Moderate/weak The function includes various not well-defined or understood subtle effect other than acute mortality.	High	Moderate	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	Moderate/weak	High	Moderate	
	t_lag (shoreline)	Moderate/weak. Due to lack of experience data, it is challenging to assign specific lag time periods for different types of shoreline habitats.	High	High/Moderate	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).

Restoration time	*R*	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)	Moderate/high	High	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	Moderate/weak. The parameter determines the strength of intraspecific competition; a simplification of real-world population dynamics.	High	High	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	Moderate/weak	High Large fluctuations of population size above and below carrying capacity is common in nature	High	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	Moderate/weak Cut off to avoid tres = ∞ in a logistical growth model.	High	High/moderate for t_{res}, Moderate/low for RDF (effect varies with percentage population loss)	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

6.1.2 Water column compartment

Results of the scoring and evaluation of water column parameters are given in Table 7.

Table 7. Summary of assessments or calculations used as basis for classification in the water column.

Function	Main parameter	Strength of knowledge (*function* where it is used)	Belief that the value may deviate from the average assumption (Natural variation of parameter)	Sensitivity of function to parameter (sensitivity index)	Comments/recommendations on handling to ensure risk is not under-estimated
Impact	Plet, THC Extracted from SSD-curve	Strong	Moderate. SSD-curve based on LC50 for 24 species	High	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	Moderate/weak. Vertical maxima, THC includes numerous components with varying toxicity	High	High	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
	Pfrac (SD and species sensitivity)	Strong. Estimated in OSCAR during the ODS using a simplified QSAR method	Moderate	High /Moderate/ Low (depending on setting)	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	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	Moderate	Moderate/high	Use the best available data to reduce uncertainty and increase the quality of the predictions. Use the same data for comparable studies.
Recovery	CritDens%	Moderate/weak	High	High (threshold level between two methods	Higher values are more conservative Expresses the threshold for when a direct relationship is modelled between larval mortality and recruitment reduction

				with different conservatism)	
	CritOilMort (%)	Moderate/weak	High	High (threshold level between two methods with different conservatism)	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 (%)	Moderate/weak	Moderate/high	Not tested	
	Annual natural mortality of matures (%)	Moderate/weak	Moderate/high	Not tested	
	Age at recruitment (year)	Moderate/weak	Low /moderate	Not tested	
	Age at first spawning (year)	Moderate/weak	Low	Not tested	
	Maximum age (year)	Moderate/weak	Low	Not tested	

6.1.3 Shoreline compartment

Results of the scoring and evaluation of shoreline parameters are given in Table 8.

Table 8. Summary of assessments or calculations used as basis for classification in the shoreline.

Equation	Main parameter	Strength of knowledge (*function* where it is used)	Belief that the value may deviate from the average assumption (Natural variation of parameter)	Sensitivity of function to parameter (sensitivity index)	Comments/recommendations on handling to ensure risk is not under-estimated
Impact	Tidal range (m)	Moderate/low	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).	Low	Lower values are more conservative. The parameter is cell specific and is used to estimate oil thickness.
	Slope (°)	Moderate/low	High/moderate. (the slope of the beach may vary considerable with a shoreline habitat type)	High	Higher values are more conservative. The parameter is ESI specific and is used to estimate oil thickness.
	OHC	Moderate/low	High/moderate. (the distribution of oil along the shoreline will also depend on factors such as current, wind, geography, that are difficult to accurate estimate outside the oil drift model)	Moderate/high	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	Moderate/low. Due to lack of experience data, it is challenging to assign a specific patchiness factor	High. Patchiness of oil may range from 1-100%	High	Lower values are more conservative Fixed look-up values
	Th	Moderate/low. It is difficult to assign a specific threshold level for lethal oil film thickness for invertebrates and vegetation	Moderate	High (threshold value)	Higher values are more conservative. Threshold level for impact, 0.1 mm for invertebrates and 1.0 mm for wetland vegetation

	Stranded mass (ton)	Moderate/low. Basis for calculating film thickness	High	High/moderate(proportional)	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)	Strong. Depends on the quality of the data received from the data provider.	Low/moderate	High (proportional)	Use the best available data to reduce uncertainty and increase the quality of the predictions. Use the same data for comparable studies.
	Shoreline rankings	Strong. Depends on the quality of the data received from the data provider.	Moderate	High for recovery (lag-time and restitution)	ESI rankings; 1 least sensitive, 10 most sensitive
Lag-time	Lag-time	Moderate/low. Due to lack of experience data, it is challenging to assign specific lag-time periods for shorelines	High/moderate. Variable and to a large degree depending on weather conditions	High	Fixed look-up values
Recovery	Recovery	Moderate/low. Due to lack of experience data, it is challenging to assign specific restitution time periods for shorelines	High Variable depending on invertebrate and flora communities	High	Fixed look-up values

6.1.4 Sea floor Compartment

Results of the scoring and evaluation of sea floor parameters are given in Table 9.

Table 9. Summary of assessments or calculations used as basis for classification in the sea floor.

Main parameter	**Strength of knowledge (*function* where it is used)**	**Belief that the value may deviate from the average assumption (Natural variation of parameter)**	**Sensitivity of function to parameter (sensitivity index)**	**Comments/recommendations on handling to ensure risk is not under-estimated**
Mixing depth	Strong/ moderate. Knowledge of what constitutes the bioturbation depth is relatively strong	High uncertainty	40.0 % HIGH	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	Strong	Low	0.5 % LOW	Marine Geochemistry gives general values. Low sensitivity, use defaults.
Water Content	Strong	Low/Moderate	2.7 % LOW	Use lower values as conservative.
Total org. Carbon	Strong (EqP accepted methodology)	High	54.9 % HIGH	Use conservative (lower) values. Lower values lead to higher toxicity and shorter restoration times. (Higher TOC sequesters THC in sed.)
KOW	Strong (EqP accepted methodology)	Moderate	1.8 % LOW	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)	Strong	High to low depending on species sensitivity	High	Conservativity already implemented by the curve being conservatively extrapolated from the LD5-value from a large and QA'ed set of data (Nilsen et al. 2006). SSD curves are accepted methodology and inherent safety factor used.
THCsed (used as input from OSCAR	Strong knowledge of place in ERA Acute function	Is calculated by the OD model. SD is low within calculations in same model, may vary a lot between inputs from different models	High (proportional)	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 sediemnt in OSCAR are undergoing improvements, e.g. by possible inclusion of marine snow.
THC (WC)	Strong knowledge of place in SSD-curve	High uncertainty and the THC concentration is a time-averaged concentration	High (proportional)	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 Assesment incl. MIZ-proposal) could improve this.

				HOWEVER 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	High strength of knowledge	Moderate	High (proportional	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.
$C_{threshold,sed}$	Moderate strength of knowledge of function	High	High	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.
$C_{benchmark\text{-}max,sed}$	Moderate strength of knowledge of function	High	High	Value representing the maximum value at equilibrium. Based on data from the MOD data base (North Sea).
20 years def value	Moderate strength of knowledge of function	High	High	The based on MOD data from North Sea, sandy bottom, few sites have data on restoration times after use of oil based drilling muds.
SF	Moderate strength of knowledge of function.	High	High (proportional)	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
Tlag (HARD)	Fixed value	High	High	Very little research available after oil spills affecting deep sea corals. Comparable incident DWH not yet restored.
Tres (Hard)	Fixed value	High	High	Very little research available after oil spills affecting deep sea corals. Comparable incident DWH not yet restored.

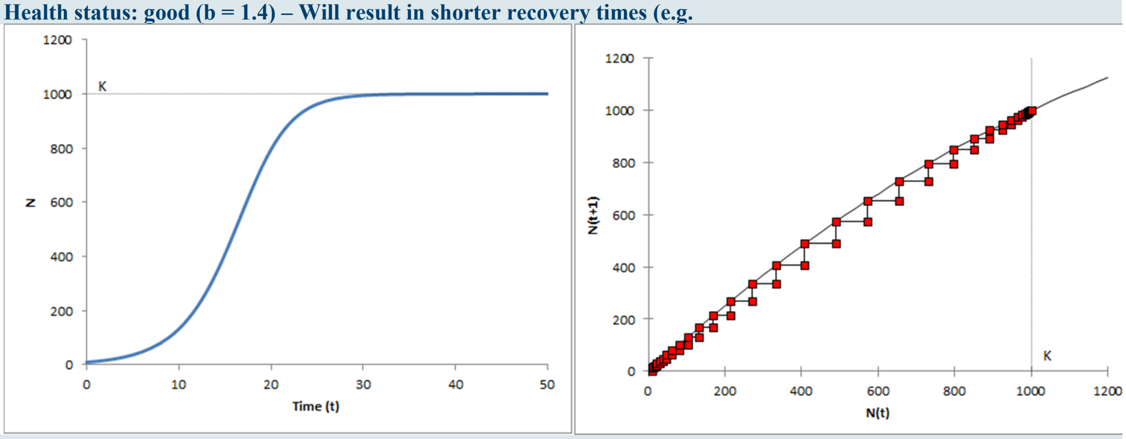
6.2 Summary table of the prioritised parameters

A summary list of the recommendations for the most important parameters to improve the certainty of, is given in Table 10. Other important parameters for the results are also listed in the compartment tables but are handled by following the best practice and guideline documents which in turn, would be subject to improvement through common experience with the model.

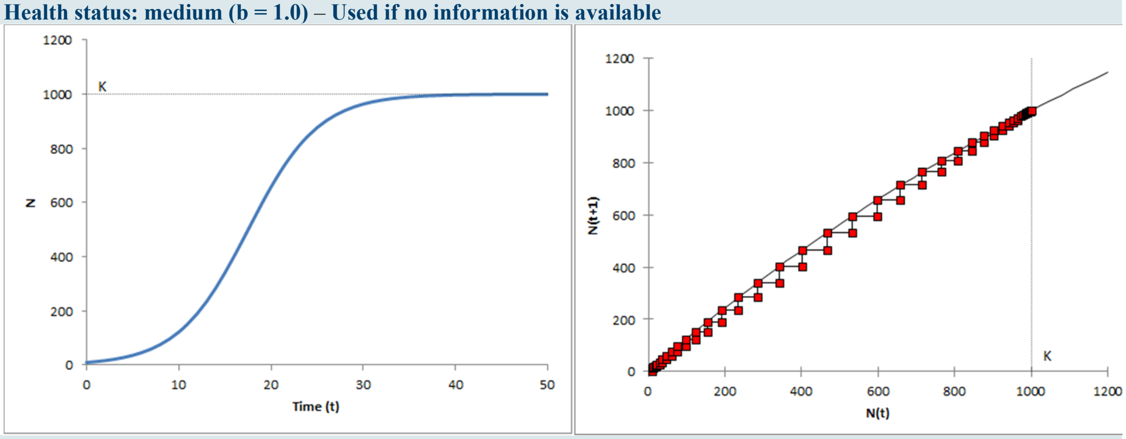
Table 10. Prioritised parameters with a potential for improvement or that have a high impact on the result.

Compartment	Parameter	Recommendation for improvement
Surface	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”). An example is given in Figure 1. The same effect may be obtained by adjusting the net fundamental growth rate R
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
	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
Sea floor	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).

Health status: good (b = 1.4) – Will result in shorter recovery times (e.g.



Health status: medium (b = 1.0) – Used if no information is available



Health status: poor (b = 0.7) – Will result in longer recovery times (e.g.)

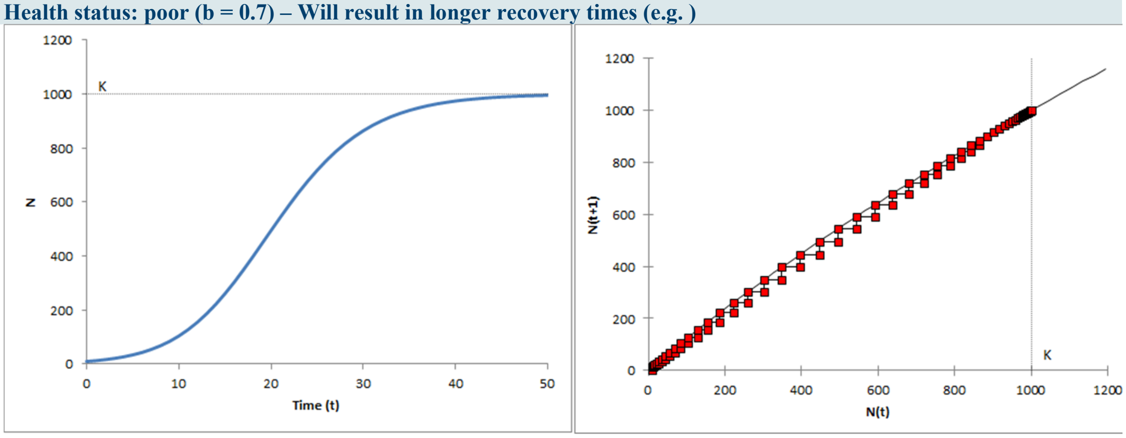


Figure 1. Illustration of the restitution curve (left) and logistic population growth (right) for different values of b, representing different "health statuses" of the population.

7 Uncertainty handling in ERA Acute

7.1 Are uncertainty factors a viable option?

ERA Acute is a complex model, consisting of a large number of complex risk functions and parameters, as well as many summarising steps, which may amplify uncertainties through the calculations. This feasibility study aimed to assess uncertainty levels, but it has not been within the scope of the project to find exact numerical values of uncertainty for each parameter. Uncertainty was assessed through the exercise of evaluating and categorising the many parameters in ERA Acute and the status of each is presented in Table 6 to Table 9, allowing users to identify parameters for which improving the accuracy of the value could be a significant increase in accuracy and quality of the result.

Uncertainty factors are often used in risk assessment models of the PEC/PNEC calculation- type, e.g. by dividing threshold values by a factor to provide a safety margin for exposure recommendations etc.. The feasibility of using uncertainty factors in ERA Acute has been discussed However, following the evaluation it was concluded not to recommend use of uncertainty factors at this point, due to the complexity of ERA Acute and that it would take much more work to establish appropriate values.

Using one uncertainty value at the end of the calculations was deemed not feasible as it would be difficult to establish such a value. Comparisons between model results and impacts from real incidents and establishing uncertainty factors based on the deviation of ERA Acute results from the "truth" would be one way to arrive at a final-endpoint uncertainty factor, the it is important to remember that the estimates of impacts from incidents are also based on modelling, and even measurements/monitoring has high levels of uncertainty. Establishing safety factors that are robust for a final endpoint risk measure was therefore deemed unfeasible (at this point).

Using uncertainty factors for each uncertain value was also deemed unfeasible as setting the factor value based on largely qualitative evaluations of uncertainty is too inaccurate and the use of factors at every point in calculation would lead to an amplification of conservativity to the level of greater inaccuracy. It was therefore concluded that it would be a large job to propose sensible safety factors and adding safety factors when the more conservative values are recommended is perhaps over-kill and "introduces uncertainty to uncertainty". The other reason for not using uncertainty factors is that the purpose of the risk assessments is decision-making in risk management. We have some important goals for handling uncertainty. The first goal is to *ensure that ERA Acute does not underestimate risk*. This could, as mentioned above, be handled by using uncertainty factors. However, as described, applying an additional factor to represent the uncertainty in each parameter used will mean that, although the risk will not be under-estimated, conservativity is amplified and the end result will be very inaccurate when the results are calculated from result "level" to result "level". This may counteract the second goal, which is to ensure that ERA Acute can be used for *illustrating the differences in risk following from risk mitigation efforts or differences between activities*. These differences could then be exaggerated or under-estimated, making it difficult to identify where reduction of risk would be optimal.

For use in risk management there are other issues related to uncertainty that are as important as finding the exact answer. *Transparency* in the method is important to ensure peer acceptance and that the method can be understood and criticised so improvements to functions and parameters can be made. Reducing *variability* is as important as providing an accurate risk measure, and this is the main, and most practicable way of ensuring *comparability* between risk assessments.

7.2 Recommended uncertainty handling at this point in model development

Ideally, as discussed above, 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, at the current point, this was not deemed possible. Instead the following general recommendations for ensuring comparability and reducing variability are given:

- Use the *conservative* values included in the method reports and current guideline
- Use the conservative "QSAR" approach to estimate larvae losses in water column impact calculations, not THC-time-weighted average
- 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

Within a region, e.g. a country etc. 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.* This study may serve as basis for identifying the parameters which could be prioritised for further work.

8 References

Reports on ERA Acute model development can be read at:

https://norskoljeoggass.no/miljo/mer-om-miljo/miljorisiko-og-miljorisikoanalyser2/era-akutt/

Brude, O.W. & Bjørgesæter A.- 2016: WP2b. Field validation feasibility study. ERA Acute Project report: ERA Acute 2b-1. EAR Acute Report ERA Acute 2C-2.

Brude, O.W. & Rudberg, A. WP2c – ERA Acute and MIRA methodology comparison

Brönner, U. (SINTEF), Nordtug, T. (SINTEF), Jonsson, H. (DNV G., Ugland, K.I. (UiO), 2015. Impact and restitution model - Water column. ERA Acute for water column exposed organisms. Trondheim, 20.05.2015.

Bjørgesæter and Damsgaard Jensen, 2015) ERA Acute Phase 3 – Surface compartment. Acona report to Statoil and Total. Report No. 37571. v.04. Oslo, 22.05.2015

Bjørgesæter A and Damsgaard Jensen, J.- 2018: WP2a. Sensitivity and Uncertainty testing. ERA Acute Project report, ERA Acute 2A-2.

Bjørgesæter A, Pedersen A., Brude, O.W., Rudberg, A. and Stephansen, C- 2018: WP2b. Validation with field data. ERA Acute Project report ERA Acute 2b-3.

Carroll, J., Juselius, J., Broch, O.J., Nepstad, R., Brönner, U., Vikebø F., Bogstad, B., Howell, D., Klok,C., Hendriks, J., de Laender, F., de Hoop, L., Viaene, K., Grøsvik, B.E., Couture, R-M. Moe, J., Langangen, Ø., Skeie, G.M. Bluhm K., & Wilson L. (2014) SYMBIOSES Final Report (1-49 pp)

Carroll, J., Vikebø F., Howell, D., Broch, O.J., Nepstad, R., Augustine, S., Skeie, G.M., Bast, R. and Juselius (2018): Assessing impacts of simulated oil spills on the Northeast Arctic cod fishery, Mar. Poll. Bull. 126 (2018): 63-73

DNV. 2007. Metodikk for miljørisiko på fisk ved akutte oljeutslipp. DNV Rapport nr. 2007-2075.

DNV GL 2018: ERA ACUTE AND SYMBIOSES COMPARISON. Northeast Arctic cod case study. Report No.: 2008-0316, Rev. 0, Document No.: 118PNWCK-2, Date: 2018-05-08.

Etkin, D.S., French-McCay, D., Michel, J., 2007. Review of the State of the Art on Modelling Interactions between Spilled Oil and Shorelines for the Development of Algorithms for Oil Spill Risk Analyses Modelling, Cortland Manor. New York, pp. p157.

French-McCay, D., 2009. State-of the -Art and Resarch Needs foor Oilspill Impact Assessment Modelling. Proceedings of the 23nd AMOP Technical seminar on Environmental Contamination and Response. Emergencies Science Division, Environment Canada, Ottawa, ON, Canada, pp. 601-653.

Jensen, F.V & Nielsen, T.D. Bayesian Networks and Decision Graphs 2007. Springer Verlag, ISBN-10: 0-387-68281-3; eISBN-10: 0-387-68282-1; ISBN-13: 978-0-387-68281-5; eISBN-13: 978-0-387-68282-2

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, I.B, Ray, C.J, and Kirchner, D.E. (2008): A Methodology For Performing Global Uncertainty And Sensitivity Analysis In Systems Biology. J Theor Biol. 2008 Sep 7; 254(1): 178–196.

Norwegian Oil and Gas (2017): Guidance on calculating blowout rates and duration for use with environmental risk analyses, NOROG guideline. https://norskoljeoggass.no/miljo/handboker-og-veiledninger/veileder-i-beregning-av-utblasningsrater-og--varigheter-for-bruk-i-miljorisikoanalyser/ https://www.norskoljeoggass.no/contentassets/d6c1d516cc4945358fea471a8610becf/guidance-blowoutrates-with-supp-report.pdf

Stephansen, C. 2018. ERA Acute and SYMBIOSES. A comparison of risk assessment models for water column. ERA Acute project report 3-6. 26.02.2018.

Stephansen, C. and Bjørgesæter A. (2017), WP2a – Seafloor Compartment Sensitivity Testing and Norwegian Sea Test Case Data. ERA Acute Project report ERA Acute 2A-3.

Stephansen, C. (2017) WP2a –Seafloor Compartment Norwegian Sea Test Case level A.1, A.2, A.3 and B ERA Acute Project report ERA Acute 2A-4.

Stephansen, C. Bjørgesæter A, Brude, O.W., Brönner, U., Rogstad, T.W., Kjeilen-Eilertsen, G., Fjeld-Nygaard, C., Libre, J.-M. and Dragsund, E., (2018) ERA Acute JIP Summary Report : Executive summary of the work carried out in ERA Acute phase 4 (2016-2018). ERA Acute Project report ERA Acute 3-14

9 Appendix

Table 11. Summary of the VEC data parameters applied in the Norwegian Sea data set adapted from Mareano substrate data. The compilation of substrate types uses Mareano sediment groups and grouping with fractions of silt/clay (~pelite), sand and/or gravel (Mareano), with TOC-values estimated from TOM/TOC-ratio in the MOD database, dry densities and water contents. (ND= No Data)

English term for Mareano data	VEC name	Fraction silt:clay	Fraction Sand	Fraction gravel	TOC (%)	DryDens	Water C	BDepth	Algorithm
Biological material	Bioclastic coarse sand	ND	ND	ND	0.4	2650	0.25	0.05	SOFT
Sand with gravel	Coarse sand	0-0.1	0.9-1	0.02-0.3	0.4	2750	0.25	0.05	SOFT
Muddy, sandy gravel	Coarse sand	0-0.1	0.2-0.7	0.3-0.8	0.4	2750	0.25	0.05	SOFT
Sandy gravel	Coarse sand	0-0.1	0.2-0.7	0.3-0.8	0.4	2750	0.25	0.05	SOFT
Sand, gravel and stones	Coarse sand				0.4	2750	0.25	0.05	SOFT
Sand	Sand	0-0.1	0.9-1	0-0.02	1	2750	0.3	0.02	SOFT
Sandy mud	Sandy mud	0.5-1	0-0.5	0-0.02	1.2	2100	0.5	0.01	SOFT
Muddy sand	Sandy mud	0-0.5	0.5-1	0-0.02	1.2	2100	0.5	0.01	SOFT
Gravel-containing muddy sand Gravel-containing sandy mud	Sandy mud	0.1-0.5	0.5-0.9	0.02-0.3	1.2	2100	0.5	0.01	SOFT
Mud	Mud	0.9-1	0-0.1	0-0.02	2.4	2100	0.65	0.005	SOFT
Thin, discontinuous layer of sediment on rock	Hard substrate							N/A	HARD
Bare rock	Hard substrate							N/A	HARD
Gravel, stones and boulders	Hard substrate							N/A	HARD
Hard sediments (sedimentary rock)	Hard substrate							N/A	HARD