Document title	Updated overall methodology to the SOM approach
Code	2-1
Category	CMNT
Agenda Item	2 - SOM approach: data collection, integration and uncertainties, projection of human activities
Submission date	3.3.2020
Submitted by	HELCOM ACTION project and HELCOM Secretariat
Reference	

Background

HOD 56-2019 endorsed the approach for analysing sufficiency of measures (SOM) and its use to support the BSAP update as contained in document 2-3 to HOD 56. The approved approach is based on a proposal presented and discussed at HELCOM SOM Platform 1-2019, intersessional adjustments by the HELCOM ACTION project, commenting by SOM Platform representatives, and the endorsement by GEAR 20-2019 (para 5.20). The method was presented in more detail to HELCOM SOM Platform 2-2019 which supported its use for the sufficiency of measures analysis (Notes of meeting, para 2.18).

The methodology for the SOM analysis has now been detailed for all its steps. This document describes the overall approach, the input data and its collection, the SOM model, and gives an overview of the format of the expected results to support the BSAP UP process. The development of human activities (step 5), format for the topic-specific background report and preliminary results of the SOM analysis are presented in separate documents.

To facilitate the understanding of the full SOM approach, this document is divided into three parts:

I Overall approach to estimate the sufficiency of measures

II Data collection and input data into the SOM model

III The SOM model: steps to analyse the sufficiency of measures.

Changes and new text compared to what was presented to SOM 2-2019 is indicated with red colour. Parts I and III are partially updated, and part II is completely new text.

Action requested

The Meeting is invited to:

- take note of the updates to the SOM methodology
- agree on the methodology for using the expert-based and literature-based data in the SOM model
- agree on the (updated) SOM model structure
- agree on the methodology on joint effects.

Table of contents

PART I OVERALL APPROACH TO ESTIMATE THE SUFFICIENCY OF MEASURES

This section is based on documents 2-2 and 2-2 Att. 1 submitted to SOM Platform 2-2019. Changes from these documents are indicated in red text. Minor editorial changes, including relocated text and removal of outdated document references, have not been indicated.

1. The proposed SOM model in the European context

The SOM approach builds on the conceptual model developed in the HELCOM SPICE project, applies state of the art methods from scientific literature (Kontogianni et al, 2015 and Oinonen et al, 2016) and is in line with how some of the HELCOM Contracting Parties analysed the effectiveness of measures for the first round of the Marine Strategy Framework Directive Programme of Measures (MSFD PoMs). Examples of the Contracting Parties' effectiveness analyses were presented and discussed in the Second Meeting of the HELCOM Expert Network on Economic and Social Analyses (EN-ESA 2-2018) (see presentations by Sweden, Latvia and Finland). The same meeting also suggested the use of national experiences to select the best practices for the regional analyses to be further elaborated by the SOM Platform for use in the BSAP update. Methodologically the approach presented here rests on previous analyses, but at the same time it takes a significant leap forward by attempting to assess measure effects through activity-pressure-state linkage chains for a sea area consisting of multiple sub-basins. The approach also considers predicted future changes in the activities that cause pressures to the Baltic Sea, whereas most of the previous analyses have assumed that these activities stay unchanged.

The main purpose of the SOM analysis in conjunction with the upcoming cost-effectiveness analysis is to support the update of the new BSAP by indicating both thematically and spatially where new measures are needed, and by suggesting how to optimize the new BSAP actions taking into account both environmental objectives and economic aspects. Further, the linkage-chain approach allows a transparent assessment of what types of measures are potentially effective in improving the state of marine environment, given the spatially / basin-specific characteristics of state, pressures and human activities. The purpose of the SOM analysis is not to retrospectively assess the implementation of BSAP and other related policies for different countries nor to take a stand on responsibilities for not reaching the good state of marine environment. The results of the presented approach move away from country level analysis to assess the measure effects on the whole Baltic Sea.

Therefore, the SOM analysis can be seen as a decision support tool on where and what types of measures should be adopted, in addition to contributing to the assessment of what is the state of the marine environment with existing measures. The results of the analyses are based mainly on expert opinion, and thus they should be interpreted only as suggestive. As such, the SOM analysis does not provide full answers on the reductions in pressures or improvements in state. All results of the SOM analysis should be reviewed in relation to other results and assessments. Generally, independent non-relative quantitative results detached from the context of the analysis do not provide sufficient means to support decisions.

2. Overall approach

The aim of the analysis of sufficiency of measures (SOM) is to assess whether existing policies are sufficient to achieve good environmental status (GES) in the Baltic Sea. It relies on estimating the status of the marine environment at some specific future point in time, given measures in existing policies, their implementation status, natural time lags, and predicted development of human activities/pressures over this time period. This is called the 'business-as-usual (BAU) status' (Figure 1). If the analysis indicates that GES is not achieved, then existing measures are not sufficient and additional measures are needed (or existing measures strengthened).

SOM analysis includes the following components:

- information on existing measures and their level of implementation, and possible time lags in their effect (Steps 1-2),
- identifying main pathways for pressures based on links between activities and pressures (Step 3),
- estimating the effect of measures on pressures and state (Step 4),
- projections of the development of human activities/pressures (Step 5),
- estimation of the changes in the state of the marine environment due to changes in pressures (Step 6),
- using the information above to assess the projected status of the marine environment (BAU status) by a specific point in time (Step 7),
- comparison of the BAU status to GES and evaluating how far we are from reaching GES, i.e. the sufficiency of measures (Step 7).
- estimation of the effect of pressure-state time-lags on state components (Step 8).

The steps are described in detail in Section 2.

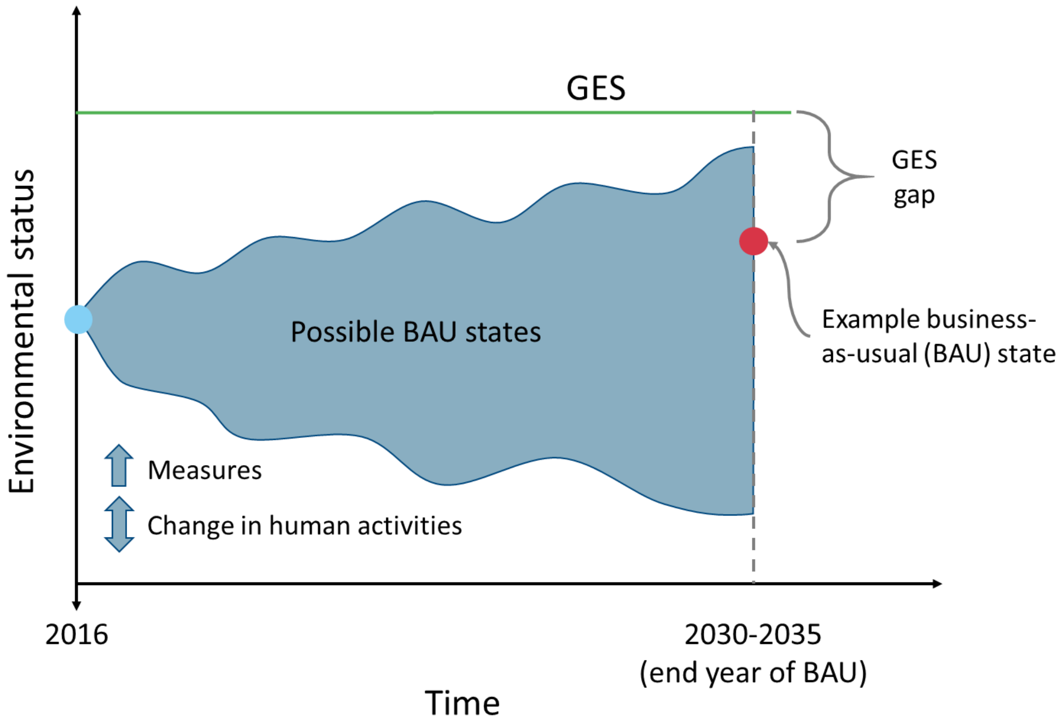

Figure 1. Illustration on the use of the BAU in the gap analysis. Source HELCOM (2018a). [Figure has been updated to include BAU end year]

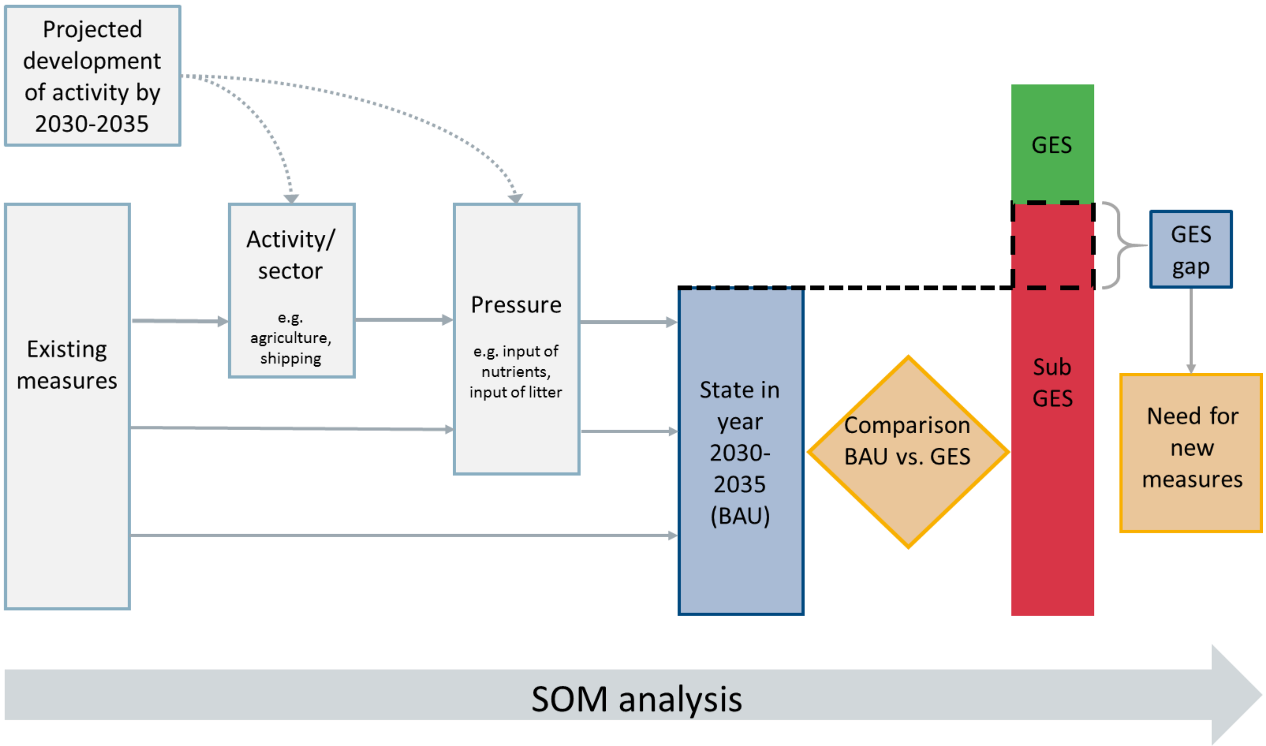

Figure 2. Main components and steps of the SOM approach. Linking measures with activities, pressures or state components; predicted changes in activities and pressures; comparison of the BAU state with GES; and estimation of the need for new measures. [Figure has been updated to include the BAU end year]

The SOM approach is based on seven main steps as outlined in the approved SOM approach. The current status of implementation is indicated below.

Step 1. Lists of existing measures have been compiled.

Step 2. Estimation of time-lags in the effect of measures on pressures is integrated into the compilation of measures lists (step 1). SOM Topic Teams have reviewed and supplemented the time-lag data following responses from the Contracting Parties. Contribution was also requested from HELCOM expert networks as needed.

Step 3. Identifying main pathways and also the contribution of activities to pressures was estimated by SOM Topic Teams and HELCOM expert networks.

Step 4. The method for estimation of the effect of measures is outlined in this document. The expert evaluation was implemented through online surveys. Literature reviews have been carried out.

Step 5. Collection of information on projection of human activities/pressures has been completed. Three to four scenarios have been developed for each activity and will be run with the SOM model.

Step 6. Methodology to link reduced pressures with state components is outlined in this document. The expert evaluation was implemented through online surveys.

Step 7. Comparison of BAU and GES and assessing sufficiency of measures is the final step of the approach, to be carried when first results are ready in early 2020.

Step 8. For increased clarity, this step has been separated from Steps 6 and 7 and is indicated as the 8th step in the model. Time-lags in recovery of the state components will affect the interpretation of the SOM results. WP5 of the HELCOM ACTION project will contribute to this step, with additional support from selected SOM topic teams. In addition, complementary information on time lags has been collected with the expert surveys.

Step 1. Existing measures, including activity-measure links
Step 2. Estimating time-lags for measure effects on pressures
Step 3. Identifying main pathways for pressures using activity-pressure linkages
Step 4. Estimation of effects of measures
Step 5. Projected development of human activities/pressures
Step 6. Linking reduced pressures with state components
Step 7. Comparison of BAU and GES and assessing sufficency of measures
Step 8. Assessment of the effect of time-lags to recovery on state components

Time frame

The time frame of the BAU should be consistent with the relevant target years of the HELCOM BSAP and the EU MSFD. The time frame should stretch beyond 2020/2021 to allow for more complete impact of existing policies and measures, but it should not stretch too far into future to avoid uncertainties in changes in the climate and policies.

The kick-off meeting of the SOM Platform suggested to use either 2030 to coincide with the target year for the majority of the UN Sustainable Development Goals (SDGs) or 2033 to coincide with the EU MSFD cycle (Item 3.13 in the Notes from SOM Platform 1-2019). HOD 56-2019 discussed the end year for the analyses and was of the view that it should be set to a year between 2030–2035, but did not come to a conclusion. This time window has therefore been adopted as the BAU end year.

Existing policies and measures

Measures that are included in the BAU status need to be clearly defined. For all existing relevant policies (e.g. current BSAP, MSFD, WFD, EU Biodiversity Strategy 2020), implemented measures with unrealized effects on base year pressure levels, on-going (or partially implemented) measures and planned measures[1] are proposed to be included in the BAU, as suggested by EN ESA 2-2018 (Item 4.3 in the Outcome of EN ESA 2-2018) and agreed by SOM Platform 1-2019. Thus, it would be assumed that all measures in existing policy frameworks are fully implemented in the time frame of the BAU, independent of their current implementation status, and their effect on reducing pressures would be realized fully in the time frame of the BAU.

~~An additional analysis was agreed to at SOM Platform 1-2019. It would use the same methodology and included measures at the standard BAU described above, except HELCOM measures would be analysed at their implementation status in the base year rather than under the assumption of full implementation by the target year as is standard. HELCOM measures refers to all measures organized under the HELCOM structure including BSAP and HELCOM recommendations. This analysis provides an illustration how the implementation of ongoing and planned measures affects the state of the sea and where distinction between the two BAU analyses is necessary, the second analysis will be referred to as the BAU implementation analysis (BAUi).~~[2]

The SOM Platform agreed with the proposal to include in the analysis all types of measures except those related to promotion of research and some administrative measures (i.e. monitoring, coordination,

[1] Note that the term *existing measures* covers implemented, partially implemented/ongoing and planned/not yet implemented measures in existing policies.

[2] Text removed due to conflicts with the proposed method for assessing the implementation of measure types. The proposed methodology does not account for the origin of existing measures. See Section III.

developing SOM indicators, setting targets, developing information systems/tools etc.), which have no direct effect on environmental status.

Environmental themes to cover

The SOM analysis will be carried out for the same environmental themes as in the State of the Baltic Sea report (Figure 3). For some themes a descriptor level evaluation could be appropriate, e.g. to compare the BAU state with the integrated status. For biodiversity, the analyses could be done by ecosystem component, groups of species (e.g. coastal fish) or in some cases by species (e.g. grey seal). For a majority of topics, the status threshold values are proposed to be used as the basis for the analyses. For eutrophication and possibly other pressure-related components, the analyses could rely on pressure targets as agreed in HELCOM. A reflection on this is provided under Section 2, Step 2. Decisions on this aspect will be made in collaboration with SOM topic teams.

For some topics there are no agreed GES threshold values or quantitative pressure reduction targets (e.g. marine litter, underwater noise) in HELCOM, and thus proper gap analysis is not possible. For these topics, it is still possible to assess how much the existing measures will contribute to reducing pressures and improving the condition of the Baltic Sea. This is further discussed in Steps 6 and 7.

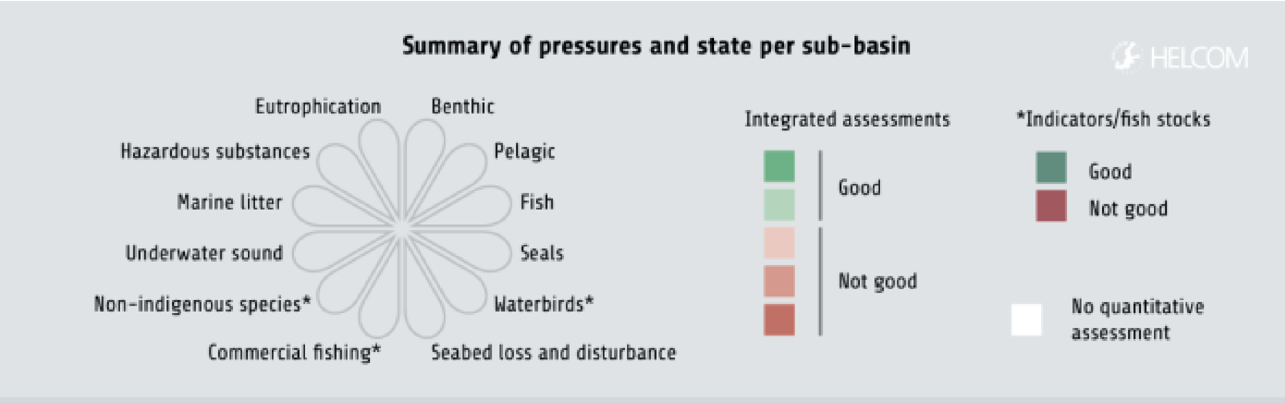

Figure 3. Proposed state components of the SOM analysis.

Geographical scale of the analysis

The geographical scale of the SOM analysis is aimed at supporting decisions from a regional Baltic Sea perspective. However, the SOM analysis will be carried out at the HELCOM scale 2 level where found relevant.

Geographic scales vary across environmental themes and reflect the fewest number of groupings required to accurately describe the activities, measures and pressures relevant to each environmental theme as identified by experts in ACTION/SOM. Different scales have been used for different stages of the analysis based on data availability or when requested by topic experts, as long as they represent HELCOM scale 2 sub-basins or aggregations thereof. Maps consisting of more than six sub-basin aggregations have been strongly discouraged for resource management purposes. These scales will inform work in steps 1, 4, 6 and 7. The geographical scales have been decided for each of the pressures based on input from the SOM Topic Teams and various HELCOM expert groups.

Data requirements

Any proposed data inputs represent places where existing data may be incorporated, they are not data requirements. Additionally, there are no requirements for the spatial or temporal coverage. Topic teams are encouraged to be as data rigorous as is possible, but data availability will vary widely between and within topics. While not the desired process, the model can function entirely on expert opinion, and as such, external data inputs are not mandatory.

3. Assumptions, simplifications and benefits of the SOM model

This section clarifies how the results of the SOM model can be interpreted and what assumptions are used in the model.

Assumptions of the SOM model

The SOM approach agreed in HELCOM has made the major assumption that all the measures (i) will be implemented by the end year of BAU period, i.e. 2030–2035, and (ii) they will have sufficient time to influence pressure reduction.

The level of implementation of some of the HELCOM agreements included in the analysis is low and this may cause a source of error in the SOM model. Due to this, the SOM model will be run also with the current implementation status to see what the urgency for implementing the remaining BSAP measures is.[3]

Implementation of the national measures under EU MSFD has been reported to EU Commission, but no summary reports have been planned to be included to the HELCOM SOM model. The assumption for full implementation will remain.

Inherent in this assumption is that the measures are implemented strongly, e.g. using best available practices or best environmental techniques. This is an assumption of the strength of measures and it is not possible to evaluate this in the SOM model. [4]

The spatial coverage of a measure covers all the marine areas in countries where it is implemented, unless there is more specific data on the spatial coverage of the measure. Total effects on sub-basins/spatial units will be estimated using spatial weights based on the proportion of national marine areas of the total areas of the basins.

Simplifications of the SOM model

The SOM model will cover a majority of human activities and pressures and the HOLAS II state components and attempts to capture all measures agreed in the Baltic Sea. This means that the model is wider than any previous model of this field and, hence, requires some generalizations.

Standard relative working units. Due to the wide coverage of measures, activities, pressures and state components, no common metrics can be found for the model. Therefore, the model builds on the principle of pressure reduction (%), which can be linked to improve the state. This also allows the comparison of results among topics and the use of results to identify where and what type of measures are potentially effective, taking into account that measures can affect multiple pressures and state variables.

Measure types instead of actual measures. It is infeasible to analyse hundreds of existing measures in the Baltic Sea region in the time allocated to the SOM analysis for the HELCOM BSAP update. To simplify the catalogue of measures, the SOM approach groups them to 'measure types' which aim to capture the main elements of the measures but still remain on relatively abstract level. This has the limitation that the measure types and real measures are not equal (i.e. the former are abstractions and the latter are closer to reality). In a hypothetical example, a measure type 'apply pingers in gillnets to reduce bycatch of harbour porpoise' does not say how many pingers are being used in gillnets, how widely this is applied in different parts of the Baltic, is this enforced or how frequently this requirement is not followed. However, a probabilistic approach is used to assess the effectiveness of measure types, which can capture these uncertainties due to abstractions. The actual measures are linked to the measure types and used in the

[3] Text removed due to conflicts with the proposed method for assessing the implementation of measure types. The proposed methodology does not account for the origin of existing measures. See Section III.
[4] Text removed as it overstates the implied assumption. The SOM model assumes fully implementation of the measure as described and its effect is estimated based on real world application of measures of this type.

SOM model in a later stage (see Steps 1 and 4). ~~Estimating the effectiveness of a measure type will, however, show that all measures underneath have certain effectiveness. The following steps then show whether the measures are sufficient for harbour porpoise (i.e. do we expect that no new measures are needed) or not. If the measures are estimated sufficient, there is the question whether they require strengthening even if there is no need for new measures in BSAP II.~~[5]

Relative scale of effectiveness of measures. ~~The estimation of the effectiveness of a measure type in reducing a pressure is not simple. Even working along a % scale is challenging for scientists. The SOM model simplifies the expert survey by asking the effectiveness on the scale from 'no effect' to 'very high effect' (in reducing the pressure). The survey covers all the measure types relevant for all selected combinations of activities and pressures and therefore the relative scale is not only simpler but also quicker.~~

~~To transform the survey responses to the % scale, the SOM approach uses existing studies as 'anchors'. These not only help in producing numbers (with uncertainties) to the model but also give validation data.~~[6]

The effectiveness of measure types will be assessed using a relative scale (no effect-highest effect). Estimating the relative effectiveness of a measure type will allow for ordering of the effectiveness of measure types in reducing pressures from certain activities or in improving certain state components, and also provide a data set applicable to the inclusion of new measures at any point in the process.

To transform the survey responses to the percent (%) scale, the expert surveys ask how much the experts estimate the most effective measure type to reduce a pressure (see step 4 b and c). The total effects of measure types on pressures will be calculated by summing the effects of individual measure types on pressures, taking joint effects into consideration. The individual percent effect of a measure type depends on the effectiveness of a measure type, activity-pressure contributions and the spatial coverage of a measure type.

Pressure –state linkage. Dependency of state on pressures is the basic assumption in environmental science. In reality, many of these links have not been established in quantitative way. In the SOM model, the expert-based pressure-state link (Step 6b, see below) is therefore essential~~, but it can in some cases be compared with the established pressure-state links (e.g. nutrient inputs, fisheries)~~.

General. The SOM model will not give the final answer with a single number of the general sufficiency, but instead all the model outputs must be interpreted. The benefits of the model use are, however, numerous.

Benefits of the SOM approach

The section above described assumptions and limitations which are good to keep in mind when interpreting outputs from the SOM model. The approach also has benefits, also beside the fact that it is feasible to conduct the analysis and run the model.

Use of effectiveness results for new measures. As the measure types are not too specific, it is possible to use them for estimating the effectiveness of the new measures. This can be done two ways: (i) if a measure is considered new but still falls under the description of the measure type, its effectiveness can be deduced based on the effectiveness of the measure type in the SOM analysis, or (ii) if the new measure is between two measure types, its effectiveness can be placed between the effectiveness of the two related types in the SOM analysis. This can support the discussions on new measures in the BSAP UP process.

Use of pressure-state linkage. The pressure-state linkage is a precondition for many environmental analyses and tools and not very often shown for marine assessments. The expert-based suggestions for these linkages

[5] Text removed due to oversimplification and potential mischaracterization of later information.

[6] Text heavily rewritten (new version below) to better reflect the applied methodology described in Section I (Step 4) and Section III.

(with uncertainty ranges) can be later validated by specific data and (if found adequate) used for further analyses. For example, preliminary analyses of the HELCOM TAPAS sensitivity scores show relatively good agreement among experts of the sensitive features of the Baltic Sea ecosystem.

Steps forward in the integrated assessment of the marine environment. The development of the methods and results represents a considerable progress in interdisciplinary research on linking measures, activities, pressures and environmental state, and improving the description of linkages between the socio-economic system and ecosystem.

Providing information for further analyses. Many of the approaches and results can be used in further analyses of the Baltic Sea environment, including the linkage framework, business-as-usual scenario (BAU), and effects of measures.

Flexibility and updating of the model. The SOM model is flexible in the sense that it can include information from both literature, studies and models and expert elicitation. In addition, the model can be updated when new information becomes available. The overall approach is general and is applicable in other contexts.

Transparency and commensurability. In principle, the SOM approach is simple as does not include complex definitions of the natural environment and it applies similar activity-pressure-state linkage chains, as well as definitions for measure effectiveness and state improvements for all topics. Therefore, it allows the comparison of results across topics and transparent analysis on the linkages and interdependencies between measures, activities, pressures and state variables. This can help identify where and what type of new measures are needed.

4. Steps of the SOM model

Step 1. Existing measures

This section gives detailed information on SOM components related to existing measures and their level of implementation.

1a. Identify measures under existing policies (i.e. existing measures) to assess their effect on the marine environment. This includes global conventions, EU directives and regulations, regional HELCOM actions and national measures.

1b. Categorize measures into common groups based on, for example, the general type of the measure (e.g. legal, technical, monitoring, knowledge and awareness), and the key type of the measure (KTM) (as in the EU MSFD and WFD). Measures are categorized into 'Measure types' which are the units used in the SOM surveys for the effectiveness of measures. The measure type is a mixture of the above-mentioned categorization (e.g. KTM) and more concrete description of the measure. An example of a measure type is 'Technical modification of fishing gears to reduce bycatch of harbour porpoise'. The categorization will allow for simplifying the analysis (i.e. by aggregating similar type of measures) and linking them with activities and/or pressures (or in case of restoration measures, to state).

A majority of measures are linked with human activities, but some may be linked to pressures (e.g. long-range transboundary pollution) and a few are directly linked to state components (e.g. restoration, restocking) (Figure 2).

- If a measure is linked to an activity, i.e. the activity is restricted or changed, then one can follow the linkage framework and estimate the consequent reduction of pressures (Steps 3-4).

- If a measure is linked to a pressure or a state component (restoration measures), then the effect in Step 4 is directly estimated.

1c. Assess the implementation status of the measure, i.e. whether the measure 1) has been fully implemented and has unrealized effects on base year pressure levels, 2) has been partially implemented or implementation is ongoing, or 3) is planned to be implemented. The implementation status of the measures may differ by countries, which needs to be taken into account. The BSAP implementation status has been assessed already in previous HELCOM processes, but some other measures (e.g. national MSFD measures) may require such an assessment on the basis of EU Member State reports. This step informs especially Step 2.

Information needed	**Data sources**	**Main contribution**
List of measures	HELCOM Explorer HELCOM Recommendations EU MSFD Programmes of measures EU WFD Other EU policies/directives as agreed	ACTION project/secretariat
Implementation status (implemented, partially implemented/ongoing, planned)	As above + EU reports on implementation of PoMs	ACTION project/secretariat, complemented as needed by CPs
Type of measure (e.g. technical, monitoring, knowledge and awareness…) and simplified description	As above	Initial sorting by secretariat/ACTION project, validation by SOM Platform
Whether a measure has an effect on activity, pressure or state	As above	Initial sorting by secretariat/ACTION project, validation by SOM Platform

Step 2. Estimating time-lags in measure-pressure links

Even fully implemented measures do not always have an immediate effect on the state due to time lags between measures and pressures (e.g. banned substance with persistent use of legacy production) and pressures and state (e.g. benthic communities after trawling).

Consideration of measure-pressure time-lags

- If a measure was fully implemented by the BAU base year, then one needs to estimate whether there could be any time-lag carrying its effect on effected pressures beyond the base year. If no time-lag is estimated to remain, then the effects of the measure should be visible in the pressure status and the measure does not need to be included in the SOM analyses. Otherwise, the measure is included.
- If a measure is only partially implemented or planned to be implemented, then the assumption is made that full implementation, including full effect on effected pressures, will take place by the BAU end year (cf. the urge by Ministerial Declaration 2018 to implement the BSAP). This assumption is partially suspended in the BAU scenario, where HELCOM measures are evaluated only at the base year implementation status and measure-pressure time-lags will need to be considered during Step 4.

Consideration of pressure-state time lags is presented in Step 8.

Information needed	Data sources	Main contribution
Data on time lags of effect of measures on state	Literature	Input from SOM Topic teams[7], ACTION project

Step 3. Identifying main pathways for pressures using activity-pressure-linkages

Assessing the effects of measures means describing how they affect pressures or state either directly or via activities. Thus, the links between activities and pressures need to be identified and quantified. Information on the linkages between activities and pressures is available, for instance, in the activity-pressure matrix of the TAPAS project, and in more detail in similar matrices of the DEVOTES project. These can be used as a starting point to identify the main pathways. A key issue is that the links should be (semi)quantitative and, hence, allow for assessing the relative contribution of the activities to the pressure. This is important for assessing the proportion of the pressure reduction attributable to each activity and for identifying potential new measures. This phase will include an expert survey which is supported by pre-filled information of significant activities for each pressure type. Experts will be asked to estimate the most likely contribution of relevant activities to specific pressures, as well as, the lower and upper bounds of contribution for each relevant activity.

Information needed	Data sources	Main contribution
Links between activities and pressures	Project results (e.g. HELCOM TAPAS linkage matrices, DEVOTES linkage matrices)	ACTION project. Anticipated that existing results can be used
Information on relative contribution to pressures from different activities	HELCOM reports, literature	Input from SOM Topic teams[7], ACTION project
	Expert-based evaluation	Survey participation by SOM Topic teams, ACTION, HELCOM ENs, EGs, WGs

Step 4. Estimation of effects of measure types

When the main pathways between activities and pressures have been identified, one will estimate how much measures will jointly reduce each pressure. In the case of restoration measures, this step will entail estimating how much measure types will affect the state components and be used in Steps 6 and 7. The information on effects of measures will be reviewed from several past and on-going projects and then surveyed from experts using probability scales.

Expert evaluation can also be used to survey for possible hidden/neglected pressures that were not identified in Step 3. The relative effects of measures on pressures and state are proposed to be defined as probability distributions that describe the probability of different reduction outcomes (e.g. using percentages (%)). The total effect of measures includes the effect of reduction in pressures on state and the direct effect on state.

4 a. **Survey**. Surveys for different topics were carried out with online surveys, with support for development gathered in physical SOM topic workshops. The surveys were implemented by Webropol software where effectiveness of the measure types was given for each activity-pressure combination (only main activities contributing to a pressure). The effectiveness of existing measures was assessed on the level of 'measure types'. The response was on a scale (no effect – highest effect), where the effectiveness of measure types was assessed relative to each other and level of certainty was asked for each response. Figure 4 illustrates the survey question for an activity causing a pressure. As several activities can contribute to same pressure,

[7] 'SOM Topic teams' refer to teams of national experts that will contribute to the analyses for topics covered by the SOM Platform, see also document 5- 7 Organization of SOM work.

the survey had several questions for a pressure and hence was limited to the main activities linked to a pressure (based on the ACTION activity-pressure survey outcome).

1. In your expert opinion, what is the relative effectiveness of each of the following measure types in reducing A PRESSURE from AN ACTIVITY, and what is the certainty of the effectiveness of each measure type?

To answer, mark one measure in the list, then click at the position in the grid where you want place it. You can adjust the position of already entered points by moving them within the grid. Begin by placing the measure type with the highest effect on the far right of the horizontal axis and continue with the remaining measure types as appropriate.

- **Measure type 1**
- **Measure type 2**
- **Measure type 3**
- **Measure type 4**
- **Measure type 5**
- **Measure type 6**
- **…**

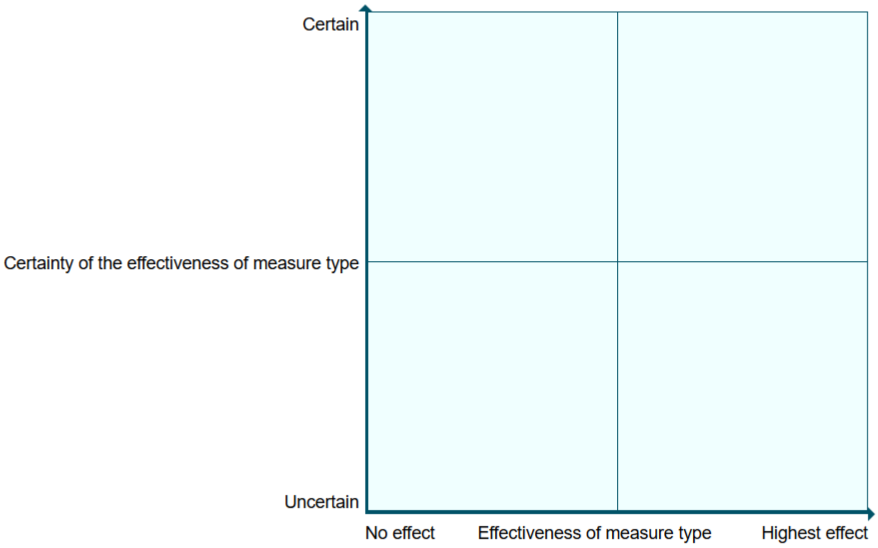

Figure 4. Grid question on the (relative) effectiveness of measure types in reducing a pressure from an activity and the certainty of the effectiveness. Effectiveness of measure types is estimated on the horizontal axis and the level of certainty on the vertical axis. [Figure has been updated]

4b. **Reference points.** The pressure reduction in Step 4a was estimated on a relative scale and this needs to be transformed into reduction %. Therefore, the survey also included a second question how much the most effective measure type is expected to reduce the given pressure (from the given activity) (see Figure 5). This estimate was requested using a sliding bar with a range from 0% to 100%. ~~This estimate can be given in categories 0%, 1-3%, 4-6%,71-10%, 11-15%, 16-20 21-30%, 31-40%, 41-60%, 61-80%, 81-99%, 100%.~~ One estimate is sufficient as it gives a reference point to the effectiveness scale (Step 4a), where effectiveness of other measures can be calculated from only one reference point.

4c. **External studies as reference points**. The effectiveness of measures can also be found from independent research outputs. These reference points are compared by the ACTION project with the ones from Step 4b.

Step 4d. **Integration**. Pressure reductions from the measure types can be summed to see the total pressure reduction, but this reduction is still from one activity. Based on Step 3, the activity-pressures contributions (%) are known, and using this information, one can integrate all the pressure reductions. This is the final figure which can be used to (i) compare against pressure targets (e.g. nutrient reduction targets) or (ii) which goes to Step 6.

The integration is further described in PART III.

2. Think of the measure type you rated as the most effective in the previous question. In your expert opinion, how much can the most effective measure type reduce A PRESSURE from AN ACTIVITY?

Provide your answer as a percent reduction. 100% means that the measure type will eliminate the specific pressure from this specific activity; 0% means the measure has no effect. The most effective measure type is the measure type furthest on the right in the previous question. When answering, assess measure effectiveness in areas where the measure could reasonably be implemented.

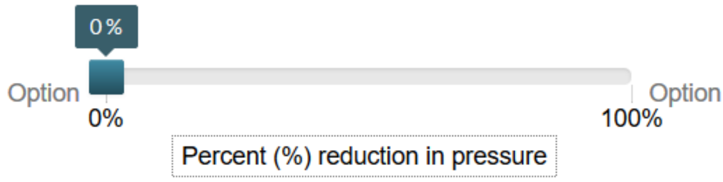

Figure 5. Question on the effectiveness of the most effective measure type in reducing pressure from an activity [Figure has been updated]

~~In addition, some deviations from the general model are planned for the following topics:~~

- ~~Spatially restricted measures: some measures are not covering the entire Baltic Sea or the assessment area. These are, *inter alia*, national MSFD measures implemented in one country only, restrictions to trawling which cover a small area of the sea, or marine protected areas. Effectiveness of these in reducing a pressure requires spatial weighting in order to not overemphasize their effectiveness. Support is asked from ACTION WP3 (MPAs).~~
- ~~Marine litter: the effectiveness will depend on litter type and this may require some modifications to the survey and the analysis.~~
- ~~Specific legal requirements over the EU Member States: some EU legal requirements are very specific in their pressure reductions. The litter-related directives set numeric pressure reduction targets for single-use plastics and packaging waste are require measures to meet the targets by 2025. As these are assumed to be implemented (as all the measures in BAU period), the litter inputs to sea can be assumed to decrease according to the requirement and not effectiveness survey is needed for these measures.~~[8]

Information needed	Data sources	Main contribution
Data on effects of measures	National data	Reporting by countries
	Research projects (e.g. BONUS, BLUE2) Scientific literature, studies and models EU MSFD Programmes of measures Sources listed in the SPICE project deliverable on Business-as-usual scenarios EC DG ENV databases (e.g. ARCADIS 2012)	Input from SOM Topic teams[7], ACTION project on existing measures Input from SOM synopses on potential new actions and measures
	Expert evaluation/validation	Working Groups, Expert Groups, ACTION project, SOM Platform

[8] Removed text refers to proposed methodological issues. Complete methodology is available in Part III.

Step 5. Projected development of human activities/pressures

The other component affecting the BAU state in addition to existing measures is the possible (external) change in activities due to changes in human behaviour in the time frame of the BAU. This may counteract the effect of existing measures if activities increase and lead to increased pressures.

This step is proposed to be run as an additional scenario on top of the effectiveness of existing measures analysis. The analysis will be limited to the predominant activities and pressures. As this component would be considered as external to the rest of the framework, the BAU status can be developed for different scenarios, e.g. assuming 1) no change, 2) the most likely change in predominant activities/pressures. This enables assessing how the future change in activities affects the BAU status.

The scenarios of projected human activities have defined following the guidelines developed in the HELCOM SPICE project.

Method for this part is described in a separate document (Document 2-2).

~~At minimum, qualitative assessment describing the trend (increasing, decreasing, no change) in the activity/pressure should be made, but quantitative information should be used when available from existing studies. For developing the BAU, the information should be converted into numerical values, e.g. 10% increase in the activity, using expert evaluation when needed. If little information is available, it would be possible to assume something about the change in activities and see how the BAU status changes.~~

~~The scenarios of projected human activities will be defined following the guidelines developed in the HELCOM SPICE project. Shared socioeconomic pathways developed in the BONUS BALTCIAPP project and alternative scenarios for blue economies created in Maritime Spatial Planning for Sustainable Blue Economies (PLAN4BLUE) project can be used in defining the scenarios of projected human activities.~~ [9]

Information needed	Data sources	Main contribution
Information on the future development of activities (qualitative/quantitative)	Literature, sectorial future outlook reports Project outputs (e.g. BONUS) National data (e.g. on EU MSFD Initial Assessments, and MSPD)	secretariat/ACTION/ Input from SOM Topic teams[2]
Converting the information into numerical values	Expert evaluation	Working Groups, Expert Groups, ACTION project, SOM Platform

Step 6. Linking reduced pressures with state components

Following the suggestions of the kick-off meeting of SOM platform, the SOM analysis will be structured using the same major pressure themes and biodiversity components as in the State of the Baltic Sea report (HOLAS II) and other HELCOM agreements. Additionally, the methodology for the SOM analysis will be adaptable to cover both topics with and without established GES thresholds or pressure targets. Information on the pressure-state linkages was collected with expert surveys.

6a. **Prioritizing pressures**. ~~ACTION project will define the most significant pressures for each state component by using the information in HELCOM core indicators and Baltic Sea Impact Index. Using a list of pressures (based on MSFD Annex III), a survey is carried out to ask the most significant pressures for each of the state components. It is possible to add other pressures to the response if necessary.~~[10] In the surveys, experts were first asked to identify 3-6 significant pressures affecting the state component in question. The

[9] Complete methodology for Step 5 is avalaible in Section III.
[10] Method revised following input from SOM topic workshops

pressure ranking was done on a relative scale (Figure 5) which allows for estimating the contributions of the pressures for the state component.

6b. **Needed pressure reduction.** After the ranking, the survey asked by how much all significant pressures need to be reduced in order to achieve state improvements, regardless of the time it takes. The survey asked for the minimum, most likely and maximum reduction (Figure 6). asks for the most significant pressure what the needed pressure reduction % is in order to reach good state assuming that other significant pressures are decreased the same amount. The ranking of pressures allows the scaling of the other significant pressures to the most significant one. It is also assumed that the uncertainty for all the other pressures is proportional to the most important one.

As the relations of pressures are known from this step, it is possible to assess what is the probability to reach good state given that the pressures are reduced the amount resulting from Step 4. It is also possible to approximate the reduction needed on one pressure to reach good state assuming that the reductions on other pressures are known.

The exact approach depends on whether the topic is described in terms of pressures or state, and on the existence of an agreed GES threshold as follows:

- When a pressure reduction target exists (eutrophication) or the indicator is defined in terms of pressures, the evaluation will stop at the pressure component and environmental state will not be evaluated in the SOM model. This is the case for eutrophication, marine litter, underwater noise and non-indigenous species. However, the pressure-state time-lags are included in the final considerations.
- When a GES threshold exists, the experts were asked how much all significant pressures need to be reduced to achieve GES. From these it is possible to define a cumulative distribution for required pressure reduction that represents the probability to reach GES for a given reduction in significant pressures. Thus, the gap to GES can be defined as a probability of not reaching GES for expected reduction in significant pressures resulting from existing measures, or based on percentiles of the pressure reduction distributions, and an improvement in state can be defined as an increase in the probability of GES. The gap can also be presented as required additional reduction in total pressure (consisting of significant pressures) to achieve GES with 100% probability. However, this reduction applies to a combined pressure that can consist of multiple pressures of varying significance from different basins.
- When a GES threshold does not exist, the experts were asked how much all significant pressures need to be reduced in order to reach a given state improvement (in %), or how much pressures needs to be reduced to reach a noticeable improvement in state. These, in combination with estimated pressure reductions resulting from existing measures, can be used in an advisory capacity to express the scale of improvement in state.
- When a GES threshold exists, contributions of pressures to state will be determined. Pressure-state response curves would then be generated through existing data and expert opinion for each identified major pressure(s), which affect the state component. These data would then allow for the calculation of a BAU status and the gap analysis.
- When a GES threshold does not exist, an approximate good status based on qualitative environmental targets set in various HELCOM documents (BSAP, topic specific action plans, ministerial declarations, etc.) is proposed. This will allow for a generalized gap analysis, which can be used in an advisory capacity to express the scale of improvement required to achieve a hypothetical GES.

Two major topics still under consideration in consultation with SOM platform and ACTION partners are (1) whether to link pressure to state improvement or to the probability of achieving GES, which will impact

~~how GES threshold values are used in the model, and (2) how to best link states to multiple pressures. The SOM platform and appropriate HELCOM bodies will be kept informed as this work progresses.~~[11]

Information needed	Data sources	Main contribution
Spatial data on pressures and impacts	HELCOM map and data service	Secretariat
Spatial data on state components	HELCOM map and data service	Secretariat
Information for selecting relevant pressures	Baltic Sea Impact Index (BSII) Core indicator reports, ODEMM framework	Secretariat/ACTION project
Responses of indicators/state components to changes in pressures	Previous research projects and reports Scientific literature Existing models	Input from SOM Topic teams[7], ACTION project
	Expert evaluation/validation	Working Groups, Expert Groups, ACTION project, SOM Platform

1. In your expert opinion, what are the most significant human-induced pressures to the <u>GIVEN STATE VARIABLE</u>?
Choose three (3) to six (6) pressures from the drop-down menus and rate their significance. Significance covers the intensity of the pressure, sensitivity of the state component to the pressure and geographic extent of the pressure.

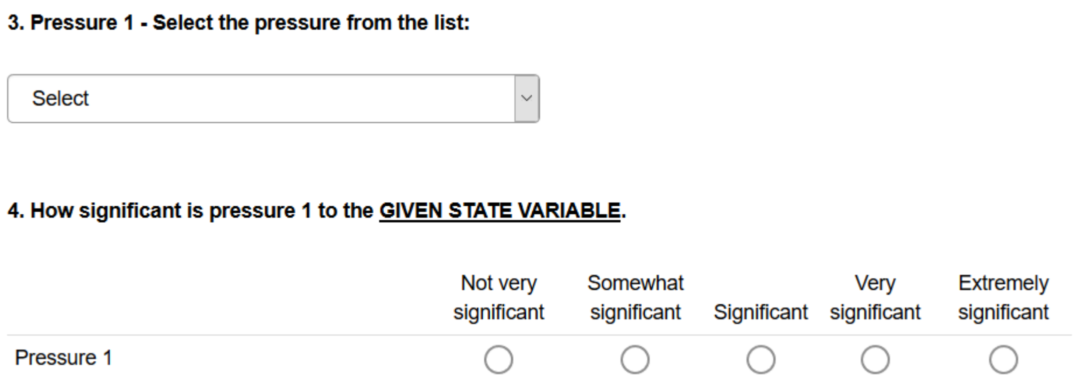

Figure 6. Question to identify the most significant pressures affecting the state variable [Figure has been updated]

[11] Rewritten based on current methodology. See Section III for the SOM model and methodological details.

2. Consider all the pressures you selected in the previous questions. In your expert opinion, by what percentage do these pressures need to be reduced in order to achieve or maintain a good state for the GIVEN STATE VARIABLE, regardless of the time it takes?

Base this assessment on the HELCOM core indicator INDICATOR NAME where good state is GOOD STATE. Current state is CURRENT STATE, i.e. the state is STATUS.

Please estimate the minimum, maximum and the most likely percent (%) reduction in pressures required to achieve or maintain a good state.

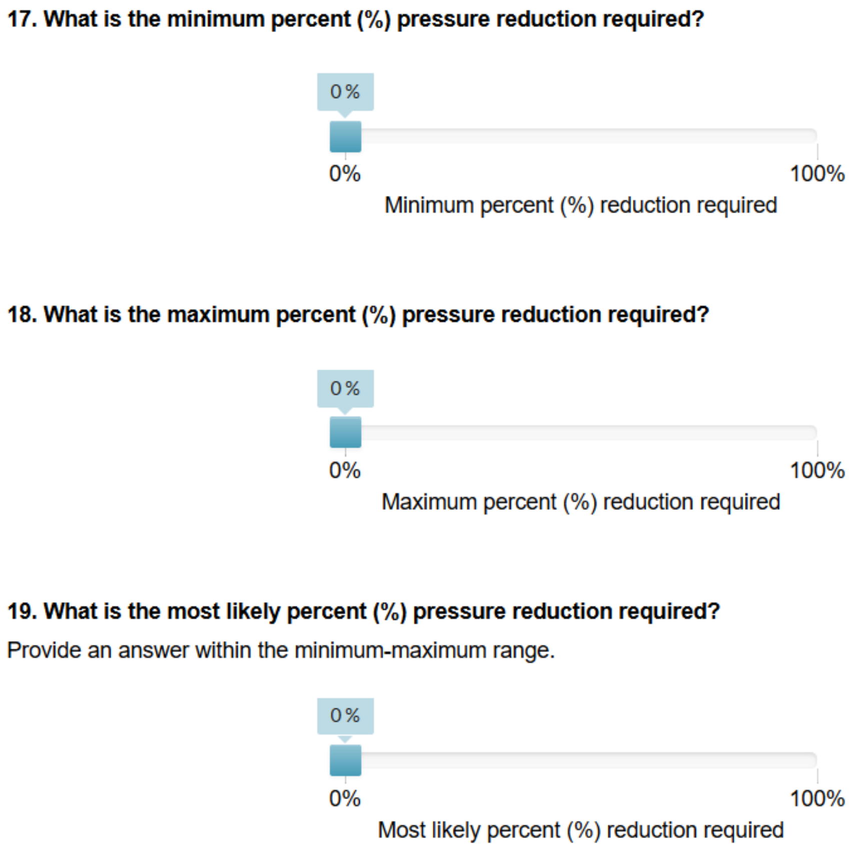

Figure 7. Question on the required pressures reductions to achieve state improvements to define three - point estimates to assess the linkage between total pressure and state. This version is for cases where there is an agreed GES threshold [Figure has been updated]

Step 7. Comparison of BAU and GES and assessing sufficiency of measures

When the BAU status has been developed, it will be compared with GES to identify whether there is a gap and new measures are needed. The total effect of measures on state is calculated as the reduction of the GES gap resulting from reductions in pressures based on the previous steps. This reduction is proposed to be measured as an increase in the probability of reaching GES for different themes and components. The probabilistic approach further enables an extensive analysis of uncertainty and risk related to the BAU outcome. In addition, the Step 5 results (projected development in human activities/pressures) will also affect the outcome of the SOM analysis. If a pressure is predicted to increase and no measures are in place to control that pressure, the gap to GES may increase.

Step 8. Time lags in state recovery

Reductions in pressures during the BAU period do not necessarily mean that the state will become good before 2030-2035. The lags in recovery may result from multiple reasons which are identified in the ACTION project.

In the context of SOM analysis, the issue with time lags could be resolved by focusing on pressure reductions and possible effect on state (even if the state recovery takes place much longer time).

Pressure-state time lags are not included in the BAU scenario. Instead they will be evaluated as additional information alongside GES thresholds as in Step 7. By separating pressure-state time lags from the BAU scenario, the effect of measures can be separated from unavoidable time-lags (e.g. population growth) and allow for the consideration of the sufficiency of measures in the case of avoidable time-lags (i.e. topic is projected to eventually reach GES under BAU conditions, but GES could be reached sooner if additional measures were implemented). Additionally, topics with the defining feature of very large pressure-state time lags (e.g. eutrophication) will only be evaluated to the level of pressure in the BAU analysis as it is already known that GES will not be achieved by the BAU target year.

Information on time lags has been collected with the expert surveys on pressure-state linkages. The formulation of the question depended on the existence of a GES threshold and the topic of the survey. Figure 8 presents an example of the time lag question for state components with a GES threshold.

Assume sufficient measures are implemented to achieve GES for the GIVEN STATE VARIABLE in an infinite time horizon. How long will it take to achieve GES?
Consider all possible time lags between changes in pressures and changes in the state variable when answering.

- 0 years (no time lag)
- 0-5 years
- 6-10 years
- 11-25 years
- 26-50 years
- 51-100 years
- More than 100 years

Figure 8. Time lag question in the pressure-state surveys

PART II DATA COLLECTION AND INPUT DATA FOR THE SOM MODEL

This section is completely new information, as data collection has mainly taken place after SOM 2-2019.

5. Data collection for the SOM analysis

All main components of the SOM analysis have entailed the collection of existing or new data. Whenever possible, data used in the analysis have been based on existing literature, studies and models. However, a preliminary investigation on data availability identified significant gaps in published information with regard to 1) activity-pressure contributions (Step 3), 2) effectiveness of measures in reducing pressures (Step 4), 3) pressure-state linkages (Step 6), and thus expert elicitation was necessary to allow for comprehensive inclusion of measures, activities, pressures and state components in the analysis.

Thus, information on activity-pressure contributions, effectiveness of measures and pressure-state linkages is mainly collected using expert elicitation. When available, expert inputs are complemented with information from existing literature, and a specific literature review is conducted for the effectiveness of measures part.

Information on existing measures (Steps 1 and 2) was compiled by the Secretariat in June – July 2019 based on available information in HELCOM reporting, EU legislation, and international conventions. The compiled measures were subsequently distributed to the Contracting Parties for review and supplementation. The compiled measures were then linked to the developed measure types to create a library of implemented measure types. SOM 2-2019 requested that a second round of review be undertaken following the development of measure types to allow countries to reassess their efforts in this new format. That process is likely to be initiated in March 2020.

Data on the projected development of human activities (Step 5) are compiled from existing national and regional reports and studies that provide assessments of how activities develop in the future in the Baltic Sea region. The data and methodology are described in detail in a separate document (Document 2-2).

The expert surveys to collect data for the SOM analysis have been implemented in 2019 – early 2020.

6. Activity-pressure contributions

Activity-pressure linkages (Step 3) are assessed as the percent (%) contribution of activities to pressures. Quantification of these linkages has been based on either expert elicitation or existing data sources.

A data-based approach has been preferred and was possible for three pressures. For loss and disturbance to the seabed (benthic habitats), the approach used in HELCOM HOLAS II has been employed, which utilizes the Baltic Sea Pressure Index (BSPI) and Baltic Sea Impact Index (BSII) to integrate data reported to the Secretariat from the Contracting Parties through regular reporting and previous data calls. For the primary introduction of non-indigenous species, data has been gathered from the AquaNIS database on reported vectors of introduction for all primary introductions into the Baltic Sea in 2005-2016. For the input of nutrients, activity-pressure linkages are being produced by HELCOM ACTION Work Package 4 and based primarily on PLC-6.

Expert elicitation has been used to estimate activity-pressure relationships for the following pressures: input of hazardous substances (further differentiated to mercury, PFOS, TBT and diclofenac), input of marine litter (further differentiated by litter item), disturbance or displacement of marine mammals by human presence (further differentiated by species), disturbance or displacement of birds by human presence (further differentiated by species), input of noise (further differentiated to continuous noise 63/125 Hz, continuous noise 2 kHz, and impulsive noise with peak energy below 10 kHz). These surveys have been distributed to relevant HELCOM expert bodies and/or nationally nominated SOM experts. Additional responses are still being sought to these surveys. Responses have been received from experts based in individual Contracting Parties as indicated in Table 1.

The remainder of the pressures do not require detailed activity-pressure linkages, as they are either by definition single-activity pressures (e.g. extraction of fish only occurs through fishing) or are not fully analysed in the SOM model context (e.g. inland habitat loss/degradation).

Table 1. Number of experts contributing to activity-pressure surveys (updated 1.3.2020)

Pressure	DE	DK	EE	FI	LT	LV	PL	RU	SE
Input of hazardous substances			1	4					
Input of marine litter	1	1	1	1			1		1
Disturbance/displacement by human presence - mammals	1	2							
Disturbance/displacement by human presence - birds							1		2
Input of underwater noise	1		1				2		2

7. Effectiveness of measures and pressure-state linkages

The expert surveys on the effectiveness of measures (Step 4) and pressure-state linkages (Step 6) were developed in the autumn 2019 in collaboration of the ACTION project and the HELCOM Secretariat. Considerable input was received also from HELCOM expert groups and networks, HELCOM SOM Platform, HELCOM SOM topic teams, as well as dedicated topic workshops to support the SOM analysis. The exact process depends on the topic in question and is outlined in Table 2.

The first versions of the general format for the effectiveness of measures and pressure-state linkages expert surveys were developed in September 2019. They were presented in the SOM Platform 2-2019 meeting, and subsequently revised, e.g. to include questions on experts' background. The surveys were then pre-tested within the HELCOM Secretariat and implemented during the first SOM topic workshops in late September (marine mammals) and early October (birds). Following feedback received in relation to these workshops, the surveys re-entered a development phase. The surveys were iteratively developed over the course of the remaining topic workshops in the fall 2019, with revisions made after each workshop to both the general structure and topic-specific contents of the surveys. For the topics not having specific SOM workshops, the survey structures were developed in collaboration with the topic teams or other topic experts. The general versions of the surveys were ready in November 2019, and the topic-specific surveys in December 2019 or January 2020, depending on the topic.

Table 2. Groups contributing to the structural development of SOM topics and/or surveys

Topic	Topic team lead	Workshop	Other groups with significant contributions
Hazardous substances	Denmark, Sweden	SOM-HZ WS 1-2019	EN HZ
Marine litter	Estonia	-	EN Litter
Underwater noise	Denmark	-	EN Noise
Benthic habitats	-	EN BENTHIC 3-2019	EN Benthic, ACTION WP 2
Migratory fish	Finland	SOM-FISH WS 1-2019	
Coastal and commercial fish	Sweden	SOM-FISH WS 1-2019	
Marine mammals	-	SOM Bio-MM WS1	
Birds	-	SOM-Birds WS 1-2019	
NIS	Secretariat	-	
Nutrients	-	-	ACTION WP4

The overall structure of the expert surveys and questions are similar across topics, but topic-specific adjustments have been made whenever needed. For example, for marine litter, a litter item-based approach was developed to better reflect the topic's regulatory environment, and for benthic habitats, a qualitative state improvement specifying a noticeable improvement was used rather than the quantitative (percent) changes in state components used for other topics.

The expert surveys on the effectiveness of measures and pressure-state linkages were implemented using an online survey tool, Webropol, in December 2019 – February 2020. The invitations to the surveys were sent via email to the experts, followed by two reminders. The structure of both surveys was similar. The first part introduced the topic of the survey, together with instructions for answering and contact information in case of questions. The second part included the main questions and the third questions on the background of the respondent(s). Both surveys included questions on experts' own evaluation of their confidence in the response they had given and allowed for saving responses and continuing later. The surveys also collect background information on respondents' field of expertise and how long they have been in that field.

As the purpose of the ACTION project and the SOM analysis is to support the HELCOM Baltic Sea Action Plan update, the identification of experts to respond to the surveys relied on existing HELCOM structures and expertise. The expert pool was formed from the representatives of the relevant HELCOM expert networks and groups, as well as any additional experts nominated by Contracting Parties specifically for the task. Altogether, 469 experts (unique cases) were identified as potential respondents to the surveys, with 35-114 experts per topic. Note that some were identified/nominated as experts to respond to multiple topics, and thus the total size of the expert pool was 518 experts.

Table 3 shows the number of contributing experts per survey and country. As group responses are allowed, one survey response may be the collaborative effort of several experts. Table 4, at the end of this section, shows the number of responses per sub-topic, taking into consideration the contribution of each expert in case of group responses.

The invitations to the surveys on benthic habitats, birds, mammals, fish, hazardous substances and non-indigenous species were sent in December 2019. Invitations to the nutrients survey were sent in January 2020, and invitations to litter and noise surveys will be sent latest early February 2020. It is worth noting that responses may not be complete: experts may have only provided answers to specific sections or questions of the survey, e.g. specific areas, species or measure types. Additionally, a lack of responses to a topic by a country should not be necessarily be interpreted as lack of engagement in the process. Several nominated experts have acknowledged the survey and declined to participate.

Table 3. Number of experts contributing to surveys on effectiveness of measures (EoM) and pressure-state linkages (P-S) (updated 1.3.2020)

Survey	DE	DK	EE	FI	LT	LV	PL	RU	SE	Total
Benthic habitats EoM	7	4	-	4	2	-	-	2	4	23
Benthic habitats P-S	7	4	-	4	1	1	-	-	2	19
Birds EoM	2	5	-	-	1	-	1	2	1	12
Birds P-S	2	6	-	1	1	-	1	2	1	14
Fish EoM	6	5	5	6	2	-	2	-	12	38
Coastal fish P-S	-	4	2	2	2	-	3	-	11	24
Commercial fish P-S	4	5	-	1	2	-	3	-	6	21
Migratory fish P-S	6	-	3	2	1	-	2	-	9	23
Hazardous substances EoM	1	-	3	6	1	-	2	-	5	18
Hazardous substances P-S	1	-	3	5	1	3	1	-	5	19
Litter EoM	2	1	2	1	1	1	1	-	-	9
Mammals EoM	2	4	2	-	4	-	-	1	-	13
Mammals P-S	1	4	1	-	-	-	-	1	-	7
NIS EoM	5	2	1	2	-	2	1	-	3	16
Noise EoM	3	-	2	-	1	2	-	-	3	11
Nutrients from agriculture EoM	1	1	3	1	*	1	3	-	*	10

Values are counts of contributing experts, not survey responses, i.e. multiple experts contributing to a single survey response are each counted individually. Responses returned by Observers are included in the value of the hosting Contracting Party. EoM = effectiveness of measures, P-S = pressure-state linkages, * indicates data submitted by correspondence.

Effectiveness of measures

The information on the effectiveness of measures comes from two sources: expert surveys and existing literature and data.

Expert surveys

In the expert surveys, the effectiveness of measures in reducing pressures was, in most cases, assessed as a percent (%) change in a specific pressure from a specific activity from implementing a measure type (generalized measure). Measure types have been designed to be general enough to apply to concrete measures applied in different parts of the Baltic Sea and to limit the number of measures to be evaluated, while also being specific enough to allow for meaningful evaluation of their effectiveness by experts.

The survey included two types of questions on the effectiveness of measures types. First, experts were asked to simultaneously assess the relative effectiveness of the measure types on a scale of no effect – highest effect and the certainty of that effectiveness on a scale of uncertain – certain. Second, experts were requested to assess the percent effectiveness of the most effective measure type in reducing the specific pressure from the specific activity. Information on the percent effectiveness of the most effective measure type from the second question and on the relative effectiveness of all measure types from the first question enables assessing the percent effectiveness of all measure types.

The only topic that deviates from this general approach is nutrient runoff from agriculture, where the expert survey was constructed somewhat differently. It enabled respondents to provide both model- and expert-based estimates of the effectiveness of measures. Model-based estimates were preferred, when available, but as this is likely not the case for all Baltic Sea countries, expert assessments were welcomed.

Literature review

A literature review on the effectiveness of measures was conducted in November 2019 – February 2020 for all pressure topics receiving an effectiveness of measures assessment. The aim was to compile information from scientific articles and reports providing estimates on the effects of measures in reducing pressures that could be used in the SOM analysis to complement the expert data, either by including the estimates in the SOM model or by providing comparison points. The literature review was conducted by topic, with the information collected into structured excel files.

In the literature searches, information from the Baltic Sea was prioritized, as well as information conforming to the structure of the expert surveys (i.e. specific state components or measure types). Literature was identified mainly via Google and Google Scholar online searches, or based on suggestions from ACTION project partners, topic team members or other collaborators. Whenever available, the topic-specific survey structure, which served as a template for building the expert surveys, was used to identify the search terms in the internet-based searches for each topic, such as specific measure types, activities, and pressures. The internet-based searches were always started in connection to the Baltic Sea area and, if no suitable articles were found, widened to a more general search.

For example, information on the effectiveness of the measure type “Improved pharmaceutical take-back schemes” for diclofenac was searched with the following terms: “baltic sea reduce diclofenac via take-back schemes”, “baltic sea mitigating diclofenac improved pharmaceutical take-back schemes”, “efficiency of measure diclofenac improved pharmaceutical take-back schemes”, “efficiency of measures management of diclofenac in the environment”.

Studies containing at least qualitative information on the effectiveness of a measure were examined further to compile a variety of data, including information about the study itself (author, year), location of the study (country, water body), certainty of assessment (either provided in the study itself or assessed by the researcher), and other attributes, whenever available, related to (dis)advantages or costs of implementing the measure. The main aim was to collect information on the measure, including a description of its extent and effectiveness, as well as links to activities and pressures. The effectiveness of a measure was recorded preferably in % pressure reduction, but also as units of total pressure reduction or qualitative descriptions, depending on availability. The final data included those studies that provided at least qualitative information on the effectiveness of a measure.

Pressure-state linkages

In the beginning of the pressure-state survey, experts were able to choose the state components and, in some cases, the geographic area they would assess the pressure-state linkages for. The surveys first asked the experts to identify up to six pressures affecting the state component in question and assess each pressure's significance to the state components in general terms. Next, the survey proceeded to asking about the pressure reductions required to achieve state improvements. The format of the pressure-state linkage questions depended on the existence of an agreed HELCOM threshold for good environmental status (GES). If an agreed GES threshold exists for the state component, the pressure-state link was assessed as the required percent pressure reduction to achieve or maintain GES. If there is no agreed GES threshold, the link was assessed as the required pressure reduction to achieve a specific percent improvement in the state component. The only exception to this was for benthic habitats, where the survey asked the required pressure reduction to achieve a "noticeable state improvement" in the benthic habitat being evaluated. Information on the pressure-state linkages were entirely based on expert elicitation. Experts were requested to assess the minimum, most likely and maximum pressure reduction required, to allow them to express possible uncertainty, as well as, to provide the model with a more realistic estimate of the pressure-state linkages.

8. Data validation

The data from the expert surveys on activity-pressure contributions, effectiveness of measures and pressure-state linkages, as well as the literature review on the effectiveness of measures will be validated by HELCOM Working Groups in the spring 2020. The validation will take place intersessionally (via correspondence or online meeting) or in the WG spring meetings (detailed information in here), depending on the timing of the meeting. Topic-specific summary statistics and distributions of the responses will be presented for validation. The data will also include summary information of the background of the respondents, i.e. their country, organization type, field and years of experience.

9. Data evaluation

Some formatting of the expert survey data is required before using it in the SOM model. Answers to the effectiveness of measures and pressure-state surveys that are based on group responses have been clarified by asking for details on the individual contributions that make up the group response, i.e. whether the experts could have answered all the questions also individually. This allows for deciding whether to treat the group response as comparable to a single expert response, or as having a higher weight than a single expert response. Answers to the activity-pressure surveys represent national responses. Thus, each country is given the same weight (one) in the analysis, and if there are several individual experts representing a country, the weight of their answers sums to one.

Initial examination of the data has revealed some issues. In some cases, the responses for the minimum, maximum and most likely are inconsistent, such that the most likely is higher than the maximum. In those cases, a simple error has been assumed, and the values for the most likely and maximum have been exchanged. In addition, there are few cases with missing answers to the relative effectiveness of measures (grid) question due to technical problems. If there are several of these missing answers for the same expert, the aim is to come back to the expert and ask them to provide a new response.

In general, missing responses to questions or parts of questions are treated as missing values.

10. Integration of the expert survey and literature data

For the effectiveness of measures (Step 4), data from expert surveys will be complemented with information from existing literature, reports and models on effectiveness of measures for all topics. These literature data can be used in different ways in the SOM analysis. The usability of such data depends on whether they can be linked to the measure types (i.e. generalized measures) employed in the expert surveys. In principle, it is possible to incorporate the literature estimates in the SOM model or use them as comparison points to the expert data. Inclusion of the literature estimates in the SOM model requires in most cases that the format of the data corresponds to the format of the expert responses, i.e. enables assessing the percent reduction in a pressure from the measure type.

Direct substitution of expert responses is not, however, straightforward, because the observational or experimental results may not be produced for the Baltic Sea, the formulation of measures, effectiveness or state components may differ from the one used in the SOM analysis, or the research question has not been directed to study the effectiveness of measures.

The integration of the literature data with the expert data depends on the format of the data (i.e. whether it is possible to use the data directly in the SOM model).

a) Cases when it is possible to use literature data in the SOM model

   It is proposed that separate SOM models are run using 1) only expert survey data and 2) both literature and expert survey data, by replacing the expert survey data points with the literature data, when available. Literature data cannot cover all data points in the model, so it is not possible to estimate a separate model using only literature data.

b) Cases when it is not possible to include the literature data in the SOM model

   It is proposed that the literature estimates are used as external points of comparison and they are reflected in the discussion of the model results.

If the literature estimates cannot be linked to measure types or existing measures, new measure types could be defined. These can be useful in assessing the effectiveness of potential new measures.

Table 4. Maximum response count by sub-topic. Actual response counts by question will be lower due to skipped questions, technical errors, and within sub-topic variation in group response contributions. Follow-up with group responses is ongoing and the potential increase in the maximum response counts is noted in the last column. Responses to the agriculture survey are not included in this table as responses are collected by country rather than expert; see Table 3 for summary of responses. EoM = effectiveness of measures, P-S = pressure-state linkages (updated 1.3.2020)

Survey	Sub-topic	Geographic area	Maximum weighted response count	Potential increase in count pending group response follow-ups
EoM Benthic		Whole Baltic	23	
P-S Benthic	hard substrate vegetation dominated community	Kattegat	5	
		Southern Baltic	9	
		Eastern Baltic	5	
		Northern Baltic	4	
	soft substrate vegetation dominated community	Kattegat	2	
		Southern Baltic	7	
		Eastern Baltic	3	
		Northern Baltic	3	
	hard substrate epifauna dominated community	Kattegat	5	
		Southern Baltic	10	
		Eastern Baltic	3	
		Northern Baltic	3	
	soft substrate infauna dominated community	Kattegat	3	
		Southern Baltic	9	
		Eastern Baltic	4	

Survey	Sub-topic	Geographic area	Maximum weighted response count	Potential increase in count pending group response follow-ups
		Northern Baltic	4	
	coarse substrate infauna dominated community	Kattegat	2	
		Southern Baltic	5	
		Eastern Baltic	2	
		Northern Baltic	2	
EoM Birds		Whole Baltic	12	
P-S Birds	Common eider - Breeding Season	Whole Baltic	11	+1
	Great cormorant - Breeding Season	Whole Baltic	10	+1
	Sandwich tern - Breeding Season	Whole Baltic	5	
	Long-tailed duck - Wintering Season	Whole Baltic	9	+1
	Red-throated diver - Wintering Season	Whole Baltic	8	+1
	Great black-backed gull - Wintering Season	Whole Baltic	6	
EoM Fish		Whole Baltic	38	
P-S Coastal Fish	Perch and other coastal piscivores	Gulf of Bothnia	10	
		Gulf of Finland	3	
		Gulf of Riga	2	
		Central (Swedish coastal areas only)	9	
		Eastern Gotland Basin (Latvian and Lithuanian coastal areas only)	2	
		South (Polish coastal areas only)	2	+1

Survey	Sub-topic	Geographic area	Maximum weighted response count	Potential increase in count pending group response follow-ups
	Cyprinids and other mesopredators	Gulf of Bothnia	7	
		Gulf of Finland	3	
		Gulf of Riga	1	
		Central (Swedish coastal areas only)	6	
		Eastern Gotland Basin (Latvian and Lithuanian coastal areas only)	2	
		South (Polish coastal areas only)	2	+1
	Flounder	Central (Swedish coastal areas only)	6	
		Eastern Gotland Basin (Latvian & Lithuanian coastal areas only)	1	
		Southwest (Danish coastal areas only)	4	
		South (Polish coastal areas only)	6	+1
P-S Commercial Fish	Herring SD 20-24, spring spawners		9	
	Herring SD 25-29, 32 (excl Gulf of Riga)		13	
	Herring SD 28.1 (Gulf of Riga)		1	
	Herring SD 30-31		8	
	Sprat SD 22-30, 32		16	
	Cod, western		10	
	Cod, eastern		18	+1
	Plaice		7	

Survey	Sub-topic	Geographic area	Maximum weighted response count	Potential increase in count pending group response follow-ups
P-S Migratory Fish	Salmon in assessment units 1-2		7	
	Salmon in assessment unit 3		7	
	Salmon in assessment unit 4		9	
	Salmon in assessment unit 5		4	
	Salmon in assessment unit 6		4	
	Seatrout - Gulf of Bothnia		6	
	Seatrout - Gulf of Finland		3	
	Seatrout - Western Baltic		7	
	Seatrout - Eastern Baltic		3	
	Seatrout - Southern Baltic		10	
	Eel - Entire Baltic Sea		9	+2
EoM Hazardous substances	mercury	Whole Baltic	10	
	TBT	Whole Baltic	8	
	PFOS	Whole Baltic	11	
	diclofenac	Whole Baltic	11	
P-S Hazardous substances	mercury	Whole Baltic	11	
	TBT	Whole Baltic	7	
	PFOS	Whole Baltic	7	
	diclofenac	Whole Baltic	10	
EoM Litter		Whole Baltic	9	

Survey	Sub-topic	Geographic area	Maximum weighted response count	Potential increase in count pending group response follow-ups
EoM Mammals	Porpoise	Whole Baltic	9	
	Seals	Whole Baltic	10	
P-S Mammals	Grey seal	Whole Baltic	5	
	Ringed seal	Northern population	1	
		Southern population	3	
	Harbour seal	Kattegat	1	
		Southern Baltic	2	
		Kalmarsund	0	
	Harbour porpoise	Western Baltic	4	
		Baltic proper	2	
EoM NIS		Whole Baltic	16	
EoM Noise	Continuous noise 63/125 Hz	Whole Baltic	7	+1
	Continuous noise 2 kHz	Whole Baltic	6	
	Impulsive noise with peak energy below 10 kHz	Whole Baltic	6	+2

PART III THE SOM MODEL: STEPS TO ANALYSE THE SUFFICIENCY OF MEASURES

This section describes the SOM model in detail and explains how the model functions. It is based partially on document 2-3 (Use of survey results from expert elicitation in the SOM model) submitted to SOM 2-2019. Changes from this document and new text are indicated in red text. Minor editorial changes, including relocated text and removal of outdated document references, have not been indicated.

11. General aim of the model

The model predicts the pressure reductions (%) from existing measures (effectiveness of measures, step 4) and the subsequent changes in environmental status (%) (pressure-state linkages, step 6), taking into consideration the activity-pressure contributions (step 3). Effectiveness of measures input data come from expert surveys, and whenever possible, is replaced with or compared to estimates from existing literature. The data on pressure-state linkages come from expert surveys. Activity-pressure contributions are either data- or expert-based, depending on the topic. For details on the data collection, see Part II.

The results enable assessing the probability of reaching GES given a reduction in total pressures affecting the state component for those components which have an existing GES threshold. When a GES threshold does not exist, it is possible to assess the probability of achieving specific state improvements given the reduction in total pressures affecting the state component.

12. Implementation status of measures in the model

The model uses measure types as the unit for assessing the BAU state. However, the actual implementation always occurs on the level of individual measures, and therefore the model links each measure type to the actual measures and the following information:

- Implementation status: only those measures which have not yet been implemented or have only been partially implemented are included in the SOM model, because they have the potential to still reduce pressures and improve the environmental state over in the timeframe of the analysis. Areas where measures are already fully implemented are not included to SOM model (see next point).
- Area of implementation: some measures are Baltic-wide, some cover EU waters, some are national (covering all the waters of those countries) and some are sub-national (covering less than all of a country's territorial waters). Thus, the area of implementation is an important factor for the implementation status.

The model considers the above two factors on the level of measure types and gives the pressure reduction potential based on the combination of these factors.

However, note that the SOM model does not operate on 'actual measures' but 'measure types', but the actual measures are the ones being implemented. Therefore, the information on implementation is aggregated to the level of measure types.

Further, it is proposed that credit for implementing a measure type only be applied once per area. This means that if more than one existing measure of a particular measure type is being implemented in any given area, the model will only consider the effect from one implementation of the corresponding measure type. Many existing measures are cross-listed under several measure frameworks, making it difficult to determine what is a distinct existing measure and increasing the risk of overcounting existing measures. This proposal would mitigate that risk. The proposal also entails that if a single existing measure is implemented under a measure type, the measure type is considered to be fully implemented.

13. Joint impacts of measure types

In principle, the effects of measures are additive (effects are added up) in the SOM model. Many measure types have joint impacts which need to be taken into account to avoid over- or underestimations of measure effectiveness. Two types of joint impacts are considered in the model:

- **Thematic overlap** in measure types due to their existence on different policy levels (global, EU, HELCOM, national) or overlapping content (e.g. MPAs in general vs. fishing closures in a specific area).
- **Chain effects of measure types in reducing pressures.** Assuming that measures take effect in a chain, a measure can only impact the pressure share that remains after the preceding measures. As the pressure reductions are in percent (%), the chain effect needs to be taken into account.

The thematic overlap is taken into account by recognizing thematically similar measure types and considering their overlaps one by one. As a result, the overlapping measure types are set in a hierarchical order where one measure type makes another one partly or completely obsolete. The effectiveness of the measure types is reduced according to the overlaps along this hierarchical order.

Identified **thematic overlaps** fall into one of four categories. The first category consists of national management plan measure types and the constituent measure types that may be present inside such management plans. If any of the identified constituent measure types are implemented, the effectiveness of the national management plan measure type is multiplied by 0.2 (i.e. the effectiveness of the national plan is assumed to be 20% of its original effectiveness). This avoids double counting of management measures while still giving credit to the coordination value of the management plan. The second category is international management plans and the constituent measure types that may be present inside such management plans. This category is identical to the first, except that the management plan measure type is multiplied by 0.4, giving less of a discount to reflect the added value of international cooperation. The third category is measure types that are completely overlapping with other measure types. These measure types typically represent a status quo situation (e.g. current wording of Stockholm convention annexes) and a more stringent wording of the same measure type (e.g. updated wording of Stockholm convention annexes). If the more stringent measure type is implemented, then the weaker measure type is multiplied by 0 (i.e. it has no effectiveness). The final category is technical overlaps, where overlap strength is assessed on a case by case basis in 20% intervals (multipliers of 0, 0.2, 0.4, 0.6, 0.8, 1). Overlap relationships can be found in document 2-1 Att 1. Measure type overlaps.

The **chain effects** are recognized by first defining individual effects (from the measure type effectiveness distributions based on survey responses and overlaps) of N actual measures affecting a pressure from an activity implemented by a country in a given basin can be defined as vector $X = [x_{i=1}, \dots, x_N]$. Assuming that measures take effect in a chain, a measure can only impact the pressure share that remains after the preceding measures. For chain effects, the joint impact of a measure ***i*** can be defined by recursive function

$$\boldsymbol{f}(x_i) = \begin{cases} x_i, & i = 1 \\ (1 - \sum_{j=1}^{j=i-1} \boldsymbol{f}(x_j))x_i, & i > 1 \end{cases}$$

The total joint impact (=sum of joint impacts of measures) of the measures until the i:th measure can also be defined as recursive function

$$\boldsymbol{F}(x_i) = \begin{cases} x_i, & i = 1 \\ (1 - \boldsymbol{F}(x_{i-1}))x_i + \boldsymbol{F}(x_{i-1}), & i > 1 \end{cases}$$

The total joint impact of all N measures is $\boldsymbol{F}(x_N)$. The total joint impact of all measures is not affected by the order of measures in the vector X, and thus the total joint impact defined for chain impacts can be used to approximate the total joint impact of effects regardless of the order of the effects. If the measure effects are between 0 and 1, the total joint impact can only have values between 0 and 1. This implies that pressures cannot be reduced more than 100%.

14. Geographical areas in the SOM model

The SOM model divides the Baltic Sea into the 17 HELCOM sub-basins. However, geographical divisions differ depending on input data: the effectiveness of measures data is for the entire Baltic Sea area, the activity-pressure contributions data for 4-17 areas (depending on the topic), and the pressure-state data has varying areal divisions, depending on the topic; the benthic survey had most areas (eight areas).

The SOM approach includes measures only from the areas where they are to be implemented during the BAU timeframe. This information is available in the list of existing measures alongside the country of implementation and specific area of implementation if different than national. The model takes the intersection of the country of implementation and sub-basin of implementation areas, and gives effectiveness on this spatial scale.

15. Model structure

Pooling of expert judgements

The probability distributions describing the views of experts on activity-pressure contributions, measure effectiveness and on the probability of achieving good state are defined based on the three-point estimates (minimum, most likely and maximum) given directly by experts (activity-pressure, probability of reaching good state), or three-point estimates derived from expert judgements (measure effectiveness). This allows the comparison of results across different topics and consistent assessment of pressure reductions for different state components that are affected by multiple activities and pressures.

The distributions are defined from the three-point estimates of individual experts to the shape of PERT distributions. The PERT distribution is a modification of the beta distribution, where a variable can take values between any minimum and maximum values, whereas for a standard beta distribution, minimum and maximum values are fixed to 0 and 1. In the PERT distribution, the expected value is defined as $\mu = \frac{min+\gamma\, ml+max}{\gamma+2}$ where γ=4, and min, ml and max are minimum, most likely and maximum value respectively. In the modified PERT distribution, the weight γ can be scaled to control the probability that is assigned to tail values of the distribution, so that a higher weight puts more emphasis on the most likely value and less to the extreme values. For all three-point estimates in the SOM analysis, this weight is set to 4, which is the standard used in unmodified PERT distributions. In symmetrical cases, where the minimum and maximum values are of equal distance from the most likely value, the shape of the PERT distribution is similar to the shape of a normal distribution, and for unsymmetrical cases the shape is often close to a log-normal distribution.

Alternatively, triangular or uniform distribution could be used to present a case where extreme values are more probable. Also, the weight of the PERT distribution could be increased to lower the probability of more extreme values. Sensitivity analyses can later be made deviating from the base case assumptions, by using alternative distribution types that are more representative for different topics than those based on three-point estimates and that can differ among topics, if such distributions based on empirical data are available. The same principal approach applies for the other distributions used in this analysis.

Finally, the above expert-specific distributions only represent the views of individual experts. The aggregated probability distributions that define the view of all experts are defined as follows. A linear pooling method is applied where equal weight is set for each individual expert or national response (see weighting description below). An equal and large number of values is drawn from each expert-specific distribution representing, for example, a certain activity-pressure contribution. Then, these drawn values (from now on drawn values are referred to as picks), are pooled together in a multiset of picks. A discrete probability distribution is applied for each multiset, where a probability is calculated for the value intervals within this multiset. For example, if the multiset has altogether 5000 picks, of which 200 fall within the value range of 1-2%, then the probability of the value range 1-2% is 4%. These discrete distributions define the combined view of all experts, and they take into account the uncertainties expressed by each individual expert. From each of these pooled discrete distributions, a large and equal number (e.g. 10 000) of picks are drawn to form ordered multiset of values, that is used in the simulations to estimate the reductions in pressure and consequential changes in state variables. It should be noted that an unequal number of experts contribute to different aggregated distributions and this affects the shape of pooled distributions and the picks drawn from them.

Responses to the majority of effectiveness of measures surveys and all pressure-state surveys are weighted on an expert level, with each contributing expert receiving equal weight. Survey responses submitted by more than one expert are being followed up to determine individual contributions. The nutrient reductions from agriculture survey is the only effectiveness of measure survey weighted differently. It is weighted on a national basis, with each country's response receiving equal weight. However, as all surveys are expected to only provide information on a single country, this is likely to have no effect on model inputs. Finally, responses to activity-pressure surveys are weighted on a national basis, with each country's response receiving equal weight.

Activity-pressure contributions

Activity-pressure contributions (Step 3) are based on responses to expert surveys (main activities contributing to a pressure) or existing empirical data (for details, see PART II). In the expert surveys, three point estimates (min-%, max-%, most likely-%) are provided by each expert for each basin or geographical assessment unit consisting of multiple basins. In the existing empirical data not all activity-pressure contributions are assessed using surveys (e.g. benthic habitats, non-indigenous species, input of nutrients), but are instead based on other data sources. The data of these sources do not always allow for three-point estimates and instead may be represented by single values or ranges.

Aggregated discrete probability distribution for an activity ($\boldsymbol{j}$) - pressure ($\boldsymbol{i}$) contribution $\boldsymbol{C}$ for a basin ($\boldsymbol{k}$) is defined as

$$\boldsymbol{fc_{i,j,k}(C)}$$

If the activity-pressure contribution is defined for an assessment unit of multiple sub-basins, then the same ordered multiset of picks is used for each basin within an assessment unit, i.e. the values are not redrawn for each sub-basin.

Effectiveness of measure types

The data on the effectiveness of measures (Step 4) come from expert surveys and from a literature review. In the expert surveys, the effectiveness is not assessed individually for actual measures, but for more aggregated measure types which are defined based on existing measures. This is done for several reasons: i) there are too many measures to assess them individually, ii) the available information on existing measures is incomplete and asymmetric, which could jeopardize equal assessment of measure effectiveness among topics, different countries and policy schemes, and iii) the measure type effectiveness can be applied to assess the effectiveness of new measures.

The effectiveness of measures survey consists of two parts. The first part is a grid question where different measure types are located based on their (relative) effectiveness (x-axis: no effect-highest effect) and certainty of their effectiveness (y-axis: uncertain-certain) to reduce a given pressure from a specific activity.

Uncertainty here means the objective uncertainty arising from the level of scientific evidence on measure type effectiveness and also on the variation of measures that belong to one measure type. Such grid-question is asked for each significant activity contributing to a certain pressure, based on the activity-pressure contributions (for an example of a grid question on the effectiveness of measure types targeting one activity-pressure combination see Figure 4 in Part I). Here, the set of measure types related to each pressure $\boldsymbol{N_i}$ includes all possible measure types affecting that pressure, but some of these might not be relevant for all activities. Thus, the measure types included in the question may be subsets of $\boldsymbol{N_i}$. It is assumed that the measure type effectiveness to reduce a given pressure from a certain activity is the same for the whole Baltic Sea. However, one has to remember that activity-pressure contributions and actual pressure levels can vary spatially, which means that the absolute or even the relative effects of certain measure type on the total pressure reduction likely differ between the basins.

The relative effects of different measure types with respect to the most effective measure type (the one the most right on the x-axis) are used to scale the measure type effects. They are defined for each measure type of each grid-question by

$$E_{i,j,n} = \frac{\boldsymbol{x_{i,j,n}}}{\boldsymbol{x_{i,j,max}}}$$

where $\boldsymbol{x_{i,j,n}}$ is the position on the x-axis and max refers to the most effective measure type (most right on the x-axis).

The uncertainty values (position on the y-axis) are used to influence the range of the effectiveness of the measure type (x-axis) in reducing the pressure from an activity. Minimum certainty (uncertain) is assumed to mean that all possible effectiveness levels from no effect to highest effect are possible and the most likely effectiveness is the place on the x-axis where an expert has placed the relative effectiveness value with respect to the other measures. Maximum certainty (certain) is assumed to mean that the effectiveness of measure type always equals the most likely value, and thus there is no range but only a point value of effectiveness. The effectiveness range is symmetrically distributed around the most likely value for all measure types, but if half of the uncertainty (position from top of the y-axis) is higher than the distance of the measure type effectiveness (position on the x-axis) from either end of the x-axis, then the rest of the effectiveness range is allocated to the other end of the x-axis (effectiveness) where there is still room. Assume for example that an expert has estimated that the most likely value of a measure type effectiveness is no effect, and that the certainty related to this effectiveness is minimum (uncertain). In this case the range of effectiveness is from no effect to highest effect with a most likely value of no effect. The relative minimum and maximum effects that different measure types can take with respect to the most effective measure type can be calculated in the same way as in formula (1): $E_{i,j,n,L} = \frac{\boldsymbol{x_{i,j,n,L}}}{\boldsymbol{x_{i,j,max}}}$ and $E_{i,j,n,H} = \frac{\boldsymbol{x_{i,j,n,H}}}{\boldsymbol{x_{i,j,max}}}$, where $\boldsymbol{x_{i,j,n,L}}$ and $\boldsymbol{x_{i,j,n,H}}$ are the lowest and highest end of the effectiveness value range for the measure type respectively.

The second part of the expert survey related to measure effectiveness (see Figure 5 in Part I) asks, in percentages, how much the most effective measure type (most right on the x-axis) reduces the pressure from the activity. The most likely effect of the most effective measure type can be defined as the mean of the given effect range or as a distribution of the values in that range. Using the mean, we can denote this effect by $\bar{R}_{i,j,max}$. The most likely effect of other measure types can be estimated as a product of the expected effect of the most effective measure type and the relative effect of a measure type with respect to the most effective measure type $\bar{R}_{i,j,n} = E_{i,j,n} \times \bar{R}_{i,j,max}$. The minimum and maximum effects for different measure types are calculated in a similar fashion but using $E_{i,j,n,L}$ and $E_{i,j,n,H}$ respectively. These effect ranges (most likely, minimum and maximum effects) define three-point estimates for each survey response. Again, the probability distributions for measure type effects as percent reduction in pressures from activities are aggregated from the PERT distributions defined for three-point estimates of individual experts. The probability distribution of a %-pressure reduction effect $\boldsymbol{R}$ of a measure type $\boldsymbol{n}$ on a pressure i from an activity j is $\boldsymbol{r}_{i,j,n}(\boldsymbol{R})$.

Calculating the total effect of measures on a pressure

The total pressure reduction effect $\boldsymbol{T}_{k,i}$ of measures on a pressure $\boldsymbol{i}$ in a basin $\boldsymbol{k}$ is calculated as a sum of all effects of measures affecting pressure $\boldsymbol{i}$ in basin $\boldsymbol{k}$ multiplied by their respective activity-pressure contributions

$$\boldsymbol{T}_{k,i} = \sum_{j \in A_{k,i}} \boldsymbol{C}_{i,j,k} \sum_{n \in N_i} \sum_{m=1}^{M_{k,j,i,n}} \boldsymbol{R}_{i,j,n}$$

where $\boldsymbol{C}_{i,j,k}$ is the contribution of an activity j on pressure i in basin k, $\boldsymbol{A}_{k,i}$ is the set of significant activities causing pressure i in basin k, $\boldsymbol{N}_i$ is the set of all measure types linked to pressure i, and $M_{k,j,i,n}$ denotes the number of measures of measure type n affecting pressure i from activity j in basin k, and $\boldsymbol{R}_{i,j,n}$ is the pressure reduction effect of the given measure type n on pressure i from activity j. Measure effects of individual measures are thus defined by the effectiveness of the measure type that they belong to.

The set of total pressure reductions $\boldsymbol{T}_{k,i}$ that is used to define the distribution of the total pressure reduction (in %) for pressure i in basin k is calculated by using large number of values $\boldsymbol{C}_{i,j,k}$ and $\boldsymbol{R}_{i,j,n}$ drawn from discrete probability distributions $\boldsymbol{fc}_{i,j,k}(\boldsymbol{C})$ and $\boldsymbol{r}_{i,j,n}(\boldsymbol{R})$ [12]as described in section Pooling of expert judgements. The measure effect values $\boldsymbol{R}_{i,j,n}$ are corrected to include the joint effects defined in section Joint impacts of measure types.

If a measure affects only certain part of some basin (for example national measures) then the effect is multiplied by the area of that part of the basin divided by the area of the whole. A probability distribution is defined for each pressure reduction in % based on the N=100 000 calculated pressure reduction effects. These distributions take into account the uncertainty in the activity-pressure contributions, as well as in the effectiveness of the measure types. These distributions allow for calculating the expected (most likely) pressure reductions, constructing percentile intervals for pressure reductions, and calculating the probability to reach a specific pressure reduction.

Pressure-state linkages

The data for the pressure-state linkages (Step 6) come from expert surveys. The first survey question on the pressure-state linkage asks the experts to identify the most significant pressures to the state variable (such as the abundance of some species) (see Figure 6 in Part I). These are asked separately for each assessed area/population. The pressures are weighted based on their proportion of the total significance of pressures:

$$W_{i,k,s} = \frac{\boldsymbol{y}_{i,k,s}}{\sum y_{k,s}}$$

where $\boldsymbol{y}_{i,k,s}$ is the sum of significance scores over the experts (0-5, 0 being “not very significant” and 5 being “extremely significant” in Figure 6) of the given pressure $\boldsymbol{i}$ in spatial assessment unit $\boldsymbol{k}$ consisting of one or multiple basins for state variable $\boldsymbol{s}$ and $\sum y_{k,s}$ is the total of all summed significance scores of all significant pressures for state $\boldsymbol{s}$ in assessment area $\boldsymbol{k}$.

The second survey question about pressure-state linkages (Figure 7 in Part I) asks how much all the pressures chosen in the first question need to be reduced in order to reach or maintain good state for the state variable. If the good state or/and current state can be quantified, these values are used when phrasing the questions for pressure-state linkage. When there is no agreed GES threshold, this question asks about pressure

[12] The values of $\boldsymbol{R}_{i,j,n}$ can also be drawn independently for each measure belonging to a certain measure type.

reductions required in order to achieve specific (%) improvements in the state variable or a noticeable improvement in the state variable.

These questions about required pressure reduction are again asked as a value range (most likely, minimum, maximum), where three-point estimates are provided by each expert. From these values, a cumulative distribution function can be defined that represents the probability of reaching a good state for different % reductions in total pressure, the probability of a specific improvement in state, or a probability to reach noticeable improvement in state. Again, a pooling method is used to define aggregate cumulative distribution from expert-specific distributions. These cumulative distributions are denoted by $\boldsymbol{FS_{k,s}(TPR)}$, where $\boldsymbol{TPR}$ is the reduction in total pressure. In principal, the reduction in total pressure means that all significant pressures are reduced by the same % amount. However, in reality it is very unlikely that all pressures are reduced by the same proportion, and thus the reduction in total pressure $\widetilde{\boldsymbol{TPR}}_{\boldsymbol{k,s}}$ can be approximated using the pressure weights based on the significance scores from Figure 6.

Reduction in total pressure ($\boldsymbol{TPR}$) for the spatial assessment unit $\boldsymbol{k}$ and state component $\boldsymbol{s}$ is:

$$\widetilde{\boldsymbol{TPR}}_{\boldsymbol{k,s}} = \sum_{\boldsymbol{i \in I_s}} W_{i,k,s} \boldsymbol{T_{k,i}}$$

where $\boldsymbol{i} \in \boldsymbol{I_s}$ is the set of significant pressures for state $\boldsymbol{s}$ (resulting from the implementation of measures). If the spatial assessment unit of given state variable consists of multiple basins, then the pressure reduction is calculated by weighing the per basin pressure reductions by the proportion of basin area of the whole assessment unit area and then summing up these weighed reductions. By plugging the approximated reduction in total pressure into the function of reaching a good state for different % reductions, one is able to estimate the probability of reaching a good state for state variable $\boldsymbol{s}$. If an expected value of the total pressure reduction is applied to study how reductions in pressures increase the probability to reach good state, then the cumulative distribution function $\boldsymbol{FS_{k,s}}$ can be used to define the expected probability to reach good environmental state. Whereas, if the total pressure reduction is defined as a distribution, then the probability distribution of reaching good state with a specific probability can be assessed, and from that it is possible to estimate what is the likelihood that the probability of reaching a good state is at least some specific percent.

Comparison of BAU and GES and sufficiency of measures

The previous sections have outlined how to determine the expected pressure reduction distributions with existing measures, allowing for calculation of expected pressure reductions, confidence intervals and the probability to achieve specific percent (%) reduction in pressures. If we know a pressure target or the threshold associated with good state and the current pressure level, we can estimate the total pressure reduction required to reach a good state. Thus, for pressures that have a GES threshold/target, we can assess whether the expected pressure reduction is sufficient to reach GES for that pressure (i.e. if the expected pressure reduction from the existing measures is as large as the required pressure reduction), or estimate the probability of reaching a the pressure target with the existing measures (=probability that given pressure reduction target is achieved from the distribution of the total pressure reduction with existing measures).

The cumulative distribution function of the total pressure reduction required to meet the good state $\boldsymbol{FS_{k,s}}$ is used to represent the probability of reaching a good state for different % reductions in total pressure affecting a state variable. If an expected value of total pressure reduction is applied to study how reductions in pressures increase the probability to reach good state, then cumulative distribution function $\boldsymbol{FS_{k,s}}$ can be used to define the expected probability to reach good environmental state. If the total pressure reduction is defined as a distribution, then the likelihood that the probability of reaching a good state is at least X% can be estimated.

When interpreting the results, one has to take into account the assumptions and generalizations that were made when defining the input distributions of activity-pressure, measure type effects and probability to reach good state, and the fact that these are based mainly on expert elicitations rather than empirical data.

Adjustments for the pressure *Input of the relevant top litter items present on the beach*

Modifications to the general approach were required in order to analyze input of litter using a by-item-approach, as requested at SOM 2-2019. Firstly, in addition to the standard activity-pressure survey, a supplementary survey linking the top beach litter items to the primary activities contributing to beach litter was also circulated to EN Litter. The supplementary survey ranks contributions on a scale from 0 to 4 linked to specific % contribution ranges (0 = <5%, 1 = 5-20%, 2 = 20-40%, 3 = 40-60%, 4 = >60%). These values can then be used to modify the results of the standard activity-pressure survey to reflect by-item contributions.

Secondly, in the evaluation of effectiveness of measures, each measure type is linked to the litter items controlled by that measure type. Measure effectiveness is then assessed as the average effectiveness across all the listed litter types. One complication to this approach is the presence of the litter item category "Plastic and polystyrene pieces" which includes otherwise unidentified pieces which may or may not belong to another top litter category if properly identified. To overcome this, two further adjustments are made. Measure types controlling the input of all top litter items were surveyed using a separate grid question (Part I, Figure 4) than those controlling a subset of the top litter items. In the grid question seeking effectiveness values for measure types controlling a subset of the top litter items, "Plastic and polystyrene pieces" was not included as a litter item due to the uncertain proportion of the category actually controlled by the measure type. Instead, effectiveness will be calculated by determining the proportion of all plastic litter items recovered during beach surveys that are controlled by the measure type and applying the measure effectiveness only to that portion.

Adjustments for the pressure *Input of nutrients*

Substantial amount of information for the input of nutrients comes from ACTION work package 4, which provides an overview of the division of activities and pressures related to eutrophication (i.e. nutrient inputs), creating an overview of source apportionment and identifying activity-pressure contributions (Step 3). This aspect is developed based on the national data reported to the HELCOM Pollution Load Compilation (PLC).

For the effectiveness of measures (Step 4), information on load reductions due to full implementation of existing measures is required. The information on effectiveness of measures is provided per activity: waste water treatment (reductions achieved by implementing the HELCOM Recommendation 28E/5 on municipal waste water treatment), atmospheric nitrogen emissions (based on EMEP data and predictions), agriculture (expert survey on the nutrient runoff from agriculture guided by HELCOM Agri group), and scattered dwellings (joint survey with the PLC-7 project). Inclusion of the estimated reduction from scattered dwellings is uncertain due to the development timeline for PLC-7.

Thus, only nutrient runoff from agriculture is based on expert elicitation. The expert survey follows the general format of the effectiveness of measures survey for the other topics in the SOM analysis, but there were also significant adjustments.

First, the survey allowed the respondents to provide assessments of the effects of measures either based on model estimates, expert evaluation, or both. The model-based estimates could be provided as the total reduction in nutrient runoff or by measure (based on HELCOM palette of measures), in tons or percent. The expert-based estimates could be provided as a total reduction in nutrient runoff (tons or percent) or the relative effectiveness of measures, as for other topics in the SOM analysis. Secondly, nationally consolidated responses were preferred.

The survey asked separately for effectiveness of measures for the input of phosphorus and nitrogen.

References

Ahlvik, L., Ekholm, P., Hyytiäinen, K. & Pitkänen, H. (2014) An economiceecological model to evaluate impacts of nutrient abatement in the Baltic Sea. Environmental Modelling & Software 55: 164-175.

ARCADIS (2012) Economic assessment of policy measures for the implementation of the MSFD. Final report and Excel database of a study for the EC DG ENV (Project No 11601). Available at http://ec.europa.eu/environment/enveco/water/pdf/report.pdf

EC project "BLUE2" on a database of policy measures for protection of inland and marine waters in Europe. Available at http://ec.europa.eu/environment/blue2_en.htm

Hasler, B., Smart, J.C.R., Fonnesbech-Wulff, A., Andersen, H.E., et al. (2014) Hydro-economic modelling of cost-effective transboundary water quality management in the Baltic Sea. Water Resources and Economics 5: 1–23

HELCOM 2018a. State of the Baltic Sea – Second HELCOM holistic assessment 2011-2016. Available at http://stateofthebalticsea.helcom.fi

HELCOM 2018b. SPICE project deliverable 3.3. Development of a regional "business-as-usual" scenario (BAU) to be used as a baseline in the integrated assessment of the marine environment.

HELCOM 2016. TAPAS Theme 1 Deliverable: Baltic Sea pressure and impact indices (BSPI/BSII) Available at http://www.helcom.fi/Documents/HELCOM%20at%20work/Projects/Completed%20projects/TAPAS/TAPAS%20Theme%201%20Deliverable.pdf (see Annex 6 for the TAPAS linkage framework between activities and pressures)

Hyytiäinen, K., Czajkowski, M., Ahlvik, L. & Zandersen, M. (2017) Dynamic cost-and-effect model under changing climate and socio-economic drivers. BONUS BALTICAPP, Deliverable 3.6, Available at https://blogs.helsinki.fi/balticapp/publications/

Knights, A. M., Piet, G. J., Jongbloed, R. H., Tamis, J. E., White, L., Akoglu, E., ... & Leppanen, J. M. 2015. An exposure-effect approach for evaluating ecosystem-wide risks from human activities. ICES Journal of Marine Science, 72(3), 1105-1115.

Kontogianni, A., Tourkolias, C., Damigos, D., Skourtos, M., & Zanou, B. (2015). Modeling expert judgment to assess cost-effectiveness of EU Marine Strategy Framework Directive programs of measures. *Marine Policy*, *62*, 203-212.

Oinonen, S., Hyytiäinen, K., Ahlvik, L., Laamanen, M., Lehtoranta, V., Salojärvi, J., & Virtanen, J. (2016). Cost-effective marine protection-a pragmatic approach. *PloS one*, *11*(1), e0147085.

Owenius, S. & van der Nat, D. (2011) Measures for water protection and nutrient reduction. Baltic COMPASS report.

Wulff, F., Humborg, C., Andersen, H. E., Blicher-Mathiesen, G., Czajkowski, M., Elofsson, K., Fonnesbech-Wulff, A., Hasler, B., Hong, B., Jansons, V., Mörth, C.-M., Smart, J. C. R., Smedberg, E., Stålnacke, P., Swaney, D. P., Thodsen, H., Was, A., and Żylicz, T., 2014. Reduction of Baltic Sea Nutrient Inputs and Allocation of Abatement Costs Within the Baltic Sea Catchment. AMBIO: A Journal of the Human Environment, 43(1):11-25.