Protocol Benthic Indicator Species Index (BISI):

Protocol BISI for generic application (BISI v2)

– Wijnhoven (2018)

This protocol describes the Benthic Indicator Species Index (BISI). The current version is the protocol for generic application: BISI v2, and comes with an Assessment tool:

- *Wijnhoven (2018). Assessment tool: 'Benthic Indicator Species Index (BISI)': Application of BISI v2 in the Dutch North Sea with consolidation of earlier identified references. v181218.*

The BISI was initially specifically developed as the Dutch national benthos indicator for evaluation of the quality status of the North Sea and the sea floor integrity in particular, focussing on areas with special ecological values (ASEVs). The BISI v1 (dated 26-09-17) is still in use for application and reporting for the Marine Strategy Framework Directive (MSFD) and delivers to the Habitat Directive Article 17 reporting and Natura 2000 fishery measures evaluations in the Netherlands.

BISI v2 is an updated version of the BISI that specifically takes:

- *ecotopes as being the basis for the derivation of the index,*
- *includes (standardized) rules for indicator species selection,*
- *the incorrect use of ^{10}log in former calculations leading to skewed results has been corrected to natural logarithms (ln),*
- *has some additions to the derivation of reference values,*
- *and involves corrections on how to calculate pooled standard deviations associated with BISI-scores,*

to allow BISI development and application in basically any area or region with sufficient data availability.

BISI v1 results and recent historic trends indicating directions of changes are still valid (significance levels at T0 might however deviate). BISI scores from v1 and v2 are not directly comparable. It is suggested to use the updated BISI (v2) in the future. Former results for the T0 of the Dutch North Sea (and recent historic trends) can relatively easy be recalculated into the scores according to BISI v2 maintaining the earlier defined indices (indicator species lists and reference levels). An update of the indices is not necesary for evaluation of quality status and developments of individual areas; It might only be desirable for a better comparison of the differences in relative quality status between areas, and comparison with areas outside the Dutch North Sea.

Background information on specifically the development of BISI v1 is described in:

- *Wijnhoven, S. & Bos, O.G. (2017). Benthic Indicator Species Index (BISI): Development process and description of the National Benthos Indicator North Sea including a protocol for application. Ecoauthor Report Series 2017 - 02, Heinkenszand, the Netherlands.*

Results of application of the BISI are presented in:

- *Wijnhoven, S. (2018). T0 evaluation of the quality status of the Dutch Exclusive Economic Zone based on the Benthic Indicator Species Index (BISI). Quality status and – developments of benthic habitats and MSFD areas of the Dutch North Sea in and prior to 2015. Report Ecoauthor & Wageningen Marine Research. Ecoauthor Report Series 2018 - 01, Heinkenszand, the Netherlands.*

which includes the factsheet: 'Benthic Indicator Species Index (BISI)': D6C3/5, as prepared for the Dutch Action plan Marine Strategy (part 1) 2018-2020:

- *Min IenW, Min LNV (2018). Mariene Strategie (deel 1). Huidige milieutoestand, goede milieutoestand, milieudoelen en indicatoren 2018-2024, Hoofddocument. Een uitgave van Ministerie van Infrastructuur en Waterstaat en Ministerie van Landbouw, Natuur en Voedselkwaliteit, februari 2018, wvl0118tp312 (in Dutch).*

This protocol should be cited as:

Wijnhoven, S. (2018). Protocol Benthic Indicator Species Index (BISI): Protocol BISI for generic application (BISI v2), v181218. Ecoauthor Report Series 2018 - 04, Heinkenszand, the Netherlands.

The protocol is based on the structure of the ICES WGBIODIV template for Indicator factsheets (version 0.1) which is developed by O. G. Bos.

Index

Benthic Indicator Species Index (BISI)

$$BISI = exp((1/S)*(\Sigma(IV_i)ln((O_i/R_i)+1)-S))$$

BISI = Benthic Indicator Species Index; S = Number of indicator species included; IV_i = Species specific Indicator Value calculated as species specific standard indicator value iv_i (value between 0-1) divided by average indicator value iv_{avg}; O_i = Observed occurrence (ratio of samples with the indicator species present) or observed numbers (average densities); R_i = Reference occurrence (set ratio of samples with indicator species present under reference condition) or observed numbers (set average densities under reference condition). ('exp' is similar to putting e to the power of the formula as indicated, which equals the inverse natural logarithm, as a back-transformation of the natural logarithm (ln) taken from the occurrence-to-reference ratios).

1 Authors

S. Wijnhoven (Ecoauthor) (sander.wijnhoven@ecoauthor.net)

2 Contributors

Contributors to BISI v1: O.G. Bos (Wageningen Marine Research; editor); the following experts commented on early versions or ideas: A.-M. Svoboda, A. Adams (Dutch Ministry of Agriculture, Nature and Food Quality; Min LNV, formerly Min EZ), D. van Schaardenburg, P. Heslenfeld, M. Platteeuw, S. Rotteveel, A. Stolk, S. Stuijfzand, W. van Loon (Ministry of Infrastructure and Water Management; Min IenW, formerly Min IenM), J. Craeymeersch, G.J. Piet, T. van Kooten (Wageningen Marine Research), Beauchard, O., Escaravage, V. (Netherlands Institute for Sea Research), E. Verduin (Eurofins), A. van Strien, M. Poot (Statistics Netherlands).

Contributors to BISI v2: P.H. van Avesaath (AMAECON) assisted in developing the methodology for BISI v2 and gave valuable comments on the concept version of the updated protocol and methodology. Early ideas for BISI v2 were presented and discussed within the OSPAR Benthic Habitats Expert Group (OBHEG); e.g. workshop of 12th – 14th of June 2018 (IEO Santander, Spain).

3 Key message

The Benthic Indicator Species Index (BISI) uses benthos data to evaluate habitat quality, sea floor integrity and ecological functioning. BISI compares temporal patterns (densities or absence/presence) of combinations of specific indicator species with predefined reference levels of these species that represent a good status. The selection of indicator species and derivation of reference levels is part of the methodology and takes place at the level of (high aggregation) ecotopes (e.g. European Nature Information System (EUNIS) levels; EMODnet, 2016). BISI can be used at different spatial scales by combining ecotope-specific results on basis of surface ratios; e.g. Exclusive Economic Zones (EEZ), Natura 2000 areas, etc.). BISI can be used to evaluate current quality status, as well as trends in quality status, e.g. to determine the effect of management measures in protected areas.

V1 - The BISI v1 has been applied to Dutch North Sea benthic data. Reference values per indicator species were estimated as maximum observed abundances and/or distributions of indicator species and have been set using available historic boxcore and dredge data from the period 1984-2014. T0 data (data of the Dutch Marine Strategy Framework Directive (MSFD) monitoring of 2015) were compared to the reference levels, and quality status developments on basis of BISI were compared in time (identification of trends). Although the BISI is typically meant to evaluate quality changes (based on benthic assemblage data) in time, e.g. relative to a T0, the T0 evaluation on its self also gives insight in spatial differences in the quality status and relative importance of potential pressures and affected functions.

In BISI v2 indicator species are selected to a standardized derivation scheme per ecotope on basis of being characteristic, or being indicative for at least one of the dominant disturbances, or being characterized by a combination of biological traits. Selected species should have opportunities to be present or return as well. Reference values based on current and maximum observed abundances and natural variation are derived according to a decision scheme taking recent historic data availability and current monitoring efforts, possible reference or alternative areas, and suitability monitoring techniques or possible alternative techniques into account.

V1 - Compared to the reference that is indicative for the quality status at low pressure levels, a significantly reduced quality status is found for 2015 based on BISI v1 for all investigated spatial areas (i.e. areas with special ecological values (ASEVs), EUNIS ecotopes level 4 and Habitat Directive (HD) habitat types). Dominant pressures defining the current quality status are related to seafloor disturbances, and are at least partly the result of seafloor disturbing fisheries. More specific: (1) Fisheries impacts are small on the Cleaver Bank (Klaverbank), (2) first quality improvements after the lowest point has been reached, can be observed in the coastal zone, (3) recovery in between disturbances is observed on the Dogger Bank (Doggersbank), and (4) at present the areas that show a decline in quality status largely consist of offshore circalittoral muddy ecotopes (EUNIS level 4) of the Central Oystergrounds (Centrale Oestergronden) and the Frisian Front (Friese Front).

Additions to the methodology in BISI v2 will amongst others be applied and tested in international case studies (e.g. international Dogger Bank within the frame of OSPAR) and in development of the BISI for evaluation of HD habitat types at national scale largely situated in the 'larger' estuarine and marine waters outside the North Sea in the Netherlands.

4 Definitions

Area specific index – Index designed/compiled for a specific area to be evaluated. Here used as the area specific BISI; indicating that each area of evaluation has its BISI, where indicator species composition with their reference occurences depends on the constitution of the area. With BISI v2 the compilation of the area specific reference is more standardized, build on ecotope presence (surface area based) with each ecotope its commonly used standard reference.

ASEVs - Areas with special ecological values. Designated areas in Dutch North Sea that include all Natura 2000 areas (and therefore the areas of the Habitat Directive), often with specific management measures; focal areas of the Dutch North Sea benthos monitoring programme.

BISI - Benthic Indicator Species Index. Benthos-based indicator for quality status assessment and evaluation of quality developments of benthic habitats and sea floor integrity in particular.

BISI-score - Score (also indicated as BISI-value) that reflects the benthic habitat quality status. In practice the score deviates between 0.372 and 37.156. A value around 0.4 means very poor quality; a value around 0.736 equals a high habitat quality similar to the internal reference (potentially within reach on the mid-long term for areas under pressure when dominant pressures are significantly lowered to a minimal level).

BISI v1 - Original version of the Benthic Indicator Species Index, specifically developed for evaluation of areas and habitats of the Dutch North Sea. V1 is especially deviating from BISI v2 in the way references were constructed area (ASEV) specific. In v2 the incorrect transformation and back-transformation leading to skewed results[1] in analyses of quality developments (trends) is corrected.

[1] Transformation in v1 was according to ^{10}log (instead of a natural logarithm in v2) while back-transformation consisted of the inverse natural logarithm, which results

BISI v2 - Updated Benthic Indicator Species Index, especially deviating from BISI v1 in the way that references are constructed ecotope specific. A reference for an area to be evaluated is built surface area based from the standard references of the present ecotopes. . In v2 the incorrect transformation and back-transformation leading to skewed results in analyses of quality developments (trends) is corrected. Other deviations are a further standardization of indicator species selection, reference level derivation and corrections in the methodology of calculating pooled standard deviations associated with the BISI.

Closed areas - A term used in management of the Dutch North Sea, indicating that there are certain fisheries restriction in the areas of concern. In general those areas are not complete closed for all human activities or potential human disturbances; in many cases only certain sea floor disturbing fisheries (i.e. beam trawling) is banned.

Ecological disturbance – Disturbance of habitat and benthic communities in particular due to sensitivity of species for toxic substances, pollutants, elevated nutrient levels, hypoxia or temperature elevations. Indicated pressures often have an impact on the same species in about similar amounts and often co-occur to some extent.

Ecotope - Relatively homogeneous, spatially explicit landscape unit described by abiotic conditions reflecting a certain biotic constitution and development. Although BISI can potentially function based on any ecotope and/or classification system, preferably widly accepted and applicated ecotope classifications like EUNIS and ZES.1 are suggested.

EUNIS - European Nature Information System (www.eunis.eea.europa.eu/index.jsp). Here specifically the EUNIS habitat types classification system is adapted, where the EUNIS classification at level 4 is suggested as a European standard of ecotopes to classify (parts of) European regional seas for application in BISI.

HD habitat types – Characteristic habitat types that are protected under the Habitats Directive (HD) in those areas specifically designated as Habitats Directive areas (part of the EU Natura 2000 network).

---

in the same direction of quality differences in v1 and v2, however the steepness of trends is different.

Hit rate - Alternate expression for 'occurrence' or spatial distribution of an indicator species in an area of evaluation.

Indicator Value – Indicated with IV_i, the indicator value is the species specific indicator value iv_i with per definition a value between 0-1, divided by the average indicator value (iv_{avg}) of all indicator species in the specific evaluation ($IV_i = iv_i / iv_{avg}$).

Internal reference or Realistic reference – A reference occurrence, abundance or biomass for an indicator species (or a combined BISI-score based on a set of indicator species) reflecting a quality status that can be reached on the mid-long term when dominant pressures are significantly reduced to a level of minimal impact. The internal references used in BISI are derivated from maximum potential observations taking natural fluctuations into account. With BISI v1 the reference BISI-score per definition equals 1; with BISI v2 this value is equal to 0,736.

MSFD - Marine Strategy Framework Directive.

Occurrence - Presence/absence of species at a monitoring location (in a standardized benthos samples). In BISI used to reflect the spatial distribution or presence of an indicator species.

Open areas - A term used in he monitoring and evaluation of Dutch North Sea areas. Monitoring sites are selected in open areas (areas with no specific restrictions) that are comparable in habitat or ecotope constitution or in constitution of the expected benthic communities, with monitoring sites situated in 'closed areas'. Herewith evaluation of efficiency of management measures (closing areas for certain sea floor disturbing fisheries) is according to a BACI design.

OSPAR - OSPAR is the mechanism by which 15 Governments and the EU cooperate to protect the marine environment of the North-East Atlantic, named after the original Oslo and Paris Conventions.

Pristine - Undisturbed condition reflecting the 'natural' situation of before substantial anthropogenic influence.

Reference - A standard value to compare with. Here a standard value of occurrence, abundance or biomass for an indicator species (or the combination of species in a BISI-score) to compare observation data with. Here a pristine and internal reference is distinguished.

Smart species – 'Smart species' is a term introduced in Wijnhoven et al. (2013) for potential indicator species (sensitive for specific pressures or representative for certain ecological functions of the habitat of concern) for which differences in spatial occurrence can be detected with a realistic monitoring effort using prevailable monitoring techniques. With BISI v2 this realistic effort is set to 40 samples as a maximum number for individual indicator species.

Species specific indicator value – Indicated with iv_i, the species specific indicator value is a validation of a characteristic of a potential indicator species per definition with a value between 0-1 with 0 = no indicator value at all (not included in a the representative specific evaluation) and 1 = a very good indicator or the specific evaluation.

Typical species – Term from the Habitats Directive; typical species are selected by countries on basis of their presence eing considered indicative for either or both a good biotic or abiotic structure and function of the habitat type of concern (e.g. Shaw & Wind, 1997).

ZES.1 - 'Zoute wateren Ecotopenstelsel' (Bouma et al., 2005: Marine environments ecotope classification system). Adapted as the basis for ecotope classification in BISI for 'transitional' (marine and estuarine waters on the transition from land to sea) larger waterbodies in the Netherlands.

%BISI - The BISI converted to a lineair (percentage) scale from 0% to 100% equalling a BISI-score from 0.372 to 37.156, with 50% (which is equal to a BISI of 0,736) representing a quality status equal to the internal reference.

5 Indicator metadata

5.1 Ecosystem component

Marine benthos: Endofauna and sessile epifauna species of which populations can be monitored with techniques giving a representative inventory like corers, grabs, dredges, trawl tows and video recordings.

5.2 MSFD Descriptor

D6 Seafloor integrity (C3 and C5 according to the Dutch Marine Strategy: Respectively 'the spatial extent and quality of habitats potentially impacted by changes in biotic and abiotic structures and functions' and 'changes in size, condition and distribution of populations of benthic macrofauna species').

5.3 Status

BISI is an indicator that has been developed in the Netherlands in 2016/2017 for the Ministry of Economic Affairs. It has been used to evaluate the quality status and the recent historic quality developments in and prior to 2015 for the MSFD (Wijnhoven, 2018), and is part of the Dutch Marine Strategy (Min IenW & Min LNV, 2018). The indicator has no status within OSPAR yet, but within OBHEG it has been agreed that the BISI will be part of the test - and comparison activities (for which BISI v2 is suggested), towards common OSPAR indicators and the Benthic Habitat indicator BH1 in particular, in which the Netherlands and Spain are co-leading.

5.4 Indicator type

State indicator.

6 Indicator description

6.1 Indicator goals/objectives

The Benthic Indicator Species Index (BISI) aims to evaluate the quality status, quality status developments and quality status differences for defined spatial areas/ecotopes.

In addition, the BISI aims to evaluate the importance of different potential pressures and the effect of the potentially decreased quality on ecosystem functions (See Table 1 in section 1.6.7 Methodology and data analyses). Overall quality indices for each of the identified areas to be evaluated are combined with specific quality indices that are based on weighted subsets of indicator species with specific characteristics and traits. The BISI intends to be a well-documented, reproducible and efficient assessment method that includes a protocol and assessment tool.

BISI v1 is specially geared for evaluation purposes of the Dutch North Sea Continental shelf. With the introduction of BISI v2, the approach has been generalized, making it applicable in areas across the world so that it will be possible for every benthic expert to construct indices (BISIs) for their own area of interest, if (recent historic or reference area and current) data availability is sufficient.

6.2 Theoretical background

Macrobenthos, macrobenthic communities and benthos indicator species in particular are expected to reflect the quality status of the (local) environment, and especially that of the benthic habitats (seafloor integrity) (e.g. Ysebaert et al., 2002; Reiss et al., 2015; Elliott et al., 2018). Most benthic species have a strong relation with the constitution of the seafloor, are typically related to the prevailing abiotic condition (that can basically be described with a limited number of abiotic parameters) and the quality status of that habitat determined by the (former) presence (or absence) of pressures of different kinds and/or severity (e.g. Hiscock et al., 2004; Dutertre et al., 2013). Additionally the biotic conditions are of importance, where the presence (or absence) of certain species or communities can provide improved habitat quality, amongst others via influencing local abiotic conditions, possibly accelerating the settlement and/or natural succession of benthic communities. Natural development of good quality benthic habitats includes habitat rejuvenation and the presence of a variety of different habitat elements in various developmental stages. This indicates that the presence of typical habitat related species[2] in their potential abundances reflect the quality status suggesting that there are opportunities for those species to arrive/return. Hampered opportunities for those typical habitat related species to return are more an indication of former presence of large disturbance levels at larger scale resulting in the lack of source populations and/or poor connectivity. Species as indicated are potential indicator species, especially when they are sensitive for specific pressures. To be suitable as an indicator species, also natural fluctuations in occurrences should be relatively small compared to fluctuations under the influence of a specific pressure, and monitoring of species natural occurrence should be possible using conventional observation techniques and relatively low monitoring efforts.

With BISI the combined occurrence of indicator species of the general quality status related to specific habitats are evaluated. Additionally, further differentiation in specific evaluations solely based on indicator species specifically sensitive and/or indicative for specific pressures and/or functions indicate the status of specific causes or effects of the observed quality status. Comparing the quality status based on BISI for different moments allows evaluation of quality status developments and identification of possible thriving factors and or resulting effects on ecosystem level.

6.3 Description of the indicator

6.3.1 Benthic Indicator Species Index (BISI):

The occurrence (spatial occurrence as presence/absence data, abundances or biomass) of an (area - and/or habitat specific) selection of indicator species at a certain moment of evaluation is compared with a defined reference value for that selection of indicator species. The methodology consists of the calculation of the weighted (species - and evaluation specific indicator values) geometric mean (i.e. ln-transformed) of observation-to-reference ratios. Testing

[2] Although Typical Species of the Habitats Directive are (also) selected as their presence is considered indicative for either or both a good biotic or abiotic structure and function of a habitat (Shaw & Wind, 1997), the list of potential indicator species considered here is more extensive. With BISI in the end not only species presence is evaluated, but optionally also differences in other occurrence data (e.g. spatial occurrence, densities, biomass) that allow quantification and statistical testing.

occurs against a (fixed) reference compiled in a standardized way for ecotopes, and for other areas, like areas with a certain protection status and/or importance, or Habitat Directive habitat types. In BISI v2 references are compiled surface-area based from the individual (most important) ecotope references in a standardized way. The BISI evaluation tool is initially developed (as BISI v1) for evaluation of the Dutch North Sea including all benthos-based specific evaluations needed for MSFD reporting, effectivity of management regulations evaluations, and providing background information on causes of and functions affected by observed changes in the quality status, amongst others of importance towards Natura 2000 and Habitat Directive evaluations.

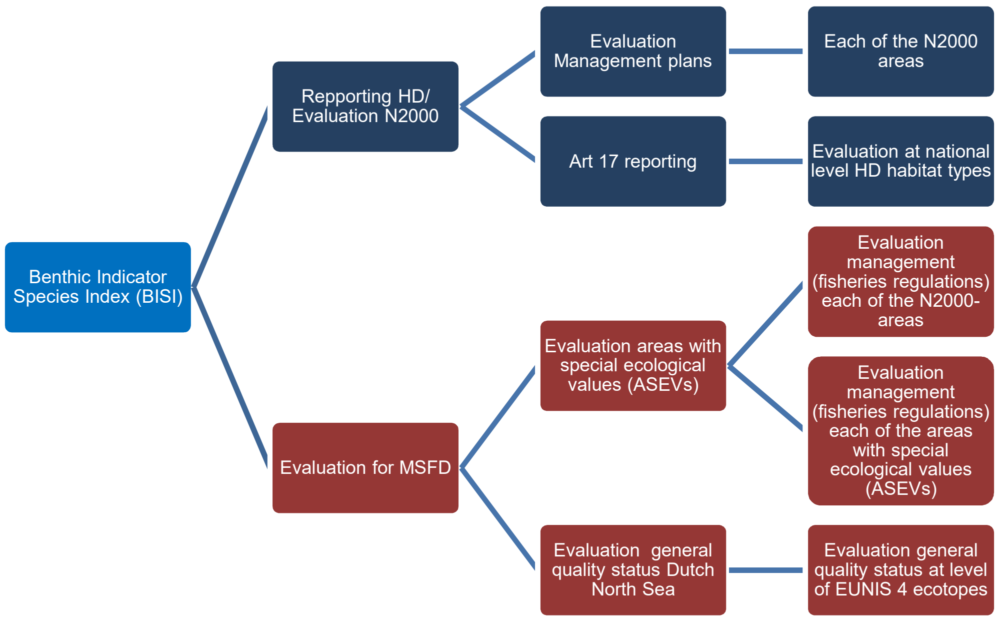


Figure 1. Schematic overview of the different Dutch evaluations for which the Benthic Indicator Species Index (BISI) is developed as BISI v1.

V1 - The BISI methodology (including reference levels, essential monitoring - and data type characteristics) is (currently) prepared according to BISI v1 for the areas with special ecological values (of which several are also Natura 2000 areas): Dogger Bank, Cleaver Bank, Central Oystergrounds, Frisian Front, Brown Bank (Bruine Bank), North Sea Coastal Zone (Noordzeekustzone), Front Delta (Voordelta), Plain of the Raan (Vlakte van de Raan), the six EUNIS level 4 ecotopes covering the Dutch North Sea; 'Offshore circalittoral coarse sediment', 'Offshore circalittoral sand', 'Offshore circalittoral mud', 'Circalittoral coarse sediment', 'Circalittoral sand', 'Circalittoral mud', and the areas indicated as being part of Habitat Directive habitat types H1170 'Reefs', H1110b 'Submersed sandbanks in the Coastal zone', H1110c 'Offshore submersed sandbanks'. A special case making use of the same reference levels as for the respective areas with special ecological values, are the evaluations of efficiency of management regulations (fisheries restrictions) for each of the areas. Specific samples (locations, methodology and data type) are selected for these evaluations in order to meet an optimized statistical design. Dependent of the type of reporting (Fig. 1) and the timing in the monitoring process (e.g. T0, T1, T2, etc.) evaluation results of above-mentioned areas, ecotopes and/or habitats (which can be seen as separate modules) will be combined, and consequently the type of significance testing can differ.

6.3.2 Evaluation relative to compiled reference levels

For each area to be evaluated, an area specific (internal) BISI reference is constructed. With BISI v2, internal references are specific for ecotopes and can potentially be used for different areas consisting of the same ecotope. Exceptions are potential indicator species that in certain regions have no opportunities of colonizing the area of evaluation in a natural way, whereas they are present in other regions. For evaluations of other types of areas and habitats, internal references are compiled from the references of the present (most important) ecotopes, surface area based.

V1 – Taking ecotopes as the basis to construct a BISI reference is different from the methodology as used for the Dutch North Sea (BISI v1). There specific references were also constructed for areas with special ecological value (based on the recent historic data from these areas). References were only compiled from existing references for larger HD habitat type areas. Internal references for areas of the Dutch North Sea are presented in Appendix 1 of Wijnhoven & Bos (2017).

Table 1 gives a short explanation based on a part of the reference of in this case the area of special ecological value Cleaver Bank, which in total consists of 30 indicator species. So also the abundance (either spatial occurrence or abundance) data for 30 species will be used for evaluation of the quality status of the ASEV Cleaver Bank. Monitoring programmes should match the purpose of the evaluations (give a representative view of the quality status of areas or aspects of concern); and selection of the suitable monitoring data might be necesary (see chapter 1.7 Determination of GES and boundaries).

In case of the Cleaver Bank it is for instance known that the sampling has not been entirely random, but focussed on the (expected presence of) the focal habitat ‘coarse sediment and area with boulders’ of the ASEV. With BISI v2 (with a reference compiled from separate ecotopes) it is more obvious to only evaluate the quality status of the focal habitat/ecotope within the ASEV and optionally only include other ecotopes when monitoring is sufficient representative (in efforts and spatial distribution of samples). The target for evaluation should ideally determine the monitoring programme in place and whether sampling is directed to a specific habitat type or the average quality status of an entire area (ASEV). Besides indication of the methodology used for species observation, it is of importance to identify which

Example of an internal reference

Evaluation in BISI occurs relative to an internal reference. The example of Table 2 shows part of the internal reference for the evaluation of the area of specific ecological value (ASEV) Cleaver Bank, There are 3 species for which the Hamon grab / boxcore monitoring data and two species for which video track data are used. For four of these species, evaluations (and reference values) will be based on observed numbers per square meter and for one species based on hit rate (per Hamon grab of 0.09 m^2). In detail, the reference for A. crassa for the ASEV Cleaver Bank is 0.526 specimens per m^2, the reference for A. opercularis is 0,016 specimens per m^2, and the reference for A. pespelecani is a presence in 15,3% of the samples taken. In case of the ASEV Cleaver Bank the complete reference consists of 21 indicator species for which Hamon grab / boxcore monitoring data and 9 indicator species for which video recordings will be used in the evaluation (as shown in Appendix 1 of Wijnhoven & Bos (2017).

Table 2. Part of internal reference showing essential information on first five indicator species to evaluate the area of special ecological value (ASEV) Cleaver Bank as part of the Dutch North Sea evaluation (BISI v1). Specified for the indicator species are which type of monitoring will be used (and the expected number of samples taken with 3-year intervals starting in 2015), the type of observations used in the evaluation (indicated with '1') and used reference values (R_i).

ASEV Klaverbank	Number of samples			Data type		**R_i**	
Indicator species	Hamon grab (0,09 m2)	Boxcore (0,078 m2)	Video tracks (20 m2)	Occurrence (Hit rate)	Densities (number/m2)	Hamon grab (0,09 m2); hit rate (per grab) or density (n/m2)*	Video tracks (20 m2); densities (n/m2)
Arcopagia crassa	17	1		1	1	0,526	
Aequipecten opercularis			16	1	1		0,016
Alcyonium digitatum			16	1	1		36,085
Aonides paucibranchiata	17	1			1	962,880	
Aporrhais pespelecani	17	1		1		0,125	
Number of species (S)	21	21	9				

* Results of Hamon grabs and boxcores are combined in the evaluation. In case of the use of hit rate, the occurrence is not compensated for slight differences in sampled surface area as methodologies also differ in sample depth (sample shape); which makes that differences in hit rate between methodologies are species specific.. In the example different types of occurrence data are combined. Hit rate is generally used in case of availability of presence/absence recordings only, or in case species have a very patchy distribution.

specimens[3] belong to the numbers per species. This is often a decision already taken in the sample identification phase (which makes that it is of importance that standard protocols for identification are used). Basically only the specimens identified to the species level as indicated in the reference list are used for evaluation (where it is indicated when especially in the historic data or in evaluation protocols other names for the same species have been used, as well). This can mean that for instance juveniles (often only identified to the genus level) are generally not considered which is perfectly fine as those are much more susceptible to seasonal or occasional fluctuations. It has to be taken into account that the sampling methodology used, and the mesh-size in particularly, determines which specimens to consider. Thiny specimens are likely undersampled and should therefore not be part of the numbers to be considered, so that selection of specimens from the observation data on basis of size (species specific as the effectivity of the methoidiology is besides size also shape related) might be necessary.

6.3.3 Calculation of BISI-values

Next step in the evaluation is defining the occurrence (Oi) of indicator species for the year(s) and area(s) of evaluation based on the occurrence data (methodology, data type, sample type) as indicated. The occurrence is evaluated taking the observed variance into account. The occurrence to reference ratio (Oi/Ri) is calculated for each of the indicator species. The observed standard deviation (as a measure of variance) is calculated into a value relative to the calculated Oi/Ri-ratio. Similarly as for the Oi/Ri ratios, the adjusted standard deviation is set to a value of 0.01 in case the observed standard deviation equals zero (Table 2).

Besides a general quality assessment (in which all area, ecotope or habitat specific indicator species equally participate, and meet the selection criteria of having sufficient power: 'Smart species'), specific evaluations are performed on weighted species subsets, in BISI v2 possibly supplemented with additional species (depending on the specific indicator value of species for certain causes of change and potential effects of change).

Therefore the natural logarithms of the adjusted O_i/R_i ratios are multiplied with the species weights (IV_i's calculated as iv_i divided by iv_{avg} as indicated in the example of Table 3). In a similar way the species specific iv_i divided by the average iv_{avg} for all included indicator species for a specific evaluation, IV_i, is multiplied with the ln of the squared truncated standard deviation (to achieve the variance that can be summed and multiplied). The square root of the product, back-transformed taking the inverse natural logarithm allows to

---

[3] Are only those specimens identified at species level included, or are optionally other taxonomic levels included in case no other related species are present (e.g. to include juveniles)?

take the standard deviation into account in the significance testing of potential differences.

Example of observation data

Evaluation in BISI is based on observation data that might have to be adjusted to make the indicator as sensitive for quality improvement a for quality deterioration nd reduce the impact of species by far transgressing reference levels. In the example of Table 2 there are two indicator species not observed during the monitoring of 2015 in the area of specific ecological value (ASEV) Cleaver Bank. As observed standard deviation therefore also equals zero, both occurrence to reference (O_i/R_i)-ratio and accompanying standard deviation are adjusted to the minimum value of 0.01. O_i/R_i-ratios and standard deviations are adjusted in case observations transgress 100 times (higher or lower) the reference value, or if the observed standard deviation equals zero.

Table 2. Part of observations for the area of special ecological value (ASEV) Cleaver Bank to be evaluated at T2015 showing the first five indicator species. Observed occurrences (O_i) including standard deviations and occurrence to reference (in numbers per square meter or hit rate as indicated in Table 2) ratios, either or not adjusted, are given for T2015.

ASEV Klaverbank	On basis of standard grab (0,09 m2) and video track (20 m2)					
Indicator species	Oi (T2015)	±stdev	Oi/Ri (T2015)	±stdev	adjusted Oi/Ri	±stdev
Arcopagia crassa	0,000	0,000	0,000	0,000	0,010	0,010
Aequipecten opercularis	0,004	0,009	0,259	0,069	0,259	0,069
Alcyonium digitatum	5,498	24,808	0,152	4,130	0,152	4,130
Aonides paucibranchiata	481,444	494,062	0,500	15,922	0,500	15,922
Aporrhais pespelecani	0,000	0,000	0,000	0,000	0,010	0,010

From these values, the general and specific Benthic Indicator Species Indices (BISIs) can be calculated according to:

BISI = $\exp((1/S)*(\Sigma(IV_i)*\ln((O_i/R_i)+1)-S))$, where

S = Number of indicator species included

IV_i = Indicator Value calculated as iv_i (species specific indicator value with a value between 0-1) divided by iv_{avg} (the average indicator value of all indicator species (with iv_i>0) in the specific evaluation.

O_i = Observed occurrence (ratio of samples with the indicator species present) or observed numbers (average densities)

R_i = Reference occurrence (set ratio of samples with indicator species present under reference condition) or observed numbers (set average densities under reference condition).

<u>Example of indicator values for indicator species</u>

In BISI, indicator values (IV$_i$*) identify the weight of the indicator species in the specific evaluations and equal the species specific indicator value (iv*$_i$*) divided by the average indicator value (iv*$_{avg}$*) for the type of evaluation (Table 3). Compensation for the average indicator value is necessary to make BISI results of different specific evaluations comparable. Ln-transformed O*$_i$*/R*$_i$*-ratios are multiplied by IV*$_i$*s in BISI calculations. This means that species without indicator value are not taken into account in the specific evaluations.*

Table 3. Part of indicator species list with designated indicator values for the area of special ecological value (ASEV) Cleaver Bank showing the first five indicator species. Indicator values *(iv*$_i$*)* indicate the relative importance (the weight) of indicator species in the evaluations. Capitals refer to the potential causes and effects of the observed quality status or quality differences as indicated in Table 1. Species with no iv$_i$ for a certain pressure or function are not expected to be indicative. The last two rows show the average indicator value (used in the calculations to make BISI results of different evaluations comparable) and the total number of indicator species present in the (specific) evaluations for the ASEV Cleaver Bank.

ASEV Klaverbank	General quality	iv for pressure				iv for recovery	national importance	iv for ecological functioning			Habitat Directive
Indicator species		A	B	C	D	E	F	G	H	I	J
Arcopagia crassa	1	1		0,5	1	0,5	1				1
Aequipecten opercularis	1	1		1	0,6	1	1				1
Alcyonium digitatum	1	1		1	1	0,5	1				1
Aonides paucibranchiata	1				0,1	1		1			
Aporrhais pespelecani	1	1		0,75	0,3	1	1				1
Average indicator value (iv_{avg}):	1,000	0,957	0,875	0,692	0,580	0,780	0,941	0,917	0,929	0,750	1,000
Number of species (S)	30	23	4	26	30	30	16	6	7	2	14

As the complete reference for the evaluation of the ASEV Cleaver Bank consists of 30 indicator species, the results of the specific evaluations are likely more distinctive in practice. The general quality evaluation is based on the entire set of indicator species for which sufficient information is available, whereas specific evaluations to identify underlying pressures and resulting effects of potential observed changes are based on subsets of indicator species whose occurrence is more or less related to the investigated pressures and effects as indicated by the iv$_i$*s. As a rule, at least 5 indicator species should be part of the (specific) evaluation before results can be considered, to minimize the potential impact of coincidental chance in species occurrences on evaluation results. In the case of the evaluation of the ASEV Cleaver Bank, there are not sufficient indicator species to evaluate ecological disturbance as a potential cause of observed changes and the potential effect of a reduced quality status on the biological activation of the sea floor top layer (which is however also a process not expected to be hampered on the Cleaver Bank due to relative large hydrodynamics and the abundant presence of coarse sediment). In case such a specific evaluation is considered to be of importance, it is suggested in BISI v2 to add additional indicator species to the specific evaluations (that are not included in the general evaluation) to a number of at least 5 indicator species for the specific evaluation, after which the realized power is calculated.*

(exp = the inverse natural logarithm (e to the power of the formula) as a back-transformation of the transformation according to the natural logarithm (ln)).

In practice first Individual Indicator Species (IIS_i) values are calculated:

$$IIS_i = IV_i * Ln\left(\frac{O_i}{R_i} + 1\right),$$

Those are averaged to come to one combined multi-species index (IIS_{avg}) according to:

$$IIS_{avg} = \frac{1}{S} * \left(\left(\sum_{i=1}^{n} IIS_i\right) - \mathrm{S}\right),$$

That is than back-transformed taking the inverse natural logarithm to achieve the BISI-score:

$$BISI = e^{IIS_{avg}}$$

(Individual occurrence-to-reference data are increased with 1 to prevent negative values, after which for the calculation of the average Individual Indicator Species score this is compensated by a deduction with the number of species in the specific evaluation).

The accompanying standard deviation as indicated in Table 4 can be calculated taking the same steps of ln-transformation, taking species specific indicator values and the total number of indicator species in the analyses into account, working with variances (the value that should be calculated with), with back-transformation in the end.

As the usual way of calculating a pooled standard deviation from a set of averages accompanied with a standard deviation is:

$$Stdev_{pooled} = \sqrt{\frac{\sum_{i=1}^{k}(n_i-1)*Stdev_i{}^2 + \sum_{i=1}^{k} n_i*(Avg_i - Avg_{avg})^2}{\sum_{i=1}^{k}(n_i-1)}},$$

With k = the number of sets to be pooled, and n the number of samples per set.

This results in the following formula to calculate the accompanying standard deviation for the BISI:

$$Stdev_{BIS} = e^{\left\{\left(\frac{1}{S}\right)*\left\{\sqrt{\frac{\sum_{i=1}^{S}(n_i-1)*(IV_i*Ln(Stdev_i+1))^2 + \sum_{i=1}^{S} n_i*(IIS_i - IIS_{avg})^2}{\sum_{i=1}^{S}(n_i-1)}} - S\right\}\right\}}$$

The general BISI is a value for the general quality status of the area, ecotope or habitat to be evaluated. Although the BISI can be tested on significance against the compiled reference, the methodology is specifically intended to evaluate potential changes in time. Related to management the future quality status is preferably compared to an initial situation; a T0 of the situation before measures are taken, or the situation at which it was decided that the area or habitat of concern should be conserved or is in need of quality improvement (in line with amongst others MSFD and HD evaluation cycles).

Example of calculation of BISI values from individual indicator species (IIS) values

Table 4 shows as an example the calculated BISI-values ± standard deviation for the general and specific evaluations for the area of specific ecological value (ASEV) Cleaver Bank on T2015 (which equals the T0). Calculated BISI values are compared (tested) against the reference with a BISI-value of 0.736, or compared with future monitoring events. BISI-values are the inverse natural logarithm of a summation of IIS-values, who are calculated as $IIS_i = (IV_i)*ln((O_i/R_i)+1)$ with $IV_i = iv_i/iv_{avg}$ (values of iv_i, iv_{avg} and O_i/R_i shown in Tables 3 and 2 respectively), divided by the total number of indicator species in the (specific) evaluation. Additionally the accompanying standard deviation can be calculated on basis of variances taking the same ln-transformation, multiplication with indicator values, the number of indicator species in the index and the total number of samples into account. Calculation of a pooled standard deviation includes the calculation of an average variance plus taking into account the resulting variance among the average values to be pooled.

Table 4. Selection of calculated individual indicator species (IIS)-values including accompanying variances for the area of special ecological value (ASEV) Cleaver Bank showing the first five indicator species. Calculated BISI values (with accompanying standard deviations) are the result of the summation of IIS_i values for all identified indicator species divided by the total number of indicator species, thats are back-transformed in the end taking the inverse natural logarithm, as indicated by the formula.

ASEV Klaverbank	General quality	iv for pressure				iv for recovery	national importance	iv for ecological functioning			Habitat Directive
Indicator species		A	B	C	D	E	F	G	H	I	J
$IIS_i=(IV_i)*ln((O_i/R_i)+1)$											
Arcopagia crassa	0,010	0,010	na	0,007	0,017	0,006	0,011	na	na	na	0,010
Aequipecten opercularis	0,231	0,241	na	0,333	0,239	0,296	0,245	na	na	na	0,231
Alcyonium digitatum	0,142	0,148	na	0,205	0,245	0,091	0,151	na	na	na	0,142
Aonides paucibranchiata	0,405	na	na	na	0,070	0,520	na	0,442	na	na	na
Aporrhais pespelecani	0,010	0,010	na	0,011	0,005	0,013	0,011	na	na	na	0,010
$BISI = exp((1/S)(\sum(IV_i)ln((O_i/R_i)+1)-S))$	**0,475**	**0,404**	**0,462**	**0,471**	**0,487**	**0,447**	**0,341**	**0,734**	**0,575**	**0,764**	**0,299**
Variance (Var_{IIS})											
Arcopagia crassa	0,020	0,021	na	0,014	0,034	0,013	0,021	na	na	na	0,020
Aequipecten opercularis	0,133	0,139	na	0,192	0,138	0,171	0,141	na	na	na	0,133
Alcyonium digitatum	3,270	3,419	na	4,723	5,638	2,096	3,474	na	na	na	3,270
Aonides paucibranchiata	5,657	na	na	na	0,975	7,253	na	6,172	na	na	na
Aporrhais pespelecani	0,020	0,021	na	0,022	0,010	0,026	0,021	na	na	na	0,020
± Stdev	*0,388*	*0,391*	*0,560*	*0,389*	*0,386*	*0,388*	*0,401*	*0,526*	*0,480*	*1,125*	*0,399*

V1 – With BISI v1 an evaluation methodology is developed to compare the quality status of the Dutch North Sea areas to a T0 situation (i.e. the quality status as calculated on basis of the monitoring data from 2015). The specific evaluations are predominantly intended to identify the relative importance of potential causes or the impact of changes in functions (effects), on basis of relative differences in specified BISI values and the significance of potential differences relative to the T0 and developments in time in (specific) BISI values.

With BISI v2 where the design of the area specific index is more standardized, the comparability of quality evaluations between different areas is improved. Compilation of the area specific index has become ecotope based, using the same references for the same ecotope in different areas. Application in various circumstances and different areas, including possible reference areas, should allow testing and possibly fine-tuning of the current methodology. It should be taken into account that results will always be impacted to some extent by aspects that are out of reach of a methodology development: E.g. data availability, representability of monitoring (techniques and efforts), habitat constitution (concealed by aggregated large-scale habitats), and etcetera.

6.3.4 Evaluation of quality status and significance testing

Basically with having the general - and specific BISI values for areas, ecotopes and habitats for moments of evaluation, the results of a quality assessment are there. It however depends on the type of reporting/evaluation (as for instance indicated in Figure 1), which quality data to test, combine or consider and what a good quality status looks like.

V1 - The different assessment areas of the Dutch North Sea according to BISI v1:

- Areas of specific ecological value (ASEV; that can be Natura 2000 areas as well),
- Habitat Directive habitat types (that include (parts of) ASEVs but generally extent outside ASEVs),
- and EUNIS (level 4) ecotopes (that cover areas partly inside and outside ASEVs),

with a special type of evaluation to test the efficiency of management measures on basis of a comparison of (partly) for specific fisheries closed and open areas (within ASEVs).

In BISI v2 the indicated assessment areas are still the core targets for evaluation. However, the EUNIS (level 4) ecotopes (or comparable ecotopes according to another ecotope classification like ZES.1 that is more accurate in for instance transitional waters) form the basis to construct area specific evaluation indices using the surface area ratio of ecotopes in the area of evaluation.

For each of the indicated areas to be evaluated the objective can be to achieve conservation of the current quality status (i.e. no decrease in the quality status) or an improvement of the quality status (for which there might be a timeline). After broader Europe-wide application including areas with different known pressure types and levels and possible application in some reference areas, certain BISI-levels could become management targets as well. A certain BISI-level, possibly ecotope related, can than indicate a good quality status.

In the Dutch situation in the meanwhile a quality status developing in the direction of a BISI-score of around 0.736 (similar to the internal reference) is a good objective for at least those areas and habitats with an improvement objective according to the Marine Strategy (Min IenW & Min LNV, 2018) applicable for the mid-long term. Also evolution in that direction for those areas and habitats with a conservation objective might be beneficiary from a nature perpective (ecological functioning) and to safeguard natural resources. In other European regions with low pressure levels a BISI-score of around 0,736 is possibly not a good objective as it might involve a decrease in the quality status; such a decrease should only be allowed for very good reasons. It is expected that future quality objectives based on BISI will arise from broad-scale application and especially application in potential reference areas.

V1 - In the Dutch situation there is an improvement objective for the EUNIS ecotope 'Deep mud' and specifically for deep low-dynamic sandy bottoms, within the frame of the MSFD. The last is part of EUNIS level 4 ecotope 'Deep sand', which makes that on national scale an improvement in quality, and especially not a transition to high dynamic sandy bottoms, should become visible. To cover these two important ecotopes, there is an improvement objective for the two ASEVs; Frisian Front and Central Oystergrounds as well. Besides, all HD areas have an improvement objective nowadays (Min IenW & Min LNV, 2018).

In all other EUNIS 4 ecotopes and ASEVs at least no decrease in quality status is allowed (except for the Brown Bank with no official status yet; situation 2018). As from 2018 on there is an improvement objective for each of the ASEVs (except for the Brown Bank) foreseen within the frame of the MSFD (Min IenW & Min LNV, 2018). Additionally, there are fisheries regulations (planned) in each of the ASEVs (except for the Brown Bank), of which the efficiency is tested. The aim is at least an improvement in the quality status of closed areas (likely more than in the open areas if there the restricted fisheries continues and is a dominant pressure affecting benthic communities), that might result in an improvement of the overall quality of the entire ASEV on the mid-long term.

Towards the evaluation within the frame of the Habitat Directive, the current indicator will provide (background) information on the causes and effects of observed developments in the quality status of the H1170 and H1110 habitats in the North Sea. It functions as an early indication of developments in quality there, as the Benthic Indicator Species Index (including all typical species indicated for the HD habitats) is much more sensitive than the estimation of expected presence of typical species in Article 17 reporting for the HD. In 2019, the quality evaluation of particular the structure and function of the marine habitat types H1170 and H1110 in the Dutch HD reporting will be based on BISI for the first time (is currently filled in). Additionally BISI is at the moment also being developed for HD habitat types H1130, H1140 and H1160 and foreseen as the evaluation tool for future reporting. It has to be noticed that benthos (and the current BISI indicator) is often one of the indicators to be evaluated in combination with others.

Knowing the targets of evaluations the suitable tests can be defined. A first evaluation, the T0 based on the monitoring data from 2015, identifies the current status with which future evaluation moments will be compared. The T0 according to the BISI does not provide an absolute quality status rating. It is true that the relative distance of BISI-values compared to the compiled reference levels will be to a certain extent indicative for the current quality status (as shown for the Cleaver Bank in Table 5). For an absolute quality status rating, the T0 should however be compared to (1) the historic reference, for as far as such reference is known. Additionally, (2) the quality status in 2015 should be put in perspective by analyses of recent (historic) developments in the BISI (showing quality developments for the last decennia if possible).

The results of both analyses for the areas to be evaluated in the Dutch North Sea are presented in Wijnhoven (2018). The T0 is compared with a realistic quality status (compiled reference according to BISI v1) using 2-sided independent t-testing. A realistic quality status, in principle is a status that can be achieved via natural development when dominant pressures are minimized, given the current habitat constitutions and species pools present. As current variation in benthic communities might be more representative for future observations than historic information often based on limited numbers of samples or even other methodologies, the observed variance at T0 is used in the calculations for the reference as well.

Future evaluations in the Netherlands will initially consist of similar 2-sided independent t-testing of years of evaluation against the reference (or 1-sided if it is obvious that the quality status is below the reference level) (Table 5) and 2-sided paired t-testing against the T0 (if the currently installed monitoring programme is unchanged as expected). In the future (indicative from 2027 onwards when at least 5 data points are available) trends in developments of BISI values can be evaluated as well.

Example of results of significance testing

Table 5 shows the results of the significance testing of BISI values at T0 for the ASEV Cleaver Bank as an example. A pooled standard deviation is calculated as the square root of 2 times the quadrat of the standard deviation (i.e. the variance) at T0 (considered representative for the natural population fluctions that can be expected, as at least monitoring efforts have been reasoble at that time) divided by the number of indicator species in the analyses. The computed t Statistic equals 0,736 minus the BISI value at T0 divided by the pooled standard deviation (as the reference BISI equals a value of 0.736). The critical value of t is determined by taking the inverse of the two-tailed Student's t distribution (=TINV(probability, degrees of freedom)), with a probability of 0.05 and two times the number of indicator species minus two, as the degrees of freedom. The probability of the computed t can be calculated according to the Student's t distribution (=TDIST(computed t Statistic, degrees of freedom and number of tails)).

The results of this example show that the general quality status is significantly different (lower) than the compiled reference levels (a realistic improved quality status for the future). It is even more interesting that the lower indicator values (BISIs) for amongst others seafloor disturbance (A), larger sized species (C: Indicative for the intensity of seafloor disturbing fisheries and longer living species (D: Indicative for frequency of seafloor disturbing fisheries) differ significantly. There is also a significant difference in the occurrence of species indicative for recovery (E). Lower values for other causes and effects compared to the reference are not significant and should be considered more or less in line with the reference.

Table 5. Results of significance testing with a two-tailed t-test of BISI results of the T0 (values ± standard deviation as indicated in Table 4) against the internal reference (with BISI values equal to 0,736) for the area of special ecological value (ASEV) Cleaver Bank. Significant differences are indicated with *** (p≤0.001), ** (p≤0.01), * (p≤0.05), and ns = not significant.

ASEV Klaverbank	General quality	iv for pressure				iv for recovery	national importance	iv for ecological functioning			Habitat Directive
Statistic		A	B	C	D	E	F	G	H	I	J
Pooled Std Dev	0,100	0,115	0,396	0,108	0,100	0,100	0,142	0,304	0,257	1,125	0,151
Computed t Statistic	2,450	2,306	0,600	2,328	2,480	2,497	1,930	0,618	0,825	0,151	1,970
Critical Value of t	2,002	2,015	2,447	2,009	2,002	2,002	2,042	2,228	2,179	4,303	2,056
Probability of Computed t	0,017	0,026	0,571	0,024	0,016	0,015	0,063	0,550	0,425	0,894	0,060
Significance	*	*	na	*	*	*	ns	ns	ns	na	ns

A special case is the testing of differences in developments of for specific fisheries closed and open areas taking potential differences in the initial benthic indicator species assemblage (at T0) into account. If differences between open and closed areas at T0 are minimal and non-significant, one can decide to evaluate whether there are differences in BISI-values using 2-sided independent t-testing at future evaluation moments. Although differences between open and closed areas are frequently not found to be significant at T0; considering them to be similar brings additional disturbance in the analyses. It is preferential to analyse findings according to a Before-After-Control-Impact design.

Therefore initially the difference in BISI-values for singular indicator species is calculated between the two areas with different treatments on T0 and the other moment of evaluation. The average difference in BISI with accompanying standard deviation is calculated from the results of individual species. Results in BISI-differences between different treated areas are compared between T0 and the other moment of evaluation using a one-sided t-test.

Example of significance testing in the comparison of different treated areas

In the Dutch situation several areas are closed for seafloor disturbing fisheries as a measure to improve benthic habitat quality in areas of specific ecological value and average quality of the Dutch North Sea. The national benthos monitoring programme is setup according to a BACI-approach for evaluation of the effectivity of taken measures. Table 5 shows the results of an evaluation of closed versus open areas in the area of specific ecological value (ASEV) Cleaver Bank, based on the data of 2015 (T0) and a fictional example for the T1 at which only the abundances of the indicator species with indicator value towards 'seafloor disturbances' (i.e. indicator species for specific evaluation A) are increased with a value equal to the in 2015 observed standard deviation, or to an abundance of 0,01 m^{-2} if not present at T1, in the closed areas, The average differences in BISI between open and closed areas with accompanying standard deviation for individual species are calculated and tested with a one-sided t-test (in this case it is clear that only potential quality improvement is found in the closed area).

Table 6. Results of significance testing of potential differences in BISI developments between for specific fisheries closed and open areas as part of the area of specific ecological value (ASEV) Cleaver Bank based on BISI-values for individual species with accompanying variances. Specific results of the first five indicator species are shown. The example includes results of the T0 and a fictional example for T1. Potential difference is tested using a one-sided t-test. Significant differences are indicated with *** (p≤0.001), ** (p≤0.01), * (p≤0.05), and ns = not significant.

ASEV Klaverbank	**General quality**													
Indicator species	Open						Closed						Difference	
T0:	Oi/Ri	± stdev	IIS	± Var	BISI	± stdev	Oi/Ri	± stdev	IIS	± Var	BISI	± stdev	in BISI	± Var
Arcopagia crassa	0,010	0,010	0,010	0,020	0,372	0,418	0,010	0,010	0,010	0,020	0,372	0,449	0,000	0,219
Aequipecten opercularis	0,078	0,094	0,075	0,179	0,397	0,411	0,378	0,371	0,321	0,631	0,507	0,445	0,110	0,173
Alcyonium digitatum	0,229	11,424	0,206	5,039	0,452	0,625	0,442	63,115	0,366	8,321	0,530	0,726	0,078	0,631
Aonides paucibranchiata	0,442	14,217	0,366	5,445	0,530	0,955	0,509	16,345	0,411	5,707	0,555	0,976	0,025	1,657
Aporrhais pespelecani	0,010	0,010	0,010	0,020	0,372	0,418	0,010	0,010	0,010	0,020	0,372	0,449	0,000	0,219
Average difference in BISI ± stdev													0,094	0,736
Indicator species	Open						Closed						Difference	
T1#:	Oi/Ri	± stdev	IIS	± Var	BISI	± stdev	Oi/Ri	± stdev	IIS	± Var	BISI	± stdev	in BISI	± Var
Arcopagia crassa	0,010	0,010	0,010	0,020	0,372	0,418	0,019	0,019	0,019	0,038	0,375	0,544	0,003	0,304
Aequipecten opercularis	0,078	0,094	0,075	0,179	0,397	0,411	1,034	0,371	0,710	0,631	0,748	0,446	0,352	0,137
Alcyonium digitatum	0,229	11,424	0,206	5,039	0,452	0,625	2,791	63,115	1,333	8,321	1,395	0,732	0,943	0,340
Aonides paucibranchiata	0,442	14,217	0,366	5,445	0,530	0,955	0,509	16,345	0,411	5,707	0,555	0,998	0,025	1,715
Aporrhais pespelecani	0,010	0,010	0,010	0,020	0,372	0,418	0,065	0,065	0,063	0,127	0,392	0,544	0,020	0,297
Average difference in BISI ± stdev													1,193	0,685

Statistic T0-T1:	
Pooled Std Dev	0,184
Computed t Statistic	5,988
Critical Value of t	2,003
Probability of Computed t	0,000
Significance	***

A fictive example of T1 is shown, with an increase of the occurrence in the closed areas of specific those species indicative for seafloor disturbance (specific evaluation A) as observed at T0 with the standard deviation or increased to 0,01 m^{-2} in case the species was absent before.

Table 6 shows a fictional example with an increase in the abundances of the indicator species with indicator value towards 'seafloor disturbances' (i.e. indicator species for specific evaluation A) equal to the in

2015 observed standard deviation or to an occurrence of 0.01 m^{-2} for those species absent at T0. Similar as for BISI values showing quality developments in time for specified areas, indicative from 2027 onwards, trends based on the results of differences in BISI values between closed and open areas (averages ± standard deviations) in time can be identified and/or analysed.

The testing with BISI v2 in other regions is the same. Whether one-on-one comparison of different years or areas or testing of possible trends and/or deviations from trends is possible or most suitable depends on the case-by-case data availability and monitoring designs.

6.4 Indicator metric (formula)

The weighted geometric mean of (species - and evaluation specific indicator values) ln-transformed observation-to-reference ratios are calculated according to:

$BISI = \exp((1/S)*(\Sigma(IV_i)*\ln((O_i/R_i)+1)-S))$, where

S = Number of indicator species included

IV_i = Indicator Value (which equals the species specific indicator value iv_i with a value between 0-1, divided by the average indicator value iv_{avg} for the specific evaluation)

O_i = Observed occurrence (ratio of samples with the indicator species present) or observed numbers (average densities)

R_i = Reference occurrence (set ratio of samples with indicator species present under reference condition) or observed numbers (set average densities under reference condition).

('exp' is similar to putting e to the power of the formula as indicated, which equals the inverse natural logarithm, as a back-transformation of the natural logarithm (ln) taken from the occurrence-to-reference ratios).

BISI = Benthic Indicator Species Index:

A value between 0.372 (very low quality) and 37.156 (very high quality) can be achieved. When the quality condition is comparable to the internal reference a BISI-score of 0.736 will be obtained. In practice a BISI-value will be around 0.4 (or significantly below 0.736) in case quality improvement is requisite, and is around 0.736 in case of rather good quality condition considering the Anthropocene is found. With a BISI-value around 0.736 it is expected that possible pressures on benthic habitats and communities in particular are significantly reduced and/or of minor importance. This is not necessarily the case for entire Europe; it is very well possible that in certain regions or for certain ecotope types (not tested yet), pressure levels are much lower, and that certain areas could be considered as reference areas. In these cases BISI-levels might transgress a value of 0.736 and there the objective might be conservation at a quality level comparable to a BISI-level >>0.736. Additionally it has to be taken into account that in regions subjected to intense human activities former disturbances could have led to a situation where the habitat constitution is completely different from pristine conditions and that certain species that potentially fit in the available habitats had no opportunities to return (Wijnhoven, 2018). BISI scores come with a calculated standard deviation to be considered for significance testing.

S, IV_i and R_i are area, ecotope or habitat specific fixed data dependent of the used sampling methodology. Fixed values and/or reference species lists can be adjusted to new insights. In that case recalculation of the T0 and all previous evaluations is necessary (which is however not a huge effort).

V1 - Similarly, the methodology including reference levels can potentially be reflected on similar ecotope and/or habitat types in the vicinity of the Dutch North Sea, taking monitoring efforts into account. For other regions it is essential to compile area specific reference lists based on area-specific historic data.

With BISI v2 the use of references from other areas, in this case ecotopes forming the bases of BISIs constructed for (other types of) areas to be evaluated, is formalized. Herewith the use of earlier constructed references on ecotope level in other areas is part of the methodology. Especially the use of possible reference areas for methodology development is encouraged.

The Benthic Indicator Species Index is dimensionless. However, the BISI can also be presented as a percentage (%BISI) relative to the potential maximum indicator value to achieve a lineair scale. A BISI between 0.372 and 0.736 equals a %BISI between 0% and 50% with 50% being the realistic reference level. The index can however transgress the realistic reference level (with 100 times the reference occurrences of each of the indicator species at maximum; that can be indicated as a %BISI of 100%).

With BISI v1 the relative comparability of results from different areas was not optimal yet, due to the lower level of standardization then with BISI v2. With the introduction of ecotope-based references that should basically be the same for different areas (with only exceptions for small differences due to differences in available species pools at European regional scale), the BISI results between different areas with BISI v2 are highly comparable.

6.5 Assessment benchmark

At present dependent of the area, ecotope or habitat to be evaluated, a consolidation of the current (T0) BISI values or an increase in BISI values indicates a good (or desirable) quality status for the Dutch situation. Such an increase has to be significant (at $p<0.05$) taking natural fluctuations into account. Good

Quality Status is when (certain) pressures are effectively reduced and lead to increasing index values. I.e. when management measures initially result in increasing BISI values in 'for specific activities closed' areas and might lead to an increase in BISI values for an entire area (e.g. ASEV) on the longer term.

After a number of monitoring events and having effective management in practice it should be evaluated whether the compiled reference (BISI = 0.736 or %BISI = 50%) could function as a target for the good quality status. With BISI v2 and a broader application in the European context, including possible comparison with reference areas, it is expected that internal references per ecotope can be validated. Thresholds and targets for quality assessments in relation to MSFD and Natura 2000 are at the moment discussed at various levels including nationally and within OSPAR and ICES.

6.6 Data source and description of data

There are two types of datasets related to the current methodology. In the first place, there has to be a dataset of 'historic' data and/or data from reference areas at ecotope level in particular, on which the compiled reference levels used in the methodology will be based. Historic data are not directly involved in the calculation of IIS's, but are required for the selection of indicator species and assessment of the reference values. Then there is the dataset that will be evaluated. These data likely include data of some kind of T0 or initial state: A situation that is evaluated or to which other evaluation moments, possibly in the future, will be compared. Characteristics and essentials of both data sets will be described with examples from the Dutch North Sea case.

6.6.1 'Historic' data used to extract the reference levels

Although 'historic' might suggest that observations from decades to centuries ago might have been used; this is not the case. If available, such data are scarce, highly scattered (not covering all essential areas), often descriptive and difficult to match with current sampling methodologies. But most important, it is highly questionable if a historic reference level is a good reference level to use, as habitats currently present are modified by centuries of anthropogenic activities. If all pressures were taken away today, habitats and benthic communities would naturally not develop to pristine state (without any help), due to nowadays largely deviating habitat constitution and different species pools present. If certain parts would develop in the direction of some kind of pristine state, this would be a long-term process. The aim of the current methodology is to show quality improvements (or deterioration) on the short- and mid-term potentially as a result of changes in management and taken measures. Therefore a more realistic reference is used, based on current habitat constitution and present species pools, potentially showing first indications of quality improvements and/or deterioration. It is expected that the used compiled reference levels might reflect a realistic target in case the dominant pressures are reduced, but we are aware that future evaluations might indicate that certain reference occurrences should be adjusted on the mid-long term. BISI estimations from the past can easily be recalculated according to new reference levels if there are.

V1 - In the Dutch North Sea situation (BISI v1), compiled reference levels are based on existing large data sets on North Sea benthos covering the period 1984 till 2014 (with an exception of some older data for the Plain of the Raan). Basically, each of the available datasets are analysed for the areas to be evaluated. Maximum observed (year) average occurrences (either hit rate or abundances) are extracted and compared. The reference level consists of the highest value for each species for a certain area if data coverage was expected to be sufficient (number of samples) and representative (spatial distribution and used methodology) for the area of investigation. Exceptions were made if highest occurrences were observed in recent years. In that case also the observations from 2015 were considered, maximum observed occurrences were either increased with the observed standard deviation or values were doubled, based on expert judgement with BISI v1.

With BISI v2 the derivation of the internal reference is more standardized according to the scheme presented in Figure 3, which includes the consideration of possible reference areas (ecotopes).

Historic data used for the Dutch North Sea are:

- The BIOMON/MWTL North Sea data covering the period 1991-2012. The aim of the MWTL programme for the North Sea commissioned by Rijkswaterstaat's Centre for Water Management is to map out the macrobenthos and monitor changes in the communities. The programme is based on 100 samples taken by a Reineck boxcorer from fixed locations on the Dutch Continental Shelf. The boxcorer samples have a surface area of 0.078 m^2 and a minimum depth of 15 cm and are sieved through a 1 mm mesh. Up until 2012, the MWTL sampling was carried out every year in spring. Currently, the sampling frequency is every three years. The MWTL sampling for 1995 included only 15 locations, with five samples (0.068 m^2) taken from each location. Data are made available by the Marine Information and Data Centre (IHM) at http://www.informatiehuismarien.nl/open-data/.
- The WOT mollusc survey data covering the period 2004-2014. The aim of the WOT mollusc survey is to map out populations of commercially important mollusc species in the North Sea Coastal Zone and to monitor the trends for these species. The survey, carried out by Wageningen Marine Research (IMARES during the 2004-2014 monitoring) on behalf of the Ministry of Economic Affairs, has been running since 1993 and sampling is mainly done with a dredge. The

survey covers 862 locations that have been selected according to a stratified design focused on areas where the highest mollusc densities are expected. Besides the drag dredge (sampling of a surface area of 15m^2), a suction dredge (30 m^2) and a Van Veen grab (3 x 0.1 m^2) are used locally. These have sampling depths of 10, 7 and 15 cm respectively and the samples are sieved through a 0.5 cm mesh (Goudswaard et al., 2012). In addition to commercially appealing species like *Mytilus edulis*, *Cerastoderma edule*, *Spisula subtruncata* and *Ensis directus*, other larger species are currently also counted (therefore only the data starting from 2004 are considered). Data are also made available by the Marine Information and Data Centre (IHM) at http://www.informatiehuismarien.nl/open-data/.

- Additional data have been used that were collected at various locations on the Dutch Continental Shelf with the NIOZ's Triple-D dredge, that were used within the frame of the project 'North Sea indicators under the Marine Strategy Framework Directive' (Wijnhoven et al., 2013), a precursor study for the current methodology and the monitoring programme currently in place. The NIOZ data were collected in 2007-2010 for various scientific programmes (BSIK, NNSM, and Atlas). The NIOZ dredge samples have a surface area of 20 m^2 and a depth of 18 cm. Samples are sieved through an 8x8 mm mesh and all organisms are sorted by species. The data cover the entire offshore areas of the Dutch Continental Shelf and the North Sea Coastal zone. Data are presented in distribution maps for the most common larger benthos species (Witbaard et al., 2013).
- Also in 1996 and 1997 a study within the frame of BEON (Beleidsgericht Ecologisch Onderzoek van de Noordzee/Waddenzee) has been performed using the Triple-D dredge sampling (approx. 30 m^2) a subset of BIOMON stations covering the Dutch Continental Shelf. Additional sampling has taken place using a fine meshed (1x1 cm) 3m beam trawl. Data are presented in Bergman & Van Santbrink (1998) and were extracted from there.
- Also underwater video footage shot by NIOZ at the Cleaver Bank in 2011/2012 (amongst others used for the same study of Wijnhoven et al., 2012). Video tracks covered an area of between 600 and 1500 m^2. The organisms found in this area were sorted by species (where possible) and counted, resulting in a dataset with densities per 20 m^2.
- Additional data for the Cleaver Bank were extracted from the report by Van Moorsel (2003), presenting results of monitoring with Hamon grab (3 x 0.2 m^2), video, beam trawl, dredge and observations using scuba diving carried out by Ecosub. Especially grab and video recordings were used to compile a reference.
- Historic data available to the authors from the former Monitor Taskforce (NIOZ) data base (Benthos Information System v230116) have been used as well. These consist of data from boxcore (0.071 m^2) and

Van Veen grab (0.1 m^2) sampling, sieved over 1 mm mesh, executed during the ICES North Sea Benthos Survey conducted in 1986, which were taken on the Dutch Continental Shelf (or just outside the border). Data are (partly) presented in Duineveld et al. (1991) and Craeymeersch et al. (1997). Data from Van Veen grab (0.1 m^2) and boxcore (0.068 m^2) sampling, sieved over 1 mm mesh, executed during the BOVO (Bodemdieren Voordelta) inventories during the years 1984-1988 in the North Sea coastal zone (Plain of the Raan and Front Delta). Data are (partly) presented in Seip & Brand (1987) and Wijnhoven et al. (2006). Data from Van Veen grab (0,1 m^2) sampling sieved over 1 mm mesh, executed during the years 1962-1966 by the DIHO (Delta Institute for Hydrobiological Research) are used for the Plain of the Raan. Data are presented in Wolff (1973) and Wijnhoven et al. (2006). An additional Van Veen sample taken in 1990 is used for the reference of the Plain of the Raan there origins from the MMP (Monitoring Master Plan) an international monitoring programme executed within the frame of ICES and OSPAR (Wijnhoven et al., 2006).

6.6.2 Essential data for evaluation

Basically any macrobenthic monitoring data should do, as long as some rules are considered:

- Sampling methodologies should be suitable for the detection of benthic macrofauna and should fit to the selected indicator species in the index. Additionally methodologies of monitoring and derivation of reference occurrences should be comparable in terms of about the same surface, to a similar depth, with a similar mesh size at sampling, and identification to about the same taxonomic level. Certain species can potentially be monitored with various quite different techniques (e.g. boxcores and dredges), but with introduction of an alternative monitoring technique, new methodology specific reference values should be calculated if not available.
- Monitoring should be representative to give a good view of the quality status of the ecotopes of concern. Ideally sampling is random within the ecotope in a certain area, or a fixed monitoring grid (initial random selection) is continued. Such a monitoring can be random stratified over the (in surface area) dominant ecotopes. It should be determined if not an ecotope but a composite area (of various ecotopes) is the target for evaluation, or whether the monitoring is sufficient to give a representative view of the entire area (see next point).
- Monitoring efforts should be sufficient to give a representative and reliable view of the quality status. The minimum number of samples can be determined using power analyses for which a certain level of accuracy (what differences should be detectable within what timeframe) and a significance level is considered. It is expected that those are reasonable (real-

istic) numbers of samples as power analyses are part of indicator species selections as well so that only 'smart' indicator species are part of the general evaluations with BISI v2. If certain specific evaluations are of importance, it might be that the number of necessary samples is higher (or one should accept a lower power of these tests), as those evaluations include a smaller number of 'smart' species and can include less distinguishing 'indicator' species as well. There might be options to spread out monitoring over several campaigns or years and combine data as one moment/period to be evaluated to increase the sample size. Ideally those samples are than not from the same locations. Similarly evaluating a certain moment or period against a trend (based on several years), or comparing trends, might increase the power of the tests.

As an example the data used for evaluation of the areas, habitats and ecotopes of the Dutch North Sea with BISI v1 are described here:

V1 - The evaluation makes use of the Dutch National Benthos Monitoring Programme (MSFD monitoring North Sea), that consists of a recurrent (every three years) boxcore monitoring (0.078 m^2) and dredge sampling programme. Dredge sampling is in the offshore areas (i.e. MSFD zones of Dogger Bank, Oystergrounds, Offshore) focussed on the areas with special ecological values, and in principle standardized to samples of 20 m^2 with a depth of 20 cm sieved over 7 mm mesh. In the coastal zone, the MSFD monitoring makes use of the WOT mollusc survey sampling predominantly with a dredge (15 m^2), but also suction dredge (30 m^2) and a Van Veen grab (3 x 0.1 m^2) are used (all sieved over 0.5 cm mesh). As indicated before, some additional sample locations are added to the WOT sampling, specifically for the MSFD monitoring. Nowadays a range of species in recorded. It has to be mentioned that subsampling (dependent of species groups and expected densities) takes place (which makes that dredge samples are unsuitable to use for evaluations based on hit rate; in the methodology hit rate is only used for boxcore data).

Although (especially in the coastal zone) a lot of samples are available, it is specifically indicated which samples will be used for evaluations of ASEVs, ecotopes and habitats, as the set of samples should be representative for the entire area, and sample sites are nowadays fixed allowing paired evaluation at least against the T0. Similarly, specific samples are indicated to be used for the evaluation of management measures, as initial (before measures were taken) habitats and/or communities sampled in the different management zones should be similar. Additionally sample locations are nowadays fixed allowing evaluation according to a BACI approach. Some samples are used for both types of evaluations, which are indicated with the data. For management evaluations only dredge samples are used.

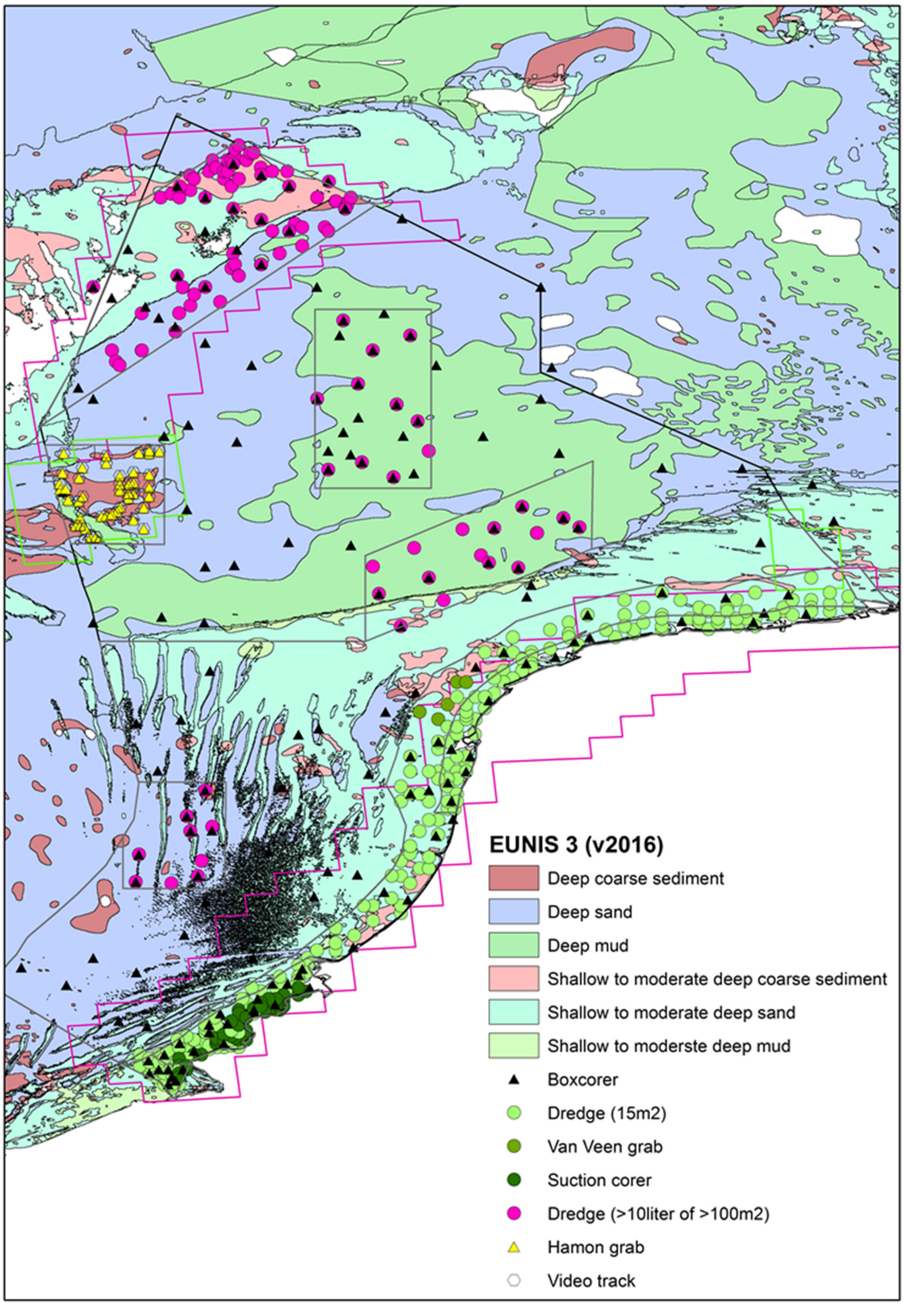


Figure 2. Map of the Dutch Continental Shelf with indication of the sample locations and used methodologies being part of the benthos monitoring within the frame of the MSFD. Locations are in principle sampled every three years from 2015 (T0) onwards. The background is a map with ecotope classification at EUNIS level 4 (v2016 available from: www.emodnet-seabedhabitats.eu), areas with special ecological values (ASEVs) are framed with grey lines, contours of Habitat Directive habitat types H1110 and H1170 are indicated with pink and green lines respectively. Each of the classifications identifies the areas to be evaluated as part of the methodology. Areas with fisheries restrictions (and for instance the associated monitoring in the ASEVs of the Frisian Front and the Central Oystergrounds) are not indicated as they are not (all) definite yet. (In the black fine-dotted areas there is a frequent alternation of ecotope types: Each pixel another ecotope).

Evaluations of developments in the quality status of ecotopes initially takes place based on boxcore samples and associated species (that cover in- and outside ASEV areas). If statistical testing indicates that there are no differences within specific ecotopes in- and outside ASEVs, then dredge samples and associated species can optionally be included in the evaluations as well.

Due to the habitat characteristics (coarse sediment and presence of boulders) there is a specific monitoring based on Hamon grab (0.09 m^2) sampling and video observations in the ASEV of the Cleaver Bank (with fixed locations for ASEV and management measure evaluations).

All data are available via http://www.informatiehuismarien.nl/open-data/, the data portal of the Marine Information and Data Centre (IHM), where also the shapes of ASEVs (and in the near future the shapes of areas with fisheries restrictions) are available. For evaluation of Habitat Directive habitat types, it is possible to fix the current situation for the coming evaluations (as done for EUNIS ecotopes as well) with the benefit that the same monitoring locations can be used in pairwise comparisons. When significant shifts in the habitat contours occur (see Article 17 evaluations) it is an option to update the shapes (which might lead to different sample stations in the analyses).

Results of the T0 evaluation of the areas, habitats and ecotopes of the Dutch North Sea, including an evaluation of the recent historic developments in quality status based on BISI v1, are presented in Wijnhoven (2018).

6.7 Methodology and data analyses

The numbers and/or occurrence of a selection of indicator species (area and/or habitat specific) is compared with a defined reference level, following the formula listed in paragraph "Indicator metric (formula)". To test for differences, the geometric mean (using ln-transformation) of weighted observation-to-reference ratios is calculated. Indicator species lists and reference abundances are ecotope specific and/or constructed from the ecotope surface area constitution of the area to be evaluated. The weighting is on basis of an indicator value (iv_i) for each of the individual species, which defines on a scale from 0 (no indicator value: Species not taken into account for the specific evaluation) to 1 (species is very good indicator for the evaluation of a certain cause or effect of observed quality status), the relative importance of species in a specific evaluation. E.g. the abundance of *Arcopagia crassa* (a bivalve) on the Cleaver Bank is considered a very good indicator (iv_i=1) for the presence of seafloor disturbance, an intermediate indicator for the recovery of benthic communities (iv_i =0,5) and has no indicator value towards ecological disturbances (iv_i =0). For significance testing of resulting BISI values, the observed within population variances are also taken into account.

6.7.1 Selection of indicator species

Indicator species are selected on the basis of the combination of 4 characteristics at the level of (high aggregation) ecotopes, preferably EUNIS 4 habitats[4]:

1) Relatedness to ecotopes: Certain species are characteristic or specific for certain habitats (e.g. sediment constitution, local hydrodynamics, tidal - and depth strata) and are therefore potential indicators for changes in habitat constitution.
2) Species traits/life histories: Benthic macrofaunal species can be characterized in terms of sensitivity, resistance (to pressures) and/or resilience (recovery after pressure) towards different pressures in their direct environment or habitat. This characterization is largely correlated to the species' life-history or to their specific traits (size, longevity, frequency and number of recruits, mobility, specific habitats of life-stages).
3) Their presence in terms of densities and distributions (e.g. equally distributed or present in aggregations with sufficient chance of detection) under natural good quality conditions.
4) Catchability with for the monitoring available and applied sampling gear.

V1 - In case of BISI v1, species selections were the result of extensive literature review, data analyses of 'historic data' from the Dutch North Sea and expert judgement of several experts who could add species to be considered. Background information on the species selection ('smart species') is described in Wijnhoven et al. (2013). The initial selections did also contain the 'typical species' as identified in the Netherlands for Habitat Directive Annex I habitats H1170 (reefs) and H1110 (Permanently submersed sandbanks: with Dutch subtypes 'H1110a', 'H1110b' and 'H1110c', where subtype 'a' is only present in a small area in the Dutch North Sea, with little monitoring, and predominantly situated outside the North Sea; i.e. in the Wadden Sea). Species selections made in 2013 consisted of species identified in Wijnhoven et al. (2013) with an indicator score of 1 or more for the formula shown there (except for those species that were hardly detectable with available monitoring techniques in the monitoring programme), supplemented with potential indicator species that were (consistently) abundant in the past, but are scarce now.

[4] Optional also other ecotope classification systems can be used, especially when EUNIS classification is not in use for the area: E.g. for Dutch transitional waters of the Water Framework Directive, ecotopes are described according to ZES.1; Bouma et al., 2005).

With the introduction of BISI v2, the selection of indicator species is still based on evidence from literature. It is suggested to work towards a database of potential indicator species for BISI with indication of relevant indicator values, where possible adapt information from widely accepted sources like WoRMS (Marine Species Traits editorial board, 2018), BIOTIC (MarLIN, 2006) and AMBI (Borja et al., 2000). Indicator species lists (and reference values) are ecotope specific; for each new area to be evaluated, references can be constructed from earlier references when identical or comparable ecotopes have been part of evaluations before. However, improving data availability might influence reference species lists and/or reference levels due to new insights.

To achieve better standardization (essential for comparison and application in a variety of European habitats and regions), the selection of indicator species is a three step procedure:

1. Selection of ecotopes of relevance

Selection of those ecotopes (EUNIS level 4 habitats) that represent (or under desired conditions[5] represent) at least 10 % of the surface of the area of evaluation[6]

2. Selection of potential indicator species per ecotope

Species are selected as potential indicator species when identified as being:

a. Either characteristic for the ecotope (common or abundant in the ecotope of concern under natural good quality conditions, whereas absent or rare in most other ecotopes).

b. Or indicative for one of the dominant disturbances that might be present in the area of evaluation (E.g. in the North Sea situation these are generally ‘seafloor disturbance’ and ‘ecological disturbance’, but it might be disturbances like increased turbidity, hydrodynamics or inundation time due to human activities in other cases).

---

[5] Especially in man-induced confined and artificial systems that are highly out of balance, the preferential ecotope surface distribution to achieve a good quality status might be different from the current situation.
[6] The monitoring should be representative for the area of evaluation: e.g. random sampling scheme, number of samples per ecotope comparable to surface area; it is possible to focus the evaluation and/or monitoring on one or a selection of ecotopes, but in that case the results do not necessarily reflect the quality status of the entire area. It is also possible to include ecotopes that are represented by a lower surface area, if those ecotopes are relevant for quality assessments as the impact of pressures might result in ecotope changes; the monitoring, including those ecotopes, should however be representative for the entire area of concern/evaluation.

c. Or scoring at least 1.5 out of 3 for the estimated species specific indicator values (iv_i); potentially growing large (1), becoming old (2) and having frequent recruits (3) (see Table 7).

And have the opportunity to return (is at least present somewhere in the vicinity).

3. Selection of indicator species with sufficient power to detect possible differences in occurrence in the area of evaluation

Selection of species for which the power to detect significant ($p<0.05$) changes of 50 % in spatial occurrence equals 0.8 with maximum 60 samples in one of the preferred habitats (ecotopes for which species are selected as potential indicator species). In practice the power to detect differences in occurrence will likely be higher in preferred than in suboptimal habitats. In practice the power is (only) found to be insufficient for species predominantly related to habitats not monitored (e.g. hard substrates is case of soft sediment monitoring) or for which applied monitoring techniques are unsuitable (e.g. monitoring of species present in low densities under natural good quality conditions with boxcores).

As the catchability of HD 'typical species' (Min EZ, 2014a,b) is not necessarily good with moderate monitoring efforts and monitoring techniques in place, compared to BISI v1, typical species are not automatically included in the selections for general evaluation at BISI v2. 'Typical species' (in the Netherlands) or other species with official status (in other countries) can always be part of specific evaluations.

A difference in constructing indicator species lists for compiled areas between BISI v1 and BISI v2, is that in case of BISI v1, indicator species were initially selected for specific areas (in the Dutch case: 'Areas with special ecological values'; ASEVs, like Natura2000 areas) from which indicator species lists for ecotopes and habitats were derived; whereas in case of BISI v2, the ecotopes are leading.

As indicated, a BISI consists of an overall quality evaluation (general quality index) for which indicator species selection is standardized as described before to make evaluations for different areas to a certain extent comparable. A general quality evaluation comes with a series of specific quality evaluations (specific indices) to detect possible causes and/or effects of observed quality levels and/or quality developments. Basically the specific evaluations consist of the same indicator species selections as the general BISI, selecting those species relevant (with an indicator value larger than zero ($iv_i > 0$) for the specific evaluation. It is allowed to add additional species to the specific evaluations (evaluations are therefore less standardised and less comparable between different areas than the general BISI-scores), to enlarge the number of indicator species in the specific evaluation. These are likely indicator species not considered general quality indicators that might however be indicative for certain very specific functions and/or pressures and certain species selections might be of specific interest as they have a certain status in management. Also the power to detect differences at individual species level might be less than defined for selection in the general quality analysis. The monitoring programme might however be sufficient extensive to detect differences (the power of BISI is expected to be larger than evaluation based on individual species, dependent of whether two monitoring moments are compared, a sitation is compared to an existing trend, or if trends for different periods are compared), or one can accept a lower power for such cases. Power analyses should give insight in the power of the tests regarding specific target evalutions and in which timeframe significant results might be expected if present, after which an option might be to change sampling efforts, frequency and/or design.

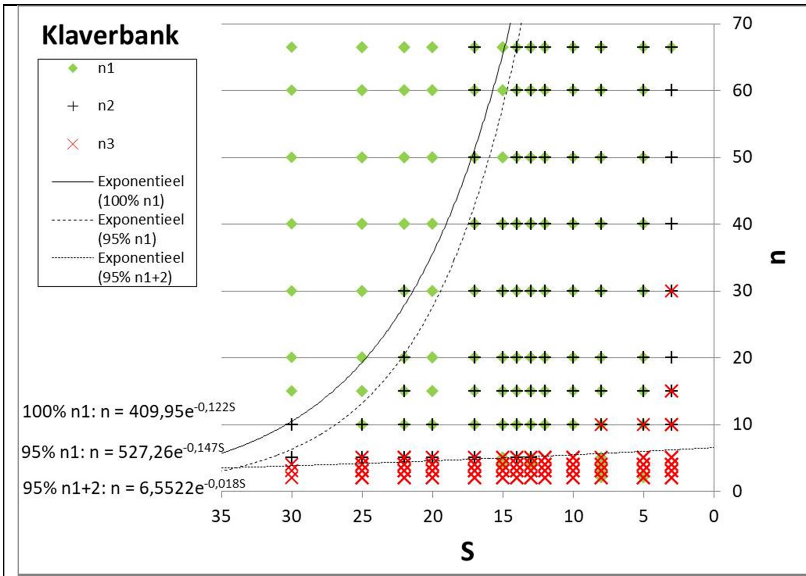
Table 7. Overview of the various BISI evaluations and recommendation for the species specific indicator values (iv). Besides the general quality evaluation, these include specific evaluations to identify potential causes for the observed quality status and potential effects of the observed quality status on ecosystem functions. Categories of specific evaluations also equal the criteria that were considered for indicator species selection at which potential indicator species should score well (Scoring an iv_i of 1 for A or B or another more important pressure in the area of concern, or an iv_i of 1 for F, or a score of at least 1,5 for the combined categories C+D+E , for BISI v2).

Code	Causes and effects (to be evaluated)	Description	Species specific indicator value (iv_i)
General quality		Selected indicator species according to the three-step procedure described before: Potential indicator species for relevant ecotope; either characteristic, indicative for a dominant disturbance or potential sufficient large, old and/or having frequent recruits and present in vicinity; specific enough towards quality status so that the power of testing is sufficient good.	1 (by definition)
A.	Sea floor disturbance	Combined indicator value for a variety of disturbances (different types, intensity and/or frequency).	3 levels (0, 0.5, 1)
B.	Ecological disturbance	Combined indicator value for effects of nutrients, pollutants and toxicants, hypoxia and temperature increases.	3 levels (0, 0.5, 1)
C.	Intensity of sea floor disturbing fisheries	Indicator value on basis of size of species (where large species can be damaged or fished away at low intensity of sea floor disturbing fisheries and smaller size classes only at high intensity of sea floor disturbing fisheries).	4 levels (0.25, 0.5, 0.75, 1)
D.	Frequency of sea floor disturbing fisher-	Indicator value on basis of age of species (species that get older are already impacted at low a frequency of sea floor disturbing fisheries, whereas species that live	10 levels (age divided by 10,

	ies	shorter are likely only impacted by frequent occurring sea floor disturbing fisheries).	value of 1 at age >10)
E.	Recovery	Indicator value on basis of frequent recruits (Species with frequent recruits are good indicators for the first phases of recovery).	4 levels (0, 0.1, 0.5, 1)
F.	Characteristic species	Species are almost exclusive or are much more abundant in the area of evaluation than elsewhere (identification of being characteristic at ecotope level is a criterion for indicator species selection).	3 levels (0, 0.5, 1)
G.	Food web structure	Species important as food sources for higher trophic levels (i.e. fish, birds, marine mammals).	3 levels (0, 0.5, 1)
H.	Habitat diversity	Species creating permanent structures providing niches for a range of additional species.	3 levels (0, 0.5, 1)
I.	Biological activation of sea floor top layer	Bioturbating and bioirrigating species with an important role towards ecological functioning (e.g. nutrient cycling, degradation of pollutants, providing suitable habitat for other species).	2 levels (0, 1)
J.	Habitat Directive typical species	Species designated as typical species are identified as important for either biotic or abiotic processes, can be characteristic or exclusive for specific habitats (species lists of Habitat Directive habitat types) if relevant for area to be evaluated.	2 levels (0, 1)

Ideally, the power of specific evaluations is calculated and presented with the results. The identified specific evaluations for BISI v1, partly in use as criteria for indicator species selection for the general quality evaluations as well, are presented with their suggested levels of scoring in Table 7. It is likely that with the development of BISIs for new areas, a need for additional specific evaluations (e.g. additional pressures of importance) arises. Those can easily be constructed and defined in a similar way based on the same (sub)set of data. Other specific evaluations presented here, are possibly less relevant, and can be omitted, except for specific evaluations C, D, E and F that have a role in indicator species selection for the general quality evaluation, and A and/or B that should than be replaced by the most important pressure(s) for the area of concern.

Another specific evaluation of relevance for certain areas could be the evaluation of indicator species characteristic for a certain preferred ecotope compared to species characteristic for alternative ecotopes (in areas where changes are expected or desirable).



n=409,95e$^{-0,1225}$	
S	n
1	362,9
2	321,2
3	284,3
4	251,6
5	222,7
6	197,2
7	174,5
8	154,5
9	136,7
10	121,0
11	107,1
12	94,8
13	83,9
14	74,3
15	65,8
16	58,2
17	51,5
18	45,6
19	40,4
20	35,7
21	31,6
22	28,0
23	24,8
24	21,9
25	19,4
26	17,2
27	15,2
28	13,5
29	11,9
30	10,5
31	9,3
32	8,3
33	7,3
34	6,5
35	5,7

Figure 3. Example of consequences of reduction of number of indicator species and/or number of samples on reliability (variability in BISI results) and power of tests. Testing consists of calculations of BISIs on basis of random selections indicator species (5 random selections per number of species) and random selections of observation data (5 random selections per number of samples) in each combination. As the realized number of samples, the average number of samples per indicator species for each of the constructed indicator speciies selections is calculated, as the (available) number of samples differs between the monitoring techniques. The testset consists of 121 video and 43 Hamon samples for the (Area of special ecological value, also Habitat Directive area) Klaverbank as collected in 2015. This is considered a good test set due to the large data availability and the (large) number of indicator species identified/selected for evaluation of this area (BISI v1). Evaluating the reduction analyses, a deviation of less than 20% of the BISI-value as calculated on basis of the the entire dataset including all 30 indicator species is considered as accurate enough. As the calculated BISI (v1 including all 30 indicator species) on basis of the entire dataset equaled 0.873; values in the range of 0.762-1.001 were considered acceptable. Additionally a power (β-1) of >0.8 to detect at least differences of 50% (which means that that delta can be calculated as (0.5*BISI)/stdev) in BISI results (with significance level = 0.05 for a two-sample t-test) was considered sufficient.
The graph shows whether or not calculated combinations of numbers of indicator species and average numbers of samples fulfill qualifications as indicated with n1 (= accuracy and power sufficient), n2 (only power sufficient) and n3 (accuracy and power insufficient). Exponential equations indicating the borders of combinations of numbers of indicator species and samples are calculated and indicated in the graph so that in 99% of the cases the accuracy and power is sufficient, so that in 95% of the cases the accuracy and power is sufficient, or that in 95% of the cases at least the power is sufficient. The shown series of numbers at the right indicates the minimum average number of samples (n) that should be taken given a number of indicator species (S) included in a BISI so that in 99% of the cases the accuracy and power of the BISI is considered sufficient as calculated for the Klaverbank and expected to be representative for other sea regions as well.

Basically such results will become visible in case areas are in change (often as a result of human related activities like large scale constructions; e.g. wind-farms, artificial islands, sand extractions and suppletions, dikes, embank-ments and construction extensions into the sea) or large-scale developments (sealevel rise or changed currents due to climate change or seabed lowering due to gas extraction), as BISI methodology is based on a static ecotope map. When large changes in the seabed habitat constitution are expected; changes in BISI scores can be compared to ecotope-difference maps (comparing eco-tope surface area changes in time).

With the construction of BISIv1, it was suggested that indicative a specific BISI should consist of at least 5 indicator species for a reliable outcome. It is however found that the reliability and power of the BISI increases drastically, especially at low indicator species numbers with the addition of species to the index. Where with an index consisting of only 5 indicator species the number of samples (average per indicator species that could deviate with monitoring technique) should indicative be at least 223 samples for a reliable quality as-sessment with sufficient power to detect differences; less than half of the samples is needed with 10 indicator species. With 20 to 25 indicator species a realistic number for general quality evaluations, indicative only 36 to 20 sam-ples are needed, which equals the amount to be taken when only indicator species inventoried with one monitoring technique are selected. In relation to specific quality assessments, for which generally not all selected species are indicative, this however means more samples might be necessary. With the design of the monitoring programme (for each new area) one should on be-forehand determine the targets of evaluation. For a general quality evaluation a limited number of samples can already be suffient, whereas when identifi-cation of the potential role of certain pressures is essential, a larger number of samples might be requisited. However, also the time-frame in which results are needed, can influence the necessary number of samples, and can be a way to reduce sample numbers (per year). One should however always determine the realized power of constructed BISI's for new areas to be evaluated, to identify the necessary number of samples in line with the current situation (quality status), as the power will deviate with indicator species relative oc-currences.

6.7.2 Indicator values (tuning BISI for the evaluation of specific aspects)

As indicated in Table 7, for each of the areas to be evaluated, an evaluation can consist of a general quality assessment and several specific evaluations. Specific evaluations will identify the relative importance of potential causes for the observed quality status and potential effects of the observed quality status on ecosystem functions. Where all classified indicator species are part of the general evaluation with the same weight (indicator value (iv_i) of 1 in the calculation of Individual Indicator Species (IIS) values), in the specific evaluations indicator species have different weights. The iv_i is a value between 0 and 1, with 0 = no indicator value at all (species not included in the specific evaluation) and 1 = good indicator. Most criteria (causes and functions to be analysed in specific evaluations) have scores between 0 and 1 at several levels as indicated in Table 7. For a few criteria, only a score of 0 or 1 is possible: species are either characteristic for the evaluated area/habitat or not. Similarly species are either or not selected as typical species for certain Habitat Directive habitat types (iv_i is either 0 or 1).

In calculation of the BISI, the species specific indicator value (iv_i) is devided by the average indicator value for all indicator species in the specific evaluation (iv_{avg}), with the resulting Indicator Value indicated as IV_i. By compensation for the average indicator value, the resulting BISI-scores are adjusted to the same scale, making BISI-scores directly comparable. (In case of specific characterizations with iv_is only being 0 or 1, $iv_i = IV_i$)

In this way, each selected species can be scored for each criterion and a matrix of the scorings of species per criterion, for each area to be evaluated can be made.

V1 - The outcomes of the scorings for each of the areas to be evaluated of the Dutch North Sea according to BISI v1 are listed in Appendix 1 of Wijnhoven & Bos (2017).

The scoring of individual indicator values for potential indicator species according to the methodology BISI v2 will not be different from BISI v1. There might however be some differences in the selected species as power and the ecotope surface constitution play a prominent role now. An example of indicator species for the ASEV Cleaver Bank with different indicator values for different specific evaluations is also given in Table 3.

There are several sources that could be used to extract species specific indicator values (that might be ecotope specific as well including platforms like WoRMS (Marine Species Traits editorial board, 2018), BIOTIC (MarLIN, 2006) and AMBI (Borja et al., 2000), but also from several initiatives within ICES and OSPAR and species specific literature. As there is and will always be some debate about certain valuations, and for several species information is lacking or only available from grey literature, it is of utmost importance that used values are centrally stored including references, so that indicator species lists and valuations can potentially be reused or if necessary adjusted.

6.7.3 Selection of samples and monitoring techniques

Dependent of size and shape of specimens with age and natural densities and distribution patterns, monitoring techniques are more - or less – suitable or not suitable at all, for qualitative or quantitative observation of species.

V1 - In the Dutch North Sea, the monitoring is (and used to be) largely based on boxcore sampling and monitoring with a benthic dredge. Although there are differences in specifics of used devices and/or applications in time, boxcore samples generally cover a small surface area (at present 0.078 m^2 in Dutch North Sea monitoring) and are sieved over 1mm mesh size. Benthic dredge samples cover larger surface areas (indicative 20 m^2; 15 m^2 in coastal zone) have become the standard for the current evaluation methodology, but in practice sampling areas vary from several to over 200 m^2. Samples are sieved usually over 7x7 mm mesh size (5x5 mm in coastal zone) but also larger mesh sizes of over 1 cm have been used. Other aspects that might deviate are: (1) Approximate sample depth (dependent of sediment type about 20 cm for boxcore, for the dredge this often has been less (including during the T0 in the Dutch North Sea) and is standard about 7-9 cm in the coastal zone), (2) potential use of subsampling with specimens identifications and (3) which specimens are identified to species level (choices towards handling small specimens, juveniles, damaged specimens, etcetera).

All kind of methodologies and characterisations might potentially be of use, but might ask for additional calculations or adjustments of the methodology.

V1 - In the Dutch North Sea at selected (fixed) sites in the coastal zone the Van Veen grab (3x0.1 m^2) or the Suction dredge (indicative 30 m^2) are used. In the area of the Cleaver Bank especially the Hamon grab (at present with a surface area of 0.09 m^2, but before other sizes have been used) is used in combination with video recordings. Hamon grabs might be more or less comparable to boxcores in the specimens that are collected. Video tracks might be more comparable to dredge sampling in a way that especially larger sized species occurring in low densities can be inventoried. A large difference is however that infauna is largely missed, and for several species (groups) identification to species level is difficult (solely from images).

Monitoring methodology: Wijnhoven et al. (2013) give for several potential indicator species the most suitable (and efficient in terms of efforts) monitoring methodology. The study forms the basis of the current monitoring programme in the Dutch North Sea. Several species can only be inventoried with one of the applied techniques, but other species can be found in each of the two devices used per area. Then it has to be decided which data are most suitable, based on which size-classes are predominantly collected and what does that tell, and how representative are the collected specimens expected to be for the populations living there. As an example, larger sized bivalves can potentially be monitored with boxcores instead of dredges when their natural abundances are not too low, but it might be especially smaller size-classes or juveniles (not part of the dredge samples due to the larger mesh size) that make up the abundances. Although for some species there are options to combine observation data with different techniques for specific purposes (e.g. Wijsman et al., 2013), it is decided in the Dutch North Sea case to always restrict to either boxcore and Hamon grab (combined in a few cases) or either dredge sampling or video recordings (where in the coastal zone Van Veen grabs and suction dredge data are combined with dredge sampling data in a few cases where it concerns fixed monitoring programmes).

For new areas, in practice all kind of other techniques can be potentially suitable as long as techniques for determination of reference and current monitoring techniques are comparable or can be calculated to each other without a resulting uncertainty becoming more important than the observation data. Monitoring techniques will determine the suitability, and especially the detectability in terms of power of tests, of potential indicator species.

6.7.4 Random sampling vs stratified sampling

It has to be taken into account that sample sites might be 'selected' randomly (or expected to give a representative view of a total defined area), or positioned in stratified way or focussed on certain habitats. One can potentially work with both types especially when, like in the case of the Dutch North Sea, sampling is according to a fixed protocol and scheme repeated periodically, but both types of samples cannot be combined and averaged without taking the areas they represent into account.

BISI v2 specifically works from the ecotope level, optionally to composite areas containing several ecotopes. This makes the evaluation methodology highly suitable for evaluation of random stratified monitoring data at the ecotope level, or a random sampling scheme (as long as the number of samples for the most important ecotopes is sufficient). In case of a lack of data, there might be options to combine information from several campaigns, or only evaluate the quality (changes) for most important ecotopes as an indication for the quality status of a certain area.

6.7.5 Selection of data

Basically all kinds of observation data (e.g. densities, weights, presence-absence data, size-frequency distributions, etc.) can be used and might eventually be combined in an indicator (e.g. Van Strien et al., 2016). Leading in this is data availability, accuracy resulting from used techniques, expected (natural) fluctuations, and specificity towards disturbances.

V1 - In case of the Dutch North Sea it is decided to focus on the use of densities (or calculated numbers per standard surface area) and spatial probability of presence in samples (hit ratio for a given set of samples). Evaluation of hit ratios might be more robust for species expected to occur on average in low densities, but when present often occur in local aggregations. In the methodology for the Dutch North Sea, hit ratios are only used based on boxcore - or Hamon grab data as these have fixed sampled surface areas and recalculating hit ratios to standardized surface areas comes with a lot of uncertainty as then natural spatial distribution determines which extrapolation methodology suits best.

6.7.6 Reference levels

There are pros and cons for using reference levels in evaluations. A pristine (pre-industrial) reference level reflecting the situation before anthropogenic disturbances will reflect an optimum quality situation. Even if we reach (or reconstruct) pre-industrial environmental conditions, it is unlikely that the original community will return, even in the long term. Taking away the pressures will likely not lead to recovery of pristine benthic communities, as habitat characteristics and present species pools have changed dramatically, which makes that natural development under high environmental quality conditions will lead to alternative stable states. Therefore, a methodology is needed which is potentially capable of showing improvements on the short and mid-long term, so that effects of management regulations can be detected within management cycles. Unrealistic reference levels would mask changes (if there are) as relative differences between reference levels and observations would be minimal.

A benefit of working with reference levels is however, that changes are put in perspective. As an example: Doubling of the observed numbers for one indicator species might be much more important than for the other, as such an increase can mean that the abundance of the indicator species is still at a poor quality level or comes at a level that no further quality improvements are expected/necessary.

The flow chart of Figure 4 indicates how reference values can be obtained, dependent of the presence of suitable reference areas, historic data availability, the area monitored, the monitoring methodology used, the within data variability, and how recent observations (year of evaluation) compare to recent historic observations.

The use of data from reference areas is an addition to the methodology in BISI v2. As the derivation of reference values and the construction of an internal reference for areas to be evaluated is ecotope based, in practice data from corresponding ecotopes are searched for in reference areas, after which a suitable reference might be constructed surface-area based. Such references should be from a comparable region, i.e. preferably part of the same regional species pool but at least from the same geographical zone, as explained before, potential indicator species that cannot return to the area of evaluation in a natural way, are excluded from the internal reference.

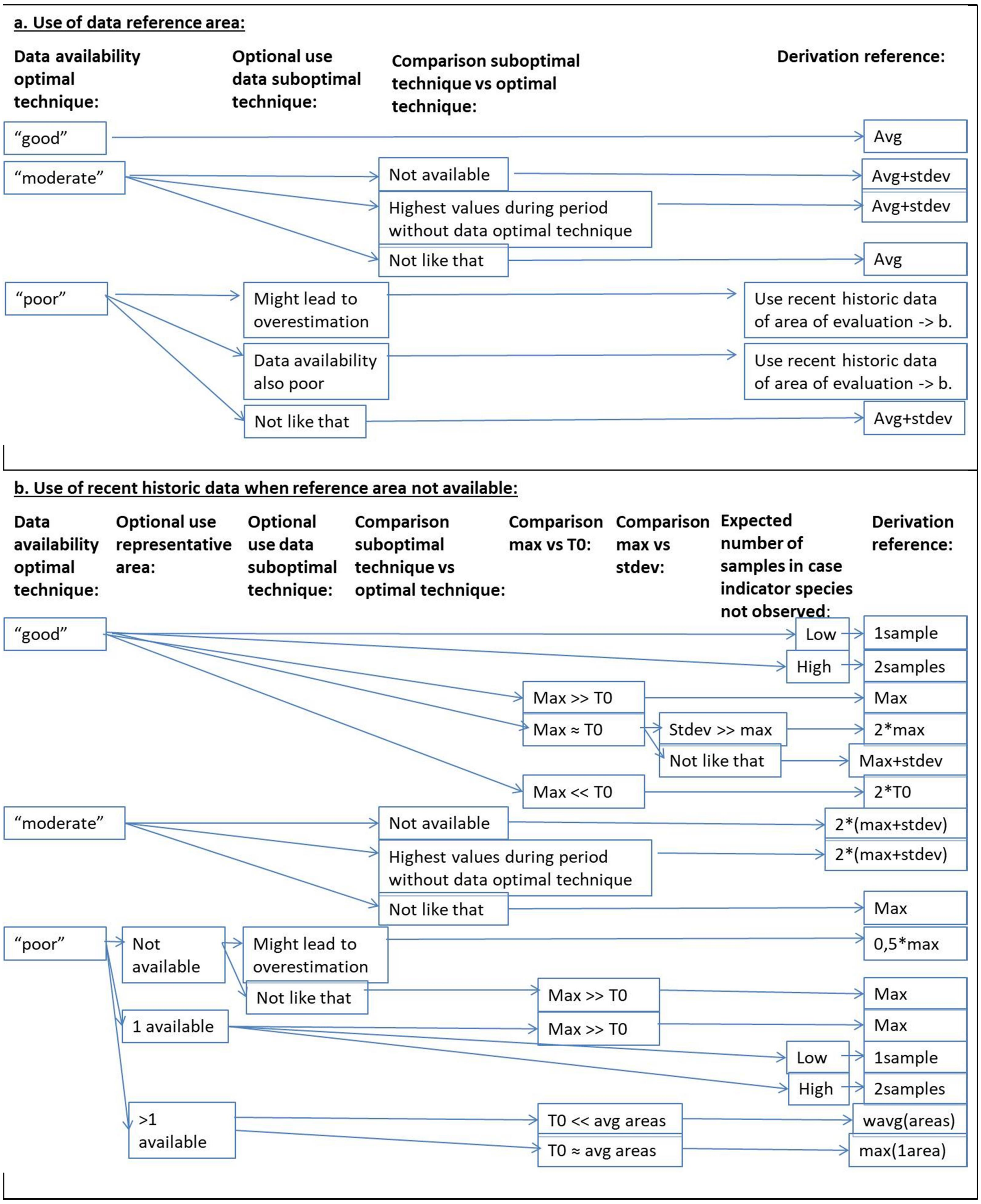


Figure 4. Flow chart showing the decision schemes of how to derive indicator species and area specific reference values. a. Scheme to use in case a reference area is available; reference values can potentially be obtained from monitoring data of a reference area. b. Scheme to use in case no suitable reference area is available, or if monitoring data for the reference area are too poor. Avg = representative year average value in this case for a reference area; Max = maximum year average value in this case in recent historic data; stdev = standard deviation; T0 = year average value as observed in the focal year for which an evaluation methodology is developed and to which future evaluations will be compared; 1sample = a density similar to an occurrence in one sample; wavg = weighted average of several areas based on the number of samples taken per area; >> = much larger; << = much smaller; ≈ = comparable values.

Therefore reference levels are either obtained from recent data of reference areas for which the possible presence of significant pressures can be ruled out, or are based on recent maximum observations of year averages that are adapted, doubled and/or increased with the standard deviation as observed from recent historic monitoring data of the area of evaluation itself. Derivation methodology depends on the historic data availability. In case of poor historic data availability, there are no better options than using maximum observations from suboptimal techniques, in which case a reference of half the observed maximum is sometimes used. There are cases that indicator species have not been present in the monitoring data of the area of evaluation (as densities were too low) during recent years. In such cases a presence of 1 specimen per square meter in only 1 or 2 of the samples is used as a reference (dependent of the expected sample intensity during monitoring the coming years).

Although there is some expert judgement involved, the derivation of the reference values is standardized as much as possible.

V1 - All reference values and derivation methodologies used per species in case of the North Sea evaluations (according to BISI v1) are indicated in the BISI Assessment Tool (Appendix 2 of Wijnhoven & Bos (2017)). Taking the current quality state and pressures into account, it is expected that the current methodology, and reference levels used, are at least suitable for detecting quality status improvements in the North Sea areas according to a very ambitious scenario the next decades without adjustment of reference values, whereas the methodology is focused on evaluating the relative importance of changes in the order of magnitude that are most likely.

It is expected that the foreseen application of BISI outside the Dutch North Sea in a larger OSPAR context and the possible identification and application to suitable reference areas can further validate the most appropriate reference occurrences and derivation methodologies in case only recent historic data are available from the area of evaluation itself. This will also be valuable towards the setting of quality targets for management purposes and MSFD- and/or Natura 2000 related evaluations. On the longer term the methodology will take full advantage from broader application, as finally reference levels for various indicator species related to monitoring techniques, will be available for a range of ecotopes which makes it easy to construct a reference surface-area based for each new area making use of available information.

6.7.7 Boundaries of the methodology and the power of the tests

The BISI methodology allows to base evaluations on variable monitoring efforts as relative differences in abundances are evaluated taking variance in the observation data into account. This is however not an encouragement to reduce the monitoring efforts, as with a reduction of the number of samples also the number of indicator species for which potential differences can be found will be reduced, leading to a very low power of the indicator (observed variability in indicator results will generally be non-significant). This loss of power of the BISI, and the effect of sample reduction and the relation with the number of indicator species that should be part of the index, is shown in Figure 3. With regards to special evaluations, there are options to add additional indicator species although these species do not reach all selection criteria for selection in de general BISI/quality assessment. This can improve the power of these specific tests, but identification of the realized power is always of importance. (Those species should not be added to the general BISI, as that will reduce the comparability with other areas).

V1 - The current monitoring programme for the Dutch North Sea (BISI v1) is based on power analyses to allow detecting at least 50% differences in abundances and/or hit ratio for a selection of the indicator species in the areas of evaluation between the T0 and an individual year of evaluation. These calculations have however been rather conservative (compared to the current evaluation methodology) as they considered an independent monitoring protocol. The evaluations for the Dutch North Sea region are however generally according to a paired- or even a Before-After-Control-Impact-design with more power. Additionally initial power estimations were often based on the variance observed from a limited number of observations (a larger sample size will generally lower the observed variance). Although 50% sounds as a huge difference, this is actually not that much considering that it is not an entire community but a set of specifically selected indicator species that are often present in low abundances in the current situation. Besides, the realized power is probably sufficient to detect smaller differences in most cases.

With BISI v2, selecting indicator species with expected sufficient power (i.e. 'smart species'), is part of the standardized methodology as described before. Potential indicator species are only selected when the calculated power, based on available recent historic and/or reference data, is sufficient (equals 80%) to significantly ($p<0.05$) detect 50% differences in spatial occurrence (presence-absence) with a maximum of 60 samples. Although this is still quite a number, herewith it is ensured that species for which monitoring techniques are inappropriate or that are predominantly related to unsampled habitats, are not part of the BISI. The realized power of the BISI can be evaluated for each of the areas for which the indicator is developed, although results will always to some extent relate to the monitoring efforts so far. These can however improve (in the future) in case calculated required effort is guaranteed in the installed monitoring programme to collect the data for quality status evaluations. Species not reaching the quality criteria (insufficient power) as described before, are not part of the general quality evaluation, but can be part of a specific evaluation. It is recommended to calculate the realized power for specific evaluations as well.

When a certain specific evaluation is considered of importance but not enough species with sufficient power could be selected, it is suggested to add potential indicator species (those not selected for the geberal BISI as they did not meet the criteria) up to a total number that together with the number of samples approaches a power of 0.8 for the specific evaluation, or at least allows a powerfull test on the longer term. In this way, at least an indication of for instance the development of the populations of typical species of the Habitats Directive can be obtained, although those species are not necessarily 'smart' species.

As the evaluation tool should be as good in the detection of possible improvement as in the detection of impairment of the quality status, the calculations make use of the geometric mean (e.g. Buckland et al., 2011; Van Strien et al., 2016) of observed - to reference abundance ratios, by ln-transformation of the individual ratios and taking the inverse of the average weighted result (e.g. Van Strien et al., 2012). To reduce the potential impact of species that become far more abundant or scarce than the reference abundance, and to handle zero-values (indicator species that are not observed) changes relative to the reference are truncated at a factor 100 (e.g. Ten Brink et al., 2000). This means that observed – to reference abundance ratios always have a value within the 0.01-100 range. Similarly, the minimum standard deviation (as the value of variance considered to calculate statistics) is truncated at a value of 0.01 (to handle possible occurrence of zero-values).

Ideally there are standard protocols for sieving, (possible subsampling), sorting and identifying specimens to species level.

V1 - In case of the Dutch North Sea this for instance includes standard rules of how to treat damaged, small and/or juvenile specimens (which might differ between the different sampling programmes). According to this standardisation, only specimens identified to the species level are considered (unless indicated else as is for instance the case with *Urticina* sp. and Porifera). This amongst others also makes that data collected with different techniques cannot directly exchanged and evaluated with reference levels belonging to other sampling methodologies.

Regarding the Dutch North Sea, sampling and laboratory procedures are described in Perdon et al. (2016) and references therein for the WOT mollusc survey, in RWSV (2017) and Leewis et al. (2017) for the boxcore sampling and sample treatment and Schellekens and Faasse (2015) for the dredge monitoring and related sample treatment (as conducted in 2015), and are abstracted in Wijnhoven (2018) as well. Methodology and procedures of monitoring in the area of the Cleaver Bank are expected to come available soon (Cuperus personal communication). Laboratory procedures are according to Cuperus & Swarte (2016).

6.7.8 Conceptual testing of the methodology

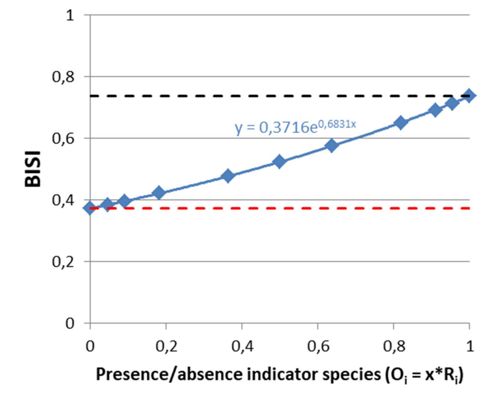
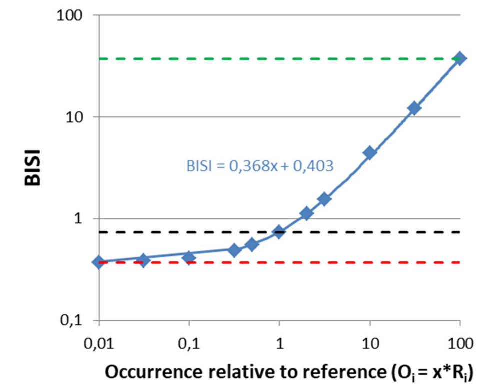
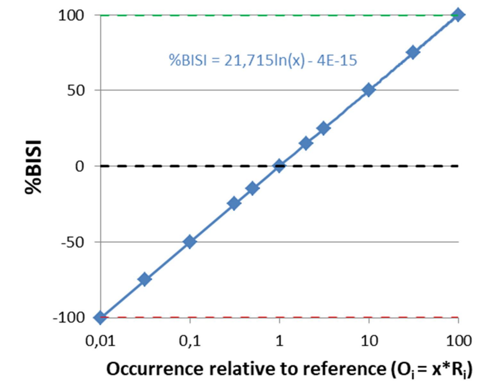


Figure 5. Response curves of BISI. a. Response of BISI-score when presence/absence of indicator species changes. The x-axe indicates the share of indicator species present relative to the number of species forming the internal reference. (An occurrence at reference levels is suggested when an indicator species is present). The BISI increases from the minimum score of 0.372 when no indicator species are observed to the reference level with a BISI-score of 0.736, when all indicator species are observed.





b. Response of BISI-score in case all indicator species are present. The graph shows an increase of all indicator species (in similar amounts relative to the reference) from an occurrence at a level 100 times lower than the reference occurrence (Oi = 0.01*Ri) via the reference occurrence for each of the indicator species (Oi = Ri) to an occurrence for all indicator species of 100 times the reference levels (Oi = 100*Ri). c. Response of BISI on a percentual scale (%BISI) relative to the reference level showing the same graph as in b. with an increase of all indicator species (in similar amounts relative to the reference). Indicated levels of the minimum BISI-score (= 0.372), reference BISI-score (= 0.736) and maximum BISI-score (= 37.156) are indicated with dotted lines in the colours red, black and green respectively. Shown trends are close approximations of the response curves slightly subjected to rounding differences.

Evaluation on basis of BISI leads to a quality score (BISI value) deviating between 0.372 and 37.156 as minimum and maximum values representing 100 times lower and 100 times higher than the internal reference occurrence of the indicator species. The index (consisting of a set of selected indicator species) responds however to the absence of indicator species, and the relative occurrence of those indicator species present.

Basically, an occurrence 100 times lower than the reference occurrence of an indicator species is equal to being absent, whereas species transgressing the maximum occurrence of 100 times the reference get the maximum score similar to 100 times the reference.

To clarify the response of BISI to differences in observations the presence and absence of species, and the impact on the BISI score, is shown separately (Fig. 5a) from the effect of changes in occurrences of species (Fig. 5b. & 5c.). In practice the BISI-score will be the result of the combination of both aspects deviating, where species occurrences differ independently. Additionally, differences in indicator values of individual species might result in a slightly larger impact of differences in occurrence of the one species on the resulting BISI score than from the other.

The graphs indicate that the presence or return of each additional species is of about similar importance (near lineair response in graph 5a) with a slightly larger impact of the last indicator species to return. It should be taken into account that initially those returning species are likely present in lower abundances. Those species are of increasing importance (towards the quality rating) when abundances approach reference levels. Indicator species transgressing set reference abundances still have an (additional) positive impact on the quality score. It is expected that the risk of indicator species becomming nuisance species is small; otherwise they were not selected as indicator species in the first place, so that their increase indeed might be a sign of quality improvement. When however such an increase in abundances might have ecological negative impacts it is expected that this will be reflected in the occurrence of several other indicator species, in the end leading to a lower BISI score.

6.8 Assessment units

V1 - In case of BISI v1, three types of units are present which were (Wijnhoven, 2018) and will be evaluated with the current methodology covering the entire Dutch part of the North Sea in the future.

The six EUNIS level 4 ecotopes covering the entire Dutch North Sea; 'Offshore circalittoral coarse sediment' (A5.15: Formerly indicated as 'deep coarse sediment'), 'Offshore circalittoral sand' (A5.27: Formerly indicated as 'Deep sand'), 'Offshore circalittoral mud' (A5.37: Formerly indicated as 'Deep mud'), 'Circalittoral coarse sediment' (A5.14: Formerly indicated as 'Shallow to moderate deep coarse sediment'), 'Circalittoral sand'(A5.25 and A.26 Formerly indicated as 'Shallow to moderate deep sand'), 'Circalittoral mud' (A5.35 and A5.36: Formerly indicated as 'Shallow to moderate deep mud'), are evaluated at the national level.

With BISI v2, ecotopes (EUNIS level 4: EMODnet, 2016) form the bases of the methodology on which level internal references are constructed, and from which BISIs for other types of areas are compiled surface area based. Preferentially, the same level of classification of ecotopes is used in other regions as in the Dutch North Sea so that eventually evaluations (and BISI development in particular) can benefit from earlier work and results will be more comparable. However, in principle there are no limitations to the aggregation level of ecotopes for the construction of BISIs, and can BISIs be developed for rather specific habitats or areas. By using reference data for ecotopes from other areas, and different geographic parts in particular, one should take into account possible differences in species pools that might include potential indicator species without options to recolonize certain areas in a nature way (in which case they should be excluded from the internal reference).

Separate evaluations of (developments in) the quality status of individual areas of special ecological value (that are often also Natura 2000 sites) are performed based on BISI v1. There will be separate evaluations comparing the quality status developments of for specific fisheries restricted areas ('closed' areas) with similar sample sites (in number and expected presence of specific habitat and/or benthic communities before management measures were taken) positioned in areas without fisheries restrictions (open areas). This involves subareas of the ASEVs, where it depends on the variability in (initial) communities between different 'closed' areas within ASEVs whether those will be evaluated together or separately. The ASEVs are the Dogger Bank, Cleaver Bank, , Frisian Front, North Sea Coastal Zone, Front Delta and Plain of the Raan (all Natura 2000 areas with (planned) fisheries restrictions) and the Central Oystergrounds (with planned fisheries restrictions as well) and the Brown Bank.

Providing background information to the Article 17 reporting, the areas indicated as being part of Habitat Directive habitat types H1170 'Reefs', H1110b 'Submersed sandbanks in the coastal zone', H1110c 'Offshore submersed sandbanks' are evaluated separately. In case of the H1110 subtypes evaluation is in two parts as the two parts have quite different species assemblages and reference levels.

BISIs for both; specific areas and HD habitat types will be constructed from BISIs for ecotopes (EUNIS level 4) surface area based, with BISI v2.

6.9 Geographic coverage

BISI v2 is specifically developed, by some adjustments of BISI v1, for generic application, including the construction of area specific internal references, in any sea region and type of area. The above mentioned types of areas might be important examples for application in other countries and regions as well, but are examples, without restrictions for application in other types of areas. At the moment development and application of BISIs (according to v2) for HD habitat types as present in Dutch coastal and transitional waters takes place. Test cases within the broad OSPAR region in cooperation with partner countries are foreseen.

BISI v1 was initially developed for application in the Dutch part of the North Sea to evaluate:

- Areas with special ecological values (ASEVs) under the MSFD,
- Natura 2000 areas,
- Habitat Directive Annex I habitats,
- Ecotopes relevant for the MSFD (i.e. EUNIS level 4),
- Areas with fishery management regulations

6.10 Temporal coverage

V1 - For the Dutch North Sea areas (BISI v1) internal reference levels used, are based on 1984-2014 monitoring data (with an exception of some older data for the Plain of the Raan). First application is on the monitoring data of 2015, which equals the T0 situation of before most of the management measures to improve the quality status in the Dutch North Sea have been implemented (Wijnhoven, 2018). The monitoring of 2015 will be repeated every three years. Depending of the area to be evaluated, the 2018 situation will be the first effect measurement or an extended T0. It is expected that from 2027 onwards besides year-to-year comparisons also trend analyses will be possible in the Dutch North Sea.

Regarding BISI v2 it is suggested to base internal references ideally on the current or recent quality status of identified reference areas, or on ecotope specific recent historic monitoring data of indicative the last 25 years. One-and-another depends on monitoring data availability, but might influence reliability of the indicator and quality assessments in particular. Poor data availability for the year of evaluation will lead to non-significant results although quality levels (BISI-scores) might seemingly deviate from the reference. Poor reference and/or recent historic data availability will lead to a limited and possible insufficient number of indicator species as part of the area specific BISI, as in several cases criteria set to the power of individual potential indicator species will not be met. A limited number of years to base reference values upon might lead to elevated internal reference levels. If a possible effect of historic data limitation is expected, a calibration of the internal reference based on new monitoring data (for the area of evaluation or potential reference areas) after some years of monitoring, is suggested. This might lead to an adjustment of the internal reference and results of evaluations, asking for a recalculation of results of earlier years. Results on quality developments are not expected to be influenced that much (although non-significant tendencies might become significant), but relative quality levels might be influenced (and become more comparable to other well-monitored areas of evaluation).

7 Determination of GES and boundaries

To evaluate the quality status (and give a representative view of an area or habitat of concern) it is essential that the samples are expected to give a representative view. This is expected to be the case at a random sampling programme, or when at least the samples were initially ‘placed’ randomly after which their positions are fixed. There is a possibility to evaluate changes in the quality status on basis of samples that do not necessarily reflect the quality of the entire area, but in that case knowledge of patterns in abiotics, pressures and functions is essential and one has to be extra careful, especially when samples are missing. In case of stratified positioning of samples used for evaluation of entire areas, this should be considered at interpretation of index and testing results. As used references for specific areas (like water bodies or Natura 2000 areas) in BISI v2 are the result of an ecotope-based compilation, evaluation on basis of a to ecotopes stratified sampling design is very well possible as long as sampling within used ecotopes is representative for those ecotopes.

V1 - For all benthos samples taken within the frame of the Dutch North Sea quality status monitoring it is indicated whether they are part of the set of samples for quality status evaluation of certain areas (initially random fixed positions) or whether they are specifically meant for evaluation of management measures (stratified sampling scheme), at which some samples can be used for both types of evaluations.

For the Dutch North Sea a good ecological status is currently defined as an improvement of the quality status based on BISI v1 (general quality evaluation) for all ASEVs (except for the Brown Bank), the earlier indicated 6 ecotopes (EUNIS 4) at national level, and the HD habitat type H1170, H1110b and H1110c at national level, compared to the situation of the year 2015 (T0) (Min IenW & Min LNV, 2018). It is expected that in the near future thresholds and targets for quality assessments in relation to MSFD and Natura 2000 in the Netherlands will be more specified at national level in consultation with international developments and agreements (e.g. within OSPAR).

GES would also be a significant increase in the quality status of ‘closed’ areas relative to open areas compared to the T0 based on the general BISI, indicating effective management, which on the mid-term would lead to a significant increase in the quality status of the entire ASEV, indicating a sufficient large area with specific measures, as well.

For other areas of evaluation a GES based on BISI could be defined as at least a consolidation of the quality status based on the general BISI, a significant improvement or specific BISI-score in case of taken measures or an improvement of a specific BISI-score (e.g. the occurrence of specifically Habitat Directive typical species in case of evaluation of HD habitats or HD areas).

At present, the targets for a GES have to be defined for each new area. It is expected that in the near future targets for a GES, especially in relation to the MSFD and Natura 2000 will be more tuned between European countries.

8 References background information on methodology

For details and background information, specifically on the development of BISI v1 (on which v2 proceeds), see:

- Wijnhoven, S., Bos, O.G. (2017). Benthic Indicator Species Index (BISI): Development process and description of the National Benthos Indicator North Sea including a protocol for application. Ecoauthor Report Series 2017 - 02, Heinkenszand, the Netherlands.

The report includes a BISI Assessment Tool (Annex 2): V260917, is the assessment tool based on BISI v1 including indicator species lists references and methodologies as applied for the Dutch North Sea. For details and results of the first application and a recent historic analyses of the quality developments based on BISI (v1) for the areas of evaluation as identified in the Dutch North Sea, see:

- Wijnhoven, S. (2018). T0 beoordeling kwaliteitstoestand NCP op basis van de Benthische Indicator Soorten Index (BISI). Toestand en ontwikkelingen van benthische habitats en KRM gebieden op de Noordzee in en voorafgaand aan 2015. Rapport Ecoauthor & Wageningen Marine Research. Ecoauthor Report Series 2018 - 01, Heinkenszand, the Netherlands (in Dutch). (a translation in English is planned).

The adjustments made to the methodology and presented as BISI v2 are implemented in the BISI Assessment Tool v181218 for the areas and habitats of evaluation in the Dutch North Sea. An exception are the indicator species selections for the areas and habitats of evaluation, that are not adjusted according to the, with v2 introduced, decision scheme, but that are consolidated as originally constructed with BISI v1. The BISI Assessment Tool v181218 may function as an extensive example of application of BISI v2 towards application in new areas.

- Wijnhoven, S. (2018). Assessment tool 'Benthic Indicator Species Index (BISI)', v181218: Application of BISI v2 in the Dutch North Sea with consolidation of earlier identified references.

All products (reports, protocols, application tools and evaluations) will be made available via the Ecoauthor website: www.ecoauthor.net, where a specific page on BISI will be created soon. Search term for information: http://ecoauthor.net/?tag=bisi.

9 Strengths and weaknesses of data

9.1 Strengths

V1 - The evaluation of the Dutch North Sea areas make use of a specifically for the current indicator installed monitoring programme as from 2015 onwards. Therefore the monitoring efforts (number, type and positioning of samples) in case sampling has been according to the programme, is sufficient to detect reasonable changes in the quality status (already within 3 years) if there are.

Besides that (general) changes in the quality status can be detected for the different areas, ecotopes and habitats under investigation, the indicator gives insight in the causes and effects of observed changes.

Even if changes are not significant (yet), a series of evaluations might indicate whether developments seem to move into the direction of a good ecological status or might indicate whether quality improvement or deterioration can be expected in the near future.

The focussing of the methodology and the monitoring programme on specific management measures gives insight in the local effectiveness and the enfluence of measures on larger scales, the relative importance of different pressures on the current quality status, and the potential of (additional) measures in the future.

As the methodology makes use of a reference based on real observations of the last three decades, a realistic reference is ensured, that is within reach with effective management of the (Dutch) North Sea and the separate areas. Although the methodology is specifically developed to evaluate change in quality status in time or between different treated areas making use of a fixed monitoring design, a comparison of the T0 quality status relative to the realistic reference is expected to be reliable concerning the relative importance of causes and effects of differences in quality status.

The BISI indicator makes use of a range of indicator species with deviating indicator value related to different pressures. This makes the methodology robust, so that it is not susceptible to coincidental changes in occurrences of singular species. Moreover, the methodology appears to be robust enough that an occasional missing of samples does not lead to another interpretation of the quality status (it can lead to a decrease of importance of especially less common or 'highly variable' species in the assessment of the quality status). It has been found that even an evaluation based on one of the two dominant sampling methodologies (evaluation only based on boxcore samples or only based on dredge samples) generally leads to comparable results for quality status assessments (i.e. Wijnhoven, 2018). Such a reduction of the monitoring efforts and the number of indicator species involved does however impact the likelihood of finding reliable results (e.g. Figure 4) for the specific indices indicating causes and effects.

With BISI v2 additional species might be added to specific evaluation, although they do not reach the criteria set for the general quality evaluation. Specific power analyses should indicate whether the resulting specific indicator is more sensitive with or without the inclusion of certain potential indicator species.

9.2 Weaknesses

As the methodology is in the first place developed to detect changes in the quality status between the T0 and future evaluations, the accurateness of the internal (realistic) reference is something that has to be confirmed in the near future by the results of the coming monitoring years (according to the national benthos monitoring programme). It is very well possible that (as expected) the internal reference in use now is a reliable reference level for a realistic GES, in terms of what can be reached naturally on the mid-term if dominant pressures are diminished. This is however something that should be tested by using the evaluation tool for future evaluation, by focussing on areas with effective management measures, possibly combining sample locations with detailed pressure mapping, and/or application of the methodology in areas outside the Dutch continental zone (NCP). It is expected that the identification of potential causes and effects of observed quality differences is reliable (see also 9.1 strengths). It is however not completely certain at present whether the internal reference should be slightly adjusted, and more important, if this is more the case in certain areas to be evaluated than in others. The reliability of the internal reference is likely dependent of the monitoring efforts in the past (that show spatial differences). Therefore the comparisons of relative quality states between different areas at T0 should be handled with care.

With BISI v2 the derivation of the internal reference is more standardized making evaluations of different areas more comparable, although certain aspects as monitoring efforts, habitat heterogeneity at a level hidden by the aggregated ecotopes and present species pools will to a certain extent have some impact on evaluation results. Additionally a broader application in various regions, including possible identification of reference areas, might give opportunities for improvement of the comparability of evaluations. Therefore recalculations of internal references as a calibration measure are suggested for areas for which after installation of a BISI methodology a series of monitoring campaigns has come available.

A weakness related to data availability is that there is variability in the methodologies used for sampling, sorting and identification. With BISI v2 efforts are made to reach a better comparability between the samples.

V1 - Aspects that will not be solved in the near future already for the Dutch North Sea situation are different dredging strategies and devices in the coastal and the offshore regions, and subsampling used to assess occurrences of species. Other aspects, like the sampling with non-random varying surface areas during dredge sampling, have now been solved from 2018 onwards. There will however always be some effect of changes in sampling strategies, efforts, companies involved, changes in protocols, etc., which are not always directly foreseen (e.g. planned monitoring for 2018 in the Dutch North Sea has partly been postponed to 2019 due to unforeseen capacity problems). It is however expected that effects on results of those various discontinuities are small, as precautions have been taken, for instance concerning the selection of indicator species (focus on rather common species), or concerning the use of hit rate (only for the better standardized boxcore sampling).

Another uncertainty is related to the debate around indicator value of some of the species. It is a fact that not all ecological aspects are known even for rather common marine species. It is expected that possible new insights in the ecology of species and their indicator value, will have minimal impact on evaluation results as the number of species under debate is kept to a minimum and the methodology is based on a large number of species to reduce possible impacts. With BISI v2, and the application within the frame of OSPAR, it is foreseen to align the characterization and qualification of potential indicator species by using the same sources and central storing of indicator characterisations of species per ecotope (possibly integrated in existing platforms).

10 Further work required

First of all the methodology will benefit from application the coming years in a variety of areas under different circumstances, to identify possible weaknesses. At present only the T0 evaluation has been performed, including recent historic analyses of quality developments for the Dutch North Sea areas on basis of BISI v1. As the methodology (with BISI v1 in particular) is actually developed to detect changes in the quality status instead of exactly defining the quality status at a certain moment, the first real test on a full-scale specifically designed monitoring programme will be on basis of the 2018 monitoring data.

It is however not to say that the methodology cannot be used to assess the actual quality status at a certain moment in time, but therefore the internal reference level has to be tested. Besides application in the future, the methodology will benefit from testing in other areas in the broad OSPAR region. Good test cases will be comparable regions with different monitoring techniques and efforts, like the international Doggerbank with differences in approaches between countries. The methodology will additionally benefit from evaluation based on detailed pressure maps, and the possible identification of low pressure regions for validation and/or alternative derivation opportunities for the internal reference as formalized in BISI v2.

The evaluation tool is developed as a living methodology, which allows adjustment of used selections of indicator species, indicator values and reference values according to new insights. Although in that case recalculation of earlier evaluations will be needed, those can be realized with limited efforts. The methodology is robust enough so that these adjustments will not dramatically change earlier outcomes.

The current protocol describes both methodologies BISI v1 and BISI v2. Herewith both the former application including the related reports (with v1) can be understood and reconstructed, but also the tools (v2) for application in new areas in the future are presented. Used examples are from the Dutch North Sea region, however already according to BISI v2. Application of BISI v2 in amongst others Dutch transitional waters and in international case studies (within the frame of OSPAR) will provide new examples of application including the ecotope-based construction of indicator species selections and references. The current protocol and accompanying BISI Assessment Tool (v181218) are specifically intended to allow application in new areas and the later can be used as a scheme for application where tables can be adjusted to fit new areas with related indicator species and reference levels. The current v2 methodology will be presented in an OSPAR CEMP-protocol with these new applications and examples in the near future (a draft generic guideline for the Coordinated Environmental Monitoring Programme –CEMP- is foreseen for 2019).

11 References

- Bergman, M.J.N. & Van Santbrink, J.W. (1998). Distribution of larger sized invertebrate species (megafauna) in the Dutch sector of the North Sea. In: Bergman, M.J.N., Van Santbrink, J.W., Buijs, J., Craeymeersch, J.A., Piet, G.J., Rijnsdorp, A.D., Laban, C., Zevenboom, W. (eds.) The distribution of benthic macrofauna in the Dutch sector of the North Sea in relation to the micro distribution of beam trawling. BEON Rapport nr. 98-2, NIOZ, NIOO-CEMO, RIVO-DLO, NITG-TNO, RWS-DNZ, p. 55-89.

- Borja, A., Franco, J., Pérez, V. (2000). A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Marine Pollution Bulletin 40, 1100-1114.

- Bouma, H., De Jong, D. J., Twisk, F., Wolfstein, K. (2005). Zoute wateren EcotopenStelsel (ZES.1). Voor het in kaart brengen van het potentiële voorkomen van levensgemeenschappen in zoute en brakke rijkswateren. (No. Rapport RIKZ/2005.024). Middelburg: RIKZ (in Dutch).

- Buckland, S.T., Studeny, A.C., Magurran, A.., Illian, J.B., Newson, S.E. (2011). The geometric mean of relative abundance indices: a biodiversity measure with a difference. *Ecosphere* 2(9):100. doi:10.1890/ES11-00186.1.

- Craeymeersch, J.A., Heip, C.H.R., Buijs, J. (1997). Atlas of North Sea benthic infauna. (Based on the 1986 North Sea Benthos Survey). ICES Cooperative Research Report, No. 218. pp. 86.

- Cuperus, J., Swarte, M. (2016). A2.107, Waterbodem marien – Uitzoeken en determineren van Macrozoöbenthos, V4. (in Dutch).

- Duineveld, G.C.A., Künitzer, A., Niermann, U., De Wilde, P.A.W.J., Gray, J.S. (1991). The macrobenthos of the North Sea. *Netherlands Journal of Sea Research* 28, 53–65.

- Dutertre, M., Hamon D., Chevalier C., Ehrhold, A. 2013. The use of the relationships between environmental factors and benthic macrofaunal distribution in the establishment of a baseline for coastal management. *ICES Journal of Marine Science* 70, 294–308.

- Elliott, S.A.M., Guérin, L., Pesch, R., Schmitt, P., Meakinsc, B., Vina-Herbon, C., González-Irusta, J.M., De la Torriente, A., Serrano, A. (2018). Integrating benthic habitat indicators: Working towards an ecosystem approach. *Marine Policy* 90, 88–94.

- EMODnet (2016). EMODnet broad-scale seabed habitat map for Europe (2016), licensed under CC-BY 4.0 from the European Marine Observation and Data Network (EMODnet) Seabed Habitats initiative (www.emodnet-seabedhabitats.eu), funded by the European Commission.

- Goudswaard, P.C., Perdon, K.J., Jol, J., Hartog, E., van Asch, M., Troost, K. (2012). Het bestand aan schelpdieren in de Nederlandse kustwateren in 2012. Rapportage IMARES Wageningen UR, Rapport CO85/12, 44 pp. (in Dutch).

- Hiscock, K., Langmead, O. & Warwick, R. (2004). Identification of seabed indicator species from time-series and other studies to support implementation of the EU Habitats and Water Framework Directives. Report to the Joint Nature Conservation Committee and the Environment Agency from the Marine Biological Association. Plymouth: Marine Biological Association. JNCC Contract F90-01-705. 109 pp.

- Marine Species Traits editorial board (2018). Marine Species Traits. Accessed at http://www.marinespecies.org/traits on 2018-07-09.

- MarLIN (2006). *BIOTIC - Biological Traits Information Catalogue.* Marine Life Information Network. Plymouth: Marine Biological Association of the United Kingdom. [9-7-2018] Available from www.marlin.ac.uk/biotic.

- Min EZ (2014a). Profiel habitattype H1110 Permanent overstroomde zandbanken (versie 2014), (www.synbiosys.alterra.nl/natura2000/.../Profiel_habitattype_1110_2014.pdf). (in Dutch).

- Min EZ (2014b). Profiel habitattype H1170 Riffen (versie 2014), (www.synbiosys.alterra.nl/natura2000/.../Profiel_habitattype_1170_2014.pdf). (in Dutch).

- Min IenW, Min LNV (2018). Mariene Strategie (deel 1). Huidige milieutoestand, goede milieutoestand, milieudoelen en indicatoren 2018-2024, Hoofddocument. Een uitgave van Ministerie van Infrastructuur en Waterstaat en Ministerie van Landbouw, Natuur en Voedselkwaliteit, februari 2018, wvl0118tp312 (in Dutch).

- Leewis, L., Verduin, E.C., Stolk, R. (2017). Macrozoobenthosonderzoek in de Rijkswateren met Boxcorer, Jaarrapportage MWTL 2015. Waterlichaam: Noordzee. Rapport Eurofins AquaSense 31 maart 2017 definitief, 75 pp. (in Dutch).

- Perdon, K.J., Troost, K., Van Asch, M., Jol, J. (2016). WOT schelpdiermonitoring in de Nederlandse kustzone in 2016. Wageningen University & Research Rapport C093/16, 34 pp. (in Dutch).

- Reiss, H., Birchenough, S., Borja, A., Buhl-Mortensen, L., Craeymeersch, J., Dannheim, J., Darr, A., Galparsoro, I., Gogina, M., Neumann, H., Populus, J., Rengstorf, A. M., Valle, M., van Hoey, G., Zettler, M. L., Degraer, S. (2015). Benthos distribution modelling and its relevance for marine ecosystem management. *ICES Journal of Marine Science* 72, 297–315.

- RWSV (2017). Voorschrift Rijkswaterstaat, Code 913.00.B200. Bemonstering van macrozoöbenthos en sediment in het litoraal en sublitoraal in mariene wateren. Methode: Reineck boxcorer, Van Veen happer, Vacuüm steekbuis, Steekbuis. Version 28-02-2017. Rijkswaterstaat – Ministerie van I & M, (https://staticresources.rijkswaterstaat.nl/binaries/RWS_913.00.B200_tcm21-103148.pdf). (in Dutch).

- Schellekens, T., Faasse, M. (2015). RWSV Bodemschaaf. Versie oktober 2015 – definitief. eCOAST report, 9pp.

- Seip, P.A., Brand, R. (1987). Inventarisatie van macrozoobenthos in de Voordelta. Rapportage BOVO-project, NIOZ rapport 1987-1, 147 pp. (in Dutch).

- Shaw, P., Wind, P. (1997). Monitoring the condition and biodiversity status of European conservation sites - a discussion paper. Report to the European Environment Agency on behalf of the European Topic Centre on Nature Conservation, Paris.
http://biodiversity.eionet.eu.int/publications/SNH_NERI_1997.pdf.

- Ten Brink, B.J.E., Van Strien, A., Van Hinsberg, A., Reijnen, M.J.S.M., Wietz, J., Alkemade, J.R.M., Van Doben, H.F., Higler, L.W.G., Koolstra, B.J.H., Ligtvoet, W., Van der Peijl, M., Semmekrot, S. (2000). Natuurgraadmeters voor de behoudoptiek. RIVM rapport 408657005, 109 pp. (in Dutch).

- Van Moorsel, G.W.N.M. (2003). Ecologie van de Klaverbank. BiotaSurvey 2002. Ecosub, Doorn, 154 pp. (in Dutch).

- Van Strien, A.J., Soldaat, L.L., Gregory, R.D. (2012). Desirable mathematical properties of indicators for biodiversity change. *Ecological Indicators* 14, 202-208.

- Van Strien, A.J., Gmelig Meyling, A.W., Herder, J.E., Hollander, H., Kalkman, V.J., Poot, M.J.M., Turnhout, S., Van der Hoorn, B., Van Strien-van Liempt, W.T.F.H., Van Swaay, C.A.M., Van Turnhout, C.A.M., Verweij, R.J.T., Oerlemans, N.J. (2016). Modest recovery of biodiversity in a western European country: The Living Planet Index for the Netherlands. *Biological Conservation* 200, 44-50.

- Wijnhoven, S., Bos, O.G. (2017). Benthic Indicator Species Index (BISI): Development process and description of the National Benthos Indicator North Sea including a protocol for application. Ecoauthor Report Series 2017 - 02, Heinkenszand, the Netherlands.

- Wijnhoven, S., Duineveld, G., Lavaleye, M., Craeymeersch, J., Troost, K., van Asch, M. (2013). Kaderrichtlijn Marien indicatoren Noordzee. Naar een uitgebalanceerde selectie van indicatorsoorten ter evaluatie van habitats en gebieden en scenario's hoe die te monitoren. Monitor Taskforce Publication Series 2013-02. NIOZ, Den Hoorn & Yerseke, 105 pp. (in Dutch).

- Wijnhoven, S., Sistermans, W., Escaravage, V. (2006). Historische waarnemingen van infauna uit het Voordelta gebied. NIOO-CEME rapport 2006-04. Monitor Taakgroep, KNAW-NIOO, Yerseke, 72 pp. (in Dutch).

- Wijnhoven, S. (2018). T0 beoordeling kwaliteitstoestand NCP op basis van de Benthische Indicator Soorten Index (BISI). Toestand en ontwikkelingen van benthische habitats en KRM gebieden op de Noordzee in en voorafgaand aan 2015. Rapport Ecoauthor & Wageningen Marine Research. Ecoauthor Report Series 2018 - 01, Heinkenszand, the Netherlands (in Dutch).

- Wijsman, J.W.M., Goudswaard, P.C., Escaravage, V., Wijnhoven, S. (2013). De macrobenthosgemeenschap van de Zeeuwse Banken na zandwinning. Een overzicht van drie T0 jaren en een eerste jaar van rekolonisatie. IMARES Rapport C164/13; NIOZ, Monitor Taskforce Publication Series 2013-17, 95 pp. (in Dutch).

- Witbaard R., Duineveld, G.C.A., Bergman, M.J.N., Lavaleye, M.S.S., Watmough, T., 2013. Atlas of the megabenthos in the Dutch Economic Zone of the North Sea. NIOZ report 2013-04, pp. 221.

- Wolff, W.J. (1973). The estuary as a habitat. An analysis of data on the soft-bottom macrofauna of the estuarine area of the rivers Rhine, Meuse, and Scheldt. Zoologische Verhandelingen 126, 1-242.

- Ysebaert, T., Meire, P., Herman, P.M.J., Verbeek, H. (2002). Macrobenthic species response surfaces along estuarine gradients: prediction by logistic regression. *Marine Ecology Progress Series* 225, 79-95.