Annex 1. Protocol Benthic Indicator Species Index (BISI)

This protocol describes the Benthic Indicator Species Index (BISI), the Dutch national benthos indicator for evaluation of the quality status of the North Sea and the sea floor integrity in particular. The BISI is developed 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.

Background information is described in the main report:

- *Wijnhoven, S. & Bos, O.G. (2017). Benthische Indicator Soorten Index (BISI): Ontwikkelingsproces en beschrijving van de Nationale Benthos Indicator Noordzee inclusief protocol voor toepassing. Ecoauthor Report Series 2017 - 02, Heinkenszand, the Netherlands.*

The BISI is summarised in the factsheet: 'Benthische Indicator Soorten Index (BISI)': D6C3/5), which is part of the Dutch Action plan Marine Strategy (part 1) 2018-2020:

- *Min IenM, Min EZ (in prep.). Mariene Strategie voor het Nederlandse deel van de Noordzee, Actualisatie 2018-2020, Deel 1. Publicatie van het Ministerie van Infrastructuur en Milieu in samenwerking met het Ministerie van Economische Zaken, Den Haag.*

The protocol is according to the structure of the ICES WGBIODIV template for Indicator factsheets (version 0.1) which is developed by O. G. Bos based upon the Helcom indicator factsheets (http://www.helcom.fi/baltic-sea-trends/biodiversity/indicators), the OSPAR technical guidance template, DEVOTES, and the ICES WGBIODIV criteria for good indicators.

1 Benthic Indicator Species Index (BISI)

1.1 Authors

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

1.2 Contributors

O.G. Bos (Wageningen Marine Research) (editor)

Valuable comments were made on earlier versions and discussions during project and expert meetings by:

- A.-M. Svoboda, A. Adams (Ministry of Economic Affairs),
- D. van Schaardenburg, P. Heslenfeld, M. Platteeuw, S. Rotteveel, A. Stolk, S. Stuijfzand, W. van Loon (Ministry of Infrastructure and Environment),
- 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)
- and others who participated in one of the expert meetings.

1.3 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 (absence/presence, or densities) of combinations of specific indicator species with predefined reference levels of these species that represent a good status. BISI can be used at different spatial scales (Exclusive Economic Zone (EEZ), European Nature Information System (EUNIS) levels, Natura 2000 areas, etc.) and can evaluate both the current quality status, as well as trends in quality status, e.g. to determine the effect of management measures in protected areas.

The BISI 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) 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. 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.

Compared to the reference that is indicative for the quality status at low pressure levels, a significantly reduced quality status is found for 2015 for all investigated spatial areas (i.e. areas with special ecological values (ASEVs), EUNIS ecotopes level 3 and Habitat Directive habitat types). Dominant pressures defining the current quality status are related to sea floor disturbances, and are at least partly the result of sea floor 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 most impacted areas are the deep muddy ecotopes (EUNIS level 3) of the Central Oystergrounds (Centrale Oestergronden) and the Frisian Front (Friese Front).

1.4 Indicator metadata

1.4.1 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 and eventual video recording techniques.

1.4.2 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').

1.4.3 Status

BISI is a new indicator that has been developed in the Netherlands in 2016/2017 for the ministry of Economic Affairs. The indicator has no status within OSPAR yet, but has been presented to the OSPAR benthic habitat expert group and is currently considered to be adjusted for common application as an OSPAR Benthic Habitats candidate indicator (BH1).

1.4.4 Indicator type

State indicator.

1.5 Indicator description

1.5.1 Indicator goals/objectives

The Benthic Indicator Species Index (BISI) aims to evaluate the quality status, quality status developments and quality status differences for predefined spatial areas and habitats in the Dutch North Sea (see paragraph 'Geographic coverage'). For each of these areas, a selection of benthic macrofauna indicator species has been made that together determine the overall quality status.

In addition, the BISI aims to evaluate the importance of different potential pressures and the effect of the potentially decreased quality on ecosystem functions (Table 1). 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 also intends to be a well-documented, reproducible and efficient assessment method that includes a protocol and assessment tool.

1.5.2 Theoretical background

The occurrence and/or numbers 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 of weighted log-transformed observation-to-reference ratios is calculated. 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 sea floor 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.

<u>Selection of indicator species</u>

Indicator species were selected on the basis of the combination of 4 characteristics:

1) 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).
2) Relatedness to habitats/areas: Certain species are typical for certain habitats (e.g. sediment constitution, local hydrodynamics, tidal - and depth strata).
3) Their presence in terms of densities and distributions (e.g. equally distributed or present in aggregations with sufficient chance of detection under natural densities).

4) Catchability with for the monitoring available and applied sampling gear.

Species selections are 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).

Although the catchability of 'typical species' is not necessarily good with moderate monitoring efforts, the presence or absence of these species is an important aspect of the 6-yearly Art. 17 Habitat Directive evaluations, which made that the species were included in the selections. The 'typical species' lists have been updated in 2014, therefore the latest more extensive species lists (Min EZ, 2014a,b) are part of the selections now.

To improve the robustness of the indicator, species selections made in 2013 were supplemented with additional species. The additions consist of two categories of species:

1. Species that were initially not selected by Wijnhoven et al. (2013) because they were not sufficiently generic, but which are good indicators for specific pressures or ecological functions (see Table 1). All species identified in Wijnhoven et al. (2013) with an indicator score of 1 or more were selected (of which only a few were withdrawn as they appeared to be difficult to observe or identify with the current monitoring techniques applied in the national benthos monitoring programme).
2. Species with good indicator values that were (consistently) abundant in the past, but are scarce now. It is less important that species are nowadays present in reasonable numbers when it is known that they have been present in reasonable numbers in the recent past. The data presented in Bergman & Van Santbrink (1998), De Bruyne et al. (2013), Lavaleye et al. (2000), and Van Moorsel (2002) were analyzed to extract additional potentially abundant and rather characteristic species.

As indicator species lists amongst others are based on the fact whether species are considered characteristic, (have been) present in reasonable numbers and can be caught with (one of) the monitoring techniques, it is clear that each of the areas of specific ecological value (ASEVs), EUNIS ecotopes and

Habitat Directive habitat types have their own list of indicator species. Initially, indicator species lists have been compiled for the ASEVs. As ASEVs are considered representative for certain Habitat Directive habitat types, those habitats have the same indicator species list, except for those cases where several ASEVs are situated in the habitat type and certain indicator species are only found in one of the ASEVs (those species are excluded). Also in case of EUNIS ecotopes, it is often the case that ASEVs are expected to be rather representative for a certain ecotope. Also for the ecotopes, indicator species lists are combined, and species present in only one of the several ASEVs of relevance are excluded. Additionally those species that are typically related to another type of ecotope (as identified from the habitat description, and the sediment preference in particular, of species in WoRMS (2017); e.g. Cleaver Bank indicator species that might be expected to be typical for coarse sediment, but appear to related more to (fine) sand substrate that is present on the Cleaver Bank as well) are excluded from the indicator species list.

Table 1. Overview of the various BISI evaluations. 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 (approaching a score of 1) for at least a number of categories (as analyzed in Wijnhoven et al. (2013)).

Code	Causes and functions (to be evaluated)	Description	Indicator value (IV_i)
General quality		Each selected indicator species has a value of 1 by definition; otherwise the species would not have been selected.	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 fisheries	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 shorter are likely only impacted by frequent occurring sea floor disturbing fisheries).	10 levels (age divided by 10, 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	4 levels (0, 0.1, 0.5, 1)

		phases of recovery).	
F.	Characteristic species	Species are almost exclusive for the area of evaluation or are much more abundant in the area of evaluation than elsewhere in the Dutch North Sea.	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 adopted.	2 levels (0, 1)

Indicator values

As indicated in Table 1, for each of the areas to be evaluated, an evaluation consists 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 indicator species are considered in 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 1. 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 this way, each selected species was scored for each criterion and a matrix was made of the scorings of species per criterion, for each area to be evaluated. The outcomes of the scorings for each of the areas to be evaluated are listed in Appendix 1. An example of indicator species for the ASEV Cleaver Bank with different indicator value for different specific evaluations is also given in Table 4.

During the development of the BISI indicator, the work of Beauchard (2016; forming the basis of Beauchard et al., 2017) was taken into account and compared to the work by Wijnhoven et al. (2013) as well. Some values of Wijnhoven et al. (2013) for criterion A (See table 1) have been adjusted when there was a striking dissimilarity with the valuation of Beauchard (valuation in terms of r-, K1-, K2- and A-strategist kindly provided for selected indicator species).

Their valuation is based on classifications of species on basis of species traits, which together determine life-history strategies (Beauchard, 2016). The proliferation of certain strategies in specific areas appears to be related to trawling intensity (frequency and density of trawl tracks). Where r-strategists represent typical pioneer species for which the distributions are poorly predictable, considered resilient species, and A-strategists represent typical stress adapted and therefore resistant species towards sea floor disturbing fisheries, the specified indicator value 'A.' is expected to be 0. K1-strategists are slow growing and therefore sensitive species and K2-strategists are slow growing, highly sensitive species for which the specified indicator value 'A.' is expected to be respectively 0.5 and 1. If an occasional species identified as r- or A- strategist initially received an indicator value of 1 (for sea floor disturbance) in our valuation, the value was adjusted to 0.5. Similarly an occasional K1- and K2-strategist with an initial indicator value of 0 was adjusted to 0.5. This shows that it is of importance to include several species in the evaluations as life-histories and sensitivities of species might be less well known as sometimes suggested and are for some species still under debate.

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.

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) which is the used standard for the current evaluation methodology, but in practice deviating from several to over 200 m^2) and are sieved indicative over 7x7 mm mesh size (5x5 mm in coastal zone) but also larger mesh sizes of over 1 cm have been used before. 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, etc.).

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

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.

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. 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 locally present are often present in aggregations. In the current methodology, 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 (especially in case of species occurring in low densities, for which hit ratios are sometimes preferred in the current methodology).

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 are able to reconstruct such a reference level, it is unlikely that it is a realistic target, even on 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.

Therefore reference levels are compiled based on recent maximum observations of year averages and/or one order of magnitude improvements and/or improvements with the standard deviation. In a few cases, when historic data availability is not optimal, reference levels can be one order of magnitude improvements of maximum plus standard deviation observations. 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

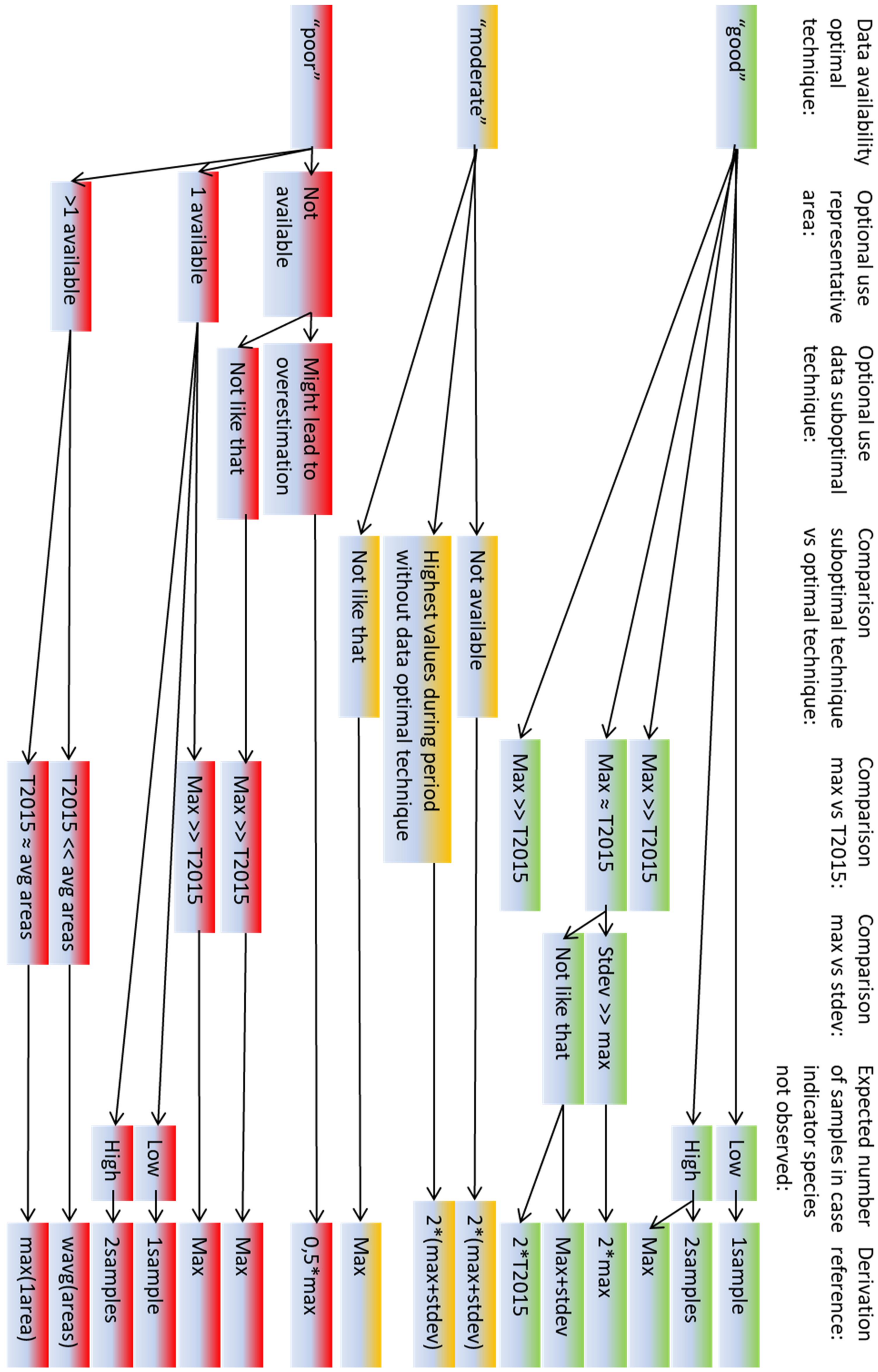

Figure 1. Flow chart showing the decision scheme of how to derive indicator species and area specific reference values. Max = maximum year average value in recent historic data; stdev = standard deviation; T2015 = the year average value as observed in 2015; 1sample = a density similar to an occurrence in one sample; wavg = weighted average of several areas based on the number of samples taken per area; >> = much larger; << = much smaller; ≈ = comparable values.

(dependent of the expected sample intensity during monitoring the coming years). Finally there are also cases where monitoring have been or is still focussed on specific areas (i.e. the ASEVs). In such cases it might be useful to extract the reference from these areas (optionally a weighted average in case more than one area is representative particularly for a part of the area of evaluation with possible limited data availability). The flow chart of Figure 1 indicates how reference values are obtained, dependent of the historic data availability, the area monitored, the monitoring methodology used, the within data variability, and how recent observations (the T2015) compare to the recent historic observations. Although there is some expert judgement involved, the derivation of the reference values is standardized as much as possible. All reference values and derivation methodologies used per species are indicated in the BISI Assessment Tool (Appendix 2).

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 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.

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). The current monitoring programme for the Dutch North Sea is based on power analyses to allow detecting at least 50% differences in abundances and/or hit ratio for a large share 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 much 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 very low abundances in the current situation. Besides, the realized power is probably sufficient to detect smaller differences in most cases.

Example of evaluation of management measures based on BISI

Figure 2 shows an example of evaluation of management regulations (comparison of development of for 'sea floor disturbing fisheries' open and closed areas according to a Before-After-Control-Impact (BACI) approach) for the area of specific ecological value (ASEV) Cleaver Bank. The T0 values show the real quality levels as measured in 2015, with a (however non-significant) tendency towards a higher quality status of the closed compared to the open areas (for each of the specific evaluations). As a fictional testing example an increase in the abundances of the indicator species with indicator value towards 'sea floor disturbing fisheries' (i.e. indicator species for specific evaluation A) equal to the in 2015 observed standard deviation is suggested. This leads to a significant increase of the general quality based on the BISI value, but besides an increase in the specific BISI-value A, also a significant increase of the specific indices C, D, E, F and J. This means that besides that the species are indicative for decreased sea floor disturbance, several of these species are also indicative for decreased intensity and decreased frequency of sea floor disturbing fisheries (as might be expected). Several species are to some extent also indicative for recovery, are found to be characteristic for the Cleaver Bank and designated as typical species for habitat type 1170 (reefs) of the Habitat Directive. Although it seems that there are improvements in other aspects (like habitat diversity 'H') as well, these changes are not found to be significant taking the 'natural' species fluctuations and the initial quality differences between open and closed areas into account.

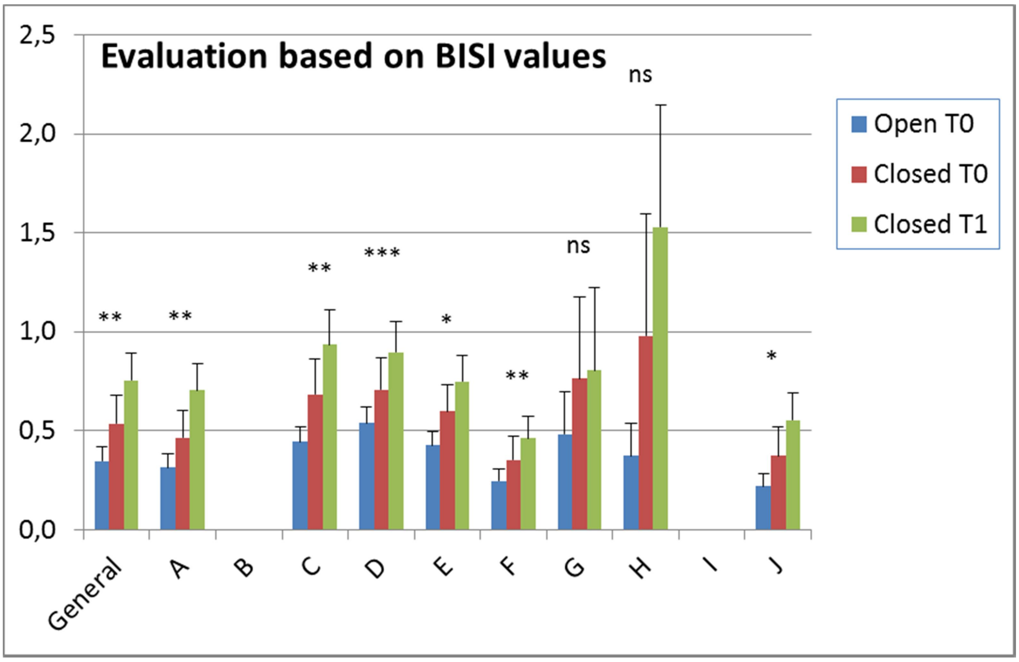

Figure 2. Fictional results of evaluation of the Cleaver Bank for seafloor disturbing fisheries closed areas compared to open areas without restrictions on T1 compared to the T0, as actually measured in 2015, with the Benthic Indicator Species Index (BISI). Besides the general quality assessment, the capitals on the x-axe refer to the specific evaluations as indicated in Table 1. Significant differences at T1 are indicated with *** (p≤0.001), ** (p≤0.01), * (p≤0.05), and ns = not significant.

As indicated by the BISI values above 1, the evaluation allows quality improvements transgressing the reference. In this example the quality status in the open areas is unchanged (this is generally unlikely but just to make the example not too complex). It has to be noticed that

Fig. 2 indicates the significant changes in quality status. One can also solely evaluate potential differences between open and closed areas at a given time. As indicated, there were no significant differences between the two treatment areas at T0; however the general quality status (as is the specific BISI value for the analyses A, C, D, E and J) is significantly higher in the closed areas at T1. So, where in case of the abundances of characteristic species (F), the observed increase is found to be significant, the quality level as such between the open and closed areas does not differ significantly at T1.

In case of the Cleaver Bank there is no evaluation of the quality related to biological activation of the sea floor top layer, as there are not enough indicator species identified for that, which is basically because bioturbation and bioirrigation processes are of less importance in the rather high dynamic, relative coarse sediment habitats of the Cleaver Bank.

It is therefore expected that with the current fixed monitoring programme for the Dutch North Sea, the sample size will not be limiting to detect quality improvements, especially in those areas where management measures are/will be taken, if they appear to be effective.

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 log-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). To reduce the risk of coincidentally observed changes in the indicator value, the minimum number of indicator species to be included in (specific) evaluations based on BISI is set at 5 (values based on less indicator species are not considered).

1.5.3 Description of the indicator

<u>Benthic Indicator Species Index (BISI):</u>

The occurrence and/or numbers of an (area - and/or habitat specific) selection of indicator species at a certain moment of evaluation is compared with a defined reference for that selection of indicator species. The methodology consists of the calculation of the geometric mean of weighted (species -and evaluation specific indicator values) log-transformed observation-to-reference ratios, taking into account the observed variances in the population of observations for testing of potential differences. Testing occurs against a (fixed) reference compiled for a series of areas with a certain protection status and/or importance, ecotopes (classification basically reflecting benthic communities for entire sea regions), and Habitat Directive habitat types as specific entities to be evaluated. The BISI evaluation tool is initially developed 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.

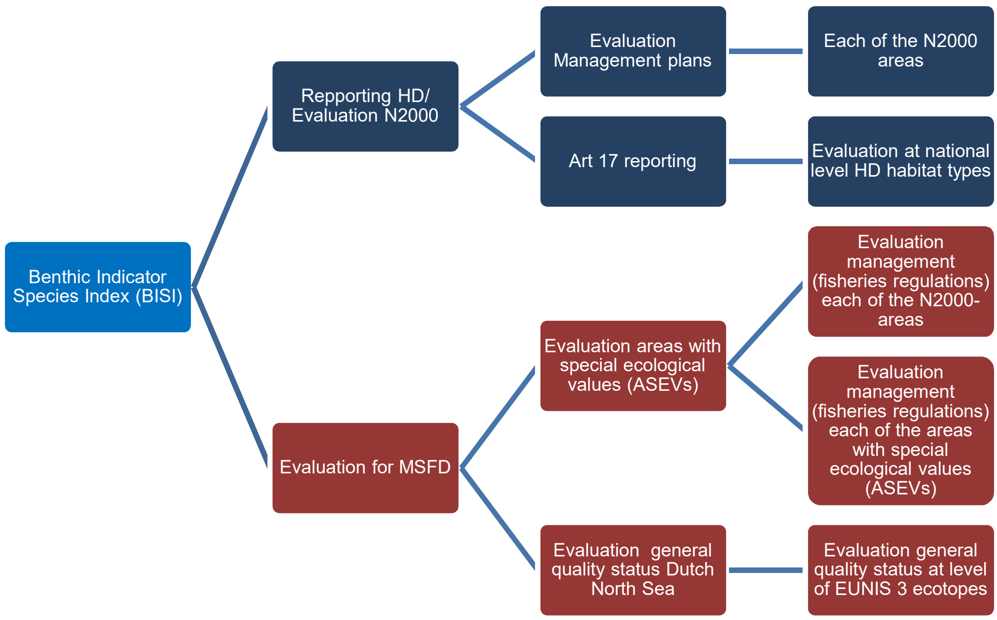

Figure 3. Schematic overview of the different Dutch evaluations for which the Benthic Indicator Species Index (BISI) is developed.

This means that the methodology (including reference levels, essential monitoring - and data type characteristics) is (currently) prepared 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 3 ecotopes covering the Dutch North Sea; 'Deep coarse sediment', 'Deep sand', 'Deep mud', 'Shallow to moderate deep coarse sediment', 'Shallow to moderate deep sand', 'Shallow to moderate deep 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 the respective areas with special ecological values, are the evaluations of efficiency of management regulations (fisheries restrictions) for each of the areas, for which specific samples (locations, methodology and data type) are selected. Dependent of the type of reporting (Fig. 3) 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.

Evaluation relative to compiled reference levels

Evaluations make use of compiled reference levels for each of the areas, ecotopes and habitats of interest as shown in Appendix 1. Table 2 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 occurrence or numbers) data for 3 species will be used for evaluation of the quality status of the ASEV Cleaver Bank. To evaluate the quality status (and therefore give a representative view of the entire area or habitat) it is essential that the samples more or less 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). 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.

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. In this case this does not seem to be problematic, as the specific aim of evaluation of the area is not necessarily improvement of sandy habitats that are also present, but specific improvement of the focal habitats.

Besides indication of the methodology used for species observation, it is of importance to identify which specimens 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 is indicated in Appendix 1 for each of the areas, ecotopes and habitats to be evaluated, if treated otherwise.

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 m²). In detail, the reference for A. crassa for the ASEV Cleaver Bank is 0,526 specimens per m², the reference for A. opercularis is 0,016 specimens per m², 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).

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. Qualified 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).

SFPA Klaverbank	Expected number of samples			Data type		**Ri**	
Indicator Species	Hamon grab (0,09 m^2)	Boxcore (0,078 m^2)	Video tracks (20 m^2)	Occurrence (Hit rate)	Numbers	Hamon grab (0,09 m^2); Hit rate (per grab) or observed numbers (m^{-2})*	Video tracks (20 m^2); numbers (per m^2)
Arcopagia crassa	17	1			1	0,526	
Aequipecten opercularis			16		1		0,016
Alcyonium digitatum			16		1		36,085
Aonides paucibranchiata	17	1			1	962,880	
Aporrhais pespelecani	17	1		1		0,153	
Number of species (S)	21	21	9				

* Results of Hamon grabs and boxcores are combined in the evaluation at which in case of the use of hit rate, the occurrence is not compensated for slight differences in sampled surface area.

Calculation of BISI-values

Next step in the evaluation is defining the occurrence (O_i) 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, in the form of a standard deviation, into account. The occurrence to reference ratio (O_i/R_i) is calculated for each of the indicator species. As log-transformation will be used (and one cannot take the log of a zero-value) and to reduce the impact of dramatically increasing population sizes of specific indicator species, occurrence to reference ratios are set to respectively 0.01 or 100 if occurrence values are less than one-hundredth or more than hundred times the reference value. The observed standard deviation is calculated into a value relative to the calculated O_i/R_i-ratio. Similarly as for the O_i/R_i ratios, the adjusted standard deviation is set to a value of 0.01 in case the observed standard deviation equals zero.

Example of observation data

Evaluation in BISI is based on observation data that might have to be adjusted to overcome problems with zero-values and make the indicator as sensitive for quality improvement and quality deterioration. In the example of Table 3 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 3. 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.

SFPA Klaverbank	Oi (T2015)	±stdev	Oi/Ri (T2015)	adjusted Oi/Ri	±stdev	adjusted stdev
Indicator Species	On basis of standard grab (0,09 m²) and video track (20 m²) surface area					
Arcopagia crassa	0,000	0,000	0,000	0,010	0,000	0,010
Aequipecten opercularis	0,004	0,009	0,259	0,259	0,544	0,544
Alcyonium digitatum	5,498	24,808	0,152	0,152	0,687	0,687
Aonides paucibranchiata	481,444	494,062	0,500	0,500	0,513	0,513
Aporrhais pespelecani	0,000	0,000	0,000	0,010	0,000	0,010

Besides a general quality assessment (in which all area, ecotope or habitat specific indicator species equally participate), specific evaluations are performed on weighted species subsets (depending on the specific indicator value of species for certain causes of change and potential effects of change).

Example of indicator values for indicator species

In BISI indicator values (IV_i*) identify the weight of the indicator species in the specific evaluations (Table 4) as* IV_i*s are multiplied with* O_i/R_i*-ratios (for example those of Table 3) in the calculations. This means that species without indicator value are not taken into account in the specific evaluations. Based on the first 5 indicated indicator species of the example of ASEV Cleaver Bank, the results of the specific evaluations to identify the possible impact of sea floor disturbances (A) and the effect of disturbances on species with national importance (F) or species of the Habitat Directive (J) will be the same (same indicator values).*

Table 4. 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 indicate the relative importance (the weight) of indicator species in the evaluations. Capitals refer to the causes and functions 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 and the total number of indicator species present in the (specific) evaluations for the ASEV Cleaver Bank.

SFPA Klaverbank	IV for Pressure					IV for recovery	National importance	IV for Ecological Functioning			Habitat Directive
Indicator Species	General quality	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 IV:	1	0,957	0,875	0,692	0,58	0,78	0,941	0,917	0,929	0,75	1
Number of species (S)	30	23	4	26	30	30	16	6	7	2	14

No specific indicator species to identify the possible impact of ecological disturbance (B) or possible changes in habitat diversity (H) or biological activation of the sea floor top layer are included in the example. 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 functions 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).*

Therefore species weights (IV_is as indicated in the example of Table 4) are multiplied with the log of the adjusted O_i/R_i ratio. In a similar way the species specific IV_i, divided by the average IV_i for all included indicator species for a specific evaluation, is multiplied with the log of the adjusted standard deviation to allow taking the variance into account in the significance testing of potential differences.

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

$$BISI = \exp((1/S)\Sigma(IV_i)\log(O_i/R_i)),$$

with an accompanying standard deviation as indicated in Table 5. 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 meant to evaluate potential changes in the future against the T0 (i.e. the quality status as calculated on basis of the monitoring data from 2015). The specific evaluations are predominantly meant to identify the relative importance of potential causes and the impact of changes in functions, 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.

<u>Evaluation of quality status and significance testing</u>

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 3), which quality data to test, combine or consider and what a good quality status looks like.

The different assessment areas:

- 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 3) 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).

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). Within the frame of the MSFD there is an improvement objective for the

EUNIS ecotope 'Deep mud' and specifically for deep low-dynamic sandy bottoms. The last is part of EUNIS level 3 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.

<u>Example of calculation of BISI values from individual indicator species (IIS) values</u>

Table 5 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 per definition a BISI-value of 1, or compared with future monitoring events. BISI-values are the inverse logarithm of a summation of IIS-values, who are calculated as $IIS_i = (IV_i)\log(O_i/R_i)$ (values of IV_i and O_i/R_i shown in Tables 4 and 3 respectively), divided by the total number of indicator species in the (specific) evaluation. Additionally the accompanying standard deviation is calculated as log(adjusted stdev)IV_i/ IV_{avg}), which is the log of the adjusted standard deviation times the species specific indicator value, divided by the average indicator value of all included indicator species for the specific evaluation.*

Table 5. Selection of calculated individual indicator species (IIS)-values including accompanying standard deviations 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 as indicated by the formula.

SFPA Klaverbank		IV for Pressure				IV for recovery	National importance	IV for Ecological Functioning			Habitat Directive
Indicator Species	General quality	A	B	C	D	E	F	G	H	I	J
Arcopagia crassa	-2,000	-2,000	na	-1,000	-2,000	-1,000	-2,000	na	na	na	-2,000
Aequipecten opercularis	-0,586	-0,586	na	-0,586	-0,352	-0,586	-0,586	na	na	na	-0,586
Alcyonium digitatum	-0,817	-0,817	na	-0,817	-0,817	-0,409	-0,817	na	na	na	-0,817
Aonides paucibranchiata	-0,301	na	na	na	-0,030	-0,301	na	-0,301	na	na	na
Aporrhais pespelecani	-2,000	-2,000	na	-1,500	-0,600	-2,000	-2,000	na	na	na	-2,000
BISI = exp((1/S)∑(IV)log(Oi/Ri))	**0,475**	**0,420**	**0,509**	**0,594**	**0,659**	**0,533**	**0,363**	**0,753**	**0,598**	**0,817**	**0,299**
Standard deviation:											
Arcopagia crassa	-2,000	-2,091	na	-1,444	-3,448	-1,282	-2,125	na	na	na	-2,000
Aequipecten opercularis	-0,265	-0,277	na	-0,382	-0,274	-0,339	-0,281	na	na	na	-0,265
Alcyonium digitatum	-0,163	-0,170	na	-0,235	-0,281	-0,104	-0,173	na	na	na	-0,163
Aonides paucibranchiata	-0,290	na	na	na	-0,050	-0,372	na	-0,316	na	na	na
Aporrhais pespelecani	-2,000	-2,091	na	-2,167	-1,034	-2,564	-2,125	na	na	na	-2,000
± stdev	0,654	0,561	0,752	0,693	0,751	0,593	0,472	1,008	0,735	0,739	0,398

In all other EUNIS 3 ecotopes and ASEVs at least no decrease in quality status is allowed (except for the Brown Bank with no official status yet; status 2017). 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

IenM & Min EZ, in prep.). 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.

Example of results of significance testing

Table 6 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 at T0 divided by the number of indicator species in the analyses. The computed t Statistic equals 1 minus the BISI value at T0 divided by the pooled standard deviation (as the reference BISI equals per definition a value of 1). 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 sea floor disturbance (A) and for characteristic species (F) and typical species of the Habitat Directive (J) differ very significantly. There is also a significant difference in the occurrence of species indicative for recovery (E) and larger sized species (C: indicative for the intensity of sea floor disturbing fisheries). 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 6. Results of significance testing of BISI results of the T0 (values ± standard deviation as indicated in Table 5) against the internal reference (per definition with BISI values equal to 1) 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.

SFPA Klaverbank	IV for Pressure					IV for recovery	National importane	IV for Ecological Functioning			Habitat Directive
Indicator Species	General quality	A	B	C	D	E	F	G	H	I	J
Pooled Std Dev	0,169	0,166	0,532	0,192	0,194	0,153	0,167	0,582	0,393	0,739	0,15
Computed t Statistic	3,114	3,501	0,924	2,114	1,757	3,046	3,817	0,425	1,023	0,247	4,665
Critcal 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	0,003	0,001	0,391	0,04	0,084	0,003	6E-04	0,68	0,327	0,828	8E-05
Significance	**	***		*	ns	**	***	ns	ns		***

Towards the evaluation within the frame of the Habitat Directive, the current indicator has no official status yet, but 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 might function 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 expected to be much more sensitive than the current estimation of expected presence of typical species in Article 17 reporting for the HD. 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 testing 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. 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). A T0 report including both aspects is one of the recommendations of the main report. The current evaluation based on BISI values compares the quality status with a realistic quality status (compiled reference) that can be achieved via natural development when dominant pressures are minimized, given the current habitat constitutions and species pools present.

As a first step, the difference between the T0 and the compiled reference will be tested using 2-sided independent t-testing. 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 will initially consist of similar 2-sided independent t-testing of years of evaluation against the reference 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.

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.

<u>Example of significance testing in the comparison of different treated areas</u>

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 provides for evaluation of the effectivity of taken measures by a monitoring setup according to a BACI-approach. Table 7 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 'sea floor disturbing fisheries' (i.e. indicator species for specific evaluation A) are increased with a value equal to the in 2015 observed standard deviation), Analyses compare average (± standard deviation) differences of individual indicator species (IIS)-values between the T0 and the T1. Depending on F-testing results, comparisons are based on paired t-tests with equal or unequal variance as monitoring sites are revisited every year of monitoring, The example is the same as shown in the results of Fig. 2, where consequences and specifics of the findings are discussed.

Table 7. Results of significance testing of potential differences in average individual indicator species (IIS)-values with accompanying standard deviation between for specific fisheries closed and open areas as part of the area of specific ecological value (ASEV) Cleaver Bank showing the first five indicator species. Shown are the results of T0 and a fictional example for T1 similar to those shown in Fig. 2. Results from paired t-tests with equal or unequal variance are shown, depending on the results of the F-tests. Significant differences are indicated with *** (p≤0.001), ** (p≤0.01), * (p≤0.05), and ns = not significant.

$\exp(IIS_{closed}) - \exp(IIS_{open})$	IV for Pressure					IV for recovery	National importane	IV for Ecological Functioning			Habitat Directive
T0 (SFPA Klaverbank)	General quality	A	B	C	D	E	F	G	H	I	J
Arcopagia crassa	0	0	na	0	0	0	0	na	na	na	0
Aequipecten opercularis	0,325	0,325	na	0,325	0,262	0,325	0,325	na	na	na	0,325
Alcyonium digitatum	0,174	0,174	na	0,174	0,174	0,112	0,174	na	na	na	0,174
Aonides paucibranchiata	0,044	na	na	na	0,006	0,044	na	0,044	na	na	na
Aporrhais pespelecani	0	0	na	0	0	0	0	na	na	na	0
T1 (fictional example)											
Arcopagia crassa	0	0	na	0	0	0	0	na	na	na	0
Aequipecten opercularis	0,684	0,684	na	0,684	0,494	0,684	0,684	na	na	na	0,684
Alcyonium digitatum	1,035	1,035	na	1,035	1,035	0,524	1,035	na	na	na	1,035
Aonides paucibranchiata	0,044	na	na	na	0,006	0,044	na	0,044	na	na	na
Aporrhais pespelecani	0	0	na	0	0	0	0	na	na	na	0
F-test	0,000	0,001	0,052	0,000	0,052	0,000	0,092	0,694	0,050	0,935	0,000
t-test	0,002	0,003	0,325	0,009	0,001	0,023	0,009	0,363	0,147	0,500	0,046
Significance	**	**		**	***	*	**	ns	ns		*

Therefore initially the difference in Individual Indicator Species (IIS) values is calculated between the two areas with different treatments on T0 and the other moment of evaluation. The difference is calculated per indicator species for each of the specific or general evaluations by taking the exponential of the IISi of closed areas minus the exponential of the IISi of open areas. The sets of IISi-differences are compared between T0 and the other moment of evaluation using paired t-testing after testing for possible differences in variance using F-tests.

Table 7 shows a fictional example with an increase in the abundances of the indicator species with indicator value towards 'sea floor disturbing fisheries' (i.e. indicator species for specific evaluation A) equal to the in 2015 observed standard deviation, of which the results are also visualised in Fig. 2). 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 IIS values between closed and open areas (averages ± standard deviations) in time can identified and/or analysed.

For each of the ASEVs with management regulations it has to be identified whether it is beneficial to combine different closed areas in one analysis or whether separate analyses are more sensitive. (Increase of sample size comes with possible increase of variance: In many cases it is expected that differences in species compositions between different closed areas are rather large, which pleas for separate analyses).

1.5.4 Indicator metric (formula)

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

$BISI = \exp((1/S)\Sigma(IV_i)\log(O_i/R_i))$, where

S = Number of indicator species included

IV_i = Species specific Indicator Value (0-1)

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).

BISI = Benthic Indicator Species Index, always a value between 0 (very low quality) and 1 (high quality), which comes 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). 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.

The Benthic Indicator Species Index is dimensionless; being a ratio varying between 0 and 1 it indicates the relative share present of an indicator assemblage reflecting a certain realistic (given the current habitat constitution and available species pools) good quality status. Due to the area specific reference levels, the BISI values of such areas with different reference levels cannot directly compared one-on-one.

1.5.5 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. Consolidation and/or increase have to be significant (at $p<0.05$) to take 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 could function as the good quality status.

1.5.6 Data source and description of data

There are two types of datasets related to the current methodology. In the first place, there is an extensive dataset of 'historic' data on which the compiled reference levels used in the methodology are based. These data are not necessary to perform evaluations using BISI. Then there is the dataset that will be evaluated. These data currently include the data for the T0 and will be supplemented with future data. Characteristics and essentials of both data sets will be described.

'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.

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.

Historic data used 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 attractive 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).

Essential data for evaluation

The evaluation methodology based on the Benthic Indicator Species Index 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).

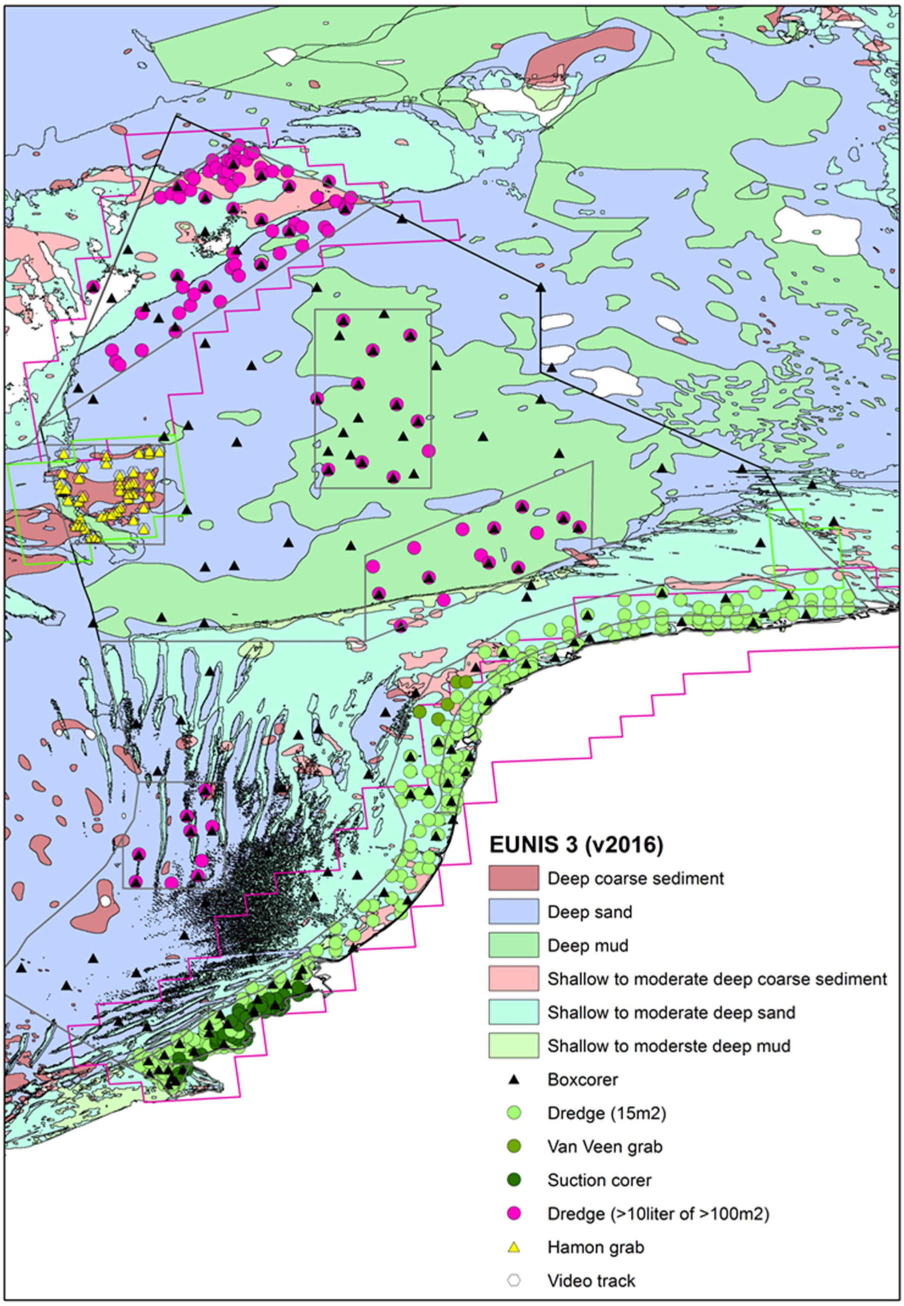

Figure 4. Map of the Dutch Continental Shelf with indication of the sample locations and used methodologies being part of the benthos monitoring within the frame of the MSFD. Locations are in principle sampled every three years from 2015 (T0) onwards. The background is a map with ecotope classification at EUNIS level 3 (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).

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. 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 come available via the data portal (http://www.informatiehuismarien.nl/open-data/) 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).

1.5.7 Methodology and data analyses

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

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). 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).

1.5.8 Assessment units

Three types of units are present that will be evaluated with the current methodology covering the entire Dutch part of the North Sea.

Separate evaluations of (developments in) the quality status of individual areas of special ecological value (that are often also Natura 2000 sites) will be performed. 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, Central Oystergrounds, Frisian Front, North Sea Coastal Zone, Front Delta and Plain of the Raan (all Natura 2000 areas with (planned) fisheries restrictions) and the Brown Bank.

The six EUNIS level 3 ecotopes covering the entire Dutch North Sea; 'Deep coarse sediment', 'Deep sand', 'Deep mud', 'Shallow to moderate deep coarse sediment', 'Shallow to moderate deep sand', 'Shallow to moderate deep mud' will be evaluated at the national level. (It might be beneficiary in the future to split up one or two of the ecotopes in different geographic parts, as benthic communities appear to be quite different in the south-western and the north-eastern part of the Dutch North Sea).

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.

1.5.9 Geographic coverage

The methodology is 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 3),
- Areas with fishery management regulations

Application in comparable areas, ecotopes and/or habitats in the vicinity (e.g. bordering areas in the UK, Belgium and Germany) should be possible without too many efforts when currently a comparable monitoring programme is in place. Application in other regions would ask for the compilation of area specific reference levels based on associated historic data.

1.5.10 Temporal coverage

At present the methodology, and the 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 will be on the monitoring data of 2015 which is the T0 situation of before most of the management measures to improve the quality status in the Dutch North Sea have been implemented. 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.

1.6 Determination of GES and boundaries

A good ecological status is currently defined as a consolidation of the quality status based on BISI values (general and specific) for all areas with special ecological values and ecotopes at EUNIS level 3, compared to the T0 situation. However, specifically for the ASEVs of the Central Oystergrounds and the Frisian Front, and the ecotopes 'Deep mud' and 'Deep sand' the GES is defined by a significant increased quality status based on the general BISI value compared to the T0. From 2018 onwards the GES for all ASEVs (except for the Brown Bank) will an improvement of the quality status on the mid-term (Min IenM & Min EZ, in prep.).

At present not a specific GES based on BISI is defined for the Habitat Directive habitat types as another system of evaluation is in place (evaluating the expected presence of all typical species). One could imagine that a GES for HD habitats would be at least the presence of each of the typical species in the monitoring of the 6-year period to be evaluated (so taking the specific monitoring efforts per habitat into account). Besides, at least a consolidation of the quality status based on specific BISI 'J' (Evaluation of the occurrence of Habitat Directive typical species) is expected to be desirable.

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.

At present, the targets for a GES have to be defined for each new area (e.g. additional evaluations of areas in neighbouring countries of the Netherlands) to be evaluated. In the future the compiled reference levels, with possibly some adjustments based on the first evaluation results, could be adapted as the targets for a GES.

1.7 References

For details and background information, see:

- Wijnhoven, S., Bos, O. (2017) Nationale Benthos Indicator Noordzee: Proces van ontwikkeling en presentatie van de 'Benthische Indicator Soorten Index (BISI)'. Ecoauthor Report Series 2017 - 02, Heinkenszand, the Netherlands.

1.8 Strengths and weaknesses of data

1.8.1 Strengths

The evaluations 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 effluence 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. Such a reduction of the monitoring efforts and the number of indicator species involved does however impact the likelihood of finding reliable results for the specific indices indicating causes and effects. These will often not give significant results (or the minimum threshold of 5 indicator species necessary to do a (specific) evaluation is not reached).

1.8.2 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 1.8.1 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. Backwards evaluation (of years before 2015) could also be useful for testing reference levels and the quality status at T0 at least for certain areas, although such an exercise should be combined with the testing of the behaviour of the methodology at significantly reduced monitoring efforts and the restriction to only one sampling methodology (e.g. for several areas only boxcore monitoring data are available for the years before 2015).

A weakness related to data availability is that there is variability in the methodologies used for sampling, sorting and identification. Efforts are made to reach a better comparability between the samples in the future. Aspects that will not be solved in the near future 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 of non-random variable surface areas during dredge sampling might be solved in 2018. There will however always be some effect of changes in sampling strategies, efforts, companies involved, changes in protocols, etc. It is however expected that effects on results 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.

1.9 Further work required

First of all the methodology will benefit from application the coming years, to identify possible weaknesses. At present only the T0 evaluation can be performed on a full-scale monitoring programme matching the methodology. As the methodology 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 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 could benefit from testing in neighbouring areas (e.g. similar ecotopes and habitats in the UK, Belgium or Germany). The methodology could benefit from evaluation based on detailed pressure maps, and the possible identification of low pressure regions.

Although some testing has been done during the development phase, it would be wise to test the robustness of the methodology for different areas in case monitoring efforts are reduced, only one of the monitoring techniques is used, or if certain species are omitted from the evaluation (which might occur at changes in sorting and/or identification protocols).

If above described testing results are promising (the methodology appears to be very robust in case of certain reductions in the data availability), backwards evaluations for years with sufficient data for the period 1984-2014 could fine-tune the methodology and especially improve the methodology as a tool to assess the quality status at a given time. How does the T0 compare to earlier years and what did the recent developments (of the last three decades) look like?

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.

1.10 References

- Beauchard, O. (2016). Life-history strategies of North Sea benthic macroinvertebrates: Typology and indication of beam-trawl fishing impacts. PowerPoint presentation part of the NIOZ colloquia given on 27-10-2016 at NIOZ-Yerseke.

- Beauchard, O., Veríssimo, H., Queirós, A.M., Herman, P.M.J. (2017). The use of multiple biological traits in marine community ecology and its potential in ecological indicator development. Ecological Indicators 76, 81-96.

- 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.

- 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).

- De Bruyne, R., Van Leeuwen, S., Gmelig Meyling, A., Daan, R. (2013). Schelpdieren van het Nederlandse Noordzeegebied. Ecologische atlas van de mariene weekdieren (Mollusca). Uitgeverij Tirion, Utrecht en Stichting ANEMOON, Lisse, 414 pp. (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.

- 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).

- Min EZ (2014a). Profiel habitattype H1110 Permanent overstroomde zand-banken (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 IenM, Min EZ (in prep). Mariene Strategie voor het Nederlandse deel van de Noordzee, Actualisatie 2018-2020, Deel 1. Publicatie van het Ministerie van Infrastructuur en Milieu in samenwerking met het Ministerie van Economische Zaken, Den Haag (in Dutch).

- Lavaleye, M.S.S., Lindeboom, H.J., Bergman, M.J.N. (2000). Macrobenthos van het NCP. Rapport Ecosysteemdoelen Noordzee. NIOZ-Rapport 2000-4, NIOZ, Texel, 66 pp. (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).

- 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).

- 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., 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).

- 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.