Methodology for calculating macro-algae indicators of abundance and composition

Research note from
DCE – Danish Centre for Environment and Energy

Date: 13 October 2017

Jacob Carstensen

Department of Bioscience

Claimant:
The Danish Environmental Protection Agency

No. of pages: 17

Referee:
Karsten Dahl

Quality assurance, DCE:
Susanne Boutrup

AARHUS
UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

Tel.: +45 8715 0000
E-mail: dce@au.dk
http://dce.au.dk

Contents

Summary

This document briefly outlines the theoretical foundation for the three macroalgae indicators that were proposed in Carstensen et al. (2014). More details on the statistical models for the indicators can be found in Carstensen et al. (2008) and Carstensen et al. (2014). For operationalising these indicators, such that they can be computed on a limited data set covering typically one water body and one assessment period, parameters generic to all such limited data sets have been compiled from a large data set with more than 8500 observations. These parameters are fixed for the operational computations, such that only a few parameters are estimated from the limited data set. This report presents the analysis of the large data set and the generic parameters and the operational routines for calculating the three macroalgae indicators.

1 Introduction

The European Water Framework Directive aims at achieving good ecological status in all European surface waters before 2017. The overall ecological status of a water body is assessed using the status of a number of Biological Quality Elements (BQE) through the one-out-all-out combination rule. For coastal water bodies the BQEs are phytoplankton, benthic vegetation and fauna. The BQE benthic vegetation should assess the status of both angiosperms and macroalgae. In previous studies, Carstensen et al. (2008) proposed six different indicators for macroalgae, and since there was some redundancy among these, Carstensen et al. (2014) suggested that macroalgae cover and composition could be described by means of three indicators: 1) cumulative algal cover, 2) fraction of opportunists, and 3) number of late-successional species.

The Danish Environmental Protection Agency (Miljøstyrelsen) has decided to operationalize these three indicators, meaning that they can be calculated for specific water bodies upon request through standardized routines extracting data from the monitoring database. This note provides the technical documentation of these calculations. First, the statistical models behind the indicators are revisited (see also Carstensen et al. 2014), before the operational versions of the indicators are presented. The statistical models for the indicators are within the framework of generalized linear models, which are computed using matrix algebra. Consequently, the indicators, and particularly the variance associated with these, should be computed with a software that can handle such statistical models. In this case, the R statistical package was chosen, since it is an open-source software. Finally, the implementation of the routines for calculating the indicators is reported.

2 Basis for macroalgae indicators

Data extracted from the national database (ODA) were collected as part of the Danish National Aquatic Monitoring and Assessment Programme (DNAMAP) and regional monitoring activities. They represent a total of more than 8500 observations, distributed along 267 sites each with a number of observations along a depth gradient in each of 72 water bodies (*Table 2.1*). Only relatively few observations represented water depths >15 m (mainly from around Bornholm, Kattegat, and Northern Belt Sea), and therefore the analysis was restricted to water depths <15 m.

Table 2.1. Overview of WFD water bodies (as defined in the national database ODA), depth range and number of sites and observations of the macroalgae variables included in the statistical analyses (see below). Number of observations is indicated in parentheses for the variables 'cumulative cover', and the number of observations used for the other indicators was similar. Sampling period: 2001-2016. Note that some waterbodies are outside the WFD baseline but have been included to increase the number of observations.

Area	Depth range (m)	#sites (#obs)
12 sømile grænse Samsø Bælt	7.2-13.8	1 (4)
Als Fjord	3.1-9.3	3 (141)
Augustenborg Fjord	1.1-5.0	7 (103)
Avnø Fjord	3.1-7.2	1 (32)
Bornholm 12 sm	5.1-14.5	3 (96)
DSØ åbne del	3.0-11.2	8 (155)
Ebeltoft Vig	5.2-9.9	3 (30)
Endelave og kystvandet fra Norsminde Fjord	3.1-13.9	6 (72)
Faaborg Fjord	3.1-3.1	1 (1)
Fakse Bugt	5.1-14.7	1 (9)
Farvandet Djursland øst	5.1-13.0	8 (168)
Femerbælt	8.2-8.2	1 (2)
Flensborg Fjord, indre del	1.0-5.8	2 (103)
Flensborg Fjord, ydre del	1.1-13.6	7 (382)
Genner Bugt	1.2-3.6	1 (13)
Haderslev Fjord	1.1-3.1	1 (22)
Horsens inderfjord	1.1-5.1	7 (131)
Horsens yderfjord	3.2-5.2	2 (12)
Indre Kalø Vig	1.3-9.1	4 (116)
Isefjord	1.1-6.0	7 (187)
Kalundborg Fjord	3.1-9.8	6 (125)
Kattegat	5.1-11.8	5 (176)
Kattegat –Læsø	5.1-14.6	3 (227)
Kattegat < 20 m (øst)	5.1-13.7	2 (85)
Kattegat < 20m (vest)	5.2-14.8	1 (111)
Kattegat > 20 m	5.1-13.7	1 (120)
Knebel Vig	1.3-1.4	2 (6)
Kolding inderfjord	1.1-1.3	1 (3)
Kolding yderfjord	3.4-3.4	1 (1)
København Havn	5.3-9.9	2 (24)
Køge Bugt	5.1-9.2	4 (36)
LBT Syd	1.2-13.0	7 (227)
Langelandssund	5.1-9.8	3 (52)
Lillebælt, Bredningen	1.1-14.7	2 (443)
Lillestrand	1.5-3.0	1 (6)

Lindelse Nor	1.2-2.0	2 (15)
Lovns, Skive, Riisgårde, Bjørnholms bugt	1.1-5.0	2 (183)
Lunkebugten	4.8-5.1	1 (3)
Musholm Bugt, indre del	5.0-12.4	4 (105)
Nakkebølle Fjord	1.1-8.8	2 (70)
Nissum, Thisted, Kås, Løgstør, Nibe, Langerak	1.1-5.9	18 (1176)
Nordlige Bælthav	5.1-11.4	3 (89)
Nordlige Kattegat 12 sm	7.8-14.9	3 (64)
Nordlige Lillebælt	5.3-13.2	1 (16)
Nordlige Roskilde Fjord	3.1-4.9	2 (45)
Nordlige Øresund	5.1-11.4	8 (137)
Nyborg Fjord	3.1-3.1	1 (1)
Odense Fjord ydre del	3.1-3.4	2 (8)
Rødsand	3.0-8.0	2 (92)
Sejerøbugt	5.1-13.9	9 (354)
Skagerrak 12 sm	7.7-14.9	4 (118)
Smålandsfarvandet Åben del	3.1-14.5	16 (447)
Smålandsfarvandet, syd	5.3-5.3	1 (2)
Stege Bugt	5.0-11.0	1 (3)
Sydlige Lillebælt Als-Ærø	5.1-14.5	1 (244)
Sydlige Lillebælt, omr. N f.Als	3.1-11.6	4 (91)
Sydlige Roskilde Fjord	1.1-5.7	5 (226)
Torø Vig	1.3-4.9	2 (11)
Vejle inderfjord	1.0-3.9	7 (90)
Vejle yderfjord	1.0-11.5	5 (259)
Åbenrå Fjord	1.1-7.8	11 (489)
Åbne STB 12 sm Nord	11.0-12.7	1 (3)
Åbne STB Nord	5.0-12.6	4 (130)
Åbne del Storebælt Øst	5.1-14.9	2 (85)
Åbne del Østersøen 12sm	5.2-13.0	1 (29)
Åbne del, Kattegat	9.0-13.0	3 (8)
Århus Bugt syd, Samsø og Djursland Syd	5.1-14.7	9 (57)
Århus Bugt, Kalø og Begtrup Vig	3.1-11.0	4 (162)
Øresundstragten	5.2-14.0	2 (137)
Østersøen	5.1-14.9	5 (155)
Østersøen 1sm. Bornholm	5.1-14.9	2 (203)
Østersøen 1sm. Christiansø	6.4-14.7	2 (8)
Total	1.0-14.9	267 (8756)

Macroalgae data were collected during summer (May-September) from 2001 to 2016. Although there were macroalgae data from before 2001, data from 2001 onwards were used rather than the entire data set dating back to 1989, because the more recent monitoring data in coastal areas has been collected according to improved procedures and represent the current monitoring standards (Krause-Jensen et al. 2001), where divers visually record the percent of cover of individual erect algal species. Data were assigned to 6-year periods to resemble the WFD assessment procedure, i.e. 2001-2006, 2007-2012, and 2013-2016 (latter only 4 years). Although these periods are not aligned with the WFD assessment periods they serve to partition the data into distinct periods of the relevant length. The purpose of this is to estimate the interannual variability within 6-year periods.

Algal cover was estimated in percent of the hard substratum at various depths along the depth gradients/sites. As a quality check, observations where the summed cover of algal species constituted < 80 % of the estimated total algal cover were excluded, because it is likely that species identification in these data sets might be incomplete.

All species were allocated to a functional group, using the system of Steneck & Dethiers (1994, Table 2.2). The functional groups 1-3: micro-algae, filamentous algae and single-layered foliose algae are dominated by opportunistic algal species with thin thalli, fast growth rates and ephemeral life forms, while the remaining groups primarily include perennial species with thick, corticated, leathery or calcareous thalli and generally slower growth rates. The following groups 2, 2.5 and 3 are referred to as 'opportunistic macroalgae' while algae belonging to groups 4, 5 and 6 are considered 'late-successional algae'. Microalgae (functional group 1) and crustose algae (functional group 7) were not consistently recorded in the entire data set and were therefore excluded from the analysis. Crustose algae are also characterized by having a much lover light demand compared to erect algae. In addition, all species were characterised as 'perennial', 'annual', and 'crustforming', where perennial is largely overlapping with the definition of late-successionals although some opportunist species are also perennial (e.g. *Polysiphonia sp*. and *Sphacelaria sp*.).

Table 2.2 Overview of functional groups (Steneck & Dethiers 1994) and our grouping of late-successional and opportunistic species in the present study. *Microalgae and crustose algae were not consistently recorded in the entire data set and were therefore omitted from the analysis.

Functional group		Examples of algal genus	Grouping in this study
1.	Microalgae (single cell)*	Cyanobacteria and diatoms	Opportunists
2.	Filamentous algae (uniseriate)	*Cladophora*, *Bangia*	Opportunists
2.5	Filamentous and thinly corticated algae	*Polysiphonia*, *Ceramium*, *Sphacelaria*	Opportunists
3.	Foliose algae (single layer)	*Monostroma*, *Ulva*, *Porphyra*	Opportunists
3.5	Foliose algae (corticated)	*Dictyota*, *Padina*	Late-successionals
4.	Corticated macrophytes	*Chondrus*, *Gigartina*	Late-successionals
5.	Leathery macrophytes	*Laminaria*, *Fucus*, *Halidrys*	Late-successionals
6.	Articulated calcareous algae	*Corallina*, *Halimeda*	Late-successionals
7.	Crustose algae*	*Lithothamnion*, *Peyssonnelia*, *Ralfsia*	Not included

The following three macroalgae indicators were analysed: cumulative cover, fraction of opportunists and number of perennial algal species. Cumulative cover was calculated by summing the cover values of all erect macroalgae species in each subsample, i.e. all algae except crust-forming algae. In some cases cumulative cover values surpassed 100 %, because algae can grow in several layers [canopy effect/multilayer]. The fraction of opportunists was calculated as the cumulative cover of opportunists divided by the cumulative cover of all erect macroalgae species and provided data in the range 0-100 %. The number of perennial algal species in each subsample was calculated as the total number of perennial species having a cover of at least 1 %.

In addition to depth, the percentage of suitable hard substratum has an effect on the recorded macroalgae cover (Carstensen et al. 2008). Composition of substratum was registered along with the collection of algal data. Divers visually recorded the total cover of suitable hard substratum as well as the cover of various substratum classes: size classes of stones, sand, mud and shells. Data on cover of suitable hard substratum were extracted from the database together with each algal data set.

This large data set serves as a basis for estimating parameters that will be used for the indicator estimation. Many of the parameters used in the statistical models (Carstensen et al. 2014) cannot be reliably estimated from a small data set (one water body and assessment period) and these parameters, that are believed to be of more generic nature (i.e. do not change between waterbodies and assessment periods), need to be estimated from a larger data set. Hence, the basic idea behind the macroalgae indicator is to estimate a number of general parameters ("global" parameters) from the large data set, and use these generic parameters when the macroalgae indicators are calculated for a specific water body and assessment period (*Figure 2.1*). For example, precise depth relationships for the three macroalgae indicators cannot be precisely estimated using a limited data set from a single water body and assessment period. Therefore, parameters describing the depth relationships are estimated from the large data set, and these parameters are used to account for the depth-dependency of observations when the macroalgae indicators are calculated from a data set comprising one water body and one assessment period.

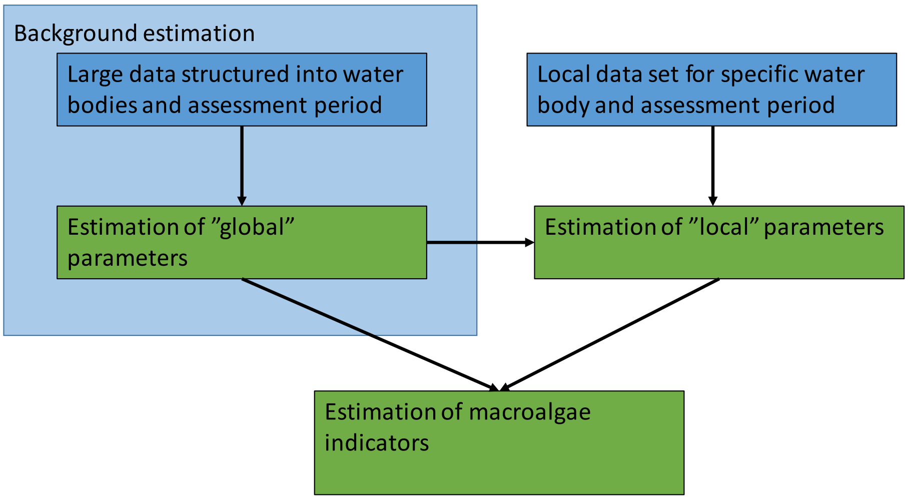

Figure 2.1 Conceptual diagram illustrating that some parameters are estimated from a large data set ("global") and that these parameters are used to estimate "local" parameters from a limited data set covering a single water body and assessment period.

2.1 Statistical analyses of algal variables

The analysis focused exclusively on algae from the depth range where physical disturbance was no longer a major controlling factor for cover (see *Figure 2.2*). Carstensen et al. (2008) employed depth cut-off values of 1 m for weakly exposed areas, 3 m for moderately exposed areas and 5 m for highly exposed areas. For the implementation of the macroalgae indicators the cut-off depth should be provided as a waterbody-specific parameter by the user.

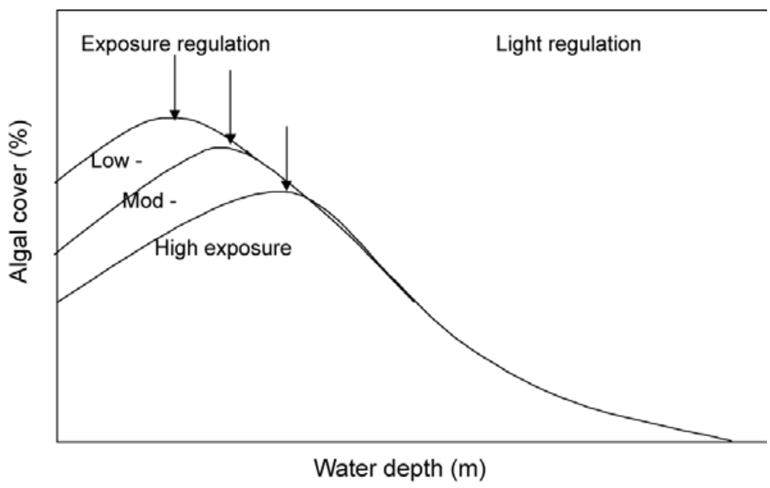

Figure 2.2 Illustration of the hypothesis that algal cover in shallow water is reduced due to physical exposure while from intermediate water depth towards deeper water algal cover is reduced in parallel to reductions in light availability. As a consequence, maximum algal cover is found at intermediate water depths and is located deeper in more exposed areas. Arrows indicate cut-off depths for different levels of exposure.

Macroalgal cover was estimated as substratum-specific cover, which should imply that cover levels were independent of substratum composition at the sampling sites. However, previous analyses have revealed that the recorded algal cover also depends on the amount of hard substratum, particularly for hard substratum in the range from 0 to 50 % (Carstensen et al. 2008). This analysis led to the formulation of a model, in which the relation between algal cover and hard substratum differed for levels of hard substratum of below and above 50%.

Macroalgal cumulative cover was assumed lognormal distributed, since random variations in cumulative cover generally increased with cover levels. The fraction of opportunists was in the range 0-100 % with expected greater variation around 50 % than at 0 % and 100 % and consequently, the angular transformation to these data (Sokal & Rohlf 1981):

$$x = \arcsin\sqrt{p} \tag{1}$$

Species number was counted as the total number of perennial macroalgal species, which covered at least 1% of the sea bottom in a given sub-sample. Macroalgal cumulative cover (*C*) and number of perennial species (*N*) were ln transformed before analysis:

$$x = \ln(C+1) \text{ or } x = \ln(N+1) \tag{2}$$

Variations in macroalgae variables (representing either ln transformed or arcsin transformed data, x) were described by the following generic model (fixed factors in small letters and random factors in capital letters), which is slightly modified from the model in Carstensen et al. (2014) that describes spatial and temporal variations as well as variation attributable to different divers:

$$\begin{aligned} x = {} & \text{waterbody} + \text{SITE (waterbody)} + \text{period} + \text{waterbody}\times\text{period} + \text{month} \\ & + \text{depth} + \text{\% hard substratum (0-50\%)} + \text{\% hard substratum (50-100\%)} \\ & + \text{DIVER} + \text{YEAR(waterbody}\times\text{period)} + \text{SITE}\times\text{YEAR(waterbody}\times\text{period)} + \text{e} \end{aligned} \tag{3}$$

where x is any of the three macroalgae variables. The model is based on the assumption that the observed level of each macroalgae variable depends on the specific water body, sites within the water body, the specific assessment period, interannual variation within the assessment period, water depth, suitable hard substratum and diver carrying out the survey. Depth was treated as a continuous variable for cumulative cover, whereas it was categorical (intervals of 2 m) for fraction of opportunists and number of perennial species because a continuous depth relationship could not be explicitly formulated for these two macroalgae variables. For estimating the macroalgae indicators for a given water body within a specific 6-year period, these sources of variation are considered fixed factors. To obtain robust estimates, generic fixed relationships for depth (linear) and suitable hard substratum (segmented linear relationship with one slope for 0-50% hard substrate joined with another slope for 50-100% hard substrate) were estimated.

Individual sites constitute a sampled subset of the infinite population of sampling locations within a water body and therefore, it is a random factor. The population of divers carrying out the survey is also infinite (particularly as new divers will enter the monitoring program in the future) and this factor also will be modelled as a random factor. The interannual variation among years within an assessment period is essentially a finite distribution (6 levels), but for estimating the variation among years it is estimated as a random factor and in the indicator implementation a correction for finite distribution will be employed (see Chapter 3). Thus, random factors include among sites, variation among divers, variation among years, changing interannual variation across years as well as residual variation expressing the variation around the depth relationship at a given surveyed transect.

2.2 Results

The mixed model for variations in algal variables (Eq. 3) was estimated using the entire data set. Variations among combinations of waterbody and period as well as depth were significant for all three macroalgal variables (Table 2.3). The relationship with suitable substrate proportion was significant for cumulative cover and number of perennial species, although cumulative cover relationship was flat for suitable substrate proportions above 50 %. The fraction of opportunists did not respond to changing proportions of suitable substrate. Monthly variations were only significant for the number of perennial species.

Table 2.3 P-values for each model component (fixed and random) for each of the modelled algal variables: cumulative algal cover, fraction of opportunists and number of perennial species. Fraction of opportunists was arcsin-transformed and the two other variables were log-transformed before the analysis. Total number of observations used for estimating each model was 8294.

Model component	Cumulative cover	Fraction of opportunists	# of perennial species
Fixed effects			
- waterbody×period	<0.0001	<0.0001	<0.0001
- Depth	<0.0001	<0.0001	<0.0001
- % hard substratum (0-50)	<0.0001	0.1908	<0.0001
- % hard substratum (50-100)	0.1337	0.0989	0.0005
- Month	0.9422	0.4385	0.0003
Random effects			
- SITE(waterbody)	<0.0001	<0.0001	<0.0001
- DIVER	<0.0001	0.0008	0.0001
- YEAR((waterbody×period)	<0.0001	0.1688	<0.0001
- SITE×YEAR(waterbody×period)	<0.0001	<0.0001	<0.0001
- Residual	<0.0001	<0.0001	<0.0001

The random factors were all significant with the exception of interannual variation within assessment periods for the fraction of opportunists (Table 2.3). Residual variation was the largest source of random variation for all three macroalgae variables, followed by spatial variation among sites within water bodies, alone and in combination with the interannual variation (Table 2.4). Variation among divers was large and comparable to the spatial random variation for cumulative cover, whereas it was smaller for the fraction of opportunists and number of perennial species. For all three macroalgae variables, the interannual variation contributed the least random variation. The depth relationships were similar to those reported in Carstensen et al. (2014).

Table 2.4 Variances of random effects for the models describing each of the algal variables: Cumulative algal cover, fraction of opportunists and number of perennial species. Fraction of opportunists was arcsin-transformed and the two other variables were log-transformed before the analysis.

Model component	Cumulative cover	Fraction of opportunists	# of perennial species
- SITE(waterbody)	0.0990	0.01785	0.07173
- DIVER	0.0609	0.00446	0.00974
- YEAR((waterbody×period)	0.0246	0.00128	0.00728
- SITE×YEAR(waterbody×period)	0.0585	0.01923	0.01439
- Residual	0.2355	0.04042	0.07910

The three indicators (macroalgae cumulative cover, fraction of opportunists and number of perennial species) have been shown to respond to levels of salinity and total nitrogen (Carstensen et al. 2008, Carstensen et al. 2014). The parameters describing the relationships with depth, suitable substrate proportion, and monthly variations from May to September, estimated from this large data set, will be used to adjust for these sources of variation when the model is applied to a smaller data set from a given water body and assessment period. Only the parameter describing the mean value of the water body and assessment period will be estimated from the smaller data set. Similarly, the variance parameters for the random effects will be held constant and used to calculate the uncertainty of the mean level of the water body and assessment period, i.e. employing a more correct description of the error structure in such small data set. Hence, it is only possible to achieve an appropriate variance partitioning using a large data set, but then these variance estimates are assumed to apply to all water bodies of similar type to those included in the large data set.

It is worth noting that the parameter estimates produced in the present study are quite similar to previous estimates (Carstensen et al. 2008, Carstensen et al. 2014) as they were based on an updated data set. This implies that these estimates are relatively robust to adding new data and consequently, re-estimating model parameters with an updated dataset is unlikely to substantially change the operational macroalgae indicator estimates.

3 Operational macroalgae indicators

In this chapter the objective is to estimate indicators for macroalgae cumulative cover, fraction of opportunists and number of perennial species from a data set covering a single water body and 6-year assessment period. Hence, the number of observations in such a data set can range from a few depth-specific observations on a single transect to multi-annual surveys of multiple transects spanning large depth gradients. The methods should produce results for any given type of data set and therefore, the approach must be robust irrespective of data availability.

Overall, the indicator calculation uses a combination of global parameters and local parameters estimated from a dataset representing one waterbody and one assessment period (*Figure 3.1*). Global parameters (from models in Chapter 2) and WFD class boundaries are stored as files and read upon initialization of the R-routine for calculating the macroalgae indicator. The R-routine produces a list with the indicator value, its standard error and probabilities for the indicator belonging to each of the five WFD status classes.

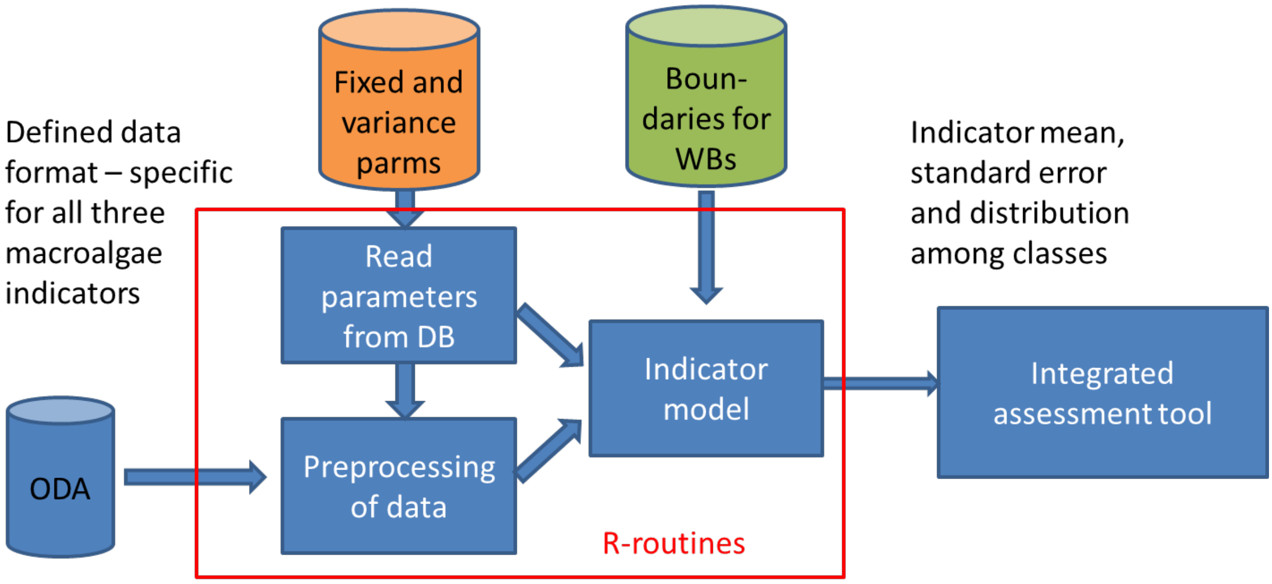

Figure 3.1 Sketch of the data-flow for the calculation of the three macroalgae indicators. Fixed and variance parameters have been found using Eq. (3) and are stored in a table or database (DB).Boundaries are similarly stored in tables or database.

3.1 Preprocessing step

The preprocessing of data involves a number of steps:

1. Aggregating species information to a depth-specific value. Cumulative cover is the sum of all species' cover (excluding crust-forming algae; see Table 2.2). Fraction of opportunist is the sum of opportunistic species' cover divided by the cumulative cover of all species. The number of perennial species is calculated as the count of species with this growth strategy having at least 1% cover. If there are no opportunists or perennial species at the depth-specific level, the observations of these are set to zero.

2. If the cumulative cover of macroalgae is more than 20% lower than the total cover of the macroalgae community, the depth-specific observation is suspect and not used for calculating the indicators (quality check).

3. Discard observations that are not sampled within May-September, deeper than 15 m or shallower than the threshold for physical exposure, which is supplied as a parameter to the routine.

4. The abovementioned transformations are applied to the depth-specific observations for cumulative cover, fraction of opportunists and number of perennial species.

5. A discrete depth class variable (intervals of 2 m) is produced from the continuous depth observation.

6. Adjust depth-specific observations (cumulative cover, fraction of opportunists and number of perennial species) for effects arising from sampling different months, depths and hard substrate. These effects are subtracted (see Eq. 3) using the global fixed parameter estimates. These adjusted observations will be used as input for estimating the local parameters.

3.2 Estimation step

The estimation procedure follows the approach outlined in Carstensen and Lindegarth (2016). After preprocessing data, the mean of the adjusted observations is calculated with a mixed model that accounts for the random effects using the global parameter estimates for the variances of the random effects. Thus, as opposed to estimating the mean by simply calculating the average of the adjusted observations, the mean is estimated as a weighted average where weights of individual observations depend on the more complex error structure in Eq. (3). Thus, in matrix notation

$$\hat{\mu} = (\boldsymbol{X}^T \widehat{\boldsymbol{V}}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \widehat{\boldsymbol{V}}^{-1} \boldsymbol{y} \quad (4)$$

where $\boldsymbol{y}$ is a vector containing the adjusted observations, $\boldsymbol{X}$ is the design matrix for the fixed factors (in this case a vector of 1's) and $\widehat{\boldsymbol{V}} = \boldsymbol{Z}\widehat{\boldsymbol{G}}\boldsymbol{Z}^T + \mathbf{R}$ is the estimated covariance matrix for the adjusted observations that expresses all the random variations affecting these, i.e. $\mathbf{R}$ contains the residual variation and $\boldsymbol{Z}\widehat{\boldsymbol{G}}\boldsymbol{Z}^T$ describes variations caused by changing divers, transects, years. The $\widehat{\boldsymbol{V}}$ matrix is computed based on the global variance estimates (Table 2.4). In practice, $\hat{\mu}$ is calculated from the mixed model (Eq. 3) by fixing the variances of all random effects to the global variance parameter estimates.

From the local estimate ($\hat{\mu}$) and global parameter estimates ($\widehat{\boldsymbol{\beta}}$) an indicator value for standardized values depth, hard substrate and month can be calculated. These standardized values can be given as parameters to the routine, but are default set to 50% for hard substrate, average over all five months, 7 m for cumulative cover or average over all depth classes for fraction of opportunists and number of perennial species. From these standardized values a vector ($\boldsymbol{L}$) can be constructed for estimating the macroalgae indicators at the standardized values

$$\hat{I} = \boldsymbol{L} \cdot [\hat{\mu};\ \widehat{\boldsymbol{\beta}}]^T \quad (5)$$

with a variance of

$$V[\hat{I}] = \boldsymbol{L} \cdot \begin{bmatrix} V[\hat{\mu}] & 0 \\ 0 & V[\widehat{\boldsymbol{\beta}}] \end{bmatrix} \cdot \boldsymbol{L}^T \quad (6)$$

The indicator will be approximately normal distributed according to the central limit theorem, and the distribution of the indicator is found by Monte Carlo simulation, calculating the proportion of simulations that fall in the different ecological status classes. These five proportions describe the probability of the indicator being in each of the five status classes.

4 Documentation of routines

MacroalgaeIndicator()

Description
Computes the mean and standard error for one of the three macroalgae indicator as well as probabilities for the five ecological status classes of the Water Framework Directive.

Usage
MacroalgaeIndicator(indicator,df,boundaries,depth_cutoff,….)

Arguments

indicator	Name of the indicator to be calculated. Possible values are "CumulativeCover", "PropOpportunist", and "NPerennials"
df	A dataframe with monitoring data for macroalgae according to the Danish National Marine Monitoring Program. The dataframe should contain the following variables: - 'kildestationsnavn' giving an identifier for the monitoring transect name - 'dato' giving the date in format yyyymmdd - 'dybde' giving the depth of the observation in m - 'hardbund_daekpct' giving the percentage cover of suitable substrate for macroalgae - 'totcover_daekpct' giving the total cover of macroalgae relative to the suitable substrate - 'art_daekpct' giving the species-specific cover relative to the suitable substrate - 'proevetager' identifier the person carrying out the monitoring transect - 'steneck' giving the Steneck number of the species - 'Growth_strategy' as 'P' for perennial, 'O' for opportunist and 'C' for crustforming algae
boundaries	A list with four elements giving the boundary values of the standardized indicator for cumulative cover. The boundary values should be given in the order high-good, good-moderate, moderate-poor, poor-bad.
depth_cutoff	The cut-off depth for physical exposure. Observations with depths shallower than this value are not used in the indicator calculation.
std_depth=7	Parameter describing the depth for which the indicator is calculated. Default is 7 m for cumulative cover; for fraction of opportunists and number of perennial species the option has no effect.
std_haardsub=50	Parameter describing the suitable substrate percentage for which the indicator is calculated. Default is 50.
n_iter=10000	Maximum number of iteration for Montecarlo simulation.
Indicator_pred_pct=0.50	The percentile of the distribution used as indicator value. Default is 0.50 (median).

Details
See documentation above.

Value
The function returns a list with the following descriptive statistics of the estimated macroalgae indicator: mean, median, P_High (probability of high status), P_Good (probability of good status), P_Moderate (probability of moderate status), P_Poor (probability of poor status), P_Bad (probability of bad status)

Authors
Jacob Carstensen, Aarhus University (jac@bios.au.dk)
Philippe Massicotte, Laval University (pmassicotte@hotmail.com)

Example
```
# Calculate cumulative cover indicator for a 6-year period
# with data from Als Fjord in Denmark

MacroalgaeIndicator (“CumulativeCover”,df=alsfjord,
c(74.2,40.8,20.4,13.4),3,10)
```

References

Carstensen, J., Krause-Jensen, D., Dahl, K., Henriksen, P. 2008. Macroalgae and phytoplankton as indicators of ecological status of Danish coastal waters. National Environmental Research Institute, University of Aarhus. 90 pp. - NERI Technical Report No. 683. http://www.dmu.dk/Pub/FR683.pdf

Carstensen, J., Krause-Jensen, D., Josefson, A. 2014. Development and testing of tools for intercalibration of phytoplankton, macrovegetation and benthic fauna in Danish coastal areas. Aarhus University, DCE – Danish Centre for Environment and Energy, 85 pp. Scientific Report from DCE – Danish Centre for Environment and Energy No. 93. http://dce2.au.dk/pub/SR93.pdf

Carstensen, J., Lindegarth, M. 2016. Confidence in ecological indicators: A framework for quantifying uncertainty components from monitoring data. Ecological Indicators 67: 306-317.

Krause-Jensen, D., Laursen, J.S., Middelboe, A.L., Stjernholm, M. 2001. NOVA – Teknisk anvisning for marin overvågning. Ch. 12. National Environmental Research Institute. Ministry of Environment and Energy, Denmark. (In Danish).

Sokal, R.R. & Rohlf, F.J. 1981: Biometry (second edition). Freeman, San Francisco.

Steneck, R.S. & Dethiers, M.N. 1994: A functional group approach to the structure of algal-dominated communities. - Oikos 69: 476-498.