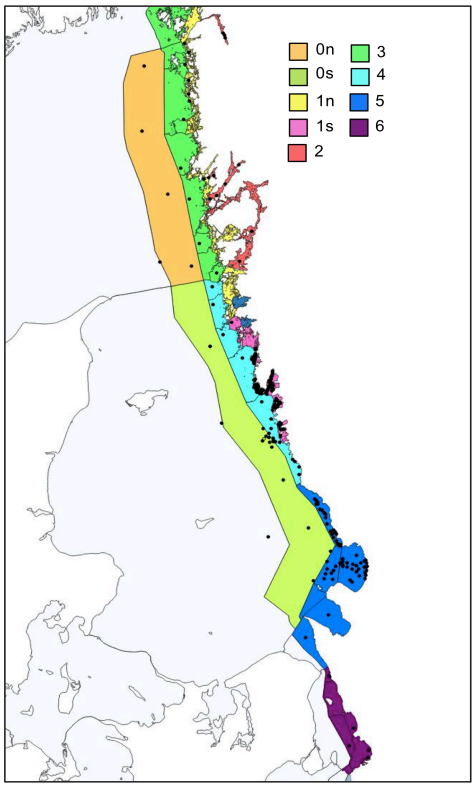
MONITORING OF BENTHIC FAUNA FOR THE MSFD ON THE SWEDISH WEST-COAST:

Modelling precision and uncertainty of current and future programs using WATERS uncertainty framework



Mats Lindegarth, Mats Blomqvist, Marina Magnusson and Rutger Rosenberg

WATERS Report no. 2014:3

WATERS Report no. 2014:3

MONITORING OF BENTHIC FAUNA FOR THE MSFD ON THE SWEDISH WEST-COAST:

Modelling precision and uncertainty of current and future programs using WATERS uncertainty framework

Mats Lindegarth, University of Gothenburg

Mats Blomqvist, Hafok AB

Marina Magnusson and Rutger Rosenberg, Marine Monitoring AB

WATERS partners:

GÖTEBORGS UNIVERSITET

Stockholms universitet

UMEÅ UNIVERSITET

AARHUS UNIVERSITY

SVERIGES VATTENEKOLOGER

Marine Monitoring RESEARCH & CONSULTING

SLU

SMHI

TOXICON AB

Hafok AB

AquaBiota WATER RESEARCH

WATERS: Waterbody Assessment Tools for Ecological Reference conditions and status in Sweden

WATERS Report no. 2014:3.
Title: Monitoring of benthic fauna for the MSFD on the Swedish west-coast: Modelling precision and uncertainty of current and future programs using WATERS uncertainty framework.
Publisher: Havsmiljöinstitutet/Swedish Institute for the Marine Environment,
P.O. Box 260, SE-405 30 Göteborg, Sweden
Published: November 2014
ISBN 978-91-9806646-9-8

Please cite document as:
Lindegarth, M. Monitoring of benthic fauna for the MSFD on the Swedish west- coast: Modelling precision and uncertainty of current and future programs using WATERS uncertainty framework., WATERS Report no. 2014:3. Havsmiljöinstitutet, Sweden.

http://www.waters.gu.se/rapporter

WATERS is a five-year research programme that started in spring 2011. The programme's objective is to develop and improve the assessment criteria used to classify the status of Swedish coastal and inland waters in accordance with the EC Water Framework Directive (WFD). WATERS research focuses on the biological quality elements used in WFD water quality assessments: i.e. macrophytes, benthic invertebrates, phytoplankton and fish; in streams, benthic diatoms are also considered. The research programme will also refine the criteria used for integrated assessments of ecological water status.

This report is the result of an additional task to WATERS WP 2.2 and 3.1 commissioned by the Swedish Agency for Marine and Water Management (contract dnr: HaV 03575-2013) dealing with uncertainty of current monitoring programmes in the perspective of the EU Marine Strategy Framework Directive and to some degree the Water Framework Directive. We analyses uncertainty of current monitoring and evaluate different options for dimensioning and sampling strategy for future monitoring programs.

WATERS is funded by the Swedish Environmental Protection Agency and coordinated by the Swedish Institute for the Marine Environment. WATERS stands for 'Waterbody Assessment Tools for Ecological Reference Conditions and Status in Sweden'. Programme details can be found at: http://www.waters.gu.se

Table of contents

Executive summary

This report presents a study of the usefulness of current monitoring of benthic invertebrate fauna in Västerhavets River Basin District (VRBD) for status assessment according to the Marine Strategy Framework Directive (the MSFD) and to some degree for the Water Framework Directive; WFD). The analyses are based on a coherent methodology proposed as a general framework for WFD assessments, developed within the research programme WATERS. The methodology involves partitioning of spatial and temporal sources of variability using existing data. Quantitative information on the importance of different sources of variability and information on the design and dimensioning of monitoring programmes can be used to model uncertainty of status assessments under different scenarios. Specific aims of the study were to:

i. Estimate spatial and temporal sources of variability in all water body types in VRBD;

ii. Estimate precision and confidence in classification for individual water body types and water bodies using existing monitoring programmes;

iii. Evaluate precision and confidence in classification for individual water body types and water bodies for a number of selected scenarios for revised monitoring programmes;

iv. Analyse statistical power for detection of trends at the level of stations and water body types.

More than 3000 samples form the years 2001-2012 were used for the first three items. The Swedish indicator BQI as well as species richness, Shannon-Wiener and Margalefs diversity indices and biomass. In initial assessment showed that (1) BQI and the different measures of diversity showed similar spatial and temporal patterns, (2) the benthic fauna differs among water body types and that (3) all variables were non-linearly related to sampling depth. The main finding the regarding the four aims (i-iv) are outlined below.

(i) The importance and size of spatial and temporal sources of variability for all types and variables are compiled in tables 4.1 – 4.3. These were used to address subsequent aims but can also be used as a ”library of uncertainties” for further assessments of monitoring designs for the WFD and MSFD. In general the analyses indicated that BQI was the most precise indicator, spatial variability among water bodies and stations were particularly dominant and that much of the spatial variability was associated with differences in sampling depth.

(ii) The precision of a monitoring programme depends on patterns of variability and on its design and dimensioning. Analyses of existing programmes showed that there are large differences in precision among (1) water body types and (2) the level of spatial and temporal aggregation. The precision (and uncertainty) of mean estimates were assessed at the level of individual years and within 6-year periods at the level of stations, water bodies and water body types. These analyses revealed that current monitoring in off-shore areas have a better precision than coastal water body types and that the standard error varies from 2- 30% of the mean BQI depending on spatial and temporal scale. Note that unlike the present Swedish WFD assessment criteria, both confidence of mean estimates and of status classification were evaluated.

(iii) Using estimates from (i) the precision of alternative strategies for monitoring were modelled in terms of (1) dimensioning (number of samples, stations, years) and (2) structure (revisiting or selecting new stations, i.e. ”crossed” vs. ”nested” designs). The aim was not to develop specific sampling designs but rather to evaluate effects of various scenarios for monitoring. Nevertheless, the precision of specific monitoring designs can be addressed graphically using figures 4.17 – 4.25. Generic conclusions can be summarised by noting that the uncertainty of mean estimates in water bodies and types are largely determined by the number of stations sampled, that the number of samples per station is only important for the precision within stations and that given a ”crossed” design, the number of years sampled has a small effect on improving the precision of 6-year means. However, if a nested design is used, the number of years has a large effect on the 6-year mean. Analyses show that the uncertainty may be halved using a similar number of samples compared to a crossed design. Alternatively a precision comparable to that of a crossed design can be achieved with substantially fewer samples. Nested designs, however, cannot be used to evaluate trends at the level of stations, which may be a concern for some purposes.

(iv) Finally, the planned analyses of statistical power of trends at stations and in water body types were not performed because (1) the dynamics of benthic communities differed strongly among stations and areas and (2) the number of stations with sufficiently long trends was small. The analyses showed strong and significant trends at many off-shore stations, which mean that programs are powerful enough to detect trends. Many coastal stations, however, showed strong cyclic patterns and linear trends were less relevant.

Svensk sammanfattning

Denna rapport sammanfattar arbete med att utvärdera övervakningen av mjukbottenfauna i Västerhavets vattendistrikt med avseende på dess användbarhet för statusbedömning enligt Havsmiljödirektivet (och i viss mån Vattendirektivet). Arbetet bygger på en metodik som utvecklats inom forskningsprogrammet WATERS som handlar om Vattendirektivets bedömningsgrunder. Denna metodik involverar beräkning av rumsliga och tidsmässiga variationskällor med hjälp av befintliga data. Denna information används tillsammans med information om provtagningens dimensionering och utformning för modellering av skattnings- och klassificeringsosäkerhet. Specifika målsättningar var att:

v. Skatta rumsliga och tidsmässiga variationskällor i västerhavets samtliga vattentyper;

vi. Beräkna precision och osäkerhet i klassning av enskilda vattentyper och vattenförekomster med befintliga program;

vii. Utvärdera precision och osäkerhet i klassning av enskilda vattentyper och vattenförekomster med ett urval av tänkbara scenarier för reviderade program;

viii. Analysera statistisk styrka för upptäckt av trender på stations- och typnivå.

För de tre första uppgifterna användes data från totalt över 3000 prover från åren 2001-2012. Analyser gjordes på den svenska indikatorn BQI men i viss mån även på artrikedom, Shannon-Wieners och Margalefs index samt på biomassa. En inledande översikt av data från dessa år visade i sammanfattning att (1) BQI och de olika diversitetsindexen visade liknande rumsliga och tidsmässiga mönster, (2) det finns skillnader mellan vattentyper och att (3) alla variabler visade icke-linjära samband med djupet.

(i) Betydelsen och storleken av rumsliga och tidsmässiga variationsbidrag för samtliga vattentyper och för alla undersökta variabler finns sammanställda i tabellerna 4.1 – 4.3. Förutom att de använts för att angripa resterande frågeställningar, kan dessa användas som ett ”osäkerhetsbibliotek” för framtida utvärderingar av alternativa övervakningsprogram inom Havsmiljödirektivet och Vattendirektivet. Andra viktiga slutsatser inom denna del var:

- En generell slutsats är att precisionen för ett medelvärde givet ett visst övervakningsprogram kommer att variera på ett förutsägbart sätt mellan indikatorer. Bäst precision kommer BQI att ha, följt av diversitetsindexen (artrikedom, Shannon-Wiener och Margalef) och sämst precision kommer biomassan att ha.

- Variationen mellan vattenförekomster och stationer dominerar över den tidsmässiga variationen och rumsliga mönster är förhållandevis stabila mellan år.
- Betydelsen av rumsliga variationskällor återfinns i alla vattentyper, men storleken varierar. Skagerraks fjordar, Kattegats kustområden och Öresund är speciellt variabla mellan stationer och / eller vattenförekomster.
- Initiala analyser visar att en stor del av variationen mellan stationer kan förklaras av skillnader i djup. Kunskap om djup kan därmed inkorporeras för att minska osäkerheten hos medelvärden. Utveckling av rutiner för att praktiskt göra detta inom ramen för statusklassning bör prioriteras.

(ii) Precisionen hos ett övervakningsprogram beror av variationsmönster och övervakningens dimensionering och utformning. Analyser av pågående program visar att precisionen varierar (1) mellan vattentyper och (2) mellan tids- och rumsskalor. Precision (och därmed osäkerhet) utvärderades för medelvärden skattade ”inom år” och ”inom 6-års perioder”, för stationer, vattenförekomster och vattentyper. Några specifika slutsatser var:

- Övervakningen i Skagerraks och Kattegats utsjöområden, samt delar av Skagerraks yttre delar har generellt högre precision än i kustområdena, speciellt Kattegat och Öresund.
- För BQI varierar osäkerheten uttryckt som standardfel (SE) mellan 2-30% av medelvärdet mellan olika skalor. Motsvarande siffror för biomassa är 2-200%. Som nämnts tidigare kan denna osäkerhet reduceras om hänsyn tas till stationsdjup.
- Givet en rumslig skala är precisionen generellt något bättre för medelvärdesskattningar över 6-års perioder jämfört med enskilda år.
- I enlighet med kraven inom vattendirektivet (men i kontrast till nuvarande bedömningsgrund), presenterades skattningar på klassificeringsosäkerhet. Analyserna visar för första gången hur en sådan process kan utformas för dess data och antyder att nuvarande program ofta har tillräcklig precision för att åstadkomma meningsfulla klassningar.

(iii) Med hjälp av skattade variansbidrag (i) modellerades precisionen hos alternativa övervakningsstrategier med avseende på (1) dimensionering (antal prover, stationer, år) och (2) övervakningsstrategier (återbesök eller nyetablering av stationer under varierande antal år). Syftet var inte att komma med konkreta förslag på dimensionering eller strategier, utan att bidra med underlag för framtida beslut genom att undersöka effekten av olika övervakningsscenarier. Frågor om precision för specifika situationer kan med fördel utvärderas grafiskt i figurerna 4.17 – 4.25. Övergripande slutsatser från dessa analyser är dock att:

- Osäkerheten hos medelvärden för vattenförekomster och vattentyper styrs till stor del av antalet stationer som provtas.
- Antalet prover per station är viktig stort sett endast för precisionen inom en station. Även vid ett litet antal prov per station är osäkerheten förhållandevis liten.
- Antalet år som provtas har en relativt liten betydelse för standardfelet av ett medelvärde skattat inom en 6-års period. Effekten på konfidensintervall är dock något större.

- Jämförelser mellan övervakning med återbesökta stationer kontra nyetablerade stationer antyder betydande skillnader i precision över en 6-års period. I enlighet med tidigare slutsatser om betydelsen av antalet stationer och den ringa effekten av antalet provtagna år i ett upplägg med återbesök, antyder analyserna att osäkerheten kan halveras om nya stationer besöks varje år. I ett upplägg med nya stationer bidrar varje provtaget år med en bättre rumslig täckning och kan därmed minska osäkerheten, medan återbesöken tillför relativt lite ”ny information”. På nivån av vattentyp och / eller vattenförekomst kan därför en önskad precision åstadkommas med ett betydligt lägre antal prover med nya stationer jämfört med återbesök. Nackdelen med denna strategi är dock att den omöjliggör trendanalyser på enskilda stationer och potentiellt ökar osäkerheten i trender på högre rumslig nivå. Även om avvägningar mot andra målsättningar måste göras antyder dessa analyser att provtagning i en lägre frekvens enligt en återbesöksstrategi förmodligen har liten negativ inverkan på precisionen men att allokering av prover till nya stationer har stor potential att förbättra statusbedömningen inom en given rumslig bedömningsenhet.

(iv) Planerade analyser av statistisk styrka av trender hos stationer och vattentyper utfördes inte eftersom (1) dynamiken varierade stort mellan lokaler och områden och (2) antalet trendlokaler per typ var litet och eftersom längden på tidsserierna varierade stort. Analyserna visade tydliga och signifikanta trender på många djupa utsjöstationer. Detta innebär att programmens styrka ofta var tillräcklig för att upptäcka förändringar. Ofta var dessa så stora som 0.1 – 0.2 BQI per år. Många kuststationer däremot uppvisade tydligt cyklisk dynamik och linjära trender var här mindre relevanta. Sammanfattningsvis antyder dessa resultat att de pågående programmen kan upptäcka, lång- och kortsiktiga, förändringar hos bottenfaunan. Dessa förändringar förefaller ha varit relativt dramatiska under de senaste 20-40 åren.

1 Introduction

1.1 Status assessment using monitoring data

The pressure on the marine environment and the demand for its valuable ecological goods and services is increasing. Partly as a response to this, we have during the last few decades witnessed world-wide development of new regulations and legislation for the protection of the marine environment (Moksnes et al. 2013). In the European countries, international development and national implementation of directives such as the "The Habitats directive" (HD), "The Waters Framework Directive" (WFD) and "The Marine Strategy Framework Directive" are particularly important examples.

For the European scene, these directives have introduced a range of new concepts and put new and potentially costly demands on management systems and authorities. These novelties involve among other things (1) international coordination of biological "indicators" (e.g. quality elements [WFD] and descriptors [MSFD]), (2) definition of desired states and thresholds, (3) systems for integrated assessment and (5) definition of spatial and temporal units for assessments. One consequence of these developments is that they define in a more specific way the requirements in terms of knowledge, data and assessment procedures, than what was previously the case. This is perhaps particularly true for the requirements on monitoring programmes.

For Sweden and other European countries, the implementation of these new directives (i.e. the WFD and the MSFD) affect what we need to measure, when we need to measure it and where we need to measure it. It is worth pointing out that the demands defined in the directives may sometimes be partly conflicting among new and old legislation and may not always be sufficient for maintaining the aims of previous environmental aims. One such area is the strong focus on status assessment in defined spatial units (e.g. water bodies, water body types, regions) and temporal units (6-year assessment period) in the WFD and MSFD. Previously, monitoring have at least in Sweden been more concerned with monitoring the effects of point sources and temporal trends at a number of more or less independent, strategically selected "stations". These two perspectives have fundamental consequences for the design of monitoring programmes, including those on benthic fauna. Nevertheless, it is important to find explore ways in which monitoring can be adjusted so that they combine different perspectives and meet demands from different policy objectives. This study is intended as a contribution in that context.

1.2 Assessing status of benthic assemblages using the BQI

1.2.1 The Water Framework Directive (WFD)

The marine benthic quality index (BQI; Rosenberg et al. 2004, Blomqvist et al. 2006, Leonardsson et al 2009) is a central indicator for assessing ecological status of soft-sediment assemblages in Swedish coastal waters (i.e. the WFD). To avoid confusion with a

limnic index it is called BQI_m in the current regulations from the Swedish Agency for Marine and Water Management "Havs- och vattenmyndighetens föreskrifter om klassificering och miljökvalitetsnormer avseende ytvatten" (HVMFS 2013:19).

The BQI is a general indicator of disturbance based on three components: the relative abundance of sensitive and tolerant species, species richness and abundance. The index is calculated as:

$$BQI = \left[\sum_i^S \left(\frac{N_i}{N_{tot}} * Sensitivity_i\right)\right] * log_{10}(S+1) * \left(\frac{N_{tot}}{N_{tot}+5}\right),$$

where S is the total number of species (for which there is a sensitivity assessment), N_i is the abundance of the *ith* species, N_{tot} is the total abundance (including only species with sensitivity assessments) and $Sensitivity_i$ is the sensitivity value for the *ith* species. The sensitivity of a species is determined by whether its distribution is associated with undisturbed, species rich areas. Tolerant species are "by definition" found in disturbed areas, and sensitive species are missing in these areas. In the Skagerrak, Kattegat and the Öresund, sensitivity has been determined quantitatively from a large number of samples, while in the Baltic sensitivity values have been adjusted using expert opinion. It is worth pointing out that the BQI is conceptually based on and consistent with the general model for responses of macrobenthic assemblages to organic enrichment and pollution (Pearson and Rosenberg 1978). This includes the definition of class boundaries, which can be related to different stages of succession according to the "Pearson-Rosenberg model" (Fig. 1.1). Furthermore, several studies have shown that the BQI is responds to a wide range of disturbances and that it is correlated to other indices used for assessing the quality of benthic habitats (e.g. Josefsson et al. 2009).

Tabell 1.1. Klassgränser för klassificering av status uppdelat per typ. Numrering av typer enligt typindelning i NFS 2006:1.

			BQI_m			
Bassäng	**Typ nr**	**Djupstrata**	**HG**	**GM**	**MO**	**OD**
Västerhavet						
	1-6 och 25	5-20 m	13,9	10,3	6,9	3,4
	1-6 och 25	> 20 m	15,7	12,0	8,0	4,0
Östersjön						
	7	5-60 m	10,7	4,0	2,7	1,8
	8	5-60 m	10,5	3,5	2,3	1,6
	9	5-60 m	10,7	4,0	2,7	1,8
	10	5-60 m	9,3	4,0	2,7	1,8
	11	5-60 m	8,0	4,0	2,7	1,8
	12	5-60 m	10,7	4,0	2,7	1,8
	13	5-60 m	9,0	3,0	2,0	1,3
	14	5-60 m	10,7	4,0	2,7	1,8
	15	5-60 m	10,7	4,0	2,7	1,8
	24	5-60 m	7,7	3,0	2,0	1,3
Bottniska viken						
	16	> 5 m	10,7	4,0	2,7	1,8
	17	> 5 m	10,0	4,0	2,7	1,8
	18	> 5 m	10,0	4,0	2,7	1,8
	19	> 5 m	10,0	4,0	2,7	1,8
	20	> 5 m	10,0	4,0	2,7	1,8
	21	> 5 m	10,0	4,0	2,7	1,8
	22	> 5 m	7,5	2,0	1,3	0,9
	23	> 5 m	6,3	1,5	1,0	0,7

Figure 1.1 Class boundaries for BQI in Swedish coastal waters (reproduced from HMVFS 2013: 19, in Swedish).

Since its implementation, the Swedish assessment criteria for BQI within the WFD has been based on a procedure, which accounts for uncertainty within a water body, by calculating a 80% one-sided (lower) confidence limit, using a bootstrap procedure, and thereafter assigning a classification depending on how this limit is related to the class boundaries (HMVFS 2013: 19). According to the criteria and the handbook, these calculations should be based on at least five independent samples within a water body and year (i.e. b=1, d=5 and n=1 according to the terminology used here [see section 3]). Note that this means that the current assessment method is only strictly applicable to single water bodies, during single years and when five independent samples are available (the latter is a condition which very rarely are met in Skagerrak and Kattegat but more so in the Baltic).

1.2.2 The Marine Strategy Water Framework Directive (MSFD)

The BQI was widely accepted as an important tool in the new Swedish assessment criteria for the WFD and when the Swedish definitions of "good environmental status" (GES) according to the MSFD was defined, the BQI was employed as an indicator to assess quality for three descriptors: biodiversity (criterion 1.6 "Livsmiljöns tillstånd"), eutrophication criterion 5.3 "Fleråriga växter uppvisar naturlig utbredning och ingen minskning av syrekoncentrationer till följd av övergödning förekommer") and seafloor

integrity (criterion 6.2 "Det bentiska samhälets tillstånd"; HMVFS 2012: 18). In coastal types, which are included both in the WFD and the MSFD, as well as in off-shore areas, only included in the MSFD, the regulations have defined the "good"-"moderate" boundary (*sensu* WFD) as the minimum threshold to achieve GES (*sensu* MSFD).

The strategy to use an established and successful indicator, such as the BQI, also for the descriptors in the MSFD is of course logical and efficient. Nevertheless, there are aspects which may complicate the transfer of procedures developed for the WFD in coastal waters into the MSFD domain. One such aspect is use of the same class-boundaries as those used in the WFD. This aspect is further investigated within WATERS WP3.1. Another aspect is the fact that assessment procedures and monitoring requirements for BQI defined in the Swedish regulations (HMVFS 2013: 19), are developed to suit the WFD typology. This typology and monitoring requirements involves sampling in water bodies, which is the main unit for status assessments in the WFD. The MSFD, on the other hand, does not involve the same spatial units and Swedish regulations and future practical implementation have instead defined water body types as the operative unit for status assessment. Thus, simply referring to the WFD regulations (e.g. HMVFS 2013: 19) as is done now in the MSFD regulations (HMVFS 2012: 18) does not fully cover all necessary steps in an assessment procedure. Thus, the aim of this study is to evaluate a methodology for aggregating data on BQI at spatial and temporal scales which fully comply with the requirements of the Swedish regulations for the MSFD.

Finally, it is worth pointing out that neighbouring countries have decided upon a range of different indicators for benthic assemblages (e.g. HELCOM Secretariat 2013). These include related indices and aspects of biodiversity and biomass. Therefore, in order to assess the robustness of conclusions reached for the BQI, some parts of this report also involve analyses of other potential indicators.

1.3 The uncertainty framework

Lindegarth et al. (2013a) proposed that uncertainty should be assessed in the Swedish assessment WFD criteria by means of framework-based estimation of variance components using mixed models (e.g., Bolker et al. 2009). The framework applies general procedures for uncertainty (or error) propagation (e.g., Cochran 1977, Taylor 1997) and is based on scientific studies demonstrating the need for the combined assessment of various sources of uncertainty (e.g., Clarke et al. 2002, 2006a,b, Clarke & Hering 2006). By explicitly adapting to temporal and spatial scales relevant to the WFD and / or the MSFD, the framework constitutes a general basis for further work in WATERS and in Swedish water quality assessment.

Details of this framework are given in Lindegarth at al. (2013) and references therein, but its main features are: (1) partitioning of variability into fixed and random components using linear models, (2) calculation of total variability by combining uncertainty components using formulae for error propagation and (3) estimation of uncertainty according to the definitions given by the WFD (or here the MSFD). The WFD and its

guidance documents define two aspects of uncertainty confidence in the mean estimate and confidence in classification (EC 2003a, b).

Thus, first the framework involves specifying a general linear model including random (CAPITAL letters) and fixed (lowercase letters) factors and interactions. These components can be categorised as temporal, spatial, and spatio–temporal interactions and variability associated with sampling and measurement. A general example of this was given in Lindegarth et al. (2013)

$$y = \mu + \underbrace{year + YEAR + season + SEASON{\times}YEAR + DIURNAL + IRREGULAR}_{temporal\ sources\ of\ uncertainty} + \underbrace{gradient + GRADIENT + PATCHINESS}_{spatial\ sources\ of\ uncertainty} + \underbrace{YEAR{\times}GRADIENT + SEASON{\times}GRADIENT}_{spatio-temporal\ interactions} + \underbrace{sampling\ devices + PERSON + instrument + REPLICATE}_{sampling\ and\ measurement\ uncertainties} \quad \text{[Eq.1]}$$

Using such comprehensive models, the importance of different random variance components, e.g. s^2_{YEAR}, $s^2_{PATCHINESS}$ and $s^2_{REPLICATE}$, are separated and estimated. Specific models and details on methods for partitioning, relevant to this study are given in section 3.

Second, because mean estimates at particular spatial and temporal scales may be affected by multiple sources of uncertainty, the total variance ($V[\bar{y}]$) associated with a certain mean estimate ($\bar{y}$) is calculated. A general formulation of the total variance ($V[\bar{y}]$) affected by three random sources of variation (i.e., A, B, and C), each with *a*, *b*, and *c* levels, is that the sampling variance of a mean ($\bar{y}$) consists of three variance components, i.e. s^2_A, s^2_B, and s^2_C. The combined total variance of the estimated mean, $\bar{y}$, is estimated from the size of the variance components and the number of levels:

$$V[\bar{y}] = \frac{s^2_A}{a} + \frac{s^2_B}{b} + \frac{s^2_C}{c} \quad \text{[Eq. 2]}$$

Third, estimate of total variability is used to estimate confidence (or uncertainty). Thus, the total variability, $V[\bar{y}]$, is transformed into the standard error of the mean, $SE_{\bar{y}} = \sqrt{V[\bar{y}]}$. The standard error provides the basis for calculation of uncertainty. The first aspect of uncertainty mentioned in the WFD, the *confidence of a mean estimate*, i.e. the confidence interval is calculated as:

$$CI\% = \left[\sqrt{V[\bar{y}]} * t_{\alpha/2,df}; \sqrt{V[\bar{y}]} * t_{1-\alpha/2,df}\right] \text{[Eq. 3]}$$

where $t_{\alpha/2,df}$ and $t_{1-\alpha/2,df}$ are the percentiles of the *t*-distribution (usually the 2.5 and 97.5 percentiles, corresponding to α = 5%) with *df* effective degrees of freedom. If the degrees of freedom for $V[\bar{y}]$ exceed 30, the percentiles of the *t*-distribution can be approximated using the standard normal deviates, i.e., $z_{\alpha/2}$ and $z_{1-\alpha/2}$.

The *confidence in classification*, on the other hand, is a measure of the probability of a certain classification being correct. The confidence of five classes can be calculated using the normal distribution. For each class boundary in turn, we calculate the probability, p_i, of observing an indicator value of x or better if the true mean quality, μ, is equal to the class boundary, L_i:

$$p_i = \Pr(X \geq x \mid \mu = L_i) = 1 - \Phi\left[(x - L_i)/\sqrt{V[\bar{y}]}\right]$$

where **Φ** denotes the cumulative normal probability. This probability statement says that $\Pr\left(X \geq \mu + z_i\sqrt{V[\bar{y}]}\right) = p_i$, where $z_i$is the standard normal deviate corresponding to 1 – p_i and $\sqrt{V[\bar{y}]}$ is the standard error of the mean. We can turn this into a confidence statement by inverting it, giving:

$$Confidence\left(\mu \leq x + z_i\sqrt{V[\bar{y}]}\right) = p_i.$$

Thus we can calculate: confidence of class 5 = p_5, confidence of class 4 = $p_4 - p_5$, confidence of class 3 = $p_3 - p_4$, confidence of class 2 = $p_2 - p_3$, and confidence of class 1 = $1 - p_2$ (note that these five quantities sum to 1).

2 Objective

The general objective of the present report was to evaluate current monitoring of benthic invertebrates on the Swedish west-coast, mainly in the light of the requirements of the Marine Strategy Framework Directive (MSFD) but also with the Water Framework Directive (WFD) in mind, and to analyse effects of different options for design and dimensioning of future programs. Four specific task were defined in the contract:

ix. Estimation of spatial and temporal variance (uncertainty) components using existing data from 2001-2012.

x. Analysis of precision and uncertainty in classification for individual water body types and water bodies for existing monitoring.

xi. Assessment of precision and uncertainty in classification of water body types and water bodies using a selection of monitoring scenarios (determined at an initial dialog meeting).

xii. Analysis of statistical power for detecting temporal trends at the level of stations and water body type.

The aim of these analyses is to provide general guidance in further discussions on a future coordinated design of benthic monitoring for the MSFD and the WFD. To propose adjustments of current monitoring or specific designs of dimensioning is beyond the scope of this report. Such discussions need to take into account economic, logistic, ecological and other aspects, needs to be initiated by relevant authorities and needs to be coordinated nationally as well as internationally.

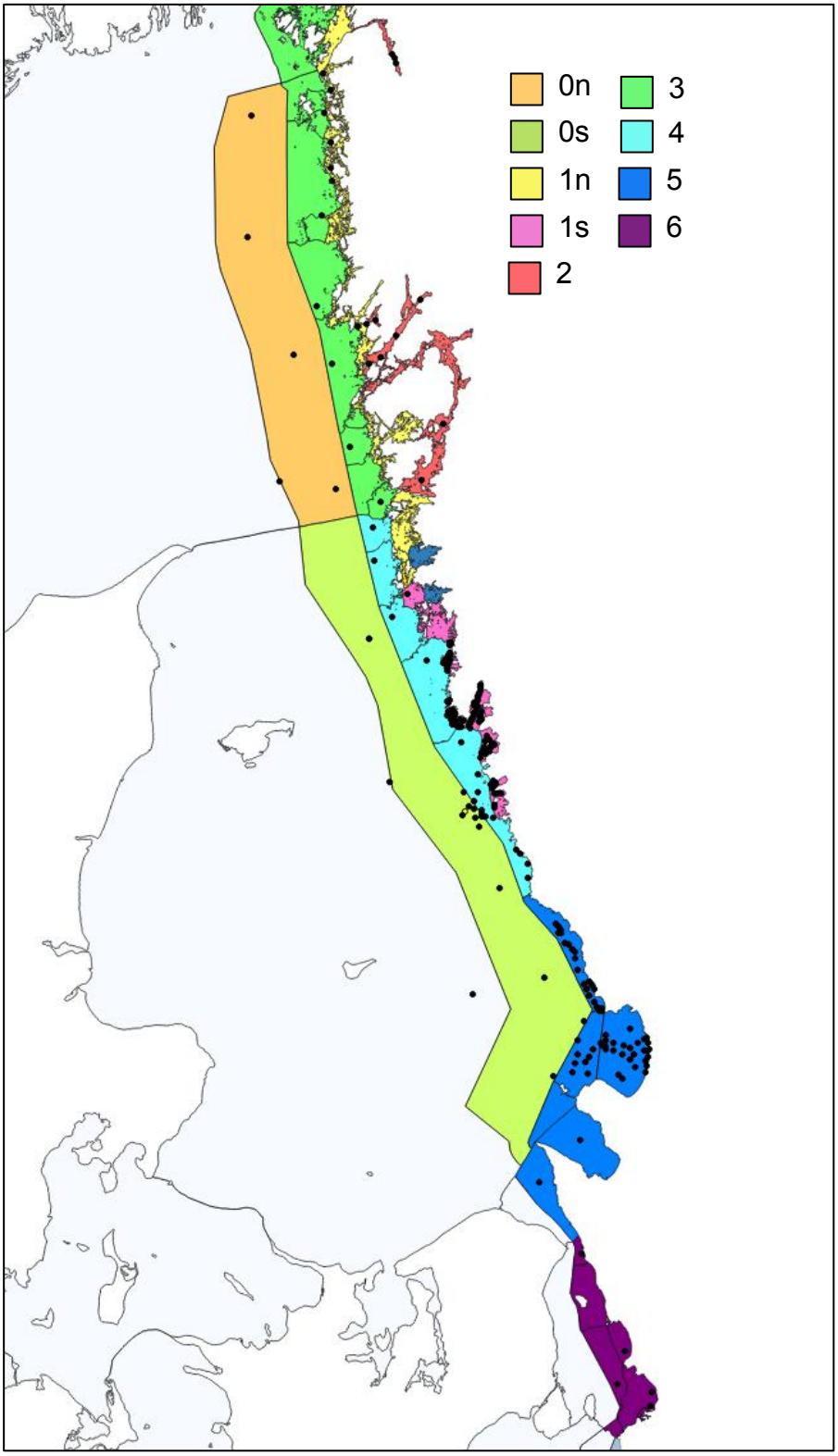


Figure 3.1. Map of water body types and sampling stations, which have been included in the study.

3 Methods

3.1 Typology and data

This study involves data from all coastal water body types on the Swedish west coast (except the transitional waters in "Göta- and Nordre älv" estuary (Fig. 3.1). Note that the off-shore "types" 0n (parts of off-shore Skagerrak) and 0s (parts of off-shore Kattegat) are not defined I the typology for the WFD. Nevertheless, for the MSFD they are appropriate spatial units. Note also that type 1 (Inner coastal Skagerrak and Kattegat have bee split up into northern (1n) and southern (1s) parts, in accordance with common practice.

For the purpose of the analyses, samples form depths shallower than 5 m were excluded. The number of stations and samples available vary among years and water body types (Table 3.1). In particular, the types in Kattegat (1s, 4 and 5), where occasional inventories were done in 2006, 2007 and 2012, vary substantially among years in the number of stations sampled. Furthermore, because the different stations originate from different water bodies and monitoring programs with partly differing purposes, the number of samples per station and year, as well as the number of years sampled vary among these stations (Table 3.2).

TABLE 3.1
Summary of number of stations per water body type and year for available data in Kattegat, Skagerrak and Öresund. Note that missing data in 2012 in some types is because they have not been reported when the data was assembled.

Water body type	Year 00	01	02	03	04	05	06	07	08	09	10	11	12
0n	3	5	4	4	4	4	4	4	4	4	4	4	0
0s	8	8	8	8	8	8	14	9	8	8	8	8	5
1n	0	0	5	5	5	5	5	5	5	5	5	5	0
1s	3	3	10	4	4	4	27	27	44	8	8	8	64
2	2	2	5	5	5	5	4	4	4	4	7	6	0
3	2	3	8	8	8	8	7	7	6	7	7	7	0
4	4	5	8	8	8	8	12	23	10	8	8	8	6
5	4	4	4	4	4	4	4	26	34	4	4	4	39
6	5	5	5	5	5	5	5	5	5	5	5	5	5

TABLE 3.2
Summary of typical number of levels for different components and depth intervals in existing monitoring programmes in Kattegat, Skagerrak and Öresund. Occasional inventories are not included. Number of stations used in section 4.3 are shown in bold. Numbers in brackets show minimum depths. Samples shallower than 5 m excluded from analyses.

Water body type	Water bodies c	Stations d	Samples n	Years b	Minimum depth	Maximum depth
0n	not relevant	**5**	**4**	**6**	67	106
0s	not relevant	**8**	5 or **4**	**6**	21	77
1n	**5**	**1**	**2**	**6**	28	60.5
1s	**4**	**1**	**5**	**6**	5 (2)	34
2	**3**	**1**, 2 or 3	**2** or 4	**6**	9.3	118
3	**5**	**1** or 2	**2** or 4	**6**	27	89
4	**5**	**1** or 4	**5** or 2	**6**	5 (2.3)	59
5	**3**	**1** or 2	**5**	**6**	5 (2)	23
6	**4**	**2** or 1	**5**	**6**	5 (3)	29

3.2 Estimation of variance components

A fundamental step in the assessment and modelling of the uncertainty of current and future monitoring designs, is the partitioning and estimation of different variance components. In order to do this we defined a number of statistical models for the partitioning of data according to different strategies. These mixed models (i.e. containing both fixed and random factors) were fitted to data on BQI, species richness, Shannon-Wiener and Margalefs diversity indices and biomass (transformed to ln[X+0.01]) and the importance of different random components were estimated using the restricted maximum likelihood (REML) method with the program R (specifically the libraries"lme4" version: 1.0-5 and "lmerTest" version 2.0-3; Bates et al. 2013, Kuznetsova et al. 2013, R Development Core Team 2008). These components are listed and explained in table 3.2.

3.2.1 Model of all water body types

To obtain a general assessment and comparison of the importance of different sources of uncertainty for the five response variables, an overall model including all water body types was defined (equation 4). The model included two fixed factors periods ("pe") and water body types (ty"), and three random factors years within periods ("YE(pe)"), water body within types ("WB(ty)") and stations within water bodies ("ST(WB,ty)") and a number of interaction terms as stations were regularly revisited during all years.

$$y = \mu + pe + ty + pe * ty + YE(pe) + ty * YE(pe) + WB(ty) + pe * WB(ty) + YE(pe) * WB(ty) + ST(WB, ty) + pe * ST(WB, ty) + YE(pe) * ST(WB, ty) + RES \quad \text{[Eq. 4]}$$

3.2.2 Models for individual water body types

Because patterns of variability may differ among water body types and because specific monitoring strategies may be developed for individual types, models were also developed fitted to data from separate types (equations 5 and 6). These models involved the same factors as the previous model but because data from only one type was included, they allowed calculation of type specific uncertainty components. One additional feature was also that equation 6 included a fixed factor depth ("de"), which was introduced to assess the potential for reducing uncertainty by accounting for sampling depth (note however that data from depths shallower than 5 m were excluded from all analyses). By introducing depth in the model we predicted that uncertainty would be reduced, but to further develop principles for monitoring designs is beyond the scope of this report. Nevertheless, interpreted sensibly conclusions about uncertainties from models not accounting for depth can still provide robust measures precision and guidance in design and dimensioning.

$$y = \mu + pe + YE(pe) + WB + pe * WB + YE(pe) * WB + ST(WB) + pe * ST(WB) + YE(pe) * ST(WB) + RES \quad \text{[Eq. 5]}$$

$$y = \mu + de + pe + YE(pe) + WB + pe * WB + YE(pe) * WB + ST(WB) + pe * ST(WB) + YE(pe) * ST(WB) + RES \quad \text{[Eq. 6]}$$

TABLE 3.2
Summary of variance components and levels of factor used for calculation of overall precision of scenarios in section 3.3.

Component	Interpretation	No. of levels.[§]
s^2_{YE}	variability among years within periods	b[#]
s^2_{WB}	variability among water bodies	c
$s^2_{ST(WB)}$	variability among stations within water bodies	d
s^2_{YE*WB}	interactive variability between years and water bodies	
$s^2_{YE*ST(WB)}$	interactive variability between years and station within water bodies	
s^2_e	residual variability among cores within stations and years	n

[§] a is reserved for the number of levels in the fixed factor *'period'*.
[#] the maximum number of levels in this factor is YE=6 (i.e. the number of years in an assessment period).

3.3 Precision of existing programmes

3.3.1 Precision of estimated means

As explained above, the standard error (SE) is a central measure of uncertainty, which is affected by the variability and the sampling design. Because variability and sampling intensity tend to differ among spatial and temporal scales, we must always assume that the uncertainty (e.g. SE) is specific for a certain combination of spatial and temporal units. The aims of this study define a number of such combinations, for which uncertainty

needs to be assessed. Using the model for individual water body types (equation 5), a number of expressions can be developed for each of these cases. Each expression, representing the error of estimated means at a certain temporal and spatial resolution, is the sum of different variance components divided by the number of measurements. The expressions needed to calculate errors are given in table 3.3. Following the initial dialogue in the beginning of the project, it was conclude that precision at the scale of water body types within MSFD assessment periods (eq. 11) and individual years (eq. 10) were of high priority. Additionally for coordination with other monitoring requirements, such as the WFD, the scale of water bodies within assessment periods (eq.9) and within years (eq. 8) as well as within stations (eq. 7) were also of interest. These combinations were therefore the main focus for analyses of precision of existing monitoring and for future scenarios.

TABLE 3.3
Summary of expressions for calculation of overall error at a number of combinations of temporal and spatial resolutions.

Resolution	Expression	Eq.
Error within stations and years	$V[\bar{y}_{ST\_YEAR}] = \frac{s_e^2}{n}$	7
Error within water bodies and years	$V[\bar{y}_{WB\_YEAR}] = \frac{s_{ST(WB)}^2}{d} + \frac{s_{YE*ST(WB)}^2}{d} + \frac{s_e^2}{dn}$	8
Error within water bodies and periods	$V[\bar{y}_{WB\_period}] = \frac{s_{YE}^2 * (1 - \frac{b}{YE})}{b} + \frac{s_{ST(WB)}^2}{d} + \frac{s_{YE*ST(WB)}^2}{bd} + \frac{s_e^2}{bdn}$	9
Error within water body types and years	$V[\bar{y}_{type\_YEAR}] = \frac{s_{WB}^2}{c} + \frac{s_{ST(WB)}^2}{cd} + \frac{s_{YE*WB}^2}{c} + \frac{s_{YE*ST(WB)}^2}{cd} + \frac{s_e^2}{cdn}$	10
Error within water body types and periods	$V[\bar{y}_{type\_period}] = \frac{s_{YE}^2 * (1 - \frac{b}{YE})}{b} + \frac{s_{WB}^2}{c} + \frac{s_{ST(WB)}^2}{cd} + \frac{s_{YE*WB}^2}{bc} + \frac{s_{YE*ST(WB)}^2}{bcd} + \frac{s_e^2}{bcdn}$	11

3.3.2 Confidence in classification

The current Swedish assessment criteria for BQI is based on a procedure which accounts for uncertainty within a water body, by calculating a 80% one-sided (lower) confidence limit, using a bootstrap procedure, and thereafter assigning a classification depending on how this limit is related to the class boundaries (HMVFS 2013: 19). This procedure implements the "precautionary principle" by applying a classification based on the lower confidence limit (i.e. not classifying according to mean, which is the most likely state), but it does not involve any probabilistic statement about the confidence in a certain classification (other than that the true mean is above the estimated limit with an 80% probability). Estimates of confidence in classification are actually required by the WFD (Clarke and Hering 2006, Clarke 2013).

Furthermore, according to the criteria and the handbook, these calculations should be based on at least five independent samples within a water body and year (i.e. b=1, d=5 and n=1 according to the terminology used here). Note that this means that the current

assessment method is only strictly applicable to single water bodies, during single years and when five independent samples are available (the latter is a condition which very rarely are met in Skagerrak and Kattegat but more so in the Baltic). Thus, at present there is no formal procedure to classify water body types at individual years, nor during whole assessment periods.

In order to address the second aim of this study, to assess confidence in classification of water body types during years and periods, procedures described in Lindegarth et al. (2013) were used. As described in section 1.3.3, these procedures are based on estimated standard errors (see 3.3.1), in combination with the standard normal distribution (i.e. $X \sim N(0,1)$) and existing class boundaries. Furthermore, to compare the result to current procedures, the 80% lower confidence limit was estimated using estimated standard error and the critical value of standard normal distribution (Z_{crit}=0.8416).

3.4 Precision of BQI under different monitoring scenarios

Having established estimates of current monitoring designs at various combinations of spatial and temporal scales, it is of great interest to evaluate scenarios for dimensioning and designs that in the future might increase the confidence and a constant resource or that might lower maintain confidence at a lower cost. Discussions at an initial meeting clarified a number of scenarios that were of interest for the Swedish Agency for Marine and Water Management. As described earlier, it was concluded that assessment of precision at the combinations of spatial and temporal scales listed in table 3.3 were of high relevance to the agency. These combinations were assessed according to two different aspects: (A) dimensioning (i.e. by evaluating effects of sample sizes in terms of samples (n) per station, stations per water body (d), stations per water body type (c*d) and number of years (b) per assessment period) and (B) whether a crossed or nested design are used.

3.4.1 Varying number of samples, stations and years within a period

Thus, using the equations in table 3.3 the absolute, relative precision and confidence intervals was assessed for BQI five different spatial and temporal resolutions for each of the nine water body types separately:

TABLE 3.4
Summary of combinations of spatial and temporal resolution modelled to assess precision in individual types.

Resolution	Range of levels
Precision within station and year	n=1-10
Precision within water body and year§	n=1, 5; d=1-30
Precision within water body and periods§	n=1; d=1-30; b=1, 6
Precision within water body types and year#	n=1, c*d=1-30
Precision within water body types and periods#	n=1, c*d=1-30; b=1,6

§ Note that the water bodies are not defined for the off-shore areas (0n and 0s). Thus estimates for these types also uncertainty due to larger scale processes.

For all coastal types (i.e. all except 0n and 0s), variability due to stations and water bodies are summed and stations are assumed randomly allocated among water bodies).

3.4.2 Crossed vs nested designs

Earlier analyses have suggested that more or less static spatial patterns of variability at the scale of stations and water bodies contribute substantially to the uncertainty of mean estimates. Temporal variability, however, contribute to a smaller degree. As a consequence of this we can expect that the most efficient way to obtain precise estimates, is to allocate resources to maximise the number of stations and / or water bodies. As a consequence it is interesting to assess different possibilities to increase spatial replication (i.e. number of stations and / or water bodies). One possible way to do this could be to opt for a nested monitoring design instead of a traditional crossed design (or rather a combination of approaches), which is currently totally dominant (Lindegarth et al. 2013b).

The crossed design is characterised by a set of "stations" (randomly or otherwise selected), which are repeatedly revisited on a yearly or monthly basis. The strength of such designs is that can be used to evaluate changes in time without confounding effects of spatial variability, and although the number of samples typically vary between n=1 to 5 per station among Swedish benthic monitoring programmes, this type of design is completely dominating. The linear model of a crossed design with *d* stations which are revisited in *b* years is defined as (using the same notation as earlier in this report):

$$y = \mu + YE + ST + YE * ST + RES \text{ [Eq. 12]},$$

and the variability around the total mean can be calculated as:

$$V[\bar{y}] = \frac{s^2_{YE}*(1-\frac{b}{Y})}{b} + \frac{s^2_{ST}}{d} + \frac{s^2_{YE*ST}}{bd} + \frac{s^2_e}{bdn} \text{ [Eq. 13]}.$$

The nested design, on the other hand, is characterised by random selection of new stations every year. Thus, individual stations are not revisited and changes at individual stations can therefore not be evaluated. If a number of stations are sampled each year in a water body or a type, trends in these spatial units can, however, be assessed. The potential strength of this approach is that a larger number of stations can be sampled within for example a water body type during an assessment period. The linear model of a nested design with *d* new stations in *b* years is defined as:

$$y = \mu + YE + ST(YE) + RES \text{ [Eq. 14]},$$

and the variability around the total mean can be calculated as:

$$V[\bar{y}] = \frac{s^2_Y*(1-\frac{b}{Y})}{b} + \frac{s^2_{ST(YE)}}{bd} + \frac{s^2_e}{bdn} \text{ [Eq. 15]}.$$

Because all current programs are based on crossed designs, component involving nested stations within years was calculated from the combined variability due to station and its interaction with years:

$$V[\bar{y}] = \frac{s^2_Y*(1-\frac{b}{Y})}{b} + \frac{(s^2_{ST}+s^2_{YE*ST})}{bd} + \frac{s^2_e}{bdn} \text{ [Eq. 16]}.$$

Differences in efficiency between crossed and nested design are likely to vary depending on the nature of spatio-temporal patterns, but initial analyses suggest that nested designs might be more precise at comparable number of samples in some circumstances (Lindegarth et al. 2013a).

Precision of crossed and within (a) water body types and (b) water bodies within 6-year assessment periods were assessed for BQI using estimates derived from the model incorporating all water body types (equation 4).

3.5 Detection of long-term trends

In order to evaluate the efficiency of monitoring stations for detecting long-term trends, thirty-two stations with particularly long time series were selected (supplementary material). These stations were distributed among all water body types except type 1n. Because these stations belong to different monitoring programs, the number of samples per station and year varies between 2 – 5 but for the majority of stations n=4 or 5. The number of years per station varies from 39-10 years and the earliest data are from 1973.

The original aim was to analyse "statistical power for detecting temporal trends at the level of stations and water body types". This task was not primarily motivated by the needs of the MSFD or the WFD which are mainly concerned with assessing status within assessment periods in larger spatial units and testing whether changes occur from one period to another. Another aspect is that initial analyses demonstrated strong and significant trends in many of the stations. This means that current sampling programs at these stations are sufficiently powerful to detect relevant trends. Therefore, potential lack of power is not generally an issue at these time-scales and with the current dimensioning.

Nevertheless, trends are clearly relevant for the interpretation of long-term changes and in contexts other than these directives. One difficulty, however, was the small number of long-term stations within water body types. Therefore aggregation of data at scales larger than stations were of limited value. Instead, the analyses were focussed on the consistency (or lack of consistency of temporal trends) among stations within types. These analyses can indicate whether trend-stations are likely to represent trends in larger areas or whether they are representative to single stations only. Finally, additional analyses were done to test whether variability in observed changes were related to the length of the time-series and whether they are explained by differences in station depth.

4 Results and discussion

4.1 General description of spatial and temporal patterns

This assessment involves data from 12 years from a total of more than 300 stations distributed among all water body types. The major trends and differences among water body types for the different variables are shown in figures 4.1-4.5. This initial overview of the 2 906 samples reveal several spatial and temporal patterns for the five response variables worth mentioning. First, it is evident that spatial and temporal patterns for all variables including species diversity, i.e. BQI, richness, Shannon-Wiener and Margalefs indices, are strongly correlated and very similar. This is very much expected but nonetheless striking and important. Biomass on the other hand, appears to show some spatial and temporal patterns, which are to some degree independent from those of the other variables. Second, there are more or less persistent differences among the different types but these differences vary among the different response variables. Third, very seldom can we observe simple unidirectional increasing or decreasing trends changes for any of the water body types or variables. Yearly averages typically display fluctuating patterns, which are often asynchronous among types and averages over 6-year assessment periods are usually small and variable among types. A fourth observation is that some water body types, appear to have similar dynamics, while others are more different. For example, the BQI in large parts of coastal Kattegat and the Öresund (4, 5 and 6) tend to increase from the first to second period, while changes in the Skagerrak and the off shore parts are less consistent. Note, however, that some of the spatial and temporal differences observed in figures 4.1 – 4.5 may be explained by differences in depth distribution of samples among types (e.g. stations in types 1s, 5 and 6 are generally placed in shallower waters than in other types) and by differences in sampling designs among years (e.g. occasional inventories in mainly types 1s and 4 during 2006-2008 and 2012 affect temporal trends in these areas).

One additional pattern worth mentioning is the relationship between indicators and depth (Fig. 4.6). The most striking pattern is shown by BQI, which show a strong but non-linearly increase with depth when all types are considered. The pattern is somewhat less clear for biomass but nevertheless the biomass is increasing and becoming less variable with depth. Whatever the processes and causes are behind these patterns, the aim of this general description is to introduce the overall patterns and trends in existing data as a background for further analyses of uncertainties and sampling designs.

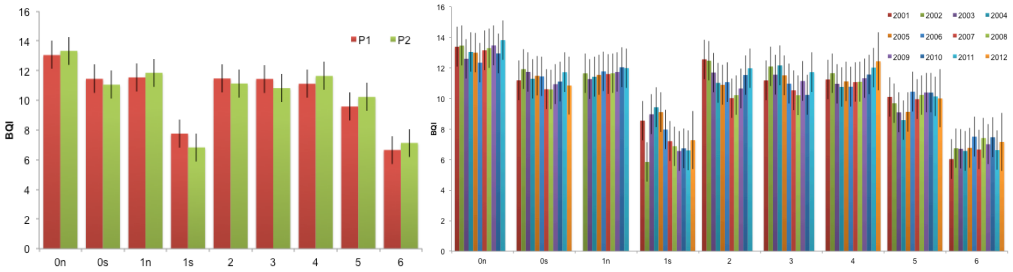


Figure 4.1. Estimated mean BQI in individual water body types in years between 2001-2012 (left) and periods (P1: 2001-2006 and P2: 2007-2012). Means include data from all available sources and depths.

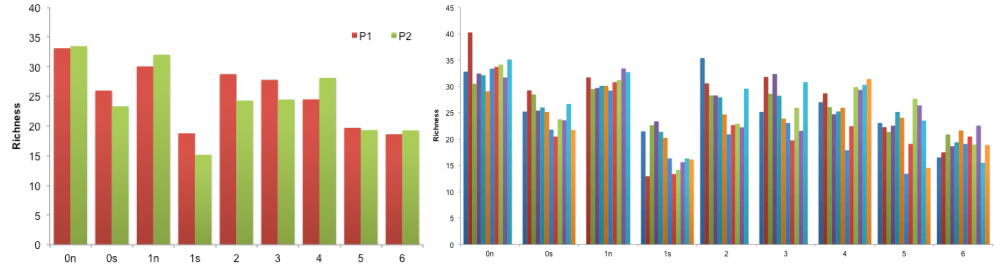


Figure 4.2. Estimated mean species richness in individual water body types in years between 2001-2012 (left) and periods (P1: 2001-2006 and P2: 2007-2012). Means include data from all available sources and depths.

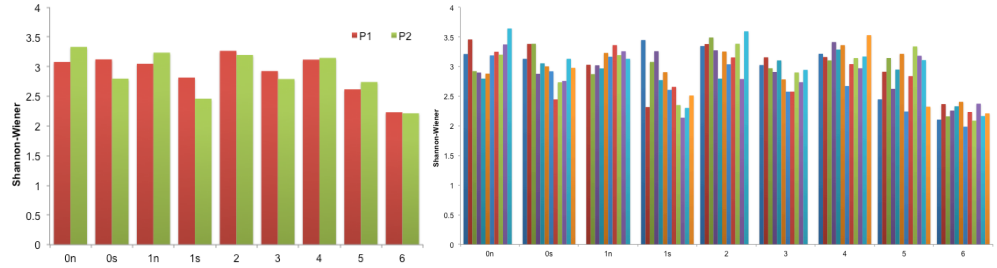


Figure 4.3. Estimated mean of Shannon-Wiener diversity in individual water body types in years between 2001-2012 (left) and periods (P1: 2001-2006 and P2: 2007-2012). Means include data from all available sources and depths.

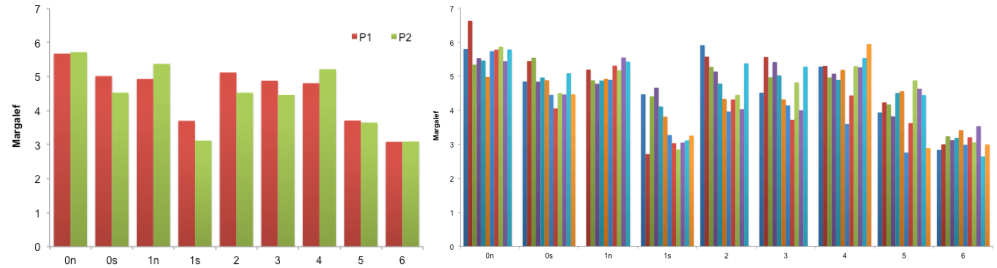


Figure 4.4. Estimated mean of Margalef index of diversity in individual water body types in years between 2001-2012 (left) and periods (P1: 2001-2006 and P2: 2007-2012). Means include data from all available sources and depths.

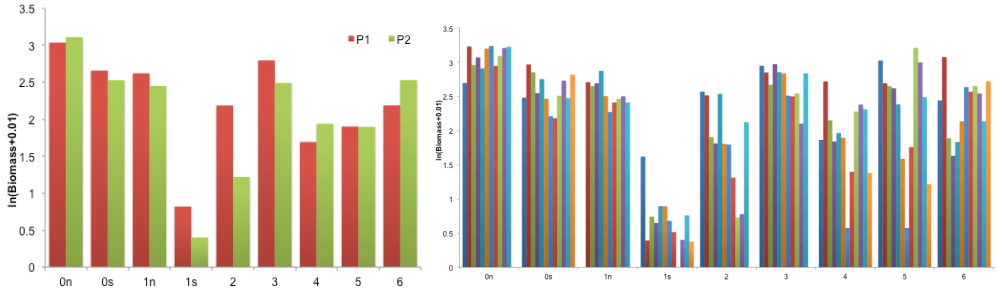


Figure 4.5. Estimated mean of ln(biomass+0.01) in individual water body types in years between 2001-2012 (left) and periods (P1: 2001-2006 and P2: 2007-2012). Means include data from all available sources and depths.

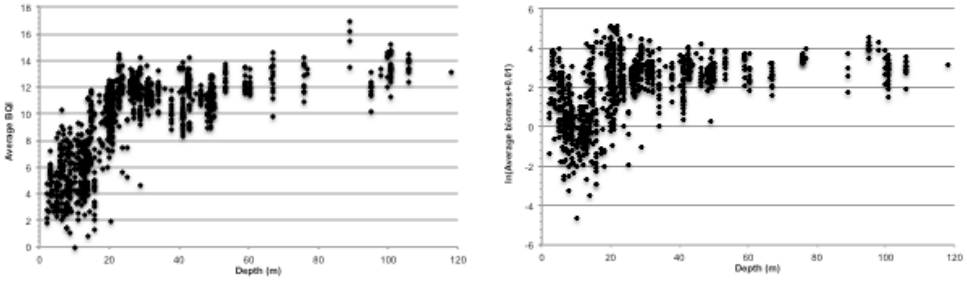


Figure 4.6. Observed relationship between the two variables BQI (left) and biomass (right) and depth. Averages represent mean of stations in individual years.

4.2 Variance components

4.2.1 Model of all water body types

Quantitative estimates of variance components and overall mean (intercept) for all response variables using the data from all water body types are shown in table 4.1. As noted earlier there are large similarities in the patterns displayed for BQI, richness, Shannon-Wiener and Margalefs index. In fact, for all of these variables the four most important components were identically ranked. The three most important sources of uncertainty were due to spatial variability among water bodies, stations and replicate samples and the fourth most important was interactive variability among years and

stations (e.g. WB>ST>RES>YE*ST). Thus, although there are some unpredictable dynamics, mainly associated with changes at individual stations, patterns of diversity appear to be fairly static and dominated by spatial components. Consistent differences among years throughout this region appears to be of little importance, as well as interactive effects involving spatial units and longer (6-year) assessment periods. Althogh the same four components dominate also for biomass, the relative sizes of components appears to differ slightly. Here the sources of small-scale variability are more important as both stations and replicate samples are larger than variability among water bodies (e.g. ST>RES>WB>YE*ST). Nevertheless, the same emphasis on spatial sources is observed also for biomass.

TABLE 4.1
Variance components for BQI, Richness, Shannon-Wiener, Margalefs index and ln(biomass+0.01) based on all types and during two assessment periods (i.e. 2001-2006 and 2007-2012). See 3.2.1 for description of model.

Component	BQI	Richness	Shannon-Wiener	Margalef	Biomass
YE(pe)*ST(WB,ty)	0.63	11.41	0.10	0.27	0.23
YE(pe)*WB(ty)	0.00	0.81	0.03	0.01	0.03
pe*ST(WB,ty)	0.22	4.38	0.03	0.10	0.00
ST(WB,ty)	2.59	21.22	0.20	0.58	1.14
YE(pe)*ty	0.06	2.05	0.00	0.05	0.03
PE*WB(ty)	0.00	0.36	0.00	0.01	0.00
WB(ty)	5.48	49.29	0.27	1.30	0.64
YE(pe)	0.03	0.38	0.00	0.00	0.00
RES	0.64	16.23	0.14	0.49	0.87
Grand mean	9.96	24.22	2.81	4.36	1.99

In order to assess the importance of these components it is informative to estimate their sizes relative to expected means of the respective variables. One way to do this is to calculate the coefficients of variation (CV=SD/Grand mean; Fig. 4.7). Using this relative measure of precision we can again conclude that biomass differs from the variables related to diversity. The general pattern is that biomass is less precise relative to its mean than the other variables. The most striking examples are variability among stations, which is 15-20% for BQI, richness, Shannon-Wiener and Margalefs index but ≈40% for biomass, and residual variability, which is 10-15% for the former and ≈35% for the latter (Fig. 4.7). An additional observation is that while biomass is least precise, the BQI tend to be the most precise variable in relation to the mean, also in comparison to the other measures of diversity. In terms of variability among replicate samples and interactive variability among years and stations, the spread for BQI is approximately half of that observed for the other measures. Because BQI and biomass represent two extremes in terms of precisions, subsequent analyses on the precision of current monitoring in specific water body types are focussed on these variables.

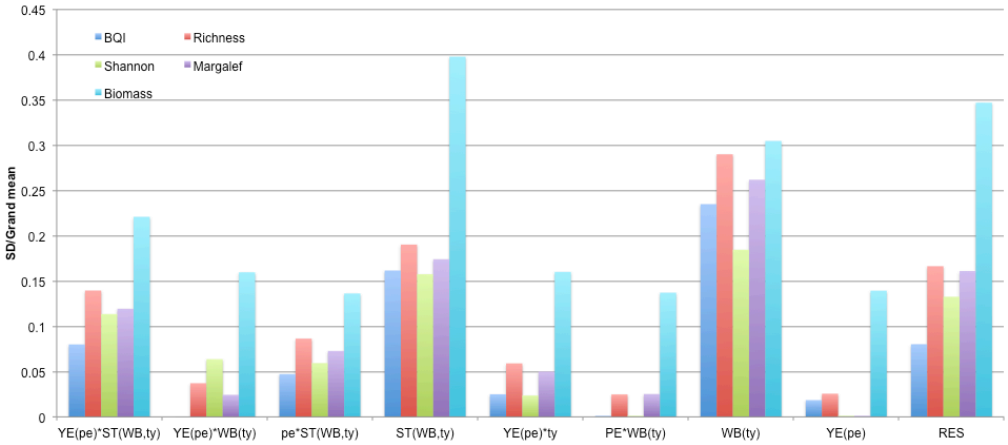


Figure 4.7 Estimated standard deviations relative to the intercept for random sources of variability using equation 4 all data from all water body types of all potential indicators from the Swedish west coast (depth >5 m).

4.2.2 Models for individual water body types

Following the results of the overall analyses including all types, components of uncertainty within individual types were partitioned for the two extremes, BQI and biomass. Variance components for BQI, which are later used to model precision and confidence, are shown in table 4.2 (see also Fig. 4.8). These analyses show that there are differences in relative importance among water body types but that some common patterns are discernible. Similarly to the overall analyses, components due to stations, water bodies and replicate samples are dominant in many types (Fig. 4.8). For example the residual variability is the most important source in three types and among the four most important in all types. Water bodies are important in all near-shore types (i.e. 1n, 1s, 2, 5 and 6), but not in outer coastal waters (3 and 4) (Table 4.2; note that water bodies are not defined in 0n and 0s). Variability due to stations are among the most important in seven out of nine types and the interaction between years and stations is important in eight of nine instances.

Measured as spread in relation to means (i.e. CV), it is evident that all interaction terms except YE*ST are on the order of <5% of the mean in most areas, residual variability is consistently 5-10% of the mean while the variability among stations and water bodies is more variable among types (Fig. 4.8). These range from 5-50% in various types.

Because water body types and stations typically differ in sampling depths, a model involving and removing fixed effects of depth was attempted for BQI. The analyses using the model involving depth as a fixed factor, showed that it is possible to substantially reduce uncertainty by accounting for depth (Fig. 4.9 and 4.10). This attempt was particularly successful in reducing uncertainty due to stations and water bodies in costal Kattegat and the Öresund (e.g. types 1s, 5 and 6), where the spread was reduced by 10-30% for stations and water bodies. These results are clearly promising attempts to

implement such covariates into a complete assessment procedure is beyond the scope of this study but such issues are to be addressed in different parts of WATERS.

TABLE 4.2
Variance components for BQI for individual types and during two assessment periods (i.e. 2001-2006 and 2007-2012). Empty cells not relevant due to non-existing water bodies in off-shore types. See 3.2.2 for description of model.

Component	0n	0s	1n	1s	2	3	4	5	6
YE(pe)*ST(WB)	0.28	0.66	0.01	0.98	0.75	0.30	0.27	1.21	0.32
YE(pe)*WB			0.14	0.00	0.28	0.00	0.00	0.00	0.11
pe*ST(WB)	0.40	0.12	0.00	0.04	0.08	0.04	0.52	0.30	0.00
pe*WB			0.11	0.00	0.07	0.01	0.00	0.22	0.00
YE(pe)	0.00	0.03	0.00	0.21	0.15	0.14	0.12	0.16	0.09
ST(WB)	0.34	0.76	0.00	1.96	0.10	2.50	6.66	2.07	11.16
WB			0.06	4.04	15.22	0.00	0.00	8.64	11.97
RES	0.82	0.80	0.30	0.66	0.55	0.41	0.50	0.81	0.42
Grand mean	13.20	11.54	11.67	6.69	8.70	11.79	9.79	8.38	7.20

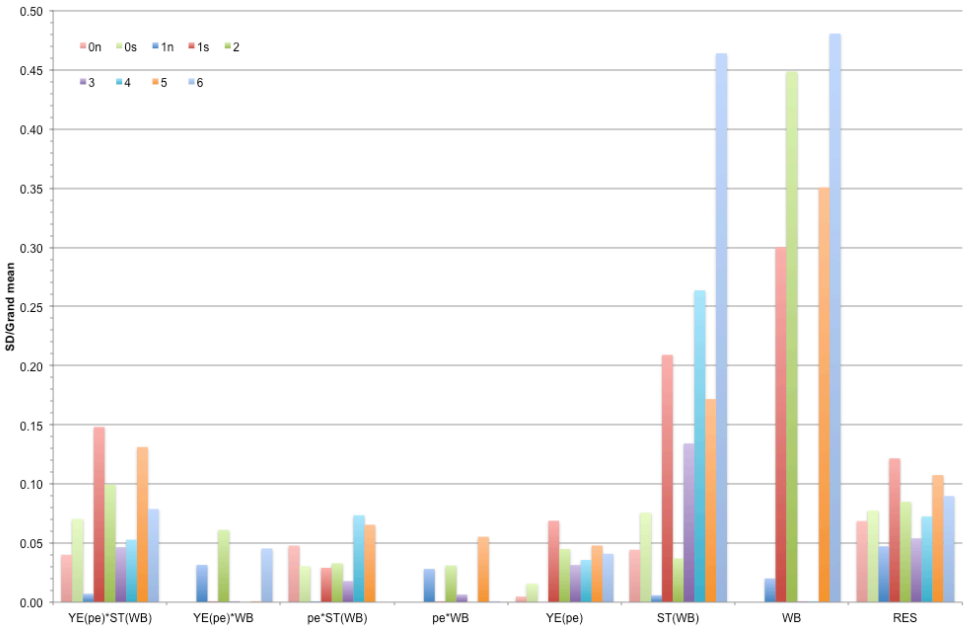


Figure 4.8. Estimated standard deviations relative to the intercept for random sources of variability using equation 5 not including depth data from individual water body types for BQI (depth >5 m).

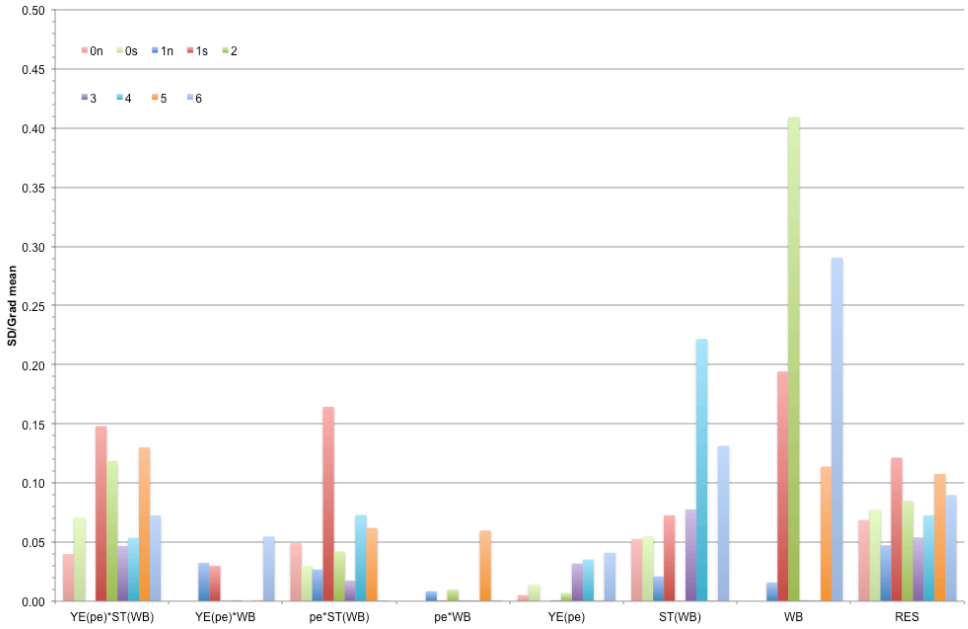


Figure 4.9 Estimated standard deviations relative to the intercept for random sources of variability using equation 6, i.e. a model including depth, and data from individual water body types for BQI (depth >5 m).

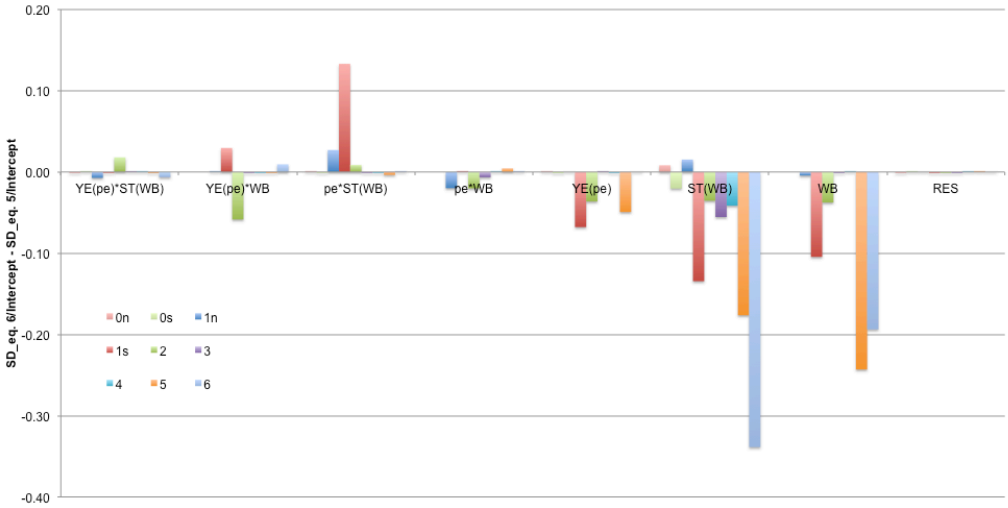


Figure 4.10 Change in standard deviation relative to mean, if depth is included in model.

Variance components for biomass are shown in table 4.3 (see also Fig. 4.11). The patterns of relative importance of different sources of variability, very much resembles that of BQI. The static, spatial components, i.e. water bodies, stations and replicates, are generally the most important (largest component in 3, 2 and 4 types respectively). It is worth pointing out that while the residual is fairly stable around 0.5 – 1.5, the estimates for

stations and water bodies vary more dramatically from 0 – 3 among types. These large fluctuations may be due to real differences in spatial variability but it is also possible that they are partly explained by unbalances and some confounding in the monitoring designs (e.g. lack of replicate stations within some water bodies), causing uncertainty in the partitioning of variability between stations and water bodies. Nevertheless, these estimates are based on practically all existing data and thus they arguably represent the best available estimates. Furthermore, any potential problems for modelling of precision caused by these issue can be properly dealt with in later stages (e.g. by pooling sources of variability).

TABLE 4.3
Variance components for ln(biomass+0.01) for individual types and during two assessment periods (i.e. 2001-2006 and 2007-2012). Empty cells not relevant due to non-existing water bodies in off-shore types. Note that values of variance components are multiplied by 100. See 3.2.2 for description of model.

Component (x 100)	0n	0s	1n	1s	2	3	4	5	6
YE(pe)*ST(WB)	2.85	6.84	0.00	40.45	0.04	8.12	7.45	92.16	15.50
YE(pe)*WB			0.00	12.89	0.00	0.00	4.62	0.00	0.00
pe*ST(WB)	0.27	0.00	0.71	0.00	0.77	0.00	4.54	0.00	8.35
pe*WB			0.01	0.00	0.00	2.62	0.00	0.00	0.01
YE(pe)	0.22	3.47	0.00	0.00	10.89	0.00	0.00	4.41	13.24
ST(WB)	46.78	18.82	0.00	35.16	52.42	8.41	265.69	166.41	0.08
WB			11.34	62.09	334.89	0.00	0.00	169.00	111.02
RES	41.40	80.59	39.65	118.81	57.30	26.52	79.92	151.29	58.73
Grand mean (x 1)	3.22	2.66	2.55	0.10	1.08	2.74	1.02	1.65	2.19

Viewed relative to the grand mean, the patterns of variability for biomass are strikingly different from those of BQI (Fig. 4.11). Here four types (1s, 2, 4 and 5) have components, which are often larger than 50% and sometimes up to 200% of the mean. These types are all located in the Kattegat or in the Skagerrak fjords. However, for biomass in other parts of Skagerrak (i.e. types 0n, 1n and 3) as well as off-shore Kattegat (0s), coefficients of variation are on the order of 5-15%, which is comparable to those of BQI. Variability is also relatively small in the Öresund (6). Many instances of large relative variability are of course related to low average biomass in some of these types, in particular in 1s, 2, and 4.

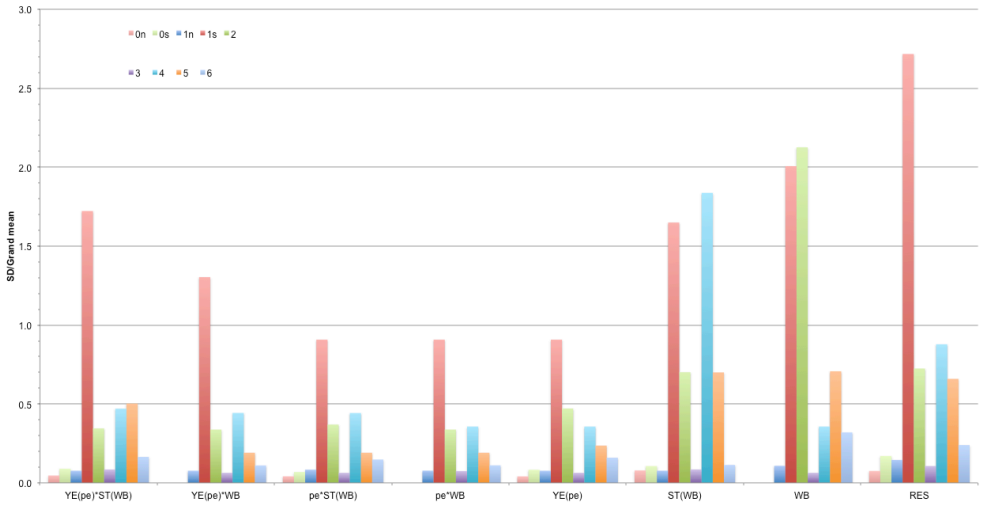


Figure 4.11. Estimated standard deviations relative to the intercept for random sources of variability using equation 5 not including depth data from individual water body types for biomass (depth >5 m). Calculated from back-transformed standard deviations and intercepts, e.g. $[\exp(SE_{\ln(x+0.01)}) - 0.01] / [\exp(intercept) - 0.01]$.

4.3 Uncertainty of existing programmes

The typical monitoring designs and dimensioning (Table 3.1), equations for error propagation (Table 3.3) and estimated components for individual water body types (Tables 4.2 [BQI] and 4.3 [biomass]) were used to assess the expected total error for individual types at a number of spatial and temporal scales (defined in section 3.3.1).

4.3.1 Precision of estimated mean BQI

The typical standard errors for means calculated of BQI at different combinations of spatial and temporal scales using current monitoring are shown in figure 4.12. It is evident that differences in patterns of variability and sampling designs cause large differences in expected precision among scales and water body types, and the different types are therefore discussed separately.

Offshore areas in Skagerrak (0n) and Kattegat (0s). In these areas we can expect that the SE is smaller than 0.5 BQI units within stations and years, within types and years, as well as within periods (i.e. at all scales investigated here). Within assessment periods SE is actually expected to be as low as ≈0.30. In relative terms this means that the error is on the order of 2-4% of the mean for all combinations of spatial and temporal scales.

Outer coastal types 3 (Skagerrak) and 4 (Kattegat) –These types are characterised by large differences in SE among scales but similar patterns among Skagerrak and Kattegat. SE was smaller than 0.5 within stations and years in both types. The error is expected to be

larger within water bodies, typically 1.5 (15% of the mean) in Skagerrak and 2.5 (25% of the mean) in Kattegat within both years and assessment periods. This is likely due to the small number replication of stations within water bodies (d=1). In contrast, the expected errors within water body types were smaller both within periods and years. In Skagerrak the SE was typically 0.75 BQI units (6% of the mean) and in Kattegat it was approximately 1.2 (12% of the mean).

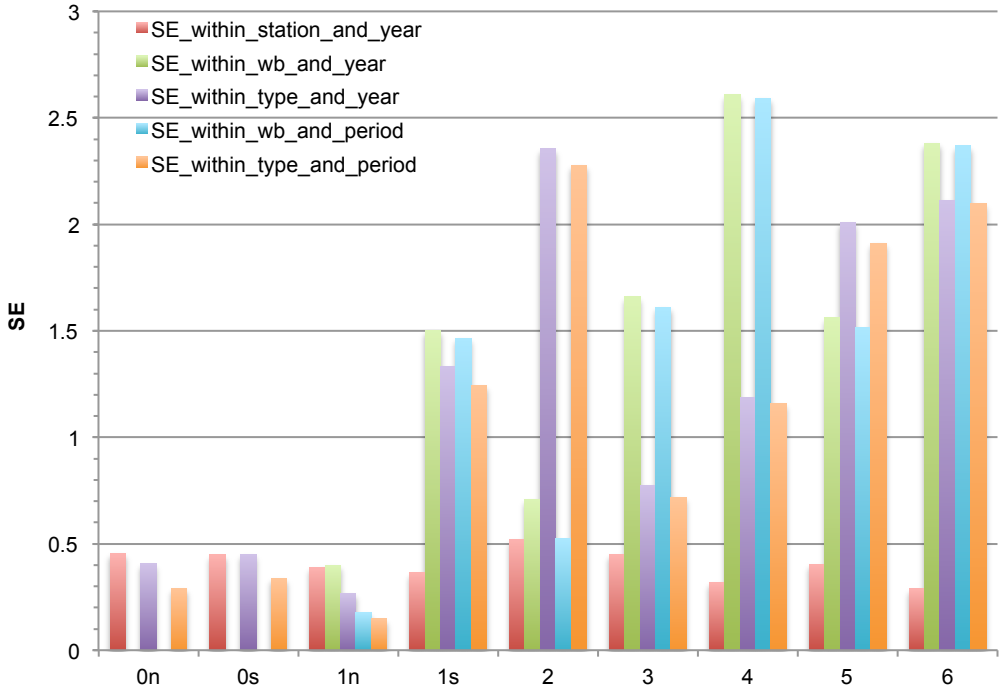


Figure 4.12. Standard errors in BQI units for individual water body types for estimated means at various spatial and temporal scales calculated using data in tables 3.1 and 4.2 and equations 7-11.

Inner coastal types 1n (Skagerrak) and 1s (Kattegat) –Similarly to in the offshore areas, SE was smaller than 0.5 at all investigated scales in the inner parts of Skagerrak (1n). Within assessment periods the relative error is expected to be on the order of 1-2% of the mean, while within years the error is 2-3%. In the inner parts of coastal Kattegat (1s) SE is also smaller than 0.5 BQI-units within years at indvidual stations, while at all the scale of water bodies and types, the error is typically around 1.5. This means that the relative errors at these scales are approximately 20% of the mean, i.e. much larger than in 1n. This difference is probably to a large extent caused by differences in depth distribution. As indicated in figure 4.10 the variability among stations and water bodies within type 1s, can be decreased substantially by accounting for depth. Initial calculations have shown that this has the potential to greatly reduce both absolute and relative errors.

The Skagerrak fjords, 2 – As expected from the large variability among water bodies (i.e. mainly among fjords), there are large errors associated with overall mean for this water body type as a whole. At this scale of aggregation the error is larger than 2 BQI-units and

the relative error is approximately 25% of the mean. At the scale of stations and water bodies, however, the situation is different with SE ≈0.5 BQI-units and relative errors around 5% within years and periods.

Southern Skagerrak (5) and the Öresund (6) – Except for the errors within individual stations and years which is smaller than 5% of the mean, the southernmost types are the ones generally showing the largest errors in absolute and relative terms. In types 4 and 5 relative error is 1.5-2 BQI (20%) and 2-2.5 (30%) BQI respectively for all combinations of types, water bodies, years and periods. Similarly to in 1s, this uncertainty can however, be greatly reduced by incorporating depth into the models. Again, preliminary analyses have shown that this can in fact reduce relative errors to levels of less than 5% of the mean in types 5 and 6 at all combinations of spatial and temporal resolution assessed here.

4.3.2 Precision of estimated mean biomass

The typical standard errors for means calculated for biomass at different combinations of spatial and temporal scales using current monitoring are shown in figure 4.13. although there are differences among types and scales, no detailed treatment of individual water body types is given here. Instead we can conclude that in most of coastal Skagerrak (0n,1n and 3; except the fjords), off-shore Kattegat (0s) and the Öresund (6) the error in stations, water bodies and types within years and periods are all in the range of 1-2 g (i.e. 5-20% of the mean). In another coastal type, 1s, errors are around 2 g for most scales, while in the Skagerrak fjords and in southern Kattegat errors vary from 2-5 depending on the scale of aggregation. In relative terms, however, types 1s, 2, 4 and 5, result in errors amounting to 50-250% of the mean. Thus, in these areas it appears that current dimensioning of monitoring is not sufficient to account for the large variability observed in figure 4.11, and to produce precise data for biomass at any of the scales.

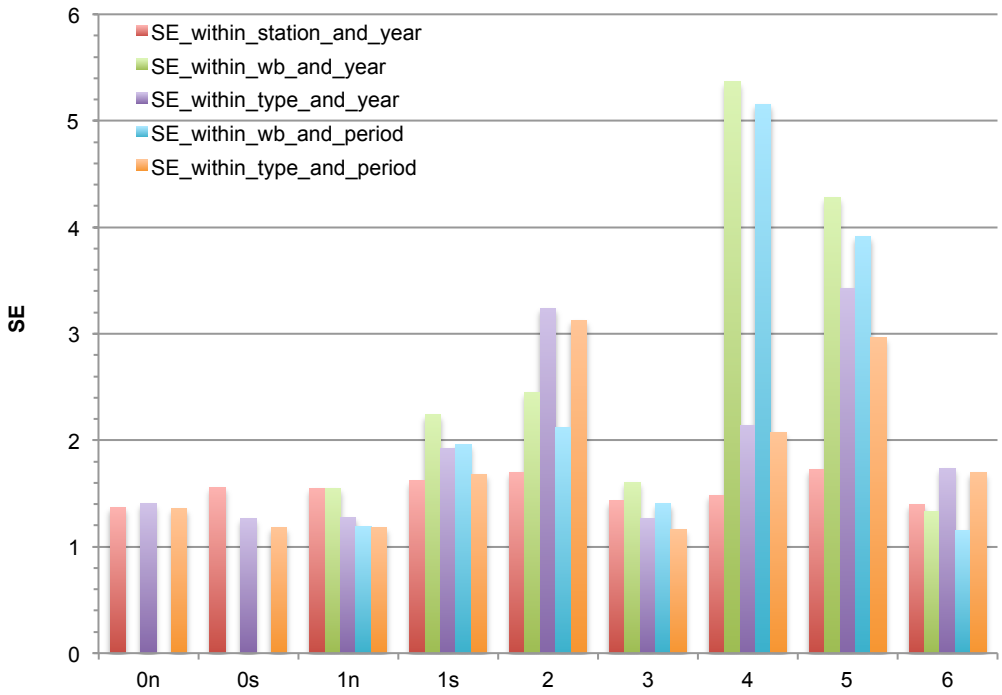


Figure 4.13. Standard errors in BQI units for individual water body types for estimated means at various spatial and temporal scales calculated using data in tables 3.1 and 4.3 and equations 7-11. Note that estimates shown are calculated from back-transformed standard errors, e.g. $SE_x = \exp\left(SE_{\ln(x+0.01)}\right) - 0.01]$

4.3.2 Confidence in classification of BQI

Because class-boundaries for BQI differ between depths 5-20 m compared to >20 m and because most of the available data come from the latter interval, confidence in classification using BQI was assessed at the scale of water body types within years and 6-year periods using data from depths >20 m (Fig. 4.14).

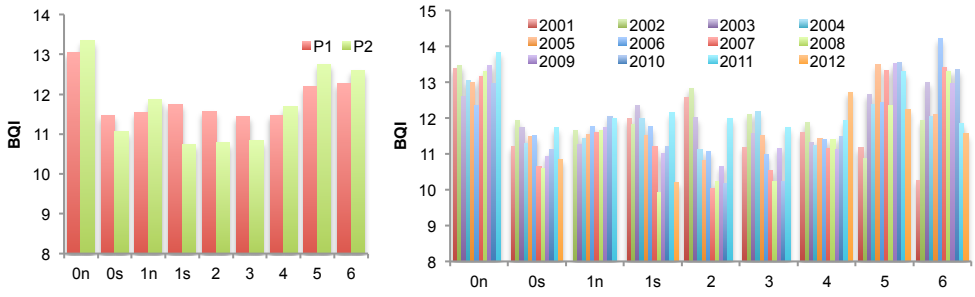


Figure 4.14. Estimated mean BQI in individual water body types in years between 2001-2012 (left) and periods (P1: 2001-2006 and P2: 2007-2012). Means include data from all available sources and depths>20m. Note that y-axes start at BQI=8.

TABLE 4.4

Standard errors used for calculation of confidence in classification for water body types within years and periods.

Scales	0n	0s	1n	1s	2	3	4	5	6
SE within type and year	0.407	0.451	0.266	1.334	2.355	0.775	1.185	2.007	2.113
SE within type and period	0.291	0.337	0.148	1.244	2.276	0.719	1.159	1.910	2.098

The calculations of confidence in classification for individual years show a large variability in confidence among years and types (Fig. 4.15). One example is the confidence in the Skagerrak off-shore areas, where all years show a high confidence (>80%) in the classification "Good" and only during two years is the confidence of "Moderate" larger than 5%. In the Öresund however, the situation is different (Fig. 4.15). Here the confidence of the dominant class is lower (50-60%), the most probable class varies among years and a total of three different classifications are at some points more probable than 20%. These patterns are to some extent a result of changes in means and proximity to the class boundaries, but in particular the with of the distribution within years is clearly a result of differences in SE among these two types (Table 4.4). At the scale of years the estimated SE in type 6 is approximately five times larger, compared to that in 0n. This is an example, which indicates that some of the differences in confidence can be dealt with by a more optimised monitoring design. Another observation is that in the majority of cases, the classification based on the 80% confidence limit corresponds to the most probable class, but in 15 of 37 cases when the most probable class is "Good", the precautionary approach identifies these as "Moderate".

At the scale of 6-year assessment periods, the patterns that emerge are similar and quite stable. This may be a consequence of smaller standard errors observed in all types at this scale of aggregation (Table 4.4). One striking feature is that despite some tendencies for change in mean between periods (Fig. 4.14), all types receive the same classification for both periods (Table 4.5).

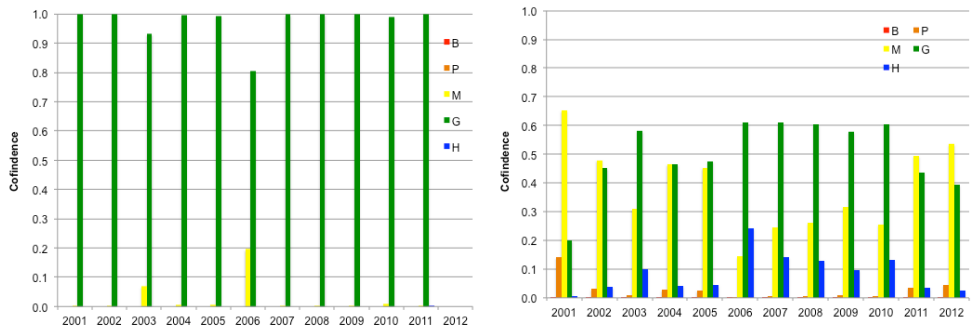


Figure 4.15. Confidence in classification for individual years in area 0n (left panel) and type 6 (right panel). Depth >20 m.

TABLE 4.5

Confidence in classification for water body types and periods (most probable class shown in bold). Also shown are classifications based on 80% confidence limits in analogy with procedures used for wfd-classifications according to HVMFS 2013:19. Depth > 20 m. Class boundaries are: 4 (B-P), 8 (P-M), 12 (M-G) and 15.7 (G-H).

Type	Period	Mean BQI	Bad	Poor	Mode-rate	Good	High	80% one-sided confidence limit	Classifi-cation
0n	P1	13.06	0.00	0.00	0.00	**1.00**	0.00	12.22	Good
	P2	13.34	0.00	0.00	0.00	**1.00**	0.00	12.50	Good
0s	P1	11.46	0.00	0.00	**0.94**	0.06	0.00	10.62	Moderate
	P2	11.06	0.00	0.00	**1.00**	0.00	0.00	10.22	Moderate
1n	P1	11.55	0.00	0.00	**1.00**	0.00	0.00	10.71	Moderate
	P2	11.86	0.00	0.00	**0.83**	0.17	0.00	11.02	Moderate
1s	P1	11.74	0.00	0.00	**0.58**	0.42	0.00	10.90	Moderate
	P2	10.75	0.00	0.01	**0.83**	0.16	0.00	9.91	Moderate
2	P1	11.57	0.00	0.06	**0.52**	0.39	0.03	10.72	Moderate
	P2	10.78	0.00	0.11	**0.59**	0.28	0.02	9.94	Moderate
3	P1	11.45	0.00	0.00	**0.78**	0.22	0.00	10.61	Moderate
	P2	10.83	0.00	0.00	**0.95**	0.05	0.00	9.98	Moderate
4	P1	11.48	0.00	0.00	**0.67**	0.33	0.00	10.64	Moderate
	P2	11.68	0.00	0.00	**0.61**	0.39	0.00	10.84	Moderate
5	P1	12.20	0.00	0.01	0.44	**0.51**	0.03	11.36	Moderate
	P2	12.76	0.00	0.01	0.34	**0.59**	0.06	11.92	Moderate
6	P1	12.27	0.00	0.02	0.43	**0.50**	0.05	11.43	Moderate
	P2	12.61	0.00	0.01	0.37	**0.54**	0.07	11.76	Moderate

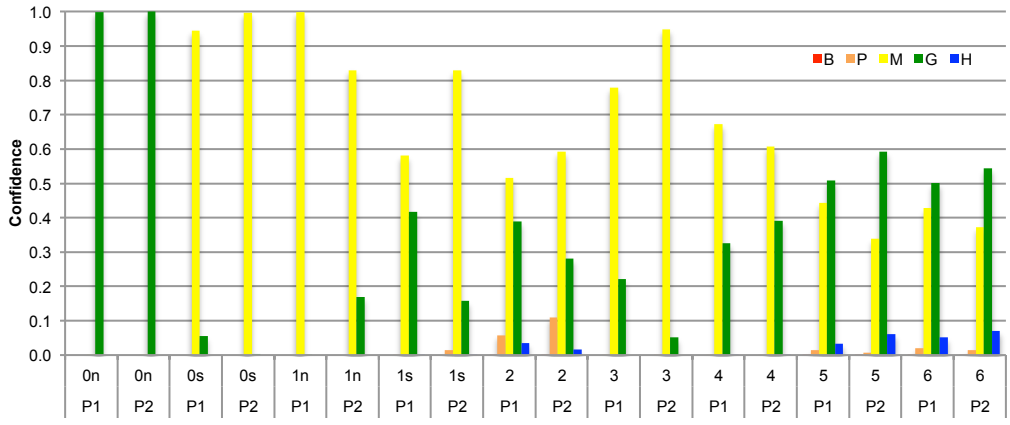


Figure 4.16. Confidence in classification for the two periods (P1: 2001-2006 and P2: 2007-2012) in all types. Depth >20 m.

Inspection of the confidence estimates, however, shows a slightly more complex situation (table 4.5 and Fig. 4.16). Overall, confidence in the dominating classes is >80% in the off-shore areas and in the Skagerrak with the exception of the fjords. In large parts of the Kattegat (1s and 4), the probability of dominating class is >60% and not more than two classes are dominant, while in 2, 5 and 6, the dominant class is ≈50% probable and up to four classes have some probability. Overall the 80% criterion identifies the most probable classification but two types 5 and 6 are identified as "Moderate", whereas the most probable class is "Good".

In conclusion, these exercises illustrate for the first time that confidence can be assigned to classifications of water body types for years and to whole assessment periods. The confidence assessments to periods, appear promising in the sense that they are stable among periods and that the confidence of dominating classes are comparable to those of individual years. This means that variability among years do not substantially decrease the quality of the classification and when there is large variability (large SE), this is properly reflected in a lower confidence. Furthermore, it is concluded that assessments can be related in a consistent way to current procedures which are develop for single years in individual water bodies.

4.4 Precision under different monitoring scenarios

4.4.1 Varying number of samples, stations and years within a period

Precision within stations and years – The analyses of precision within stations and years is not particularly important from a perspective of status assessment at a single station. Nevertheless, as varying numbers of samples per station are used in various programmes (e.g. n= 2, 4 or 5) on the Swedish west coast. The analyses show some differences among types but the main tendency is (as expected) a large, non-linear increase in precision with increasing number of samples. The expected error at n=2 varies between 0.4 – 0.65 BQI-units in the different types while at n=5 this has decreased to 0.25-0.4 (Fig. 4.17). Accounting for differences in mean, these errors roughly correspond to 4 – 10% for n=2 and 2 – 5% for n=5 (not shown). Although, these errors are quite small, it is important to note that the confidence of an estimate accounts for sample size also by the number of degrees of freedom (here df=n-1). Thus, 95% confidence interval is ±5-8 BQI-units at n=2 and ≈±1 BQI at n=5 (Fig. 4.17).

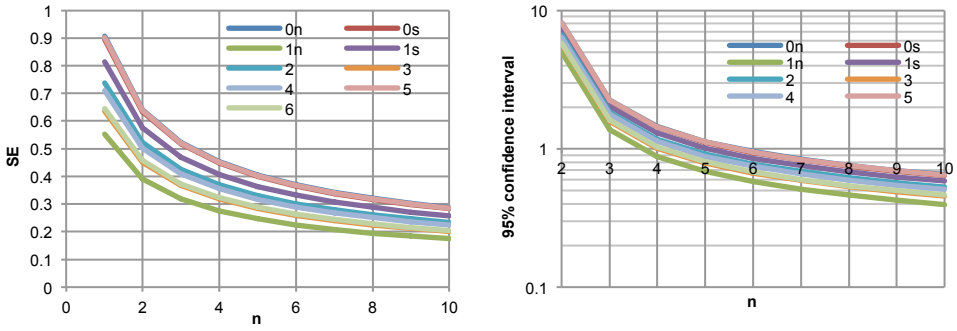
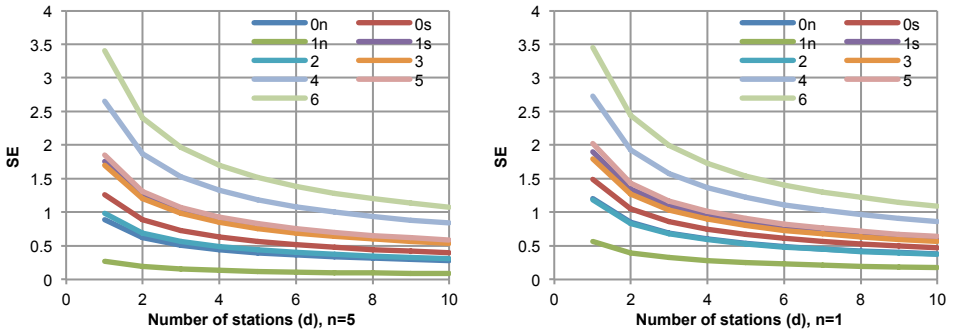


Figure 4.17. Standard errors and 95% confidence intervals at individual stations and years in specific types as functions of sample size.

Precision within water bodies and years – The uncertainty of mean estimates within years and water bodies is in general substantially larger than that within stations. This is because is influenced not only by variability among samples (s_e^2), but also by variability among stations within water bodies (s_{ST}^2) and its interaction with time (s_{ST*YE}^2). Because the latter of these components are larger than that among samples, and because the number of samples can only reduce uncertainty among samples, the effect of increasing number of stations is much more profound than that of larger sample sizes (Fig. 4.18). The difference in precision appears to be substantial among water body types, particularly when the number of stations is low. Within stations an error of <1 BQI-unit (CV≈0.1) was achieved at n≈1 for all types, while within water bodies 5-10 stations required in most types. Although not too much emphasis should be put on differences among types (the estimates themselves are associated with errors), there is a definite tendency for larger uncertainty in the Öresund and coastal Kattegat (6, 5, 4 and 1s), which all have larger SE and CV compared to Skagerrak and off-shore Kattegat (0n, 0s, 1, 2 and 3) at comparable number of stations and replicates.

In terms of 95% confidence intervals Fig. 4.19 shows that there are likely differences among types, but in general 15-20 stations are needed to achieve intervals that are around ±1 BQI. As expected smaller number of stations in the Skagerrak compared to the Kattegat and the Öresund.



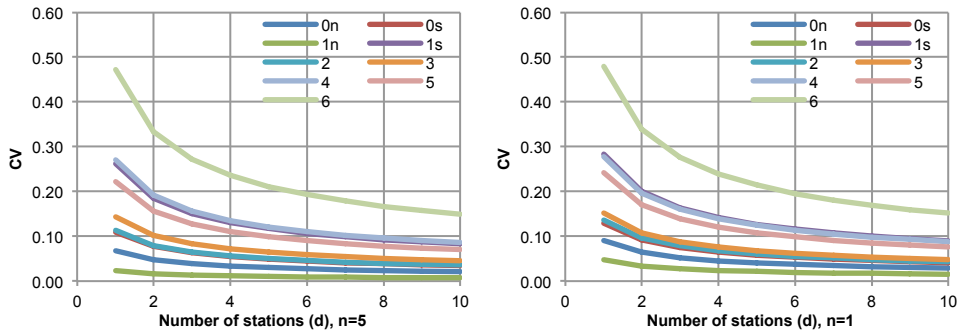


Figure 4.18. Standard errors and coefficients of variation in water bodies and years in specific types as functions of number of stations (d) for n=5 (left) and n=1 (right).

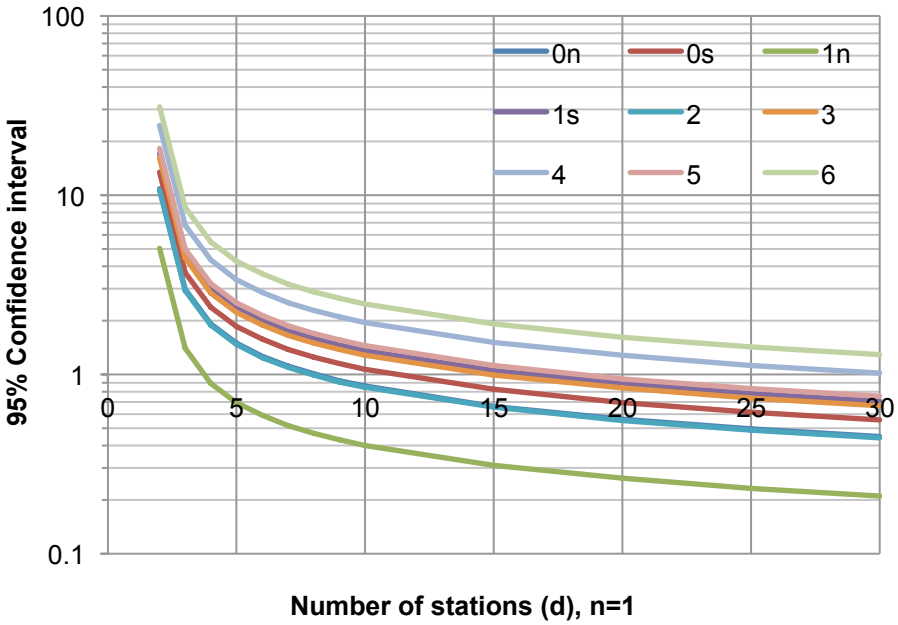


Figure 4.19. 95% confidence intervals in water bodies and years in specific types as functions of number of stations (d) for n=1.

Precision within water bodies and periods – The uncertainty of mean estimates within periods and water bodies is influenced not only by variability among samples (s_e^2), variability among stations within water bodies (s_{ST}^2), the interaction with time (s_{ST*YE}^2) but also by variability among years (s_{YE}^2). The importance of the latter is determined by the number of years sampled within an assessment period.

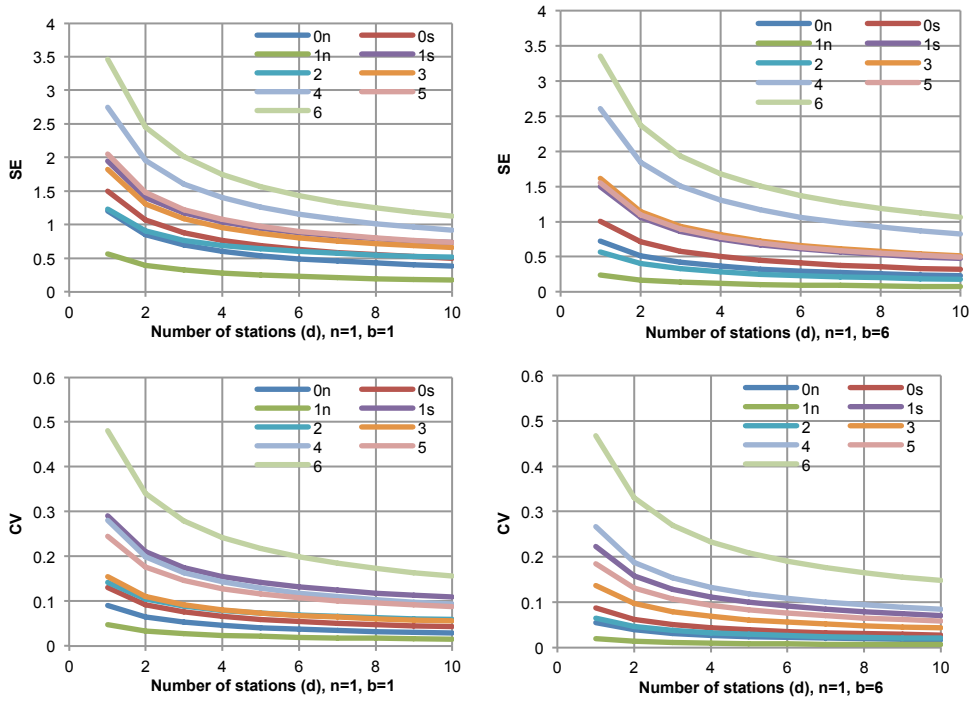


Figure 4.20. Standard errors and coefficients of variation in water bodies and periods in specific types as functions of number of stations (d) for b=1 (left) and b=6 (right).

As expected the analyses show that the absolute and relative errors decrease dramatically, with increasing number of stations (Fig. 4.20). The patterns are very similar to those within water bodies and years (Fig. 4.18). This is due to the fact that variance components due to years are generally small (Table 4.2). Types with relatively large s^2_{YE} are 1s, 2 and 5. These are also the ones showing the largest reductions in uncertainty when all years within a period are sampled (b=6) compared to when only one year is sampled (b=1) (differences in SE is 0.3-07). Overall, however, the effect of increasing the number of years within a period on SE and CV is not particularly strong. Note that when b=6, the total number of samples is six times that of b=1!

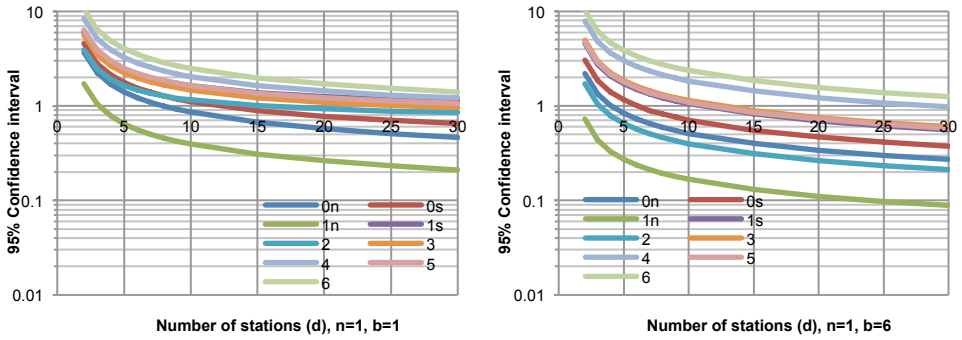


Figure 4.21. 95% confidence intervals in water bodies and periods in specific types as functions of number of stations (d) for b=1 (left) b=6 (right). As a conservative measure the degrees of freedom are approximated by df=d-1.

The effect on the width of confidence intervals is larger (Fig. 4.21). As an example we can observe that a 95% interval of ±1 is reached at d=15 stations in most types when 6 years are sampled but at d≥30 when one samples are taken in one year (note however that these intervals are likely to be overestimated as df was approximated as b-1 and thus do not fully account for samples in multiple years.)

Precision within water body types and years – The uncertainty of mean estimates within periods and water body types is influenced by uncertainties within water bodies, but importantly also by variability among water bodies (s^2_{WB}). As shown before such variability is often substantial but it also differs among types. Consequently, the errors associated with mean BQI in water body types is generally larger than those in water bodies (Fig. 4.22, c.f. Fig. 17). This is particularly true in the Skagerrak fjords (2), where SE is 3 to 0.5 BQI units larger and in the Kattegat and the Öresund (5, 6 and 1s) where the difference is 1.5 – 0.25 BQI units depending on type and number of stations (the differences are larger whn there are few stations). The pattern that emerges is that the outer and inner Skagerrak (0n, 1, and 3) and off-shore Kattegat group together with an SE = 0.2 – 0.5 and CV=2 - 5% when 10 stations are sampled, outer coastal Kattegat (4) is approximately half as precise (SE≈1 and CV≈10%), while the Skagerrak fjords (2) and the inner and southern coastal parts of Kattegat (1s and 5) have SE≈1.25 and CV≈15%. The Öresund is the least precise with SE≈1.5 and CV≈20%. All in all these patterns are quite different from those within singe water bodies, but considering what we know about the biology and components of variability they appear reasonable.

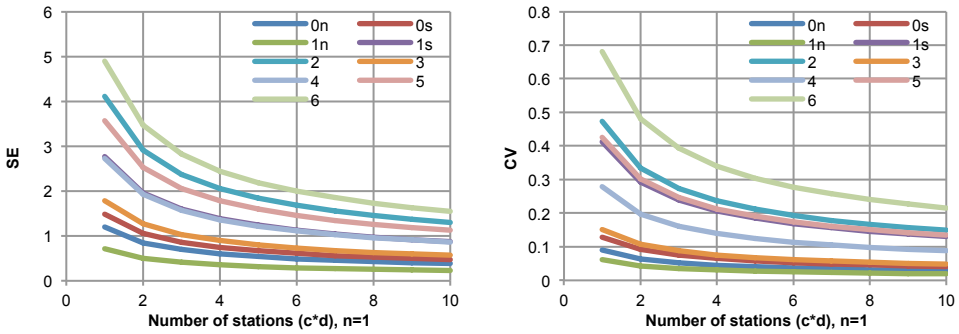


Figure 4.22. Standard errors and coefficients of variation in water body types and years in specific types as functions of total number of stations within a type, assuming that stations are randomly allocated within type (i.e. not clustered within a few water bodies).

The sizes of 95% confidence intervals as well as one-sided 20% confidence limits are shown in figure 4.23 (the latter in analogy with the Swedish assessment criteria). The patterns revealed in these figures are qualitatively similar to those of the standard errors. The group of high-precision types (0n, 0s, 1n and 3) require 5-15 stations to achieve a 95% interval of ±1 BQI, while the others (1s, 2, 4, 5 and 6) require ≥30 stations. For a one-sided 20% limit of -1 BQI <5 stations are required for the former group and 5-20 stations are needed in the other types.

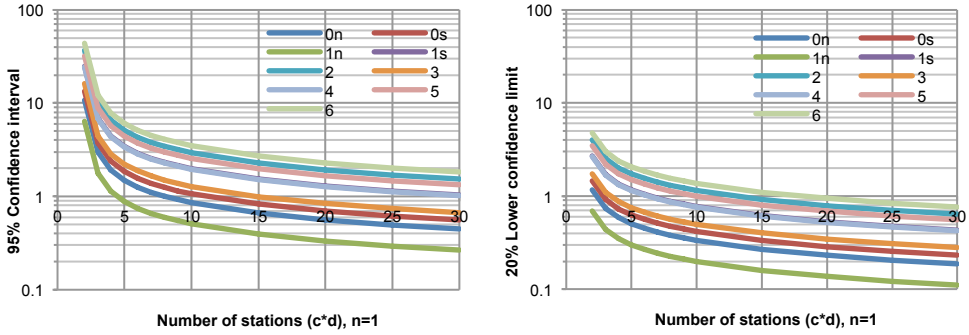


Figure 4.23. Confidence limits of mean BQI in water body types and years as functions of total number of stations within a type, assuming that stations are randomly allocated within type (i.e. not clustered within a few water bodies). Two-sided ±95% interval (left) and one-sided -20% limit (right).

Precision within water body types and periods – Similarly to the scenarios on precision within water bodies, the precision within water body types differ very little among years and periods and is marginally affected by the number of years sampled (Fig. 4.24). This is of course due to the fact that the variability among years within periods is relatively small. The qualitative patterns are also familiar, the Skagerrak (except the fjords) and off-shore Kattegat are most precise and the Öresund, the fjords and Kattegat are less precise.

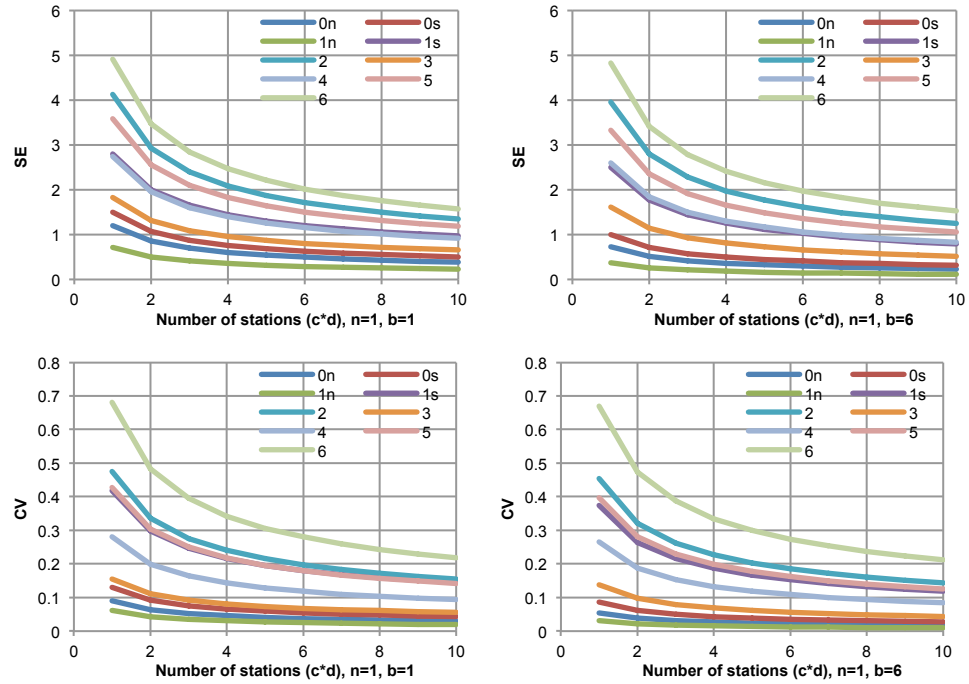


Figure 4.24. Standard errors and coefficients of variation in water body types and periods as functions of number of stations (d) for b=1 (left) and b=6 (right).

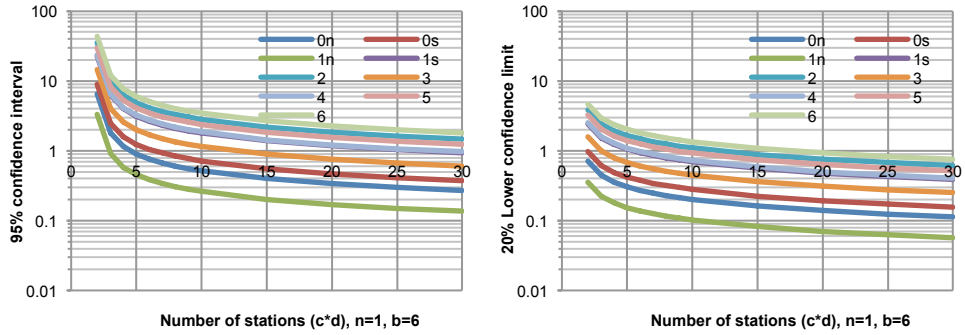


Figure 4.25. Confidence limits of mean BQI in water body types and periods as functions of total number of stations within a type, assuming that stations are randomly allocated within type (i.e. not clustered within a few water bodies). Two-sided ±95% interval (left) and one-sided -20% limit (right).

The sizes of 95% confidence intervals and the one-sided 20% confidence limits (Fig. 4.25) are similar to those of individual years shown in figure 4.23. Nevertheless, the size of intervals are consistently somewhat smaller than within years. Thus if 30 stations are sampled, the limit of the 20% interval is smaller than 0.3 BQI-units in the four most precise types (0n, 0s, 1n and 3) These require 5-15 stations to achieve a 95% interval of ±1 BQI, while the others (1s, 2, 4, 5 and 6) require ≥30 stations. For a one-sided 20% limit of -1 BQI <5 stations are required for the former group and 5-20 stations are needed in the other types. Again, however, it should be noted that the number of df used to calculate these intervals probably cause a slight overestimation of the size of these intervals when all years are sampled (i.e. b=6).

4.4.2 Crossed vs nested designs

Estimated variance components – The assessment of differences in precision of mean BQI between crossed and nested designs were done using components estimated from the most extensive dataset involving all types (Table 4.1). In order to estimate relevant standard errors we thus used s^2_{YE}=0.03, s^2_{WB}=5.48, $s^2_{ST(WB)}$=2.59, s^2_{YE*WB}=0.00, $s^2_{YE*ST(WB)}$=0.63 and s^2_{RES}=0.64.

Precision within water body types and periods – The modelling of precision of crossed and nested designs within water body types and periods show as predicted that the number of stations is crucial for both designs (Fig. 4.26). However, it is also evident that there are substantial effects of the number of years sampled and importantly of the way stations are allocated. First, when a crossed design is used, i.e. when stations are selected (randomly) at the start of the period and revisited repeatedly, it is evident that the number of years it is sampled can be expected to have a very small effect on precision (Fig. 4.26). This makes intuitive sense because variability due to years is very small and, thus once a certain set of

stations are selected, sampling these at additional years do not add substantial information. Second, the precision of monitoring using a nested design, i.e. where new stations are selected (randomly) at each year, appears to be strongly influenced by the number of years sampled (Fig. 4.26). For example, when five stations are sampled at six years SE≈0.5 compared to SE≈1.2 when only one year is sampled. Again this make intuitive sense as the total number of stations in a nested design is determined by the number of yearly stations and the number of years. Thus, when five stations are sampled during six years the total number of stations is 30 compared to 5 if only one year sampled. As a consequence, if only one year is sampled, the precision of the two designs is identical.

Another aspect of this is that, it appears clear that when a nested design is used a smaller number of yearly stations are needed to achieve a certain precision when a crossed design is used (Fig. 4.26). For example, when samples are taken at all six years, approximately 30 stations (i.e. a total of 30 stations) need to be sampled each year to achieve SE≈0.5 BQI units. Using a nested design, however, only approximately five yearly stations need to be sampled (note however that the total number of stations is 6*5=30). Although these numbers should note be taken too literally, there may be several other aspects that need to be taken into account, these analyses suggest that the choice of design may have profound impacts on the precision and costs associated with achieving a certain precision. And importantly, these indications are fully comprehensible with respect to what we know about patterns of variability. Maximising the number of stations is crucial.

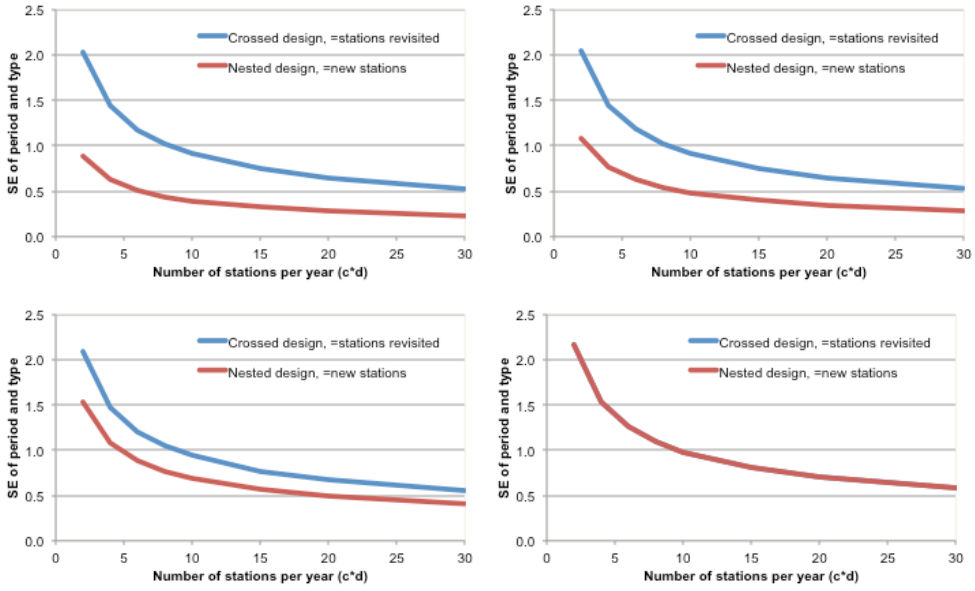


Figure 4.26. Standard error of mean BQI in water body types and periods as functions of the number of stations (c*d) sampled per year when samples are taken using crossed or nested designs in all years (b=6; top left), four years (top right), two years (bottom left) and one year (b=1; bottom right).

The resulting confidence intervals of crossed and nested designs when samples are taken at all six years are shown in figure 4.27. This situation represents the maximum difference

between the two strategies. The results indicate that the intervals for the crossed design is approximately 2.3 times larger than that of the nested design at a given number of yearly stations.

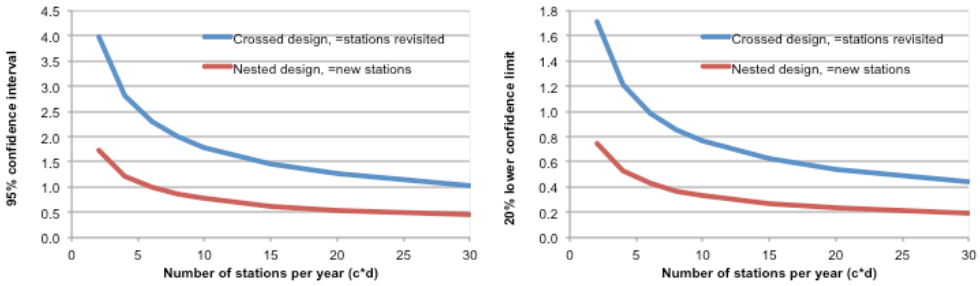


Figure 4.27. 95% two-sided (left) and 20% one-sided (right) confidence limits of BQI in water body types and periods as functions of the number of stations (c*d) sampled at six years when samples are taken using crossed or nested designs.

Precision within water bodies and periods – The results of the modelling of precision of crossed and nested designs within water bodies display qualitatively identical patterns compared to those of the water body types (Fig. 4.28). At the scale of water bodies, we can however expect that errors and confidence intervals are slightly smaller. Furthermore, the difference between designs is also slightly smaller (i.e. for the crossed design intervals are 2.1 times larger than those of the nested).

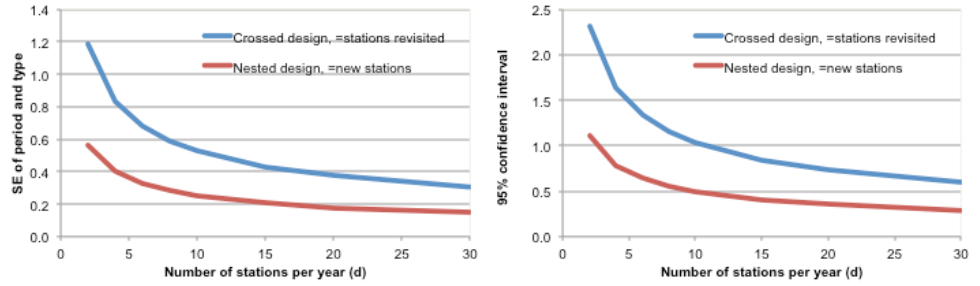


Figure 4.28. Standard error (left) and 95% two-sided confidence limits (right) of BQI in water bodies and periods as functions of the number of stations (d) sampled at six years when samples are taken using crossed or nested designs.

5 Detection of long-term trends

Trends at individual stations for BQI, richness, Shannon-Wiener, Margalefs index and biomass, are shown in the supplementary material (in Swedish). A qualitative observation, which is consistent with previous conclusions is that temporal trends and fluctuations of BQI and different indicators of diversity are largely co-varying and that the uncertainty of estimates of biomass are more variable. Another observation is that the dynamics of benthic fauna differ qualitatively among different stations. While some stations change in a more or less monotonic way, other stations fluctuate in a cyclic (≈10 years) way (Fig. 5.1). Thus lack of simple linear trends does not always indicate stable conditions but may in fact be the result of short-term fluctuations.

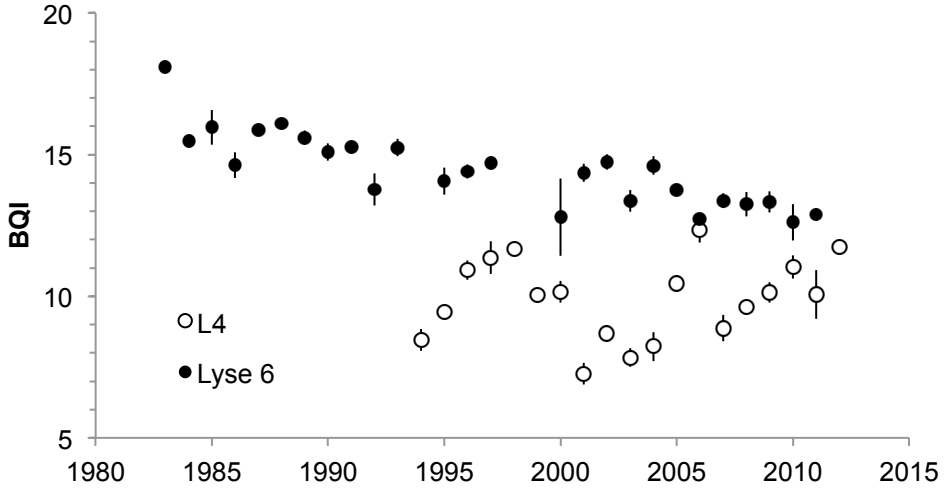


Figure 5.1. Illustration of two types of temporal changes observed in long-term data (mean±SE). Lyse 6 is an offshore station in the Skagerrak (Type 0n) and L4 is an offshore station in Kattegat (0s) which is located closely to the Laholm bay, which is well known for problematic oxygen conditions.

Linear regressions of average BQI for individual stations in show strong and significant negative trend at many stations (Table 5.1). In the offshore areas (0n and 0s) all stations except two in the southern parts of the Kattegat (L4 and N15 located on the border to the inshore type 5) show strong negative and statistically significant trends (0.10 – 0.20 BQI yr^{-1}; Table 5.1). Also in the outer parts of the Skagerrak and Kattegat (3 and 4) are there stations, which show strong negative trends. Types 4, 5 and 6 in the southern parts of Kattegat and Öresund are the trends variable among stations and often weak (Fig. 5.2)

TABLE 5.1
Number of sampled years, slope, coefficient of determintation and p-value of linear trends in stations and total average of different water body types.

Type	Station	Years	Slope	r^2	p
0n	MARS 7	26	-0.107	0.552	**0.000**
	VADE 7	27	-0.085	0.487	**0.000**
	LYSE 6	26	-0.121	0.703	**0.000**
	STRO 6	13	-0.115	0.286	0.060
	Total	27	-0.104	0.595	**0.000**
0s	ANHOLT	30	-0.107	0.509	**0.000**
	FLADEN	21	-0.160	0.738	**0.000**
	L4	19	0.044	0.031	0.471
	N10	19	-0.227	0.776	**0.000**
	N12	39	-0.175	0.702	**0.000**
	N14	20	-0.155	0.731	**0.000**
	N15	19	-0.024	0.029	0.486
	VINGA SW	25	-0.167	0.782	**0.000**
	Total	39	-0.146	0.711	**0.000**
1n	-	-			
1s	N8	19	-0.010	0.005	0.782
	N6	19	-0.068	0.145	0.107
	N5	19	-0.243	0.387	**0.004**
	DANAFJORD	17	-0.112	0.414	**0.005**
	Total	20	-0.141	0.462	**0.001**
2	HAGAR	29	-0.057	0.274	**0.004**
	Total	-			
3	VADE 4	10	-0.225	0.589	**0.010**
	MARSTRANDFJ	13	-0.130	0.574	**0.003**
	Total	13	-0.148	0.855	**0.000**
4	N11	19	0.084	0.168	0.081
	N13	19	0.034	0.087	0.220
	N7	19	-0.126	0.730	**0.000**
	N9	35	-0.067	0.358	**0.000**
	Total	35	-0.049	0.307	**0.001**
5	ÖVF 1:3	16	0.169	0.489	**0.003**
	S5	16	-0.054	0.021	0.589
	L9	19	0.052	0.029	0.490
	Ly	13	-0.147	0.373	**0.027**
	Total	19	0.057	0.049	0.361
6	ÖVF 2:3	16	0.054	0.063	0.349
	ÖVF 3:2	16	-0.012	0.013	0.670
	ÖVF 4:8	16	-0.008	0.015	0.654
	ÖVF 4:9	16	0.002	0.000	0.966
	ÖVF 4:11	14	0.104	0.246	0.071
	Total	16	0.005	0.002	0.857

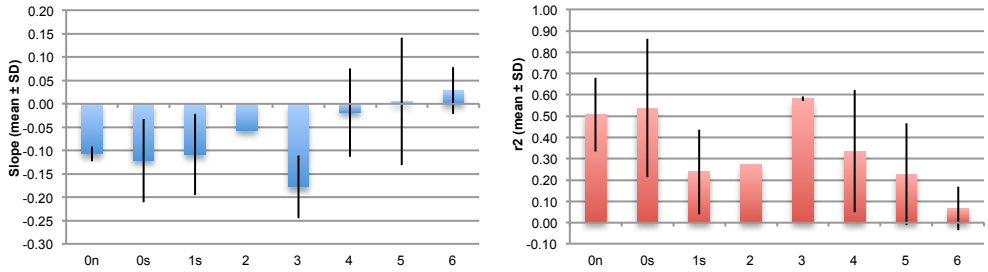


Figure 5.2. Average slope and coefficient of determination in different types (mean±SD).

Thus, in some areas (i.e. 0n, 0s, 1s and 3) long-term negative trends in BQI are generally strong and highly significant (r^2_{crit}=0.13 and 0.26 for N=30 and 15 respectively). Nevertheless, despite predominantly negative trends in large parts of the area, there is substantial variability in the strength of trends. Some of this variability appear to be related to depth. At depths below 30 m, trends are consistently negative with slopes between 0.1 – 0.2 BQI yr^{-1}. At shallower depths, trends are more variable (Fig. 5.3). Some of these patterns coincide with the length of time series, but in general it appears that stong negative changes occur over large areas in deeper regions.

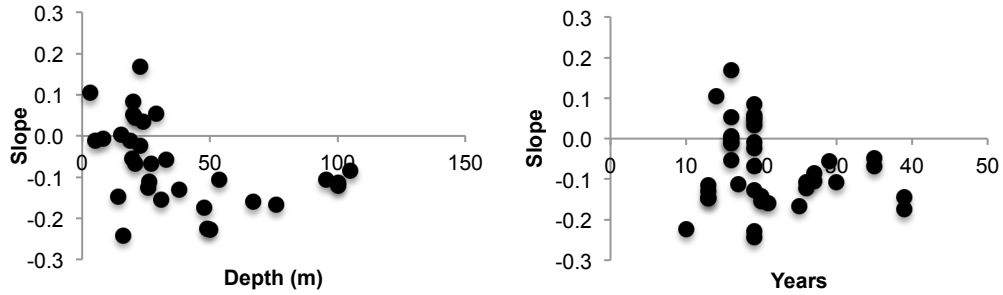


Figure 5.3. Strength of trends (slope) as a function of depth and number of sampled years.

In summary, these analyses show that there have been substantial decreases in BQI and thus ecological status in the deeper parts of Skagerrak and Kattegat. In shallower areas trends are more variable, although negative trends are more common also here.

The existence of these patterns demonstrate that current programs are capable of detecting trends but the low number of long-term stations in the respective water body types make a quantitative synthesis including power analyses at the scale of water body types difficult. Thus, despite the fact that conventional power analyses evaluating the probability of detecting linear trends (perhaps under different monitoring scenarios) were not attempted as originally planned, there can be no doubt about whether programs have a sufficient power to detect relevant trends at the scale of stations. Such trends were detected with 100% probability!

Finally, all available experience including the analyses presented here suggest that of trends of these relatively long-lived benthic assemblages are temporally autocorrelated and often have strong non-linear components. These circumstances mean that tests using ordinary regressions may be unreliable but more importantly may not be relevant. Detecting linear negative (or positive) trends to data, which are fluctuating in a cyclic way may not be very useful and may not reveal the important patterns. Devising appropriate tests for autocorrelated data and non-linear relationships is not straightforward but it can be done. However, to define relevant effect sizes and alternative hypotheses, which is an integral part of power analyses, these test procedures need to be defined a priori. This was clearly beyond the scope of this project.

6 Summary and conclusions

Summarised below is the main conclusions from the large number of analyses outlined above. Conclusions are structured according to the different aims of the project and each section is concluded with a short statement focussing on consequences for monitoring of the MSFD (shaded boxes).

6.1 Spatial and temporal patterns

The main aim of this study was not to describe spatial and temporal patterns of BQI or other indices. Nevertheless, the extensive dataset covering twelve years in all relevant water body types on the Swedish west coast (a total of almost 3000 samples), provided some general insights into the status of benthic fauna during these years. Some of the main conclusions were that:

- All variables involving biodiversity (BQI, richness, Shannon-Wiener and Margalefs index) showed similar spatial and temporal patterns. Patterns of biomass are generally different;
- There are more or less persistent differences among water body types;
- None of the variables show monotonous changes during the investigated twelve-year period;
- Changes in averages among two six-year assessment periods are small and not consistent among types;
- A general pattern taken over types is that of increasing diversity and to some extent biomass with increasing depth.

These observations describe some general tendencies in the data and briefly summarise differences among variables but are mainly to be considered qualitative observations. Analyses with explicit consequences for monitoring and status assessment of the MSFD (and sometimes the WFD) are summarised in the coming sections.

6.2. Variance components

Estimates of variance components. One fundamental tool for assessing uncertainty of current monitoring programs and modelling of uncertainty of any alternative program is the estimation of variance components. The tables showing quantitative estimates of variance

components (i.e. tables 4.1-4.3), represent a valuable source of information which can be used directly to optimise and assess the quality of future programs (c.f. "uncertainty library" as suggested by Lindegarth et al. 2013a). Issues to do with optimisation of monitoring can be addressed both in a generic sense (using results of analyses using all water body types) and for individual types (using analyses of individual types). Thus, the information contained in these tables can be used to address a wide range of design related questions, some of which are described below. Another, potentially important area of use is the possibility to use the estimated components of variability to account for variability and uncertainty that are not estimated properly in the monitoring programmes (e.g. lack of replicate samples, or replicate stations within a water body. This possibility is not further pursued here but as such these data are very valuable (Lindegarth et al. 2013b).

Analyses of all water body types. The analyses of the model including all water body types revealed that (1) there are large differences in precision among the investigated indicators and that (2) spatial components of variability (among samples, stations and water bodies) are generally more important in the monitoring data than the temporal component (indicating variability among years within assessment periods) or interactive components (involving both temporal and spatial components).

Spatial and temporal variability were assessed for five potential indicators (BQI, richness, Shannon-Wiener, Margalefs index and biomass). The analyses revealed that BQI was "the most precise" indicator in relation to its mean. This means that the standard deviation relative to the mean was generally smaller compared to the other indicators (in a strict sense precision also accounts for sample size). This was true for most variance components. There were however also large similarities among the four first of these (all involving aspects of diversity) in relation to the fifth, measuring biomass. Biomass was substantially more variable and differed somewhat in relative importance of variance components. As an example, coefficient of variation among samples within stations and years was ≈5% for BQI, ≈35% for biomass and ≈15% for the remaining indicators.

Comparisons of the relative importance of variance components showed that spatial components of variability were the most important. The general pattern for all indicators involving biodiversity was that variability among water bodies (CV≈0.25), stations within water bodies (15%) and samples within stations (5-10%). These are all static components, which suggest that there are strong and consistent patterns related to stations and water bodies. This fact can be expected to have important consequences for optimisation of monitoring programmes, i.e. spatial replication can be expected to be relatively more important for minimising sampling errors.

Analyses of individual water body types. Estimates of variance components using models for individual water body types were obtained for the most precise (BQI) and the least precise indicators (biomass). These analyses indicated that there were differences in the size of variance components among water body types, but there were also substantial, qualitative consistencies among types in the relative importance of components. In consistency with estimates obtained from the analyses involving all types, spatial components were dominant compared to those of temporal and interactive sources. Both for BQI and

biomass, variability associated with stations and water bodies were particularly important in the southern coastal types in the Kattegat (1s and 5), in Öresund (6) and in the fjords of Skagerrak (2). One likely explanation for the exceptional spatial variability among stations and water bodies in these types is that these involve sampling at a wide range of depths. Earlier analyses had also indicated that our indicators are related to depth in a non-linear way (Fig. 4.6).

The view that depth is an important factor causing spatial variability in some of these types, was supported by analyses involving depth as a fixed factor. Particularly in Kattegat and Öresund (1s, 5 and 6), the relative dispersion was reduced by 10-30% for some components. One interesting observation is that the similarly large variability among water bodies in the Skagerrak fjords (2) cannot be explained by differences in depth. Instead this suggests that other factors associated with local conditions and perhaps impacts are the main drivers for the status in these semi-enclosed water bodies. Nevertheless, these analyses indicate that further development aiming at reducing uncertainty by accounting for depth is worth pursuing further.

Accounting for depth (or other environmental factors) in an operative way within future new assessment criteria for the WFD or the MSFD can be done in several different ways. One aspect that might be attempted would be to adjust class-boundaries in relation to depth. Several options exist to do that. As a first example, current assessment criteria for benthic fauna on the Swedish west coast, employ a strategy where class boundaries are adjusted for in two depth strata. This strategy is practical due to its simplicity but may be also be overly simplistic because the relationship between depth and BQI may not be accurately described by a "step function". One potential argument, however, is the fact that these areas are often characterised by a salinity gradient and halocline at ≈20 m. Second, similarly to other quality elements (e.g. fish in inland waters), class boundaries may be defined by a model capturing the relationship between depth and BQI. This may be more difficult to communicate but on the other hand it may be more accurate (provided that the relationship is strong and robust). Finally, another option may be to standardise estimates of BQI by calculating the expected value at a standardised depth and compare this value to a fixed class-boundary. Similarly to the second option, this would require a strong and known relationship between depth and BQI. This approach is used for the Danish criteria for macrovegatation (Carstensen, pers. comm.). In summary, the analyses presented here suggest that the uncertainty of estimates and classifications may be reduced by accounting for depth in a any of the ways described above. To develop such routines is, however, beyond the scope of this project.

Estimates of variance components for BQI, Richness, Shannon-Wiener, Margalefs index and biomass are listed in tables 4.1 – 4.3. In combination with appropriate formulae, these can be used as a "library" to address issues to do with precision of current or alternative monitoring designs and dimensioning for the MSFD.

A comprehensive analysis of data from all water body types, show that under any given monitoring design, the precision of different indicators will rank as BQI > Richness ≈ Shannon-Wiener ≈ Margalef > Biomass at all temporal and spatial scales. Spatial patterns tend to be fairly consistent among years. Therefore, replication at the level of stations and water bodies are likely to be fundamental for the precision of status assessments of water body types.

Analyses of individual water body types confirm conclusions about the importance of spatial sources of variability, but the size of components vary among types. Fjords in Skagerrak, coastal Kattegat and Öresund are particularly variable among stations and / or water bodies. Uncertainty due to such spatial variability can be substantially reduced by accounting for sampling depth. Different options for such adjustments are described conceptually.

6.3. Uncertainty of existing programmes

Precision of estimated means. Variance components for BQI and biomass in combination with analyses of prevailing sampling designs and dimensioning with individual types were used to estimate expected precision of current monitoring programmes. These analyses showed that expected precision varied among types and indicators. These differences were caused both by differences in the size of variance components and differences in dimensioning of monitoring. One important aspect, however, is that precision is highly specific for particular combinations of spatial and temporal resolution (i.e. stations, water bodies, types, years and 6-year (assessment) periods).

Nevertheless, for BQI it was concluded that the error was small (SE<0.5 BQI) for all resolutions in the off-shore types in Skagerrak and Kattegat (0n and 0s) as well as in inner Skagerrak (1n). Similar errors are expected in all types at the finest spatial and temporal resolution (i.e. within stations in individual years). In outer Skagerrak the expected error at the scale of types is <1 BQI within years and periods, while the precision within water bodies is smaller. In the Skagerrak fjords (2), coastal Kattegat (1s, 4 and 5) and Öresund (6), a small number and a small number of stations result in large errors (1.5 <SE<2.5). In relative terms, the general picture is that the SE range from 2-3% to 30% of the mean for BQI while for biomass the range is considerably larger (SE = 2-200% of the mean).

It is important to remember that these estimates of precision are based on the assumption that the monitoring stations are representative to the water bodies and or types. In this or any other sampling exercise in the field, it is not possible to know whether this is the case. It is possible to assess representativity with respect to geographic location, depth etc. but in practice we can never be certain that we have a representative sample (i.e. one which results in unbiased estimates for parameters of the statistical population of interest). The best way to achieve representativity is usually to employ some sort of random selection of sites. This is, however, not an accurate description of the selection procedures used for these data. Although the process for selection of stations is not completely clear in all instances, it is likely that a variety of criteria have been used in different programs, but to

describe the process as completely random (even within potential depth and substrate criteria) is probably seldom correct. This means that these results need to be interpreted with some degree of caution. Nevertheless, it is clear that these estimates of variability and precision are based on the best available data, and given some considerations, they can arguably provide important guidance on variability and precision. First, even though stations are not randomly selected with respect to location, information on benthic assemblages has often been limited. Therefore, it may be argued that the selection process is largely unpredictable (≈random) with respect to benthos. Second, stations are sometimes (often?) selected with the aim to standardise to a certain type of environment (i.e. deep, undisturbed, favourable substrate etc.), this means that the estimated components of variability are unlikely to be over-estimates, but rather under-estimates. Therefore, from that perspective calculated precision can in some cases be considered best-case scenarios. Third, as described in previous sections, it may be possible to account for factors such as depth and thus remove the effect of a biased selection of stations. Such procedures may in fact improve precision considerably. In summary, it is likely that there are problems with biased selection of stations in these data, but at the same time they represent the only conceivable source of information, which can be used to address the questions at hand.

Confidence in classification of BQI. This study represents the first attempt to provide estimates of confidence in classification for benthic fauna in Sweden. This was done using the estimated variance components, procedures for error propagation and typical monitoring designs, class boundaries of developed for water bodies within the current assessment criteria for the WFD. Here these boundaries were applied for water body types at the scale of individual years and 6-year assessment periods. These examples were largely compatible with classifications based on the 80% one-sided confidence intervals used in the WFD but also demonstrated some interesting aspects associated with differences among types and temporal resolutions.

In general, classifications at the scale of assessment period were characterised by a high degree of confidence. Confidence for the most probable classification, "moderate" or "good" was usually 0.6-1.0, but in a few instances, the majority classification was slightly lower than the one resulting form the current assessment criteria which applies a precautionary approach (i.e. by defining a 80% one-sided interval). In a similar way, calculations of confidence in classifications for individual years were largely consistent with the existing WFD routines. Again the confidence for the dominant class was large in many water body types but in some types confidence in classifications was below 50% and the most probable classification was variable among years within assessment periods.

Precision of monitoring is affected by patterns of variability and sampling design. The precision of current monitoring typically vary among types and choice of spatial and temporal scale. At a certain spatial scale, the precision is generally slightly better within periods compared within years. Monitoring in off-shore types (0n and 0s), and parts of Skagerrak (1n and partly 3) is generally more precise than

in coastal areas, particularly in Kattegat and Öresund. SE for BQI vary from 2-30% of the mean among scales, while corresponding figures for biomass is 2-200%. Potential problems with and remedies for biased selection of stations are discussed.

Using estimates of variability, information on current monitoring designs and class boundaries from the WFD, procedures for estimation of confidence in classification was demonstrated for the first time at the scale of water body types within years and assessment periods. These exercises demonstrated that current programs usually are sufficiently accurate to produce classifications with high confidence and which are largely compatible with current assessment procedures. The latter does however not provide guidelines for the calculation of confidence in classification.

6.4 Modelling precision of BQI under different monitoring scenarios

Varying number of samples, stations and years within a period. One of the main aims of this study was to assess the precision of monitoring under different monitoring scenarios in different water body types. Previous sections have demonstrated that any statement about uncertainty and precision must be related to a certain spatial and temporal resolution. Therefore, the precision of different monitoring scenarios were modelled at a number of combinations of spatial and temporal resolution within individual types. Thus, although the SE, CV and confidence intervals can easily be calculated from the various equations given in this report, many questions can be addressed graphically from figures 4.17 - 4.25. Furthermore, these analyses were initially done according to a monitoring strategy based on a crossed (orthogonal) design were a limited number of monitoring stations are revisited at consecutive years, which is the prevailing monitoring strategy.

In terms of the MSFD, the most important analyses are those performed at the largest spatial and temporal scale, i.e. within water body types and 6-year assessment periods. Nevertheless, in order to provide guidance for the dimensioning at smaller scales, which will be implicit in the larger scales, and to provide guidance for the WFD, analyses of the importance of the number of samples within stations (n) and stations within water bodies (d) were done.

First, analyses of at *individual monitoring stations* showed that a small number of samples were needed to achieve a relatively small sampling error. Thus, at n=1 the expected error is SE<0.9 BQI in all water body types and already when two samples are taken, the expected error is as small as 0.65-0.40 BQI-units.

Second, *at the scale of water bodies*, analyses showed that the number of samples within stations were of little importance. Instead the crucial determinant of precision is the number of stations within a water body. These analyses indicated substantial differences in precision among water body types. At a comparable number of stations, the types in

Skagerrak can generally be expected to be more precise in terms of SE than those in the Kattegat and Öresund. As an example, for the most precise type (1n), SE within water bodies can be expected to be ≈0.15 BQI when five stations are sampled while in the least precise type (6), the corresponding figure is SE≈1.5 BQI. In relation to the means these figures represent approximately 2 and 20% of the mean respectively. Another significant conclusion is that the precision within water bodies and years is surprisingly similar to that within 6-year assessment periods. Furthermore, the number of years sampled within periods has a relatively small effect on the error (SE). However, effects vary among types, are more important when a smaller number of stations are sampled and are more pronounced in terms of confidence intervals (as compared to SE and CV).

Third, as expected *at the scale of water body types*, the total number of stations was of main importance for the precision of mean estimates. A general pattern was that precision of BQI estimates were more precise in Skagerrak (except the fjords) and in off-shore Kattegat compared to coastal Kattegat and Öresund. As an example, for the former SE was 0.3-0.75 BQI when five stations were sampled while the latter varied between 1.2 – 2.2 at similar sampling intensity. Again the difference in precision between years and periods is relatively small in terms of SE. Similarly the number of years within periods is of little importance at the scale of periods. In terms of confidence intervals, however, the number of years is more important. Many questions about various aspects of precision can be addressed from these analyses. As an example the location of 95% two-sided and 20% lower confidence limits can be extracted for all types. These show that the number of stations needed to achieve intervals <1 BQI vary from <5 to >30 among types for 95% intervals, while the correspknding number for the 20% interval is <5 and <15 stations. Although these numbers shall not be taken literally, they can be expected to be indicative of the range of monitoring requirements.

Assessment of crossed versus nested designs. As described above, the number of years sampled within an assessment period has surprisingly small effects on the error of mean estimates within periods. This can be explained by the fact that variability among stations and water bodies dominate and that spatial patterns appear to be reasonably stable among years. In the light of this it is expected that sampling of "new" stations as opposed to those already sampled during previous years, can substantially reduce uncertainty in the status assessment for periods. This proposition is tested in the comparison of crossed and nested designs.

In consistency, with the arguments above, the analyses of BQI presented here suggest that we can indeed expect large differences in precision of status assessments depending on whether a crossed or nested design is used, but the effect depends on how many years are sampled within a 6-year period. If samples are taken in all years, the SE and confidence interval for a crossed design is more than twice as large compared to a nested design (both at the scale of water bodies and types). The difference becomes successively smaller when fewer years are sampled and if samples are taken only every sixth year, the precision is identical between the two approaches.

A striking feature is that the number of years has a marginal effect on precision in a crossed design, while for a nested design additional years substantially reduce uncertainty and when six years are sampled, the uncertainty is halved. One consequence of this is that fewer yearly stations (and thus samples) are needed to achieve a similar precision for a given type. As an example, to achieve a 95% confidence interval of ±1 BQI at the scale of water body types, approximately 30 stations are needed per year. With the nested design, approximately 5 stations are needed per year.

These examples are based on estimates of variability derived from all water body types. This means that they represent an average situation, which is an under-estimate in some types and an over-estimate in others. Therefore, these conclusions need to be taken as indicative of the magnitude of differences. Nevertheless, it s likely that the choice of design strategy can strongly affect (a) the precision of a program given a specific cost and conversely (b) the cost given a certain desired level of precision.

The ramifications of these results for the design and dimensioning of future sampling programs, however, need to be balanced with other considerations and possible negative impacts. For example, a nested design cannot for obvious reasons be used to evaluate trends at the scale of stations and it is possible that trend-detection at aggregated scales may be negatively affected by a change in strategy. It is also possible that the costs associated with sampling new stations is larger compared to sampling previously sampled stations. This may be due to difficulties in finding suitable substrate etc. Although these analyses suggest that at least a partial re-allocation of resources may improve the confidence of status assessments, it is clear that a nested approach may be perceived as by many as counter-intuitive and negative in some aspects. Therefore, any changes that are suggested need to be carefully considered and evaluated from a range of perspectives based on various policy contexts.

Effects of varying number of samples, stations and years of sampling on the precision of mean estimates are evaluated at a number of spatial and temporal scales. Figures 4.17 – 4.25 can be used to deduce expected precision under a number of conditions.

A general conclusion is that the error in terms of SE and CV is strongly affected by the number of stations at the scale of water bodies and types. The number of samples per station is only important at the scale of stations but here the error is small also at small number of samples. The number of years sampled has a relatively small effect of the error (SE) within a 6-year assessment period but may be more important to achieve a small confidence interval.

The comparison between a monitoring design based on repeated sampling of a set of stations (crossed design) to one based on sampling of "new" stations each year (nested design) indicate that the latter is likely to be substantially more precise at the scale of water body types and assessment periods. In a crossed design, sampling of additional years has a marginal effect on precision, while in a nested precision can be expected to improve considerably. These conclusions are

consistent with and explained by the conclusion that spatial variability are more dominant than temporal and interactive sources. Furthermore, these results provide opportunities for improving the precision and cost-efficiency of monitoring programs, provided that proper care is taken to maintain their efficiency in relation to other environmental objectives.

6.5. Long-term trends

According to the original plan, power analyses were to be done on trend analyses at the level of stations and water body types. Two aspects made these tasks difficult and arguably less interesting, which meant that these plans were changed. First, at the level of water body types, the number of stations was quite small and the length of time series was very different among stations. Therefore the prospects of trends aggregated at the scale of types appeared relatively uncertain. Second, initial inspections of trends at individual stations showed that the dynamics differed strongly among stations and areas. Benthic fauna are relatively long-lived means in one year is often strongly correlated to the previous year. Furthermore, some stations showed marked cyclic fluctuations, for which linear trends provide a poor description of the dynamics. All in all, these arguments suggested that simple power analyses of linear trends may not be particularly helpful.

Instead the analyses were focussed on describing and testing trends at individual stations, while at the same time analysing geographic patterns in the strength and nature of trends, and finally evaluating whether variability in the strength of trend were in any way related to depth or caused by differences in the amount of data.

The results suggested that there had been large-scale deterioration of benthic fauna, as measured by BQI in parts of Skagerrak and Kattegat during the last 20-30 years. Many of the off-shore stations had decreased at an average rate of 0.1 – 0.2 BQI per year, while many of the coastal stations, particularly in coastal Kattegat and in the Öresund, typically showed a cyclic dynamic. The analyses showed that stations in deeper areas in general had consistently strong negative trends, while at depths <30 m trends were more variable. Similarly, and not entirely independently stations with longer time series (>20 years) showed stronger trends. This may also be due to the fact that time series from deeper stations are generally longer. Nevertheless, the pattern that emerges from these analyses is one of strong declines in outer Skagerrak and Kattegat, fluctuation but not static trends in some coastal areas. One station in the Öresund shows (non-significant) tendencies for increased BQI.

The planned power analyses were not further pursued due to the nature of data. These showed strong, significant (i.e. power 100%) negative trends in many deep off-shore areas, while many coastal areas were dominated by fluctuating dynamics.

7 References

Bates D., Maechler, M., Bolker, B., Walker, S., 2013 lme4: Linear mixed-effects models using Eigen and S4. Version: 1.0-5

Blomqvist, M., Cederwall, H., Leonardsson, K. & Rosenberg, R. 2006. Bedömningsgrunder för kust och hav – Bentiska evertebrater [Assessment criteria for coastal and marine waters - Benthic invertebrates].

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H., White, J.S. (2009) Generalized linear mixed models: A practical guide for ecology and evolution. Trends in Ecology & Evolution, 24:127–135.

Clarke, R.T., Hering, D. (2006) Errors and uncertainty in bioassessment methods: Major results and conclusions from the STAR project and their application using STARBUGS. Hydrobiologia, 566:433–440.

Clarke, R.T. (2013) Estimating confidence of European WFD ecological status class and WISER Bioassessment Uncertainty Guidance Software (WISERBUGS). Hydrobiologia, 704: 39–56.

Cochran, W.G. (1977) *Sampling techniques* (3rd ed.). Wiley, Hoboken, NJ, USA.

EC (2003a) WFD CIS Guidance Document No 7. Guidance on the Intercalibration Monitoring under the Water Framework Directive. Produced by Working Group 2.7 – Monitoring. European Commission, Brussels. http://circa.europa.eu/

EC (2003b) WFD CIS Guidance Document No 13. Overall Approach to the Classification of Ecological Status and Ecological Potential. Produced by Working Group 2A. Available from http://circa.europa.eu/

European Commission (2000) Directive 2000/60/EC of the European Parliament and of the Council: Establishing a framework for Community action in the field of water policy. 23 October 2000. European Commission, Brussels.

HELCOM Secretariat (2013) HELCOM Core Indicator of Biodiversity State of the soft-bottom macrofauna communities http://helcom.fi/Core%20Indicators/HELCOM-CoreIndicator_State_of_the_soft-bottom_macrofauna_communities.pdf

Josefson, A.B., Blomqvist, M., Hansen, J.L., Rosenberg, R. Rygg, B. (2009) Assessment of marine benthic quality change in gradients of disturbance: Comparison of different Scandinavian multi-metric indices. Marine Pollution Bulletin 58: 1263–1277.

Kuznetsova, A., Bruun Brockhoff, P., Haubo Bojesen Christensen, R. 2012. lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package) Version: 2.0-3

Leonardsson K., Blomqvist M. and Rosenberg R. (2009) Theoretical and practical aspects on benthic quality assessment according to the EU-Water Framework Directive – examples from Swedish waters. Mar. Poll. Bull. 58:1286–1296

Lindegarth, M., Carstensen, J., Johnson, R.K. (2013a) Uncertainty of biological indicators for the WFD in Swedish water bodies: Current procedures and a proposed framework for the future. Deliverable 2.2-1, WATERS Report no. 2013:1. Havsmiljöinstitutet, Göteborg, Sweden.

Lindegarth, M., Carstensen, J., Johnson, R.K. (2013b) Monitoring biological indicators for the WFD in Swedish water bodies: Current designs and practical solutions for quantifying overall uncertainty and its components. Deliverable 2.2-2, WATERS Report no. 2013:6. Havsmiljöinstitutet, Sweden.

Moksnes, P., Albertsson, J., Elfwing, T., Hansen, J., Lindegarth, M., Nilsson, J., Rolff, C., and Wikner, J. (2013) Sammanvägd bedömning av miljötillståndet i havet, Havsmiljöinstitutets Rapport 2013:3. Havsmiljöinstitutet, Göteborg, Sweden.

Pearson, T.H., Rosenberg, R. (1978) Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. Oceanography and Marine Biology: An Annual Review 16: 229–311.

Rosenberg R., Blomqvist M., Nilsson H.C., Cederwall H. and Dimming A. (2004) Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive. Mar. Poll. Bull. 49: 728–739

Taylor, J.R. (1997) *An introduction to error analysis: The study of uncertainties in physical measurements*. University Science Books, Sausalito, CA, USA.

Monitoring of benthic fauna for the MSFD on the swedish west-coast:

Modelling precision and uncertainty of current and future programs using WATERS uncertainty framework

In this report we evaluate the usefulness of current monitoring of benthic invertebrate fauna in Västerhavets River Basin District for status assessment according to the MSFD and the WFD. The analyses are based on methodologies developed within the research programme WATERS. Main aims of the study were to (1) estimate spatial and temporal sources of variability in all water body types, (2) estimate precision for individual water body types and water bodies using existing monitoring programmes and to (3) evaluate precision and confidence in classification for individual water body types and water bodies for a number of selected scenarios for revised monitoring programmes.

WATERS is coordinated by:

Havsmiljöinstitutet
Swedish Institute for the Marine Environment

WATERS is funded by:

NATUR VÅRDS VERKET
SWEDISH ENVIRONMENTAL PROTECTION AGENCY