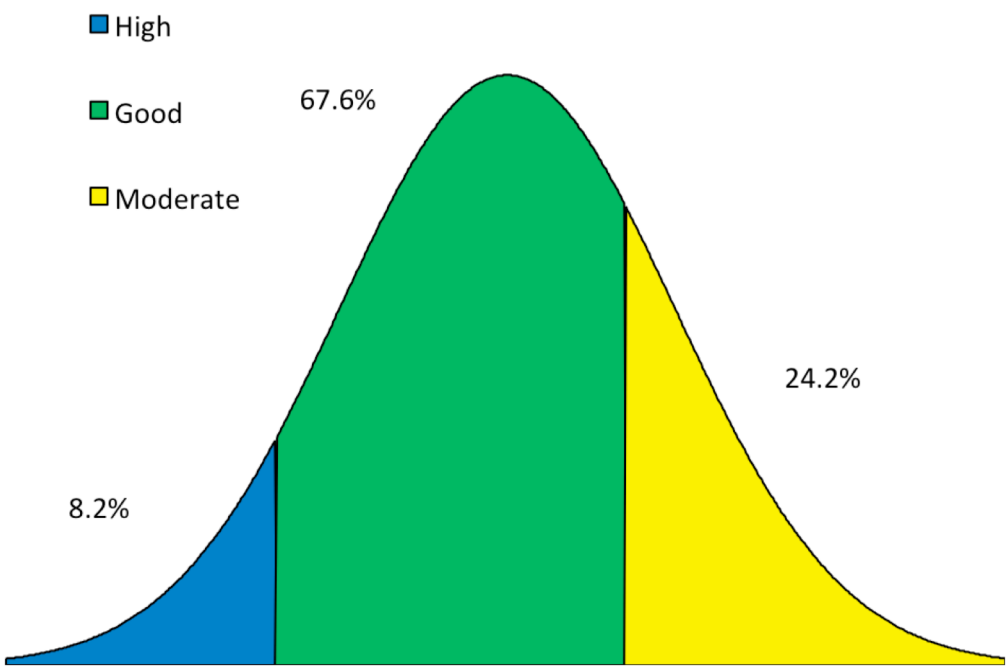

UNCERTAINTY OF BIOLOGICAL INDICATORS FOR THE WFD IN SWEDISH WATER BODIES:

current procedures and a proposed framework for the future

Mats Lindegarth, Jacob Carstensen, Richard K. Johnson

WATERS Report no. 2013:1

WATERS Report no. 2013:1
Deliverable 2.2-1

Uncertainty of biological indicators for the WFD in Swedish water bodies: current procedures and a proposed framework for the future

Mats Lindegarth, Gothenburg University
Jacob Carstensen, Aarhus University
Richard K. Johnson, Swedish University of Agricultural Sciences

WATERS partners:

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

WATERS Report no. 2013:1. Deliverable 2.2-1
Title: Uncertainty of biological indicators for the WFD in Swedish water bodies:
current procedures and a proposed framework for the future
Publisher: Havsmiljöinstitutet/Swedish Institute for the Marine Environment,
P.O. Box 260, SE-405 30 Göteborg, Sweden
Published: January 2013
ISBN 978-91-980646-3-6

Please cite document as:
Lindegarth, M., Carstensen J., Johnson, R.K. 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, 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 a deliverable of one of the scientific sub-projects of WATERS focusing on uncertainties of WFD classifications for biological quality elements. The report presents reviews of WFD requirements and current Swedish assessment criteria and proposes a coherent framework for handling uncertainty for all biological quality elements. These results will be further elaborated in coming work, thus providing a framework for a more harmonised treatment of measurement uncertainties and a tool for optimisation of monitoring programmes.

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

Assessments of ecological status according to the principles devised by the Water Framework Directive (WFD) are always associated with some degree of uncertainty. This uncertainty stems from the inevitable imperfection of the assessment criteria and from the uncertainty of measurements. While the assessment criteria (i.e. development of indicators of ecological status, refinement of reference conditions and class boundaries, and routines for integrated assessment) are dealt with in other parts of WATERS, this report aims to provide a general framework for analysing the uncertainty of measurements in Swedish inland and coastal waters.

The underlying basis for this work is that: 1) the WFD requires that member states assess and report aspects of uncertainty, 2) the Swedish assessment routines and their practical application can be further developed to better accommodate WFD requirements and to use available monitoring data more efficiently, and 3) a general uncertainty framework based on fundamental statistical principles is necessary to improve the consistency and transparency of assessments. These topics, including practical examples using Swedish data, are covered in different chapters of this report.

Review of existing policy and guidance documents reveals that two mutually related aspects of uncertainty are defined in the WFD, relating to precision and to confidence in classification. Precision, defined as the "half-width of the confidence interval", is a measure of the uncertainty of an estimated mean status. How large is the interval within which the true mean is located with a given level of confidence (e.g. 95 or 80%)? Confidence in classification is a measure of how confident we can be in a certain status classification. If the estimated average classification falls within the "good" interval, how certain can we be that this is the correct classification? In particular, the Directive stresses the importance of confidence in the "better than moderate" classification, because this marks the boundary that usually requires that actions be taken. Although the technical definitions of these concepts are well known, we conclude that issues concerning acceptable levels of confidence and burden of proof are still open to debate and value judgement.

Although the Swedish assessment procedures as developed in the "handbook" (SEPA 2010) provide some recommendations about the treatment of both precision of estimates and confidence in classification, the approach to and practical application of uncertainty assessments differ greatly among biological quality elements (BQEs). This is partly because of issues related to differences in biology and sampling methods, but partly also

due to seemingly arbitrary differences in the approaches used to assess uncertainty. Another conclusion is that none of the BQEs in the handbook provides a comprehensive treatment of spatial and temporal sources of uncertainty in a way that reflects uncertainties associated with assessment throughout a six-year cycle. Consequently, the uncertainty, in terms of both precision and confidence, likely differs greatly among BQEs, water bodies, and water body types, and there is often a substantial risk that the uncertainty of an estimate or a classification is unknown.

To improve assessments of uncertainty and to achieve better harmonisation among quality elements, a general framework is needed. Such a framework would involve conceptual identification, quantitative estimation of relevant components of variability, and estimation of total variability by combining information on variability and on the structure of the sampling design. This framework would provide tentative, qualitative assessments of the importance of spatial, temporal, interactive, and sampling-related sources of variability. It would distinguish fixed, potentially predictable components of variability from random, unpredictable components, which have different consequences for the estimation of uncertainty. We illustrate how different sources of variability are combined into a total variability measure using fundamentally different designs in terms of spatial and temporal replication, and how existing patterns of spatio-temporal variability may influence the optimisation of sampling designs. We also briefly review existing methods for the estimation of variance components and the calculation of precision and confidence.

Finally, we used the developed framework, estimation methods, and routines for combining uncertainties to analyse uncertainty in two datasets on marine benthic vegetation and fauna. These analyses demonstrate how the framework can be used to estimate sources of uncertainty and to assess overall uncertainty. Among other matters, the examples illustrate that: 1) many sources, both fixed and random, contribute to uncertainty, 2) coherent analyses of larger datasets produce more reliable estimates of sources of uncertainty, 3) the uncertainty of status assessment can be reduced by accounting for fixed components, 4) the size and relative importance of different sources of uncertainty can differ greatly among areas and regions, within the same BQE and, 5) despite considerable uncertainty, it is realistically possible to obtain precise status assessment if the full potential of spatial and temporal replication is realised.

The general conclusion from the analyses presented in this report is that the uncertainty framework can contribute substantially to improving the consistency and transparency of uncertainty assessments of Swedish coastal and inland waters. The possibility of developing a catalogue of uncertainty estimates for Swedish indicators based on extensive, quality-controlled datasets should be contemplated. Such a library could provide an important tool for future status assessments, particularly in instances in which monitoring programmes are insufficient for reliable estimation of uncertainty. Finally, the framework developed here will provide an invaluable basis for the further development of monitoring designs in subsequent work within WATERS.

Svensk sammanfattning

Alla klassificeringar av ekologisk status enligt EU:s vattendirektiv är behäftade med någon grad av osäkerhet. Osäkerheten uppkommer som ett resultat av brister i bedömningsgrunderna och på grund av osäkerhet i mätningarna. Medan utveckling av bedömningsgrunderna, exempelvis indikatorer, modifiering av referenstillstånd och klassgränser samt rutiner för sammanvägd bedömning sker i andra delar av WATERS, är syftet med denna rapport att presentera ett generellt arbetssätt för hantering av osäkerhet i beräkningar av ekologisk status i Svenska inlands- och kustvatten.

Bakgrunden för denna ansats är att: 1) genomförande av vattendirektivets intentioner kräver att medlemsstaterna utvärderar och rapporterar olika aspekter av osäkerhet i statusklassningen, 2) det finns utrymme för utveckling av de svenska bedömningsgrunderna och rutinerna för deras praktiska tillämpning så att de bättre tillgodoser direktivets krav och så att befintlig övervakningsdata kan utnyttjas på ett effektivare sätt, och 3) ett enhetligt arbetssätt grundat på välkända statistiska principer kan förbättra enhetligheten och transparensen hos statusbedömningarna och osäkerhetshanteringen.

Analys av direktivstext och vägledande dokument visar att vattendirektivet definierar två olika, sinsemellan relaterade, aspekter av osäkerhet: precision och sannolikhet för korrekt klassificering. Precision, definierat som bredden på halva konfidensintervallet, är ett mått på osäkerheten i en skattad medelstatus. Hur stort är intervallet inom vilket det sanna medelvärdet är beläget givet en viss önskad säkerhet (t.ex. 95 eller 80 %)? Sannolikheten för korrekt klassificering är ett mått på hur säkra vi kan vara på att en viss klassificering är korrekt. Till exempel, om den ekologiska statusen faller inom ramen för intervallet som klassas ”god”, hur säkra kan vi vara på att den sanna statusen inte är ”dålig”, ”måttlig” eller ”hög”? Direktivet fäster speciell vikt vid sannolikheten för korrekt klassificering av bedömningen ”bättre än måttlig” eftersom klassificering ”sämre än måttlig” föranleder åtgärder för att rätta till miljöproblem. Dessa två begrepp definieras på ett tillfredsställande sätt inom direktivet och dess vägledande dokument. Däremot specificeras i dessa dokument inte några definitioner av vad som är en acceptabel nivå för precision eller sannolikhet för korrekt klassificering. Inte heller ger direktivet några tydliga rekommendationer för hur osäkerheten skall påverka fördelningen av bevisbördan mellan olika intressen.

Handboken för hur vattendirektivets bedömningsgrunder skall tillämpas i svenska kust- och inlandsvatten innehåller vissa rekommendationer om hur precision och sannolikheten

för korrekt klassificering kan utvärderas (Naturvårdsverket 2007). Trots detta skiljer sig rutinerna och metoderna starkt mellan de olika biologiska kvalitetsfaktorerna. Detta kan delvis förklaras av ekologiska skillnader och övervakningsmetoder, men även av till synes godtyckliga skillnader in sättet att hantera osäkerhet. En annan slutsats är att handboken inte för någon av kvalitetsfaktorerna ger en sammanhållen strategi för hantering av osäkerhet orsakad av rumslig och tidsmässig variation inom ramen för direktivets 6-åriga bedömningscykel. En trolig konsekvens av detta är att osäkerheten, både i termer av precision och sannolikhet för korrekt klassificering, skiljer sig på ett betydande sätt mellan kvalitetsfaktorer, mellan vattenförekomster och –typer. Dessutom finns det en stor risk för att osäkerheten i skattningar och klassificeringar i själva verket är okänd.

För att förbättra bedömningen av osäkerhet och åstadkomma bättre samstämmighet mellan kvalitetsfaktorer, krävs det ett gemensamt arbetssätt för hantering av osäkerhet. Ett sådant arbetssätt innefattar konceptuella definitioner och kvantitativa skattningar av relevanta källor till variation (variationskomponenter), samt att den övergripande osäkerheten beräknas genom att kunskap om variationskomponenter kombineras med information om övervakningsprogrammens rumsliga och tidsmässiga struktur. Det föreslagna arbetssättet ger en preliminär, kvalitativ bedömning av betydelsen av rumsliga, tidsmässiga, interaktiva och metod-relaterade osäkerhetskällor. Arbetssättet skiljer mellan förutsägbara (”fixerad”) och slumpmässiga variationskällor, eftersom sådana komponenter påverkar osäkerheten på olika sätt. Vi visar också hur olika variationskällor kombineras till en övergripande osäkerhet när två fundamentalt olika sätt att utforma övervakningsprogram med avseende på rumslig och tidsmässig replikering tillämpas. Effektiviteten hos sådana program bestäms delvis av hur de faktiska variationsmönstren ser ut. Vi ger även en kort översikt av metoder för beräkning av variationskomponenter och beräkning av precision och osäkerhet hos klassificeringar.

Slutligen använder vi det föreslagna arbetssättet för att analysera osäkerheten hos två dataset över marin flora och fauna. Dessa analyser visar hur arbetssättet kan användas för att skatta enskilda osäkerhetskällor och hur de kan kombineras för att utvärdera övergripande osäkerhet. Dessa exempel illustrerar bland annat att: 1) den övergripande osäkerheten påverkas typiskt av flera slumpmässiga och delvis förutsägbara källor, 2) övergripande analyser av stora dataset ger mer tillförlitliga skattningar av olika källor till osäkerhet, 3) osäkerheten hos statusbedömningar kan minskas genom att fixerade faktorer inkorporeras i analyserna, 4) storleken på och den relativa betydelsen av olika variationskomponenter kan skilja sig mellan regioner för enskilda kvalitetsfaktorer, och 5) trots att det finns betydande osäkerheter, är det ofta möjligt att åstadkomma tillräckligt precisa statusbedömningar om den fulla potentialen hos den rumsliga och tidsmässiga replikationen hos övervakningsprogrammen utnyttjas fullt ut.

En övergripande slutsats från dessa analyser är att ett gemensamt arbetssätt för att hantera osäkerhet kan bidra till att förbättra enhetligheten och transparensen i sättet på vilket osäkerhet hanteras i svenska kust- och inlandsvatten. Möjligheten att utveckla ett bibliotek av skattade osäkerhetskällor för svenska indikatorer för biologiska kvalitetsfaktorer, baserat på stora, kvalitetssäkrade dataset bör övervägas. Ett sådant bibliotek skulle kunna

vara ett viktigt redskap för i framtida statusbedömningar, speciellt i fall där relativt lite övervakningsdata är tillgängliga. Slutligen kan vi konstatera att det arbetssätt som utvecklats här kommer att kunna utgöra en viktig grund för det framtida arbetet med utformning av övervakningsprogram i WATERS.

1 Introduction

Assessments of ecological quality according to the Water Framework Directive (WFD) (2000/60/EC) are based on four biological quality elements (BQEs): phytoplankton, macrophytes, benthic invertebrates, and fish. Using indicators (or metrics) responsive to human stressors for each of these BQEs of ecological status allows the status of individual water bodies to be assessed using data sampled in monitoring programmes. This process typically involves calculating a mean (or median) status that represents an estimate of the true mean status during the assessment period. Because this estimate is based on samples, and not on complete knowledge of the status throughout the assessment period and in all parts of the waterbody, it is very unlikely to correspond perfectly to the true mean status. Consequently, estimates of ecological status using indicators are always associated with some degree of error.

Such errors may be small or large, and may be caused by numerous processes that may differ among BQEs, but they always introduce some level of uncertainty to decisions based on the data. These facts are, of course, well known in the ecological and management literature (e.g. Green 1979, Underwood 1992), and robust methods for dealing with such uncertainty have largely been developed (e.g. Cochran 1977, Taylor 1997, Clarke et al. 2006a). Nevertheless, several reviews have identified the need for a more coherent treatment of uncertainties in BQE estimates and classification (e.g. Noges et al. 2009, Hering et al. 2010, Birk et al. 2012), and the scientific literature raising issues of uncertainty in connection with the WFD has grown at an increasing rate since the adoption of the Directive (Figure 1.1). Thus, to develop and harmonise current Swedish assessment criteria with respect to uncertainty and to illustrate how monitoring can be optimised, WATERS has devoted a specific work package to reviewing, analysing, and comparing current practices with what was intended in the WFD and to the fundamental scientific principles for dealing with various types of uncertainty.

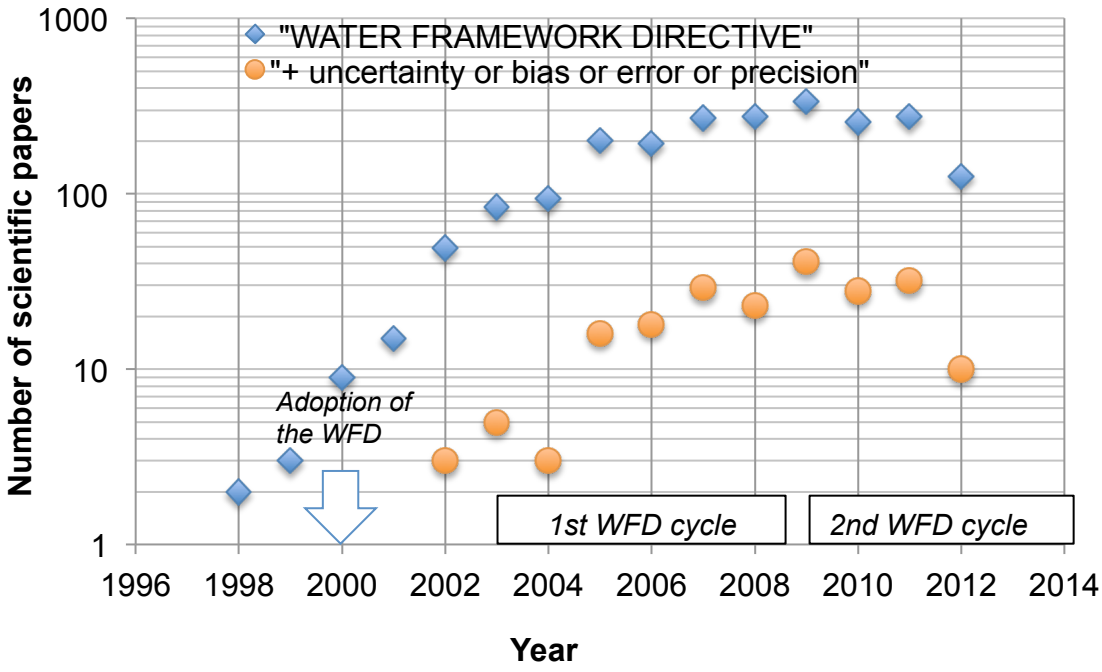

FIGURE 1.1
Amount of research focussing on the WFD since its adoption. Number of scientific papers per year mentioning the "Water Framework Directive" with and without references to "uncertainty", "bias", "error", or "precision". Search criteria: (Topic = ("WATER FRAMEWORK DIRECTIVE" AND (uncertainty OR bias OR error OR precision)) Timespan = All Years. Databases = SCI-EXPANDED, SSCI, A&HCI, CPCI-S, and CPCI-SSH).

1.1 Uncertainty in the WFD and its guidance documents

1.1.1 Definitions of uncertainty

One important component of the WFD and the resulting management cycle is the development of monitoring programmes. In this context, notions related to uncertainty are introduced; for example, Annex V states that:

> "Member States shall monitor parameters which are indicative of the status of each relevant quality element … Estimates of the level of *confidence* and *precision* of the results provided by the monitoring programmes shall be given." (p. 53)

> "Frequencies [of sampling] shall be chosen so as to achieve an acceptable level of confidence and precision. Estimates of the *confidence* and *precision* attained by the monitoring system used shall be stated in the river basin management plan." (p. 55)

The central aspects of uncertainty identified by the WFD are thus *precision* and *confidence*. These concepts are elaborated on in CIS Guidance Documents nos. 7 and 13 and the definitions provided by these documents are fully consistent with basic statistical principles. Nevertheless, as will be demonstrated in the following chapters, their application to real-world problems related to the WFD is not always straightforward.

Precision is a concept that refers to the uncertainty of an estimated parameter, usually the mean (though estimates of the precision of the medians or slopes of a regression are occasionally required). The operational definition adopted by the guidance documents is that precision equals "the half-width of the C% confidence interval". The confidence interval is the interval in which the true value of the estimated mean is located with C% probability (Figure 1.2). If a mean is estimated from a number of samples, the width of the confidence interval depends on the variability among samples and the number of samples (see Annex A for a technical definition). Little variability and many samples result in small confidence intervals. It is also well known that C% is, by convention, usually 95% in the scientific literature, though it must be noted that the Directive does not stipulate a specific confidence level.

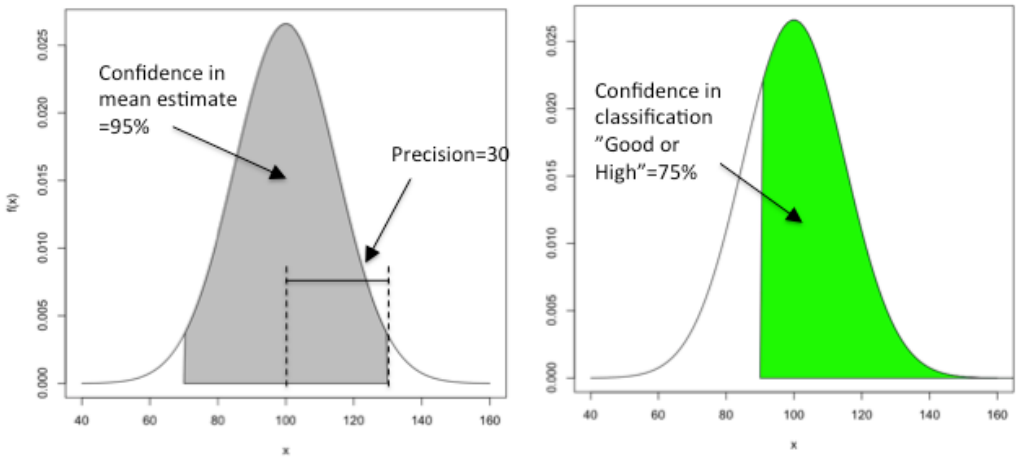

FIGURE 1.2
Schematics of the terms *precision* and *confidence*. Left: mean ±95% confidence interval; right: class boundaries (G–M = 90) and confidence of classification.

Confidence, on the other hand, is a concept related to precision but, unlike precision, is not a measure of the "goodness" of an estimate but a measure of the confidence associated with a classification (Figure 1.2), i.e. a probabilistic assessment of a statement (e.g. "75% confident that the status is good or high" or "the probability of the status being good or high is 75%"). The guidance documents state that confidence is:

> "The long-run probability (expressed as a percentage) that the true value of a statistical parameter (e.g. the population mean) does in fact lie within calculated and quoted limits placed around the answer actually obtained from the monitoring programme (e.g. the sample mean)".

Thus, confidence is an estimate of the probability that a certain classification is correct (the probability of an incorrect classification equals 1 – confidence). Although the term confidence may refer to the confidence interval, this is often trivial because it is defined by convention when the desired precision is defined. More important in terms of the WFD is the confidence in status classifications. The confidence in a classification depends on the precision of the estimated mean and the location of the mean in relation to class boundaries (Figure 1.2; see Annex A for technical definitions of confidence). Large confidence intervals and small deviations from class boundaries lead to poor confidence. One consequence of this is that any uncertainty about the location of class boundaries will introduce additional uncertainty in terms of reduced confidence. Because of its implications in terms of mitigation actions, the Directive also stipulates that the most important class boundary for classification is that between "Good" and "Moderate" (Guidance Document No. 10, p. 42); therefore, the confidence in classification at this boundary is crucial (Figure 1.2).

It should be noted that, although the Directive and its guidance documents provide conceptual definitions of precision and confidence, they do not provide quantitative rules or targets for acceptable levels of uncertainty. In the scientific community, the use of a conventional level of probability, i.e. $\alpha = 0.05$ and 95% confidence intervals, has long been dominant. For various reasons, this approach has in fact been strongly debated in the scientific community in recent decades (see Quinn & Keough 2002 for an accessible discussion), and it is clear that naïve use of such "rules" may be misleading. Furthermore, in the context of environmental impact assessment and status assessment according to the WFD, it is important to consider the risks and costs associated with various types of decision errors based on statistical arguments (e.g. Mapstone 1995). Nevertheless, the lack of guidance in matters concerning the level of confidence clearly introduces risks of arbitrariness and lack of coherence among countries and BQEs in status assessment procedures.

In conclusion, uncertainty in the estimation and classification of biological indicators is clearly unavoidable in WFD status assessment procedures. The Directive and its guidance documents acknowledge this, provide useful definitions of uncertainty, and stipulate requirements for reporting, specifying that all assessments should be associated with estimates of the precision of and confidence in classification.

1.1.2 Monitoring and uncertainty

Because variability and sample size are important determinants of precision and confidence, the design and dimensioning of monitoring programmes is crucial for the amount of uncertainty in any status classification according to the WFD. The Directive and its guidance documents provide some definitions and recommendation on these matters.

First, the Directive distinguishes among three types of monitoring: surveillance, operational, and investigative monitoring (WFD Annex V section 1.3). Briefly stated, the

purpose of surveillance monitoring is to provide data for an overall status assessment and for the detection of long-term trends and effects of human pressures on the BQEs in lakes, streams, and coastal and transitional waters. Operational monitoring is designed to assess the status of water bodies at particular risk of being classified below target and to assess whether mitigation actions have the desired effects. The aim of investigative monitoring is to disentangle the causes of any deviations from the desired status. Although these types of monitoring are sometimes difficult to distinguish in practice (e.g. with respect to sources of funding and division of responsibilities among authorities), the framework developed in this report is concerned mainly with surveillance and, to some extent, with operational monitoring.

Second, the Directive provides certain definitions as to the spatial and temporal context of monitoring and assessment. For example, it states that:

> "The monitoring network shall be designed so as to provide a coherent and comprehensive overview of ecological and chemical status *within each river basin* and shall permit classification of *water bodies* into five classes consistent with the normative definitions …
>
> On the basis of the characterisation …, Member States shall for each *period to which a river basin management plan applies* [i.e. six years], establish a surveillance monitoring programme and an operational monitoring programme." (p. 53)
>
> "Monitoring *frequencies* shall be selected which take account of the variability in parameters resulting from both natural and anthropogenic conditions. The *times* at which monitoring is undertaken shall be selected so as to minimise the impact of *seasonal variation* on the results, and thus ensure that the results reflect changes in the water body as a result of changes due to anthropogenic pressure." (pp. 55–56)

These sections identify particular spatial units for monitoring river basins and water bodies and recognise the importance of temporal aspects of monitoring, such as periods, frequency, and timing. No specific guidelines are provided as to the minimum number of samples or spatial units to be sampled, but regarding the sampling frequency, the Directive identifies the need for multiple sampling times during an assessment period (Figure 1.3). It should be stressed that the prescribed minimum frequencies are very likely to produce extremely uncertain estimates and consequently a high risk of misclassifications. It is also noteworthy that these sections: 1) attempt to differentiate among quality elements, based on differences in temporal variability, and 2) acknowledge the importance of temporal variability as a source of uncertainty regarding the overall status during an assessment period. The inadequate sampling frequencies suggested in Annex V were acknowledged in CIS Guidance Document No. 7, regarding monitoring, which recommended higher sampling frequencies for most BQEs and supporting elements.

Quality element	Rivers	Lakes	Transitional	Coastal
Biological				
Phytoplankton	6 months	6 months	6 months	6 months
Other aquatic flora	3 years	3 years	3 years	3 years
Macro invertebrates	3 years	3 years	3 years	3 years
Fish	3 years	3 years	3 years	
Hydromorphological				
Continuity	6 years			
Hydrology	continuous	1 month		
Morphology	6 years	6 years	6 years	6 years
Physico-chemical				
Thermal conditions	3 months	3 months	3 months	3 months
Oxygenation	3 months	3 months	3 months	3 months
Salinity	3 months	3 months	3 months	
Nutrient status	3 months	3 months	3 months	3 months
Acidification status	3 months	3 months		
Other pollutants	3 months	3 months	3 months	3 months
Priority substances	1 month	1 month	1 month	1 month

FIGURE 1.3
Minimum sampling frequency requirements for all of the quality elements defined in the WFD (from p. 56 of the WFD Annex V).

In summary, the Directive provides certain more or less specific definitions as to the spatial and temporal scope when monitoring the various BQEs. The primary spatial unit for assessment is the **water body** and the fundamental temporal unit is the management plan period, i.e. six years. These definitions have fundamental implications for the amount of uncertainty that we can expect and, more importantly, for how uncertainty is to be quantified and accounted for. This is because any estimate of variability and thus uncertainty always must be accompanied by a specific spatial and temporal context (e.g. Wiens 1989, Levin 1992, Schneider 2001).

Finally, it is also worth pointing out that one important aim of these definitions is to ensure that measures are taken to reduce uncertainty in estimation and classification. Two fundamental strategies for doing this can be identified. The first is based on the dimensioning and optimisation of spatial and temporal replication; this strategy will reduce uncertainty by adjusting the sample size. Second, uncertainty can be reduced by minimising and/or accounting for predictable sources of variability. Both these

approaches will be explored in WATERS. The framework developed in the present work will provide a solid foundation for both efforts.

1.2 Current procedures for handling uncertainty in Sweden

In Sweden the WFD is implemented by chapter 5 in the Environmental Code, the Ordinance on Water Quality Management (Vattenförvaltningsförordningen, (SFS 2004:660)) and regulations from the Environmental Protection Agency (Naturvårdsverkets föreskrifter och allmänna råd om klassificering och miljökvalitetsnormer avseende ytvatten, (NFS 2008:1)). Further guidance and advice on the application of assessment criteria are provided in the handbook "Status, potential och kvalitetskrav för sjöar, vattendrag, kustvatten och vatten i övergångszon" (2007:4) (see SEPA 2010 for a version in English). The latter contains general guidelines as well as BQE-specific information on how uncertainty should be assessed and dealt with within the Swedish assessment procedures.

The general guidelines provided in the handbook (chapter 4) stress that all classifications are assessed with respect to uncertainty and observe that how uncertainty is assessed differs among the BQEs. In accordance with CIS Guidance Documents nos. 7 and 13 (EC 2003, 2005), the handbook clarifies the importance of precision and confidence as central concepts in relation to uncertainty. Procedures for calculating precision and confidence based on replicate samples are presented. For situations in which replicate samples are missing but prior information on sampling variability is available, alternative routines based on the normal distribution assumption are given (see Annex A).

The table in Annex B summarises the recommendation and requirements for sampling and routines dealing with uncertainty for all existing indicators of BQEs in lakes, streams, and coastal waters (summarised from annexes A and B of the handbook). Several patterns emerge from a coherent analysis of these routines. First, the analysis indicates that, for all BQEs, sampling can be done using certain criteria defined in order to reduce variability and thus uncertainty in estimates. These criteria are often defined in methodological standards developed for monitoring programmes (Annex B). Sampling is typically done at standardised depths, substrates, and times of the year and guidelines are occasionally provided as to maximum distances among samples. These restrictions, which are based on ecological knowledge, have the effect of narrowing down the statistical population to be estimated and in most instances probably substantially reduce the uncertainty of estimated means. Nevertheless, we can still expect to encounter considerable uncertainty due to a number of "uncontrolled" factors.

Second, the intensity and resolution of spatial and temporal replication differ among BQEs. Some BQEs require replicate samples at individual sites, while the replication of sites varies among BQEs. Similarly, some BQEs require sampling several times per year while others do not. Clearly, these differences are often justified by sound ecological knowledge. One particular aspect related to the overall assessment procedures is that some BQE status assessment prescribe that data from a number of years (typically ≥3)

should be incorporated in estimates of status (e.g. phytoplankton in lakes and coastal waters). For other BQEs, it is stated that estimates are preferably based on "several measurements", but it is not clear whether these involve several years of data, and a minimum number is not prescribed. This means that the status of some BQEs is assessed based on individual years, while the status of others is averaged over a number of years, thus including uncertainty due to year-to-year variations.

Finally, it is also evident that recommendations about how to deal with uncertainty in estimates in the assessment procedure differ among BQEs and in some respects do not completely cover all aspects of uncertainty defined in the WFD guidance documents. One striking difference is the use of different measures of precision. All the metrics for lakes and streams employ the standard deviation as a measure of precision, while the metrics for coastal waters (a) make no mention of measures of precision for coastal phytoplankton, (b) use a one-sided bootstrap confidence interval (80%) for coastal macrofauna, and (c) use the standard deviation for macrophytes (note that recommendations as to the level of precision are given for only one BQE). Furthermore, in lakes and streams, standard deviations indicative of methodological uncertainties are given in tables (except in the case of fish, for which an empirical formula is given). Some of these differences may relate to differences in sampling designs (i.e. differences in spatial and temporal replication). Nevertheless, this diversity in recommendations and routines clearly could obscure a coherent assessment of uncertainty, lead to arbitrariness in the handling of uncertainty in whole-system assessment, and cause confusion in practical application. The analysis also indicates that considerations related to the confidence of estimates and the precision of reference conditions or class boundaries are not covered, even though these are central concepts in the Directive and its guidance documents.

These initial analyses have identified several issues in relation to uncertainty in estimation and classification that are largely or partly unresolved in the Swedish assessment criteria. Specifically, issues related to: 1) the quantification of uncertainties at different temporal and spatial scales (e.g. a whole assessment cycle, individual years, water bodies or water body types) in situations in which spatial and temporal replication is sub-optimal, and 2) decision rules in statistical tests of deviations from the good–moderate boundary. To address these issues, we propose a framework-based identification, estimation, and combination of various components of variability. This allows the more realistic estimation of precision and confidence than is permitted by current procedures and creates the prospect of harmonising assessment criteria in relation to the handling of uncertainty. Furthermore, this framework provides a solid foundation for attempts to reduce uncertainty by optimising monitoring designs and incorporating important environmental factors as covariates.

2 Objective

The objective of this report is to review important concepts, routines, and scientific tools for assessing and treating uncertainty in the estimation and classification of ecological status indicators. Special attention is paid to central concepts defined in the WFD and to the routines described in Swedish assessment criteria and routines.

This review will form the basis for developing a general framework for treating measurement uncertainty across all BQEs. The framework needs to be flexible in that it acknowledges fundamental differences in spatio-temporal variability, sampling costs, and other specific ecological differences among BQEs, but also general in that it can be applied to all BQEs and possibly also to other quality elements relevant to the WFD and related directives. To assess its general applicability, we test this framework by applying it to two examples using real monitoring data.

The aim is to improve the treatment of uncertainty in revised versions of the Swedish assessment criteria and routines. The main benefits will be better appraisal of the uncertainties associated with the estimation and classification of ecological status, more user-friendly routines through better harmonisation and transparency, and ultimately reduced uncertainty due to more appropriate sampling designs and by proposing various ways to account for important factors contributing to the overall uncertainty.

3 Sources of uncertainty

Assessing the ecological status of water bodies is inherently associated with uncertainty; accordingly, the WFD Common Implementation Strategy guidelines (no. 13) stipulate that the confidence of the classification should be reported. To report the confidence of an indicator referring to different status classes, the uncertainty of the indicator must be quantified. As outlined in CIS guideline no. 13, many sources of uncertainty can contribute to the overall uncertainty of the indicator.

The observations used for calculating an indicator contain elements of uncertainty stemming from the sampling and analysis of the sample. The various sources of uncertainty can be grouped into three categories of variation: temporal components, spatial components, and components associated with sampling and analyses. These sources of variability can be completely or partly *fixed* and completely or partly *random*. The distinction between fixed and random factors is sometimes difficult to make, but may greatly affect how they contribute to the uncertainty of a status assessment and how the uncertainty is calculated (e.g. Clarke 2012). For the purposes of the present report, we define fixed (i.e. predictable) and random (i.e. unpredictable) components as follows:

- **Fixed components** are either continuous variables displaying a linear or otherwise predictable relationship with the response variable, or a categorical variable with a limited number of classes for which means of the response variable differ. The component is completely fixed if the continuous variable completely explains the variability (not relevant) or if all levels of the categorical variable are sampled (i.e. all years within an assessment cycle). In this report, fixed components are denoted by *lowercase* letters.
- **Random components** are spatial and temporal components of variability that cannot be attributable to any continuous variable in a predictable way, using currently existing data or models. In this report, random components are denoted by *CAPITAL* letters.

3.1 Uncertainties associated with temporal variability

According to the WFD, the classification should be carried out for six-year periods, and the indicator should characterize the overall mean conditions for that six-year period. Since the water body cannot be continuously monitored throughout the assessment period, the overall temporal mean should be assessed based on discrete samples in time.

Variations in environmental time series can generally be partitioned into trends (i.e. interannual variation), seasonal variation, diurnal variation, and irregular fluctuations.

- **Interannual variation** describes the variation between years, and this variation is partly fixed and partly random. The fixed variation can be described by external factors influencing the environmental time series, such as temperature, freshwater, and discharge, whereas the random variation describes the remaining unexplained interannual variation.
- **Seasonal variation** describes the variation within a year that is of a cyclic character. It can also be partitioned into a fixed component, which is the mean seasonal variation repeating itself every year, and a random component, which describes fluctuations around the mean seasonal variation.
- **Diurnal variation** describes the variation within a day with a cyclic character. It can also be partitioned into a fixed component, which is the mean diurnal variation repeating itself every day or in response to changes in day length or another external factor with a diurnal pattern, and a random component, which describes additional fluctuations around the fixed diurnal variation.
- **Irregular fluctuations** are random short-term variations between samples taken within a time interval that is short relative to the other factors.

3.2 Uncertainties associated with spatial variability

The ecological status assessment of a water body should apply to the entire water body and not just consider a single monitoring station. Since it is impossible to monitor every parcel of water or every square meter of the bottom, the status of the water body should be assessed from a few spatially distinctive monitoring stations. The spatial variation can be partitioned into large-scale gradients within the water body and small-scale fluctuations.

- **Large-scale gradients** describe the spatial variation within a water body, and this variation is partly fixed and partly random. The fixed variation can be explained by differences in depth, sediment, substrate, salinity, etc., whereas the random variation describes the remaining unexplained spatial variation.
- **Small-scale fluctuations** describe random variations between samples taken near each other, for example, benthic samples from the same station.

3.3 Uncertainties associated with sampling and analysis

These uncertainties relate to the methods, materials, and people used when sampling and measuring the variable in question and, as such, are highly specific to the actual type of monitoring.

- **Variation between sampling devices** describes the variation between different sampling methods. For example, in the case of water samples, there could be differences between Niskin bottle and hose samples; for benthic vegetation cover,

there could be differences between video recording and diver assessment; and for benthic fauna, there could be differences between van Veen grab and Smith-McIntyre grab sampling. Since the number of sampling methods is fairly limited and the methods are assumed to be intercalibrated, this factor would normally be considered fixed.

- **Person(s) conducting the sampling and analysis** accounts for the human factor affecting the measurement. For example, there are differences between taxonomists counting the phytoplankton samples, between divers assessing macroalgal cover and species-specific depth limits, and between people operating HPLCs and other devices. This source of uncertainty is random, since it must account for all people potentially involved in the sampling and analysis.
- **Analytical variation between instruments** describes the variation caused by using different types of instruments (i.e. different brands and models) to measure constituents such as chlorophyll a. Since the number of different types of instruments is limited, this factor would be considered fixed.
- **Replicate and sub-sampling uncertainty** accounts for the random variation occurring if a sample measurement is replicated or a sample is subdivided into several samples that are analysed separately.

3.4 Uncertainties due to interactive variability

These ten different sources of uncertainty as well as their interactions may significantly affect the various indicators used for ecological status classification in the WFD. Although most of the interactions can be considered irrelevant and are often set to zero, two of them warrant more consideration. First, the interaction between interannual variation and large-scale spatial variation could be a significant source of random variation, in that there could be large-scale shifts in spatial distributions across years. Second, there could also be differences in seasonal variation across the large-scale spatial gradient. However, interactions between the large-scale spatial gradient and the diurnal variation as well as irregular fluctuations are more difficult to interpret and hence could be assumed to be irrelevant and set to zero for practical purposes. Similarly, the possible interactions between small-scale spatial variation and temporal variation can be considered small and assumed to be negligible. Finally, it can also be justifiable to assume that the methodological uncertainty associated with sampling and analysis is independent of the sampling in time and space, so all interactions between these factors can be set to zero.

3.5 Combining uncertainties of BQE monitoring data

A measurement variable can be assumed to be governed by the following sources of variation:

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

where fixed effects are shown in lowercase letters and random effects in capital letters. However, it is difficult to quantify all of these separately, since this would require an unrealistically large monitoring programme combining the factors at different levels. In practice, it is only possible to estimate a few of these factors from monitoring data, and the ones that can be estimated are specific to each type of monitoring data and programme. Another issue is that several of the above factors may contribute relatively little variation to the observations and therefore not merit inclusion.

In the following tables, the interpretation and possible relevance of the various factors are assessed. The relevance is assessed in relation to any uncertainty the factors may add to the estimation of the ecological status of single water bodies throughout a six-year assessment period. Therefore, they do not primarily assess the importance of large spatial (e.g. biogeographic and among water bodies or water body types) or temporal (e.g. climatic trends or decadal shifts) scales. Similar qualitative and quantitative assessments of several quality elements, including attempts to synthesise information from various parts of Europe, were conducted in the WISER project (www.wiser.eu; e.g. Neto et al. 2012, Courrat et al. 2012, Dudley et al. 2012, Thackeray et al. 2012). Furthermore, it must be stressed that these assessments are qualitative and relative to other components: there may be circumstances in which components deemed irrelevant here may add some uncertainty to estimated means. To assess uncertainty and to design sampling programmes, it is crucial to obtain quantitative estimates of these components. Procedures for estimating variability and uncertainty are described elsewhere in this report, and realistic, numerical examples of how variability and uncertainty are calculated are given in chapter 5.

3.6 Uncertainty in estimates of phytoplankton

Temporal variations in phytoplankton characteristics are dynamic and will contribute substantial variation (Table 3.1) on the interannual scale (as both predictable, *year*, and unpredictable, *YEAR*) and seasonal scale (as both predictable, *season*, and unpredictable, *SEASON*). However, it is more difficult to assess the relevance of diurnal variation (DIURNAL) for phytoplankton, as vertical migration can be important for some

communities in some waters, whereas it could be irrelevant in other cases. In addition, large fluctuations in the short-term dynamics will be included in the IRREGULAR component. It is also believed that most water bodies will exhibit a pronounced and predictable spatial gradient (gradient) in response to depth, nutrient conditions, salinity, etc., but this represents only a fraction of the explicable large-scale spatial variation, so the remaining variation is termed random (GRADIENT). The relative spatial distribution of phytoplankton may not necessarily be static, i.e. changing similarly over time in a given spatial segment, since the spatial patterns may change between years (*YEAR × GRADIENT*) and over the season (*SEASON × GRADIENT*). Finally, the sampling and analytical uncertainties associated with measuring chlorophyll a are considered small and likely not relevant, as using different water samplers, people, and instruments is believed to change the measurement results only marginally, and the variation between duplicate or triplicate measurements (REPLICATE) is small. However, for phytoplankton counts, the human factor (i.e. taxonomical skills of the person identifying and enumerating the specimens) is substantial, and there can be large variations between sub-samples, even when analysed by the same person.

TABLE 3.1
Importance of different sources of variation in phytoplankton characteristics assessed using chlorophyll a or pigment analyses as well as in phytoplankton counts for estimating the phytoplankton volume/biomass and composition.

Type of uncertainty	Uncertainty component	Chla/pigments	Phytoplankton volume/biomass	Composition
Temporal sampling	*Year*	Relevant	Relevant	Relevant
	YEAR	Relevant	Relevant	Relevant
	season	Relevant	Relevant	Relevant
	SEASON	Relevant	Relevant	Relevant
	DIURNAL	Maybe relevant	Maybe relevant	Maybe relevant
	IRREGULAR	Relevant	Relevant	Relevant
Spatial sampling	*gradient*	Relevant	Relevant	Relevant
	GRADIENT	Relevant	Relevant	Relevant
	PATCHINESS	Relevant	Relevant	Relevant
Spatio-temporal interaction	*YEAR × GRADIENT*	Relevant	Relevant	Relevant
	SEASON × GRADIENT	Relevant	Relevant	Relevant
Sampling method	*sampling device*	Not relevant	Not relevant	Not relevant
	PERSON	Not relevant	Relevant	Relevant
	instrument	Not relevant	Maybe relevant	Maybe relevant
	REPLICATE	Small	Relevant	Relevant

3.7 Uncertainty in estimates of benthic vegetation

Temporal variations in benthic vegetation can be large, although not nearly as large as those of phytoplankton, so it is not relevant to consider short-term irregular variations (*IRREGULAR*) or diurnal variations (*DIURNAL*) (Table 3.2). Benthic vegetation typically has a characteristic unimodal seasonal variation, with the largest biomass, cover, and shoot density occurring in late summer. Hence, seasonal patterns, both fixed (*season*) and random (*SEASON*), are important sources of variation. However, benthic vegetation is normally sampled only during the summer–early autumn period, to allow for comparison across years and sampling locations without taking the seasonal variation into account. This approach of excluding the seasonal variation is permissible, provided that sampling is carried out only during a fairly invariant seasonal window and that the sampling methodology stays the same in the future. Interannual variations can be large in response to changing light conditions, nutrient levels, physical disturbances, etc., and these variations are partly predictable (*year*) and unpredictable (*YEAR*). Benthic vegetation exhibits a pronounced and partly predictable spatial pattern (gradient) in response to depth, sediment/substrate characteristics, salinity, and nutrient levels. However, a relatively large spatial variation (*GRADIENT*), unexplainable by other governing factors, is believed to remain. This large random spatial variation may even vary substantially between years (*YEAR* × *GRADIENT*). If the benthic vegetation is sampled, for example, using a frame to collect data, the actual choice of sampling device (*sampling device*) may influence the outcome, though this possibility is not well documented. In most cases, the same sampling device is used over time, so it may not be relevant to consider this source of uncertainty. On the other hand, the uncertainty associated with the person (*PERSON*) analysing the sample or monitoring the benthic vegetation can be quite substantial, particularly at the taxonomical level. The choice of monitoring instrument, for example, underwater cameras, can contribute some uncertainty to the observations, but this still needs to be further investigated. Replicated monitoring of the exact same location or of the same sample is not carried out in benthic vegetation monitoring, because it is considered too expensive or impossible (e.g. to ask a diver to repeat the exact same transect) or because sampling is destructive and replicated measurement is impossible. The uncertainty associated with replicated measurement (*REPLICATE*) cannot be assessed and thus will be confounded with other sources of variation.

TABLE 3.2
Importance of different sources of variation in benthic vegetation characteristics assessed using cover and depth limits for the community as a whole and for specific key species as well as the taxonomical composition.

Type of uncertainty	Uncertainty component	Community abundance (cover & depth limits)	Abundance of key species (cover & depth limits)	Composition
Temporal sampling	*year*	Relevant	Relevant	Relevant
	YEAR	Relevant	Relevant	Relevant
	season	Maybe relevant	Maybe relevant	Maybe relevant
	SEASON	Maybe relevant	Maybe relevant	Maybe relevant
	DIURNAL	Not relevant	Not relevant	Not relevant
	IRREGULAR	Not relevant	Not relevant	Not relevant
Spatial sampling	*gradient*	Relevant	Relevant	Relevant
	GRADIENT	Relevant	Relevant	Relevant
	PATCHINESS	Relevant	Relevant	Relevant
Spatio-temporal interaction	*YEAR × GRADIENT*	Relevant	Relevant	Relevant
	SEASON × GRADIENT	Relevant	Relevant	Relevant
Sampling method	*sampling device*	Maybe relevant	Maybe relevant	Maybe relevant
	PERSON	Relevant	Relevant	Relevant
	instrument	Maybe relevant	Maybe relevant	Maybe relevant
	REPLICATE	Maybe relevant	Maybe relevant	Maybe relevant

3.8 Uncertainty in estimates of benthic diatoms

Benthic diatoms are sessile microscopic algae, primary producers often dominating periphytic communities on stones and other substrates in freshwater streams and lakes. It is a very taxa-rich group, and more than one hundred taxa can often be found within a few cm^2. Because different taxa have different tolerances and sensitivities to environmental and human stressors, the taxon composition and the relative abundance of each taxon are considered suitable for detecting human impacts. The time scale of the response of benthic diatom indices to short- and long-term changes is poorly studied, but some studies suggest that temporal variability is usually smaller than that among sites or due to human stressors. Therefore, temporal variations in benthic vegetation are probably only of medium importance to random (e.g. *YEAR* and *SEASON*) as well as fixed (e.g. *year* and *season*) components (Table 3.3). In regular monitoring, some of this variability is accounted for by restricting sampling to late summer–autumn. Because of their relatively short generation time and sensitivity to temporary disturbances, such as high water flow, short-term irregular random variations (*IRREGULAR*) may be more important than predictable diurnal variations (*DIURNAL*). The spatial components of uncertainty involve both predictable factors (*gradient*), such as depth, water flow, and bottom

substrate, as well as larger-scale *GRADIENTs* and small-scale *PATCHINESS*, which are mainly unpredictable and potentially important. To eliminate the uncertainty due to small-scale *PATCHINESS*, the required monitoring standard specifies that diatoms should be sampled from at least five stones (or macrophyte leaves if stones are absent) from a 10-m section of the stream, and then pooled into one sample per site. Furthermore, because of the unpredictability of yearly and seasonal fluctuations, that random spatio-temporal sources of variability (*YEAR* × *GRADIENT* and *SEASON* × *GRADIENT*) are likely important for the abundance and composition of benthic diatom communities. The main source of variability potentially associated with sampling and analytical methods is the difference in diatom identification in the laboratory among *PERSON*s. Diatom communities are diverse, and any metric constructed from these assemblages is dependent on accurate and reliable taxa identification. The importance of this uncertainty is underlined in the Swedish handbook (2007:4, p. 63), which stresses that 80% of the method-bound uncertainty is due to taxa identification. Finally, because the number of diatoms in one five-stone sample of diatom communities collected in the field is enormous, it is not feasible to count and identify every single cell. Therefore, uncertainties due to sub-sampling (*REPLICATE*) cannot be ruled out in the case of benthic diatoms.

TABLE 3.3
Importance of different sources of variation in benthic diatom characteristics assessed to determine the relative abundance of all diatom taxa in a sample.

Type of uncertainty	Uncertainty component	Relative abundance of taxa
Temporal sampling	*year*	Maybe relevant
	YEAR	Maybe relevant
	season	Maybe relevant
	SEASON	Maybe relevant
	DIURNAL	Not relevant
	IRREGULAR	Relevant
Spatial sampling	*gradient*	Relevant
	GRADIENT	Relevant
	PATCHINESS	Relevant
Spatio-temporal interaction	*YEAR × GRADIENT*	Relevant
	SEASON × GRADIENT	Relevant
Sampling method*	*sampling device*	Maybe relevant
	PERSON	Relevant**
	instrument	Not relevant
	REPLICATE	Maybe relevant

* "Sampling method" includes field sampling, preparation of permanent slides in the laboratory, and identification under the microscope. For the benthic diatom method, studies have demonstrated that the importance of identification (*PERSON*) is the largest source of uncertainty, followed by field sampling; slide preparation is only a small source of uncertainty (M. Kahlert, pers. comm.).

3.9 Uncertainty in estimates of benthic fauna

Benthic fauna in Swedish coastal and inland waters consists of a wide range of taxonomic and functional groups. The species composition and thus the ecological traits of benthic fauna differ dramatically between inland and coastal waters. The fauna of inland waters is typically dominated by a great variety of insect families, while that of coastal waters is more diverse at the level of phyla. Coastal sediments are typically dominated by a rich fauna of polychaetes, molluscs, crustaceans, and, on the west coast, echinoderms. Despite these differences, particularly across the dominant salinity gradient extending from inland waters and the low saline areas in the Bothnian Bay, through the Baltic Sea, to the oceanic conditions in the northern Skagerrak, certain common features can be identified in terms of the spatial and temporal components of variability.

In comparison with planktonic algae and many species of benthic vegetation, the temporal variability of macroscopic benthic fauna is usually less pronounced. This is both due to a less dynamic environment and because these organisms are more long-lived. Short-term *DIURNAL* and *IRREGULAR* components can therefore generally be neglected when estimating the biomass, abundance, and composition of benthic fauna (Table 3.4). Variability at larger time scales, however, is generally more important. Predictable and random yearly components (*year* and *YEAR*) can clearly be very important. The exact causes of these components are often very complex, but processes involving recruitment, food supply, and other biological interactions, partly those influenced by differences in meteorological and climatic factors, are likely to be important. It is also clear that predictable and random variability associated with seasonality (*season* and *SEASON*) may be important. Nevertheless, because benthic invertebrates are sampled during the same fixed periods of the year (i.e. spring in coastal waters and autumn in inland waters), the fixed component does not in practice add to the uncertainty of monitoring results.

The abundance and composition of benthic fauna are also highly variable in space. The variability within a single water body may be due to relatively large-scale and predictable *gradient*s such as depth, salinity, substrate, or wave exposure. Other large- (*GRADIENT*) and small-scale (*PATCHINESS*) sources of spatial variability may be more difficult to understand, and for all practical purposes can be considered random. This also applies to a host of spatio-temporal interactions (e.g. *YEAR* × *GRADIENT* and *SEASON* × *GRADIENT*). These interactions involve components associated with the varying strength of the effects of gradients among years and seasons. Finally, uncertainty may also be associated with the sampling procedures. In general, *sampling device* and *instrument* standardisation renders these fixed sources of variability less relevant. In coastal environments, the van Veen grab sampler is the dominant sampling device, whereas in lakes and watercourses, different sampling methods may be used. However, one potential source of uncertainty regarding benthic fauna is that due to species identification. As the benthic fauna consists of a very large number of species, considerable and specialised skills are required from the personnel. Thus, despite rigorous routines for quality assurance, *PERSON*-dependent variability is an issue, particularly in the case of long-term data series. In comparison with large-scale gradients, it is probably safe to conclude that

variability due to *REPLICATE* is smaller. Nevertheless, estimates of replicate variability from the Skagerrak suggest that this type of variability can account for approximately 15% of the variability observed within a water body. Examples from coastal areas are examined in more detail in section 6.2.

TABLE 3.4
Importance of different sources of variation in benthic fauna characteristics assessed using biomass, abundance of species weighted by their sensitivity to pollution (e.g. BQI), and overall taxonomical composition.

Type of uncertainty	Uncertainty component	Biomass	Abundance of sensitive/tolerant species	Composition
Temporal sampling	*year*	Relevant	Relevant	Relevant
	YEAR	Relevant	Relevant	Relevant
	season	Maybe relevant	Maybe relevant	Maybe relevant
	SEASON	Maybe relevant	Maybe relevant	Maybe relevant
	DIURNAL	Not relevant	Not relevant	Not relevant
	IRREGULAR	Relevant	Relevant	Relevant
Spatial sampling	*gradient*	Relevant	Relevant	Relevant
	GRADIENT	Relevant	Relevant	Relevant
	PATCHINESS	Relevant	Relevant	Relevant
Spatio-temporal interaction	*YEAR × GRADIENT*	Relevant	Relevant	Relevant
	SEASON × GRADIENT	Not relevant	Not relevant	Not relevant
Sampling method	*sampling device*	Maybe relevant	Maybe relevant	Maybe relevant
	PERSON	Not relevant	Relevant	Relevant
	instrument	Not relevant	Not relevant	Not relevant
	REPLICATE	Relevant	Relevant	Relevant

3.10 Uncertainty in estimates of fish

For practical reasons, monitoring of fish in lakes, streams, and coastal areas in Swedish national programmes is done using four different methods (see http://www.havochvatten.se/kunskap-om-vara-vatten/miljo--och-resursovervakning/provfiske-i-kust---sotvatten.html): in lakes and some coastal areas, various nets are used; in the coastal Skagerrak and Kattegat, fyke nets are used; and in streams, sampling is done using electrofishing. However, in all of these sampling programmes, the abundance, species composition, and weight of fish are measured.

In a pan-European study, temporal and spatial sources of variation in transitional areas were assessed qualitatively and quantitatively (Courrat et al. 2012). Similar studies of lakes and streams are unavailable, but the fact that the Swedish assessment criteria for these BQEs use models based on environmental (fixed) factors to predict reference conditions

and that they report uncertainty due to methodological errors (Annex B) suggest that information on these issues is available. Furthermore, taken together, this information suggests that fixed components due to temporal (*year* and *season*) as well as spatial *gradient*s are likely very relevant (Table 3.5). These patterns are ultimately related to temporal and spatial variability due to temperature, salinity, depth, etc. (likely differing among lakes, streams, and coastal areas and also dependent on temporal and spatial scale). Of course, it is very likely that fish assemblages vary both predictably and unpredictably among seasons (*season* and *SEASON*), but because monitoring is usually restricted to certain times of the year, these components may not be particularly relevant to the uncertainty of monitoring data in practice. The abundance and composition of fish assemblages may vary depending on time of day in a particular habitat (or at least what is caught in nets) because of migration and differences in activity levels. However, because most sampling techniques are standardised and integrate over a substantial part of the day, DIURNAL variability is not deemed relevant to monitoring data. Components of the sampling and analytical processes, particularly components having to do with differences in methods among areas (e.g. sampling device), may also be relevant to the uncertainty of fish measurements. *PERSON* dependence is mainly considered important in the case of measurements involving species identification, although uncertainty from this source can probably be substantially reduced by education. Finally, because net sampling is dependent on fish movements and, accordingly, its efficiency may vary, many samples are often necessary to achieve sufficient precision within a water body, suggesting that uncertainty due to *REPLICATE* is usually very relevant.

TABLE 3.5
Importance of different sources of variation in fish characteristics assessed using biomass, abundance of various key species, and overall taxonomical composition.

Type of uncertainty	Uncertainty component	Abundance of key species	Composition	Weight
Temporal sampling	*Year*	Relevant	Relevant	Relevant
	YEAR	Relevant	Relevant	Relevant
	Season	Maybe relevant	Maybe relevant	Maybe relevant
	SEASON	Maybe relevant	Maybe relevant	Maybe relevant
	DIURNAL	Not relevant	Not relevant	Not relevant
	IRREGULAR	Relevant	Relevant	Relevant
Spatial sampling	*gradient*	Relevant	Relevant	Relevant
	GRADIENT	Relevant	Relevant	Relevant
	PATCHINESS	Relevant	Relevant	Relevant
Spatio-temporal interaction	*YEAR × GRADIENT*	Relevant	Relevant	Relevant
	SEASON × GRADIENT	Not relevant	Not relevant	Not relevant
Sampling method	*sampling device*	Relevant	Relevant	Relevant
	PERSON	Relevant	Relevant	Not relevant
	instrument	Not relevant	Not relevant	Not relevant
	REPLICATE	Relevant	Relevant	Relevant

4 Methods for quantifying sources of variability

To estimate and model the uncertainty of status assessments involving data on BQEs collected from monitoring programmes, it is fundamental that the different components of variability can be separated and quantified. This can be done using various methods and approaches, ranging from simple analysis of variance (ANOVA), through general/generalised linear models (i.e. GLMs, GLZs), to Bayesian frameworks and simulation methods. In some circumstances, some of these approaches may converge, while in others a particular method may be preferable. Nevertheless, the quality of estimates of different components of variability is clearly central to the reliable analysis of precision and confidence. This requires that the best available data be combined with flexible and reliable methods for quantifying the components of variability.

The uncertainty framework described in the previous chapter involved many sources of variation, including both fixed and random factors. A statistical model that includes both fixed and random factors is termed a mixed model (e.g. Bolker et al. 2009). The aim of these models is to estimate ("parameterise") the separate fixed effects and random components of variability. The parameterisation of the fixed effects can be linear or non-linear. The distribution assumption of the dependent variable (i.e. the response variable) is not restricted to the normal distribution. Therefore, these models can be expanded to estimate random components from all distributions belonging to the exponential family (e.g. Poisson, binomial, multinomial, gamma, and negative binomial distributions). This expanded framework is called generalised mixed models. However, it is quite common in environmental science to use linear models and assume the data to be normally distributed, and in such cases the models are termed general linear mixed models.

Linear mixed models are usually formulated in matrix notation as:

$$\underline{y} = \underline{\underline{X}} \times \underline{\beta} + \underline{\underline{Z}} \times \underline{u} + \underline{\epsilon}$$

where $\underline{y}$ is a vector of all the observations, $\underline{\underline{X}}$ is a matrix of the variables used in the fixed effects, $\underline{\beta}$ is a vector of unknown parameters for the fixed effects, $\underline{\underline{Z}}$ is a matrix of the variables used in the random effects, $\underline{u}$ is a vector of unknown random-effect errors having a zero mean and dispersion matrix, $\underline{\underline{G}}$, and $\underline{\epsilon}$ is the residual error term of dispersion matrix $\underline{\underline{R}}$. The mean of $\underline{y}$ is given by the fixed effect $\underline{\underline{X}} \times \underline{\beta}$, and the variance of $\underline{y}$ is given by $\underline{\underline{V}} = \underline{\underline{Z}} \times \underline{\underline{G}} \times \underline{\underline{Z}}' + \underline{\underline{R}}$. Thus, in simple terms, the fixed effects explain the mean of the observations and the random effects explain the variance of the observations. However,

we will not explore the underlying mathematics of mixed models any further, but simply note that many statistical textbooks deal with mixed models.

Several methods are available for estimating the parameters of mixed models, but the most common is to maximise the likelihood function – either full information maximum likelihood (FIML) or restricted maximum likelihood (REML). However, the main difference between general linear models (without random effects) and mixed models is that the parameters of a general linear model can be estimated explicitly by means of least squares, whereas the parameters of a mixed model are found by iteratively improving a goodness-of-fit function with respect to the parameters until the optimum value is reached. In addition to the classical statistical estimation techniques, several computer-intensive algorithms, mostly using Monte Carlo simulation techniques, have been developed in recent years (e.g. GLUE: Beven & Binley 1992; MCMC: Christian & Casella 2004).

5 Combining uncertainties

The previous chapters have demonstrated that estimates of the mean status and classification of a BQE are associated with a certain amount of uncertainty and that this uncertainty is attributable to a diverse set of components. The magnitudes of these components can often be determined using statistical methods; using this information, it is possible to calculate the overall uncertainty of a given estimate and to model and minimise the uncertainty of future estimates from alternative monitoring designs.

To assess the overall uncertainty of an estimated mean, it is important to identify different temporal, spatial, and analytical sources of variability and to combine these using general procedures for uncertainty (or error) propagation (e.g. Cochran 1977, Taylor 1997). Several accounts of uncertainty propagation exist in relation to bioassessment methods in general and to the WFD in particular (e.g. Clarke et al. 2002, 2006a,b, Clarke & Hering 2006, Bennet et al. 2011, Mascaró et al. 2012). These studies have demonstrated the need for the combined assessment of various sources of uncertainty, of the spatial and temporal context of uncertainties, and of the benefits of reducing uncertainty by optimising sampling designs.

The fundamental temporal and spatial units for the assessment of ecological status according to the WFD are six-year periods and water bodies. These are the units for which precision and confidence in classification primarily need to be assessed. At present, no routines implemented for BQEs in the Swedish assessment criteria address all uncertainties at these temporal and spatial scales. The consequences of this deficiency may vary among BQEs. Using "method-bound" variability, as in the case of many freshwater BQEs in individual years, may lead to the underestimation of the indicator uncertainty. On the other hand, assessing status on the basis of individual years instead of combining measurements from several years may lead to the overestimation of uncertainty, because fewer observations are used for the assessment, i.e. the full potential of yearly sampling programmes is not used.

To bridge the gap between: 1) the temporal and spatial scales of assessment defined in the WFD, and 2) those implemented in the Swedish assessment routines, two fundamental scenarios based on common monitoring strategies are developed below. The aim is to provide a framework for combining different sources of uncertainty and to illustrate how sampling design and dimensioning affect precision and confidence. For the sake of simplicity, the examples below assume that all spatial variability is random, i.e. there are currently no fixed factors that can explain variability among or within sites. Similarly,

samples are taken at fixed times once a year and the seasonal components of variability can be disregarded. Variability among years is essentially random, but because there is a finite number of years to be sampled in each assessment period (≤6 years), we introduce a correction term (1 – the number of sampled years/6) reducing (and finally removing, if all years are monitored) the uncertainty associated with interannual variation (e.g. Cochran 1977, Clarke 2012). The scenarios illustrate sampling within an individual coastal, lake, or stream water body where a number of sites (S) are sampled for a number of years (Y). In each of *a* years, *n* samples are taken at *b* sites. The sites may be identical from year to year ("orthogonal design"); alternatively, new sites may be selected each year ("nested design"). As will be demonstrated below, the choice of sampling design has fundamental consequences for estimates of uncertainty within a water body over an assessment period.

5.1 Precision of the estimated metric

5.1.1 Orthogonal (crossed) design

One monitoring design representative of most current programmes in aquatic environments in Sweden is a design in which sites ("stations") are revisited and sampled repeatedly year after year (Figure 5.1). The sites may have been selected completely at random within the water body or using criteria such as a narrow depth range, substrate, or distance from shore (see Annex B). The important thing is that the sites are selected to "represent", to some degree, the water body or a defined stratum thereof. Note also that the numbers of sites (*b*) and samples (*n*) vary strongly among monitoring programmes.

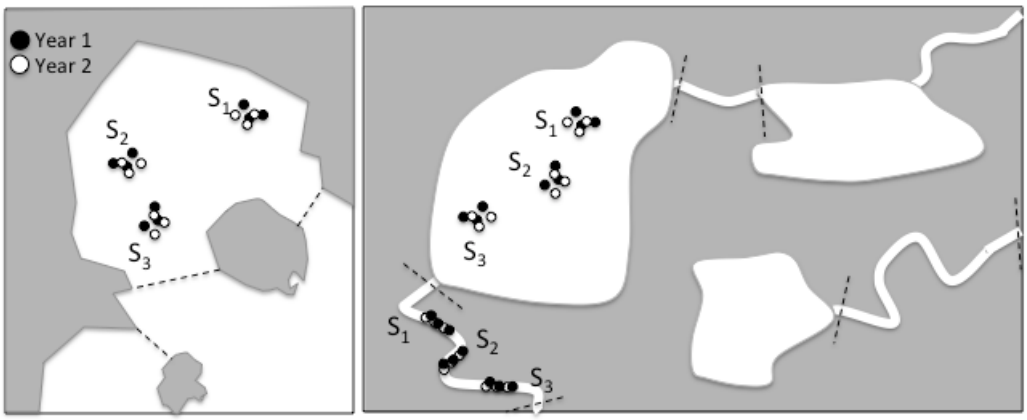

FIGURE 5.1
Schematics of orthogonal monitoring designs in a coastal water body (left) and in a lake and stream (right). In the examples, *a* = 2, *b* = 3, and *n* = 3.

Each measurement made in such a programme may be expressed using a linear model, in which the measured value, *y*, is the sum of the overall mean, μ, and deviations due to the other sources of variability.

$$y = \mu + YEAR + SITES + YEAR * SITE + PATCHINESS$$

Thus, the variability of the overall mean for such a sampling design consists of several variance components, i.e., s_Y^2, s_S^2, s_{Y*S}^2, and s_e^2 (unknown but estimated from data), each associated with a different source of variability in the linear model. The total variance of the estimated mean, $\bar{y}$, resulting from these components in an orthogonal design can be calculated as:

$$V[\bar{y}] = \frac{s_Y^2 * (1 - \frac{a}{Y})}{a} + \frac{s_S^2}{b} + \frac{s_{Y*S}^2}{ab} + \frac{s_e^2}{abn}$$

This formula for error propagation indicates how individual uncertainty components are combined into a total variance estimate and, importantly, how the numbers of samples, sites, and years affect the variance and uncertainty. Increasing the number of samples reduces the uncertainty due to small-scale variability within sites and years, but does not affect the uncertainty caused by variability among years or sites. Monitoring at many sites reduces the uncertainty due to sites and samples, but does not cause any reduction in the uncertainty due to years. Similarly, sampling at a number of years reduces the uncertainty due to years, interactive variability, and patchiness, but not among sites. Note also that if all years within an assessment period are sampled, i.e. $a = Y = 6$, all possible levels of the factor are sampled, which implies that the distribution over the six years (constituting the entire relevant population) is known (estimated) and therefore does not contribute any random variation.

To represent uncertainty, however, the total variability, $V[\bar{y}]$, needs to be transformed into a measure of the standard error of the mean, SE_{Tot}, and finally into a confidence interval according to:

$$CI\% = \left[\sqrt{V_{Tot}} * t_{\alpha/2,df}; \sqrt{V_{Tot}} * t_{1-\alpha/2,df}\right]$$

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 $\alpha = 5\%$) with *df* effective degrees of freedom. Since the variance of the mean ($V[\bar{y}]$) is calculated from several variance components, each likely determined from a different number of replications and thus different degrees of freedom, an exact value cannot be computed for *df*, which has to be approximated using Satterthwaite's formula (Satterthwaite 1946, Cochran 1977, p. 96). The effective degrees of freedom always lie somewhere between the *df* of the component with the smallest *df* and the sum of the *df*s of all components. Note that the tails of the *t*-distribution narrow as the numbers of samples, sites, and years sampled increase to approximate the normal distribution. If the degrees of freedom for $V[\bar{y}]$, as computed using Satterthwaite's approximation, exceed 30, the percentiles of the *t*-distribution can be approximated using the standard normal deviates, i.e. $z_{\alpha/2}$ and $z_{1-\alpha/2}$.

For example, the precision based on the 95% confidence interval for the mean ($\bar{y}$) resulting from an orthogonal design with a realistic number of measurements and $s_Y^2 = s_S^2 = s_{Y*S}^2 = s_e^2 = 1$ is shown in Figure 5.2. The results illustrate: 1) a generally

improved precision (i.e. smaller intervals) if several sites are sampled, 2) improved precision if several years are sampled, and 3) relative insensitivity to the number of samples, provided that several sites and/or years are sampled.

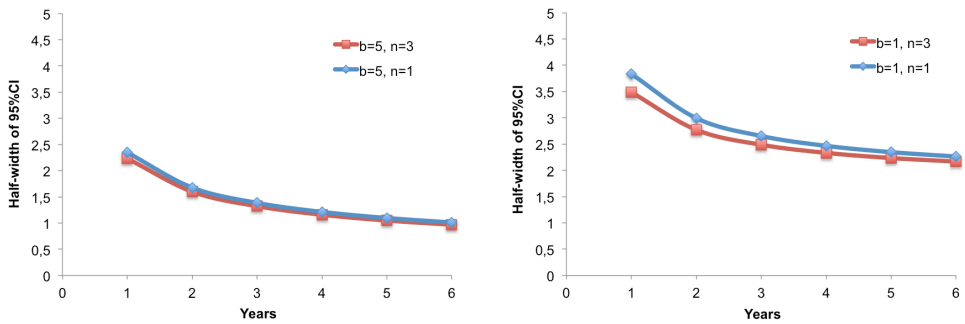

FIGURE 5.2
Half-width of 95% confidence intervals for varying sampling years and sites using an orthogonal design. Number of samples, *n* = 3 or 1. $s_Y^2 = s_S^2 = s_{Y*S}^2 = s_e^2 = 1$. Intervals are calculated using the standard normal distribution (see text).

One crucial aspect in this context is that, to realistically calculate the precision of an estimate used to classify a water body during a six-year period, it is necessary to account for all sources of variability, even if some of the variance components cannot be estimated from the monitoring data. In such cases, information on the magnitude of the variance components from other similar water bodies may provide the best estimates for calculating the indicator uncertainty, and it is assumed that the dataset used for estimating the variance components is sufficiently large that confidence intervals can be estimated using the normal distribution rather than the *t*-distribution. Using the example from Figure 5.2, if only one site is sampled (and the spatial variation cannot be estimated directly, but s_S^2 and s_{Y*S}^2 are assumed = 1) with *n* = 3 replications in all six years, appropriate estimates of variability and 95% confidence intervals at the scales of water bodies and assessment periods is calculated as (note that the component due to years disappears because *a* = 6):

$$V[\bar{y}] = \frac{1 * (1 - {}^{6}/_{6})}{6} + \frac{1}{1} + \frac{1}{6 * 1} + \frac{1}{6 * 1 * 3} = 1.22$$

$$CI95\% = \sqrt{1.22} * z_{0.975} = \sqrt{1.22} * 1.96 = 2.2$$

Ignoring the spatial and temporal variance components (i.e. setting s_Y^2, s_S^2, and s_{Y*S}^2 to zero) and assuming that the uncertainty was due only to small-scale methodological errors would have resulted in a substantially smaller, although unrealistic estimate of the uncertainty: $V[\bar{y}] = \frac{1}{6*1*3} = 0.06$ and a precision of 0.48. Thus, the precision calculated from estimates of small-scale variability would have been almost a five-fold underestimate of the actual uncertainty at the water body scale. Clearly, and as will be demonstrated later,

ignoring important variance components will have severe consequences for the confidence of status classifications.

This example illustrates the risks associated with monitoring programmes that do not take account of key sources of variability. Just because they are impossible to estimate from the monitoring programme, it does not follow that they can be ignored in assessments of uncertainty. Furthermore, the example illustrates the need for quantitative information on the size of the different components of variability. The construction of a "library" of such estimates would enable the reliable assessment of uncertainty even in situations in which sampling designs are incomplete. More importantly, it would form the basis for reducing uncertainty by modifying sampling programmes and finding relevant covariates, reducing the uncertainty associated with the most important sources of variability.

5.1.2 Nested design

Another fundamental type of design that is potentially useful but not commonly used in aquatic environments in Sweden is a design in which new sites ("stations") are sampled each year (Figure 5.3). As in the previous example, the sites may have been selected completely at random within the water body or using criteria such as a narrow depth range, substrate, or distance from shore (see Annex B). The important thing is that the sites are selected to "represent", to some degree, the water body or a defined stratum thereof. Note also that the number of sites (*b*) and samples (*n*) may vary greatly among monitoring programmes.

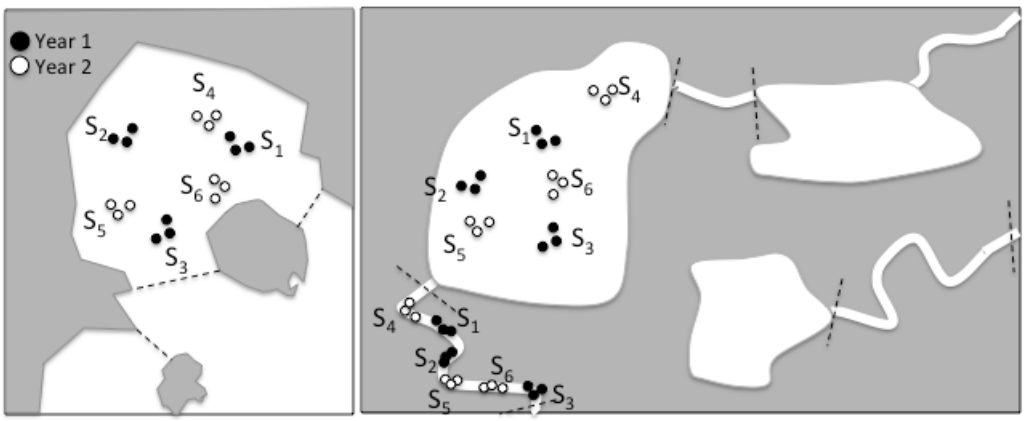

FIGURE 5.3
Schematics of monitoring designs in a coastal water body (left) and in a lake and stream (right). New sites are selected each year, so sites are nested within years. In the examples, a = 2, b = 3, and n = 3.

Each measurement can be expressed using a linear model in which the measured value, y, is the sum of the overall mean, μ, and deviations due to the other sources of variability.

$$y = \mu + YEAR + SITES(YEAR) + PATCHINESS$$

The variability of the overall mean for such a sampling design consists of three variance components, i.e., s_Y^2, $s_{S(Y)}^2$, and s_e^2, each associated with a different source of variability in the linear model. The total variance resulting from these components in an orthogonal design can be calculated as:

$$V[\bar{y}] = \frac{s_Y^2 * (1 - \frac{a}{Y})}{a} + \frac{s_{S(Y)}^2}{ab} + \frac{s_e^2}{abn}$$

This formula summarises how the components are combined into a total variance and, importantly, how the numbers of samples, sites, and years affect the variance and uncertainty. Again, we can see that increasing the number of samples per site and year reduces the uncertainty due to small-scale variability, but does not affect the uncertainty caused by variability among years or sites. As in the previous example, sampling over a number of years will reduce the uncertainty due to years; if all years are sampled, the factor is considered completely fixed and the component due to years is removed. One important difference from the orthogonal design is that, because sites are nested within years, the number of both sites and of years will contribute to reducing the uncertainty due to sites. This may substantially reduce uncertainty if the variability due to sites is dominant in relation to interactive variability, i.e. if there are consistent rather than transient differences among sites (i.e. $s_S^2 > s_{Y*S}^2$, see paragraph below). Another important difference is that the variance component, $s_{S(Y)}^2$, captures both the spatial variation across sites at any given time (s_Y^2) and the difference in this spatial variation across years (s_{Y*S}^2) Thus, given the design, it is impossible to partition $s_{S(Y)}^2$ further into the two other components.

The precision of the overall mean resulting from a nested design with a realistic number of measurements and $s_Y^2 = s_e^2 = 1, s_{S(Y)}^2 = 2$ (the latter in order to preserve the total variability in relation to the previous example) is shown in Figure 5.4. The results are qualitatively similar to those of the orthogonal design for the same combinations of years, sites, and samples. Closer examination of the numbers reveals that the nested design results in increasingly narrower confidence intervals as the number of sampled years increases. For example, with six years of sampling at five sites and three samples (i.e. a = 6, b = 5, and n = 3), the orthogonal and nested designs have precisions of 0.97 and 0.55, respectively.

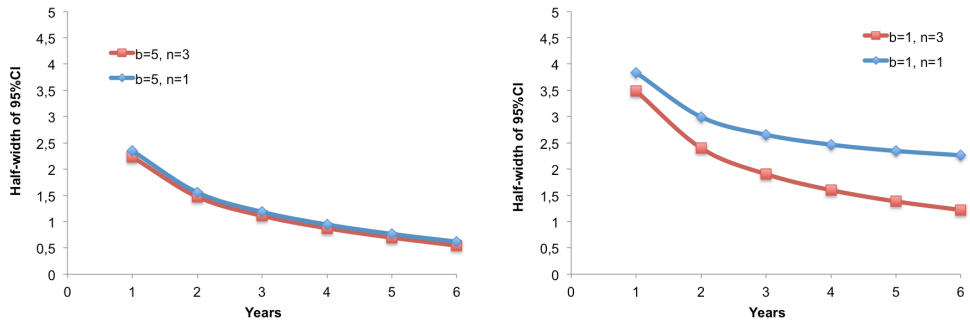

FIGURE 5.4
Half-width of 95% confidence intervals for varying sampling years and sites using a nested design. Number of samples, *n* = 3 or 1. $s_Y^2 = s_e^2 = 1, s_{S(Y)}^2 = 2$. Intervals are calculated using the standard normal distribution (see text).

One interesting difference between the orthogonal and nested designs is that the difference in precision depends on the nature of the spatio-temporal variability (Figure 5.5). If we assume constant total spatio-temporal components, i.e. the sum of the static spatial variability and the interactive variability is constant, the nested design gives identical confidence intervals irrespective of whether the static or interactive components dominate. The orthogonal design, however, has increasingly larger confidence intervals the more dominant the spatial variance components become. Furthermore, this effect is increasingly important with an increasing number of years of sampling. Technically, this effect can be understood by studying the error propagation formulae, in which the uncertainty contribution to $V[\bar{y}]$ associated with the spatial component ($s_{S(Y)}^2$) is reduced by a larger number of sites and of monitoring years in the nested design (ab), while in the orthogonal design, the uncertainty contribution associated with spatial variability is reduced by a larger number of sites only. A less theoretical explanation is that any error introduced in the original selection of sites in the orthogonal design is maintained throughout the assessment period. This error is particularly large if spatial patterns remain relatively unaltered across years, i.e. a static spatial pattern dominates, $s_{S(Y)}^2$. Under such circumstances, the independent selection of new sites, as in the nested design, gives a more representative sample of the water body and thus a smaller variance and confidence interval for the mean. However, it can also be added that the difference in performance between orthogonal and nested designs is likely to decrease as the number of sites increases. Again, this effect can be understood technically but also intuitively, as the risk of substantial error decreases with the increasing number of sites.

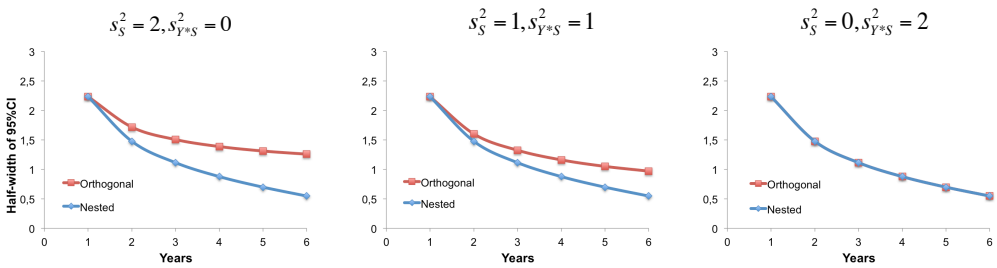

FIGURE 5.5
Half-width of 95% confidence intervals for varying years for orthogonal and nested designs at varying degrees of interactive variability. Left: static spatial variability>>interactive variability. Middle: interactive component = spatial component. Right: interactive variability>>spatial variability. n = 3, b = 5. Intervals are calculated using the standard normal distribution (see text).

In summary, these analyses indicate that not only the total number of samples, but also the allocation of samples among spatial and temporal units and the nature of the spatio-temporal variability affect the precision of mean estimates. This in turn may have a large impact on the confidence that can be placed in status classifications.

5.2 Confidence of status classification

Apart from precision, the second central aspect of uncertainty in the WFD is the confidence of a status classification. As explained earlier, the Directive and its guidance documents recommend that the confidence of a classification be reported, i.e. the probability that the particular classification is correct. Of particular importance is the confidence about classifications better or worse than the "good"–"moderate" (G–M) boundary.

The confidence of a classification depends on the estimated mean status, $\bar{X}$, the location of class boundaries, L_i, and the standard error of the mean, *SEM*. As *SEM* is a central component in estimates of precision, this means that there is a strong link between the two aspects of uncertainty and that any limitations set by a monitoring or sampling design will automatically affect confidence.

As in the case of precision, there are methods for calculating confidence in classification based on traditional statistical principles. According to these, confidence should be calculated using the *t*-distribution at small sample sizes (i.e. $n < 30$). However, it can easily be demonstrated that confidence can be approximately estimated using a simpler method based on the normal distribution even at small sample sizes (Figure 5.6). For example, at $n = 6$ and thus df = 5, the difference in confidence for any single class is <0.02 and ≈0.02 for the classification "better than moderate". These small deviations decrease further at larger sample sizes. Thus, it appears safe to recommend the use of the normal distribution as an approximation of confidence at sample sizes >5. Formulae for obtaining estimates

of confidence using the normal distribution were described by Ellis and Adrianssens (2006) and later reproduced in the Swedish handbook (2007:4) (see Annex A).

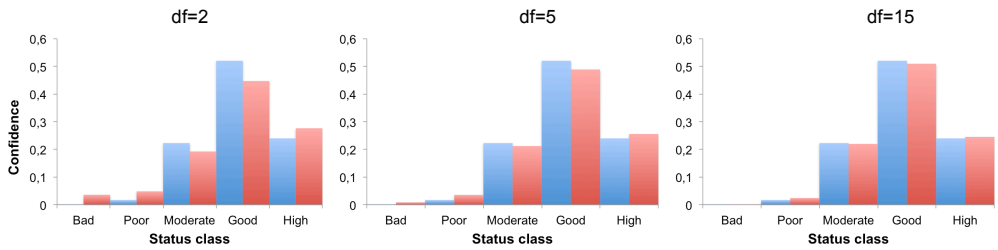

FIGURE 5.6
Confidence in classifications using the normal distribution (blue) and the *t*-distribution (red) at df = 2, 5, and 15. Estimated mean $\bar{X}$= 6 and L_i= 2, 4, 6, and 8. Differences between methods diminish with an increasing number of samples (df).

An alternative and potentially more convenient way to estimate confidence in classification is to use purpose-built software. At least two generations of user-friendly packages based on Monte Carlo simulation have been developed for the WFD context: STARBUGS and WISERBUGS ("BUGS" = bioassessment uncertainty guidance software; Clarke & Hering 2006, Clarke 2011). Both of these were developed within EU-funded projects on WFD implementation.

WISERBUGS was developed in the WISER project (Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery, http://www.wiser.eu/results/software/). The main aim of WISERBUGS is to provide a general tool for simulating and assessing uncertainty in estimates of single metrics, multi-metric indices, and multi-metric rules for combining WFD quality elements. As such, it can also be used to assess the effects of new indices, class limits, and combination rules. As pointed out in the user manual (Clarke 2011), however, it is fundamental to realize that the estimates of precision and confidence obtained using WISERBUGS are completely dependent on representative estimates of variability for the appropriate spatial and temporal context.[1]

> "The error assessment software must, of necessity, be based on the best available estimates of the various sources of variation and errors in observed metric values and EQRs … Sources of variation for which no estimates are currently available are ignored in the error assessment program (and effectively treated as zero). In

[1] Note that Clarke (2011) consistently uses SD ("sampling standard deviation") to denote the standard error of the mean (SE).

> such cases, the software system will over-estimate the precision and under-estimate the true uncertainty in the assessment of status classes." (p. 9)

Thus, the WISERBUGS (and STARBUGS) tool is very useful for assessing confidence in status classification, but the quality of its assessment is completely dependent on the quality of the estimated variability. As discussed at length in section 5.1, accounting for all relevant sources of variability affecting estimates of mean ecological status based on various types of monitoring programmes, at the scale of individual water bodies and for whole assessment periods, is a great and crucial challenge.

In an earlier example (section 5.1.1), the effects of ignoring important components of variability led to nearly a five-fold underestimation of the uncertainty, i.e. the apparent precision was unrealistically high. Applying the same input data in an example of effects on confidence indicates similar dramatic consequences (Figure 5.7). Appropriate treatment of the components of variability would have resulted in a confidence of 63%, whereas the apparent confidence considering only methodological errors would have been 100%.

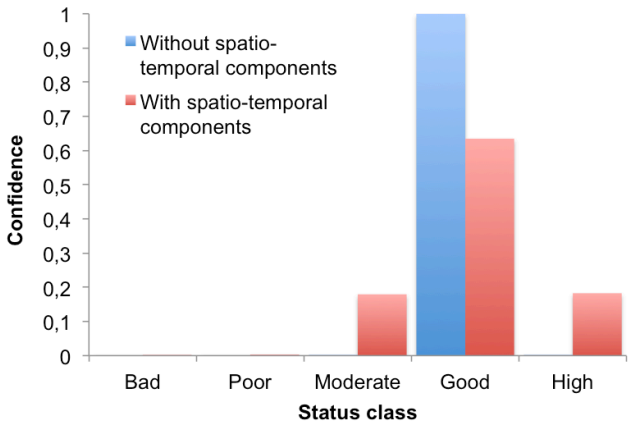

FIGURE 5.7
Confidence in classifications using the normal distribution with ($SE = \sqrt{1.22}$) and without ($SE = \sqrt{0.06}$) important spatial and temporal components of variability (see section 5.1.1 for details). Estimated mean $\bar{X}$= 6 and L_i= 2, 4, 6, and 8. Confidence in classification is grossly overestimated if spatio-temporal components are ignored.

In conclusion, these examples indicate that there are efficient methodological frameworks for estimating confidence in the way that is prescribed by the WFD. However, it is clear that frameworks for handling diverse sources of variability are necessary, not only to assess precision, but also to estimate confidence in classification.

6 Estimating uncertainty for selected quality elements

In this chapter, the estimation of variance components in the uncertainty framework introduced in chapter 3 will be exemplified using data from the Swedish monitoring programme. It should be stressed that the framework cannot be directly applied to a monitoring dataset, because the possible statistical analyses are constrained by the availability and structure of the data. It is therefore necessary to modify the uncertainty framework given the dataset in question. The aim of this chapter is to illustrate how this is achieved with specific datasets.

6.1 Uncertainty analysis of eelgrass shoot density in Öresund

Eelgrass has been monitored since 1995 along the Swedish coast of Skåne County from the island of Hallands Väderö to Falsterbo (Figure 6.1). Sampling was carried out using a frame with an area of 0.0625 m^2, which was randomly placed on top of the sediments near the monitoring station. Within each frame, the shoot density, length, biomass, and sugar content of the leaves were measured; for some stations and years, the biomass of the rhizomes was also measured. For monitoring at each station, either 6 or 12 frames were sampled and the overall coverage of eelgrass at the station was assessed. The monitoring programme included 13 sites, each of which could be represented by one or more stations typically located at different depths. Information on the identity of the diver conducting the sampling was included in the dataset.

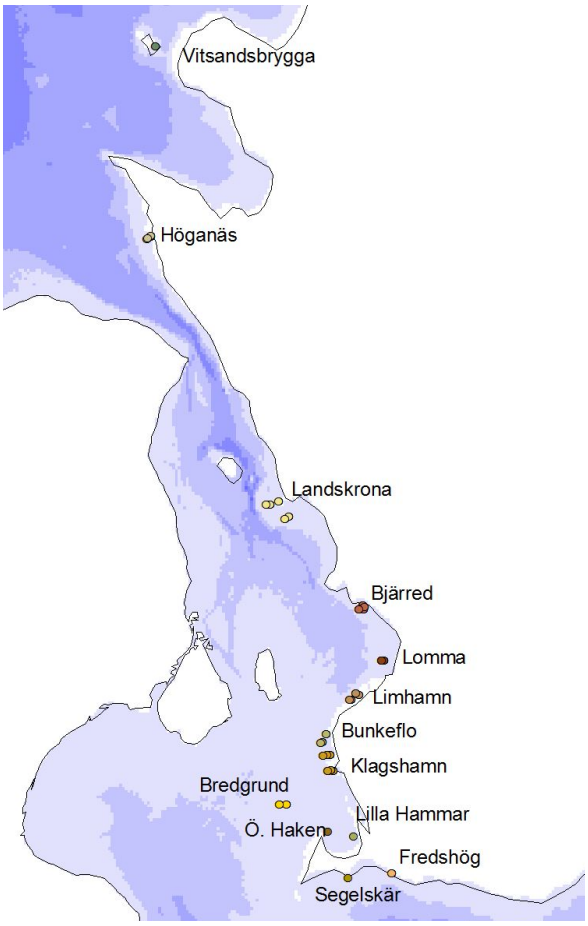

FIGURE 6.1
Location of eelgrass monitoring sites in the Öresund. Some sites had several monitoring stations.

For exemplifying the uncertainty framework, eelgrass shoot density will be used, and for the purpose of estimating the variance components, the data will be log-transformed. The implication of the log-transformation is that the random factors estimate relative uncertainties, i.e. the uncertainty due to a component relative to the mean value. A total of 1919 observations were analysed, but these were very unevenly distributed over the sites (Table 6.1). The eelgrass monitoring was conducted between July and October every year, but, with few exceptions, the stations were sampled only once per year. It is therefore difficult to assess the seasonality of the eelgrass variables, and it is assumed that the sampling months within a year as well as the sampling years represent a relatively stable period without net losses or gains in eelgrass.

TABLE 6.1
Sampling efforts at the 13 sites in the Öresund (1995–2011). Only Landskrona and Klagshamn were sampled every year. The total number of divers is six, and one diver carried out almost half of the sampling.

Area	Site	No. of stations	No. of years	No. of divers	No. of observations
Northern Öresund	Vitsandsbrygga	1	2	2	12
	Höganäs	2	12	2	144
Central Öresund	Landskrona	2	17	4	288
	Bjärred	4	16	3	288
	Lomma	3	13	3	168
	Limhamn	2	4	2	120
Southern Öresund	Bunkeflo	3	11	3	204
	Klagshamn	5	17	4	450
	Bredgrund	3	11	2	167
	Ö. Haken	1	1	1	6
	Lilla Hammar	1	1	1	6
South coast	Segelskär	1	1	1	6
	Fredshög	1	11	3	60

Model 1 – individual sites; all factors random

The first model of shoot density is a simplified version of the general model presented in section 3.5 and includes a mean value for the site and random variation between stations, between years, and between divers as well as random interannual variation between stations. The mean was estimated for each site separately.

$$y = \mu + YEAR + GRADIENT + YEAR \times GRADIENT + PERSON + PATCHINESS$$

Note that *GRADIENT* represents the random variation between stations and *PATCHINESS* represents the random variation between replicates, i.e. small-scale spatial variation. For most sites, the small-scale spatial variability (*V[PATCHINESS]*) was the largest (Table 6.2), with a relative variation ($\frac{\sqrt{V[PATCHINESS]}}{\bar{X}}$) of 10–100%. The large-scale variation between stations could be just as large, whereas the interannual variation and the variation between divers were generally smaller (<39% and <56%, respectively). The smallest random component for all sites was the large-scale variation across years (*V[GRADIENT × YEAR]*), suggesting that the spatial variation between stations changed up to 25% between years. It should be stressed that many of the variance component estimates were uncertain due to the small amount of data available for their estimation (Table 6.1).

TABLE 6.2
Variance components estimated using REML from the mixed model above. Not all variance components could be estimated for each site, as indicated by "-", due to data limitations. The standard deviation of the components is calculated as the square root of the variance.

Area	Site	V[GRADIENT]	V[YEAR]	V[PERSON]	V[G × Y]	V[PATCHINESS]
Northern	Vitsandsbrygga	-	0.1102	-	-	0.0511
Öresund	Höganäs	0.2119	0.0164	-	0.0027	0.1761
Central	Landskrona	0.0262	0.0082	0.0697	0.0082	0.2275
Öresund	Bjärred	0.0520	0.0292	0.0718	0.0111	0.1370
	Lomma	0.0777	0.0067	0.0123	-	0.1432
	Limhamn	-	0.0177	0.0167	0.0121	0.4577
Southern	Bunkeflo	0.5013	-	0.0987	0.0508	0.1132
Öresund	Klagshamn	0.1142	0.0368	0.0137	-	0.3550
	Bredgrund	0.0734	0.0388	0.1986	0.0117	0.0391
	Ö. Haken	-	-	-	-	0.0646
	Lilla Hammar	-	-	-	-	0.0089
South	Segelskär	-	-	-	-	0.0568
coast	Fredshög	-	0.0648	-	0.0012	0.0917

Model 2 – all sites combined; sites and years fixed factors

In the second model, the variance components were estimated for the dataset as a whole, and not for individual sites separately. Given that most of the variance component estimates were of similar magnitude and that those that deviated substantially were typically from sites with poorer datasets, a general model was applied to the entire dataset that included a site-specific mean for the shoot density.

$$y = \mu(site) + year + YEAR{\times}SITE + GRADIENT(SITE) + YEAR{\times}GRADIENT(SITE) + PERSON + PATCHINESS$$

Another extension to this model is that it is assumed that there is a common fixed trend (*year*) to all sites, whereas the random variation in trends among sites is described by the random factor ($YEAR{\times}SITE$). Thus, it is assumed that there are some common, although unknown, mechanisms governing the eelgrass shoot density that act on all sites in the same manner. However, there can be random interannual fluctuations at each site around this common trend. Note that inclusion of the factor *SITE* introduces a hierarchical structure, in which stations are nested within sites. In fact, *PATCHINESS* is also nested within stations, but since it represents the residual variation, the nesting is not indicated.

There was significant variation between sites ($F_{12,26} = 2.59$; $p = 0.0207$) and between years ($F_{16,36} = 3.19$; $p = 0.0019$), indicating a fairly variable north–south gradient and similarly for the trend (Figure 6.2). The estimated variance components were 0.0098 for $YEAR{\times}SITE$ (corresponding to 10% uncertainty), 0.1273 for $GRADIENT$ (~43%), 0.0133 for $YEAR{\times}GRADIENT$ (~12%), 0.0599 for $PERSON$ (~28%), and 0.2151 for

$PATCHINESS$ (~59%). These results of the more general model confirm that small-scale spatial variation is the largest source of uncertainty, followed by large-scale spatial variation and then diver-specific variation. Temporal random variation was low relative to the other components.

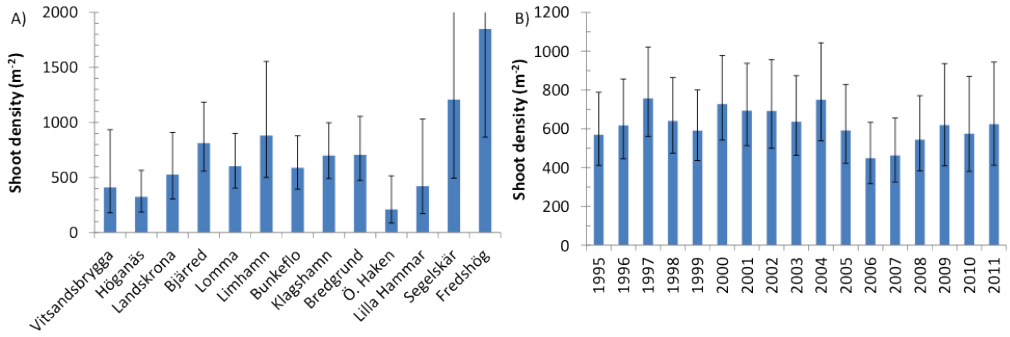

FIGURE 6.2
Mean shoot densities obtained from model 2 (after back-transformation) for the 13 sites (from north to south) and 17 years. Error bars show the 95% confidence interval of the mean shoot densities.

Model 3 – all sites combined; sites, years, and depth fixed factors

A third model examines whether the variation between stations can be explained by including depth as an explanatory variable, acknowledging that part of the variation between stations could be due to differences in depths, which vary from 1.4 to 5.6 m across stations. The depth dependency was modelled as a linear function for the log-transformed shoot densities, corresponding to an exponential decline on the original scale.

$$y = \mu(site) + year + YEAR{\times}SITE + depth + GRADIENT(SITE) + YEAR{\times}GRADIENT(SITE) + PERSON + PATCHINESS$$

The fixed effects of the model became more significant as a result of explaining more of the random variation. There was a significant variation between sites ($F_{12,26} = 11.55$; $p < 0.0001$) and between years ($F_{16,36} = 4.87$; $p < 0.0001$), as well as a decrease with depth ($F_{1,1745} = 2347.74$; $p < 0.0001$), suggesting that shoot density would decline by 27% for each additional meter in depth. The precision of the site-specific and annual means improved and the spatial pattern, in particular, was altered by including the depth dependency (Figure 6.3). Mean shoot density displayed an increasing spatial trend from north to south, except that Ö. Haken and Lilla Hammar, having just one station and one year of data (Table 6.1), deviated strongly from this trend.

The estimated variance components were 0.0063 for $YEAR{\times}SITE$ (corresponding to 8% uncertainty), 0.0181 for $GRADIENT$ (~14%), 0.0222 for $YEAR{\times}GRADIENT$ (~16%), 0.0768 for $PERSON$ (~32%), and 0.0944 for $PATCHINESS$ (~36%). Although the small-scale spatial variation is still large, the inclusion of depth as an explanatory variable

substantially reduced the large-scale spatial variation. The other components were moderately changed. Actually, the estimated variance component, $YEAR \times SITE$, was not significant (Z = 1.06; p = 0.1438), whereas $YEAR \times GRADIENT$ was (Z = 3.35; p = 0.0004). This suggests that the sites follow the same overall pattern (Figure 6.3), but that there can be large differences in the trends among stations within a site. Thus, random temporal variations at the small scale are very important.

Analysis of the residuals of model 3 with respect to depth suggested that the linear relationship was adequate for describing the declining shoot density with depth. Replicate number was also included in an additional analysis to test whether there was a potential systematic effect of either increasing or decreasing shoot density with replicates, but this additional regression variable was not significant ($F_{1,1744}$ = 3.09; p = 0.0791) when added to the model.

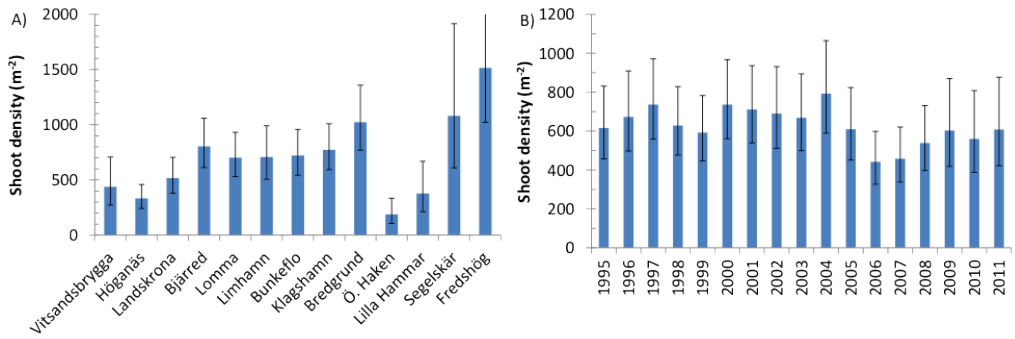

FIGURE 6.3
Mean shoot densities obtained from Model 3 (after back-transformation) for the 13 sites (from north to south) and 17 years. Differences in observation depths between sites and years were accounted for. Error bars show the 95% confidence interval of the mean shoot densities.

6.2 Uncertainty analysis of benthic fauna in the Skagerrak and the Gulf of Bothnia

The benthic fauna in soft sediments has been regularly and increasingly monitored in the Swedish coastal zone since the 1970s. The current programme is partly coordinated between national and regional authorities. The structure of monitoring programmes differs slightly among regions within Sweden, but a total of 500 grab samples (0.1 m^2) are collected around the coast in the spring each year.

To assess the ecological status of benthic fauna according to the WFD, Swedish authorities and scientists have developed a benthic quality index (BQI; SEPA 2010, Rosenberg et al. 2004, Leonardsson et al. 2009) based on: 1) tolerance and sensitivity to eutrophication and increased organic load, 2) species richness, and 3) abundance. The

index is adapted to specific conditions in the Baltic Sea and the Kattegat/Skagerrak and assessment involves type-specific class boundaries (see also Annex B).

Data from the Skagerrak and from the "coastal clusters" of the Gulf of Bothnia were used to illustrate the uncertainty framework. In both regions, data from three years were used and no information on person dependency was available. The samples from the Skagerrak were clustered in three areas, each represented by eight stations with two replicate samples per station each year. The samples from the Gulf of Bothnia were clustered in five areas, each represented by approximately 20 stations with one replicate sample per station each year. In both programmes, stations were revisited each year (note that this is only a selection of the available data).

TABLE 6.3
Sampling efforts in the eight areas used for evaluating sources of uncertainty in the BQI of benthic fauna in the Skagerrak (2002–2004) and the Gulf of Bothnia (1996–1998).

Region	Area	# stations	# years	# samples	# observations
Skagerrak	Area 1	8	3	2	48
	Area 2	8	3	2	48
	Area 3	8	3	2	48
Gulf of Bothnia	NAT 1	22	3	1	66
	REG 2	20	3	1	60
	REG 4	20	3	1	60
	REG 6	20	3	1	60
	REG 8	20	3	1	60

The aim of these examples was to estimate different sources of variability and to illustrate how these affect the uncertainty of an assessment of ecological status within a WFD assessment period (i.e. six years). Furthermore, the analysis provided opportunities to qualitatively compare differences in relative importance among and within the Skagerrak and the Gulf of Bothnia. This was done using modelling within individual areas and for whole regions. No attempts were made to reduce uncertainty by incorporating fixed covariates, as was done for eelgrass shoot density in section 6.1. Nevertheless, initial tests indicate that factors such as depth and substrate characteristics can substantially reduce variability among stations (*GRADIENTS*).

As a general observation, it is evident that the average values of the BQI were higher in the Skagerrak than in the Gulf of Bothnia (Figure 6.4). Furthermore, the overall average standard error of the mean within areas and years was ~0.65 and ~0.45 for the Gulf of Bothnia and Skagerrak, respectively. Thus, for both spatial and temporal variation, though without a deeper quantitative analysis of the sources of variability, the uncertainty generally appeared larger in the Gulf of Bothnia, in terms of absolute deviations and even more so relative to the means.

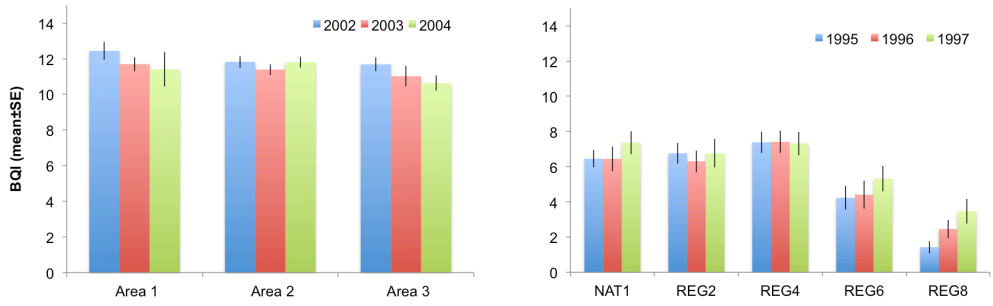

FIGURE 6.4
Mean BQI for areas in the Skagerrak (left) and the Gulf of Bothnia (right). Error bars show the standard error of the mean estimate.

Model 1 – individual areas; all factors random

Despite similarities among sampling programmes, the models for assessing uncertainties within the Skagerrak and the Gulf of Bothnia differed slightly. In the Skagerrak, the following components can be estimated:

$$y = \mu + YEAR + GRADIENT + YEAR \times GRADIENT + PATCHINESS$$

As mentioned earlier, this model contains only random factors, where *GRADIENT* represents unpredictable variability among stations and *PATCHINESS* similarly represents unpredictable variability among samples within a station. In the Gulf of Bothnia, however, only one sample is taken at each station each year. This means that the variability due to interactive variability between *YEAR* and *GRADIENT* cannot be separated from that due to *PATCHINESS*. Therefore, the model is reduced and any existing variability due to *PATCHINESS* among samples will be included in the component *YEAR* × *GRADIENT*:

$$y = \mu + YEAR + GRADIENT + YEAR \times GRADIENT$$

Analyses of separate areas in the Skagerrak and the Gulf of Bothnia revealed a number of interesting patterns (Table 6.4). First, in all areas the variability due to YEAR was the least important component. This does not necessarily imply that there were no changes among years during the two periods, as indicated by the occasional large variation of *YEAR* × *GRADIENT*; there may well be changes within individual stations that were not consistent across all stations. Second, the most important source of variability in all areas (except for REG 2) was the random component GRADIENT. In the Skagerrak, where all components could be separated, *V[GRADIENT]* was 2–8 times larger than *V[PATCHINESS]*. In the Gulf of Bothnia, where such partitioning was impossible, *V[GRADIENT]* was on average twice as large as *V[YEAR* × *GRADIENT]*. Finally, it is also evident that there were some differences in the estimated components among areas within regions. These may reflect true differences, but it is also likely that substantial parts were associated with sampling errors due to the small number of stations and years. One

exception was the estimates of *V[PATCHINESS]* in the Skagerrak, which all fall within a fairly narrow range. Nevertheless, despite these shortcomings, the overall patterns discussed above appear qualitatively robust.

The overall uncertainty of mean estimates in individual areas, using the current monitoring design, can be assessed using the formulae presented in section 5.1. In absolute numbers, the area-specific precision of BQI obtained with these programmes was similar in the two regions (Table 6.4). In the Skagerrak, the total variance was 0.15–0.68; using the normal distribution to calculate confidence intervals, the resulting precision was 0.75–1.6. In the Gulf of Bothnia, the total variance was 0.26–0.45 and the confidence intervals were 1–1.3. In relation to the average BQIs, most deviations in individual areas were generally in the range of 5–10% of the mean. In the northernmost parts of the Gulf of Bothnia (REG 6 & 8), however, relative uncertainties were slightly larger (15–25%), probably as a result of larger variability and a lower mean BQI (Figure 6.4).

TABLE 6.4
Variance components estimated using REML from the mixed models for BQI in the Skagerrak and the Gulf of Bothnia. Not all variance components could be estimated for each area, as indicated by "-", due to data limitations. The standard deviation of the components is calculated as the square root of the variance. Total variability was calculated correcting for sampling three years out of six in an assessment cycle.

REGION	AREA	V[YEAR]	V[GRADIENT]	V[YEARx GRADIENT]	V[PATCHI-NESS]	V[$\bar{y}$]	SD/ Mean
Skagerrak	AREA 1	0.000	4.445	2.682	0.541	0.68	0.07
	AREA 2	0.017	1.064	0.112	0.459	0.15	0.03
	AREA 3	0.178	2.549	0.619	0.428	0.38	0.06
Gulf of Bothnia	NAT 1	0.101	4.580	3.960	-	0.29	0.08
	REG 2	0.000	3.881	5.224	-	0.26	0.08
	REG 4	0.000	6.244	1.619	-	0.31	0.08
	REG 6	0.231	8.242	2.608	-	0.45	0.15
	REG 8	0.903	3.126	2.775	-	0.33	0.24

Model 2 – all areas within regions combined; all factors random

In the second set of analyses, the variance components were estimated for entire datasets from each of the two regions. Given that most of the variance component estimates were of similar magnitude in the Skagerrak, a general model (model 2) was applied to the entire region. For the Gulf of Bothnia, however, there were large differences, particularly between the datasets in the northern areas of the Bothnian Bay (i.e. REG6 and REG8) and those in the Bothnian Sea (i.e. NAT1, REG2, and REG4). Therefore, these two basins of the Gulf of Bothnia were treated separately in model 2:

$$y = \mu + AREA + YEAR + YEAR \times AREA + GRADIENT(AREA) + YEAR \times GRADIENT(AREA) + PATCHINESS$$

Note that, unlike in model 2 in section 6.1, *AREA* is considered a random source of variation, which means that the estimated total uncertainties represent the overall uncertainties of assessments in coastal areas of the Skagerrak, the Bothnian Sea, and the Bothnian Bay (Table 6.5). The analyses indicate that the most important source of variability in all areas is that due to *GRADIENT(AREA)*, the values of which ranged from 2.5 to 5.7. In relative terms, the error due to *GRADIENT* was ~20% in the Skagerrak, but in the two basins of the Gulf of Bothnia it was 70% and 160% of the mean. Another common pattern is that the interaction *YEAR×AREA* was not important in any of the areas. This means that interannual variation tends to be similar across areas within regions. However, the contribution from *YEAR×GRADIENT(AREA)* is generally also substantial, indicating that the small-scale spatio-temporal variation within areas (note, however, that this variability cannot be separated from that due to PATCHINESS in the Gulf of Bothnia) is more important than the large-scale spatio-temporal variation. Little variability is attributed to *YEAR* and *AREA* in the Skagerrak and the Bothnian Sea, but especially that due to *YEAR* is substantial in the Bothnian Bay.

TABLE 6.5
Variance components estimated using REML from the mixed models for BQI in the Skagerrak, the Bothnian Sea (NAT1, REG2, and REG4), and the Bothnian Bay (REG6 and REG8) using model 2. SD/$\bar{X}$ captures the uncertainty in relation to the means in the three regions (11.5, 6.9, and 3.6 for the Skagerrak, the Bothnian Sea, and Bothnian Bay, respectively).

	Skagerrak		**Bothnian Sea**		**Bothnian Bay**	
	V	**SD/$\bar{X}$**	**V**	**SD/$\bar{X}$**	**V**	**SD/$\bar{X}$**
V[YEAR]	0.090	0.008	0.000	0.000	2.059	0.581
V[AREA]	0.000	0.000	0.000	0.000	0.574	0.162
V[YEAR×AREA]	0.000	0.000	0.000	0.000	0.000	0.000
V[GRADIENT(AREA)]	2.524	0.219	4.788	0.694	5.699	1.609
V[YEAR×GRADIENT(AREA)]	1.113	0.096	3.649	0.529	2.684	0.758
Y[PATCHINESS]	0.476	0.041	-	-	-	-
V[$\bar{y}$]	0.139	0.032	0.100	0.014	0.795	0.224

Although differences in sampling designs make certain comparisons difficult, we can conclude from the analyses using model 2 that there are similarities in the relative importance of the temporal and spatial components of variability. In general, it appears that BQI is more variable in the Gulf of Bothnia and particularly so if sources of uncertainty are considered relative to the mean. Nevertheless, the overall uncertainty, incorporating several levels of spatial and temporal replication, suggest that overall means are estimated with a precision of a ~1–3% error in the Skagerrak and the Bothnian Sea. In the Bothnian Bay, however, mean estimates for the period analysed were considerably more uncertain (~22%).

6.3 Lessons from uncertainty analyses

These examples have demonstrated how the uncertainty framework can be applied to shed light on various sources of uncertainty in monitoring data in general and in relation to the WFD in particular. The framework was used successfully with monitoring datasets for vegetation and benthic fauna in coastal areas, and there is no reason to believe that it could not be applied to other BQEs in pelagic and inland environments. Important conclusions from these analyses are as follows:

- **Estimates of mean status of a BQE will likely be associated with a large number of sources of uncertainty.** These sources of uncertainty stem from unpredictable (random) or partly predictable (fixed) spatial and temporal processes and from processes associated with sampling and analyses. To properly assess the uncertainty of estimates and classifications, a coherent framework is necessary.
- **Estimating variance components with reasonable precision requires a large dataset with a structure that allows these components to be identified.** Estimates from single years and sites are often very uncertain. To obtain reliable estimates of uncertainty, it is often advisable to conduct comprehensive analyses of data from several years, sites, and areas, rather than only using the actual data from a particular water body and year.
- **The uncertainty associated with the various variance components can be reduced by including fixed, explanatory factors in the model.** For example, the variance components for shoot density associated with spatial variation (i.e. between stations and samples) were reduced by more than 50% by including depth as an explanatory variable. However, small-scale spatial variation was still the largest source of uncertainty, stressing the importance of using many samples when assessing the mean shoot density.
- **The size and relative importance of different sources of uncertainty may differ greatly among areas and regions, for the same BQE.** For example, some components of variability (and their size relative to the mean) were generally larger in BQI estimates for the Gulf of Bothnia than for the Skagerrak.
- **Proper use of replication at different spatial and temporal scales allows for the precise estimation and classification of ecological status**, despite the existence of relatively large sources of variability.

7 Conclusions: implications for future work

The aims of this report were to review important concepts related to uncertainty as defined in the WFD and its guidance documents, to analyse how these concepts are implemented in the Swedish assessment criteria, and to develop a comprehensive framework for how uncertainty can be assessed systematically.

In relation to these aims, we conclude that the Directive provides appropriate conceptual definitions of uncertainty, precision, and confidence in classification, but that guidelines for quantitative targets for acceptable levels of uncertainty are not provided and issues concerning precautions and burden of proof are not given. These may be important reasons for the lack of coherence and occasional absence of appropriate treatment of uncertainty observed in current Swedish assessment criteria. A review of methods for describing, quantifying, and incorporating uncertainty in existing routines revealed substantial differences among BQEs. Some of these differences may be caused by specific properties of each BQE, but from the perspectives of transparency and reliability, such differences appear unfortunate. Therefore, it is suggested that a common framework could offer considerable advantages.

A general uncertainty framework was developed that partitions variation into four categories: 1) temporal variation, 2) spatial variation, 3) spatio-temporal interaction, and 4) methodological variation. Several of these sources of variation contain both fixed and random components, where fixed components describe variations that are predictable, such as patterns of variation that are repeatable or can be modelled by means of explanatory variables. Potential sources of variation for these four categories are discussed with reference to the different BQEs and to their relative importance. An important aspect of the uncertainty framework is that sources of variation that may contribute great uncertainty can be disregarded by appropriately limiting their ranges. For example, an indicator can be chosen to reflect a certain seasonal window only, in which the variation is believed to be small.

The framework was exemplified using monitoring data on vegetation and benthic fauna, quantifying those sources that could be estimated given the datasets. However, since the outcomes of analyses using this general framework will differ greatly among BQEs, there is a need in WATERS to populate this framework with additional studies using other datasets and indicators. Thus, based on these realisations, we propose the construction of a library or catalogue of sources of uncertainty for Swedish indicators. An expected future outcome of WATERS will thus be a catalogue of uncertainty analyses within the general

framework that allows for the assessment of the uncertainty of various indicators to be used for ecological status assessment.

The examples used to demonstrate the uncertainty framework also highlighted the importance of estimating uncertainty components from larger datasets. Variance components estimated from small datasets will often be associated with substantial uncertainty. More consistent estimates are obtained when the uncertainty components are estimated from a larger pooled dataset (e.g. from several years and water bodies), assuming the uncertainty components to be of similar magnitude across a range of similar water bodies. Furthermore, a larger dataset also allows for the formulation of generic models to account for parts of the variation, thus reducing the random variation. Finally, a larger dataset delivers more degrees of freedom for the calculation of variation components, which reduces the confidence intervals of an indicator (but note that this still means that the standard error of the mean and thus the uncertainty is calculated using the actual number of samples, times, etc., from the unit of interest). Thus, in line with the idea of a catalogue of uncertainty components, it is recommended that, for the estimation of variance components, data be combined into larger datasets, albeit for similar ecosystems only. The implications and basis for this recommendation will be further evaluated in the project.

Finally, the conceptual framework for uncertainty, in combination with existing and future estimates of different variance components, has important consequences for the design and optimisation of monitoring programmes. For example, strong and consistent patterns of random spatial and temporal components of variability were identified in the benthic fauna. The relative and absolute size of these components, in particular, the importance of static and interactive spatial patterns, has strong implications for the design of monitoring programmes. Some of these results indicate that there may be important trade-offs between nested and crossed designs, but the generality of these results needs to be further analysed and understood. Therefore, analyses of the practical consequences of estimated patterns of variability for monitoring designs will follow.

In conclusion, we have developed a general framework for assessing the uncertainty of biological indicators and illustrated some of its benefits. We believe that such a framework can substantially improve the transparency and reliability of ecological status assessments according to the WFD in Swedish coastal and inland waters.

8 References

2000/60/EC. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy.

Bennet S., Roca G., Romero J., Alcoverro T. (2011) Ecological status of seagrass ecosystems: an uncertainty analysis of the meadow classification based on the Posidonia oceanica multivariate index (POMI). Marine Pollution Bulletin, 62: 1616–1621.

Beven, K.J., Binley, A.M. (1992) The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes, 6: 279–298.

Birk S., Bonne W., Borja A., Brucet S., Courrat A., Poikane S., Solimini A., van de Bund W., Zampoukas N., Hering D. (2012) Three hundred ways to assess Europe's surface waters: an almost complete overview of biological methods to implement the Water Framework Directive. Ecological Indicators, 18: 31–41

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(3): 127–135

Christian, P.R. and Casella G. (2004) Monte Carlo statistical methods (2nd ed.). New York: Springer.

Clarke R.T. (2011) WISERBUGS (WISER Bioassessment Uncertainty Guidance Software) tool for assessing confidence of WFD ecological status class. Deliverable D6.1-3: User Manual and software: Release 1.2 (November 2011)

Clarke R.T. (2012) Assessing uncertainty of indicators of Water Framework Directive ecological status class. WISER deliverable 6.1-2. http://www.wiser.eu/download/D6.1-2.pdf

Clarke R.T., Davy-Bowker J., Sandin L., Friberg N., Johnson R.K., Bis B. (2006a) Estimates and comparisons of the effects of sampling variation using "national" macroinvertebrate sampling protocols on the precision of metrics used to assess ecological status. Hydrobiologia, 566: 477–503.

Clarke R.T., Furse M.T., Gunn R.J.M., Winder J.M., Wright J.F. (2002) Sampling variation in macroinvertebrate data and implications for river quality indices. Freshwater Biology, 47: 1735–1751.

Clarke R.T., Lorenz A., Sandin L., Schmidt-Kloiber A., Strackbein J., Kneebone N.T., Haase P. (2006b) Effects of sampling and sub-sampling variation using the STAR-AQEM sampling protocol on the precision of macroinvertebrate metrics. Hydrobiologia, 566: 441–460.

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.

Cochran, W.G. (1977) Sampling techniques (3rd ed.). New York: John Wiley & Sons.

Courrat A., Lepage M., Alvarez M.C., Borja A., Cabral H., Elliott M., Gamito R., Neto J.M., Pérez-Domínguez R., Raykov V., Uriarte A. (2012) The ecological status assessment of transitional waters: an uncertainty analysis for the most commonly used fish metrics in Europe. WISER deliverable 4.4-2, part 2. http://www.wiser.eu/download/D4.4-2_part2.pdf

Dudley A., Dunbar M., Penning E., Kolada A., Hellsten S., Kanninen A. (2012) Report on uncertainty in macrophyte Metrics. WISER deliverable 3.2-2. http://www.wiser.eu/download/D3.2-2.pdf

Ellis J., Adriaenssens V. (2006) Uncertainty estimation for monitoring results by the WFD biological classification tools. Bristol, UK: Environment Agency.

EC (2003) WFD CIS Guidance Document No. 7. Guidance on the Intercalibration Monitoring under the Water Framework Directive. Produced by Working Group 2.7: Monitoring. Available from http://circa.europa.eu/

EC (2005) 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.

Green R.H. (1979) Sampling design and statistical methods for environmental biologists. New York: John Wiley & Sons.

Hering D., Borja A., Carstensen J., Carvalho L., Elliott M., Feld C.K., Heiskanen A.S., Johnson R.K., Moe J., Pont D., Solheim A.L., van de Bund W. (2010) The European Water Framework Directive at the age of 10: a critical review of the achievements with recommendations for the future. Science of the Total Environment, 408: 4007–4019.

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

Levin S.A. (1992) The problem of pattern and scale in ecology. Ecology 73(6), 1943–1967.

Mapstone B.D. (1995) Scalable decision rules for environmental impact studies: effect size, type I and type II errors. Ecological Applications, 5(2): 402–410.

Mascaró O., Bennett S., Marbà N., Nikolic V., Romero J., Duarte C.M., Alcoverro T. (2012) Uncertainty analysis along the ecological quality status of water bodies: the response of the Posidonia oceanica multivariate index (POMI) in three Mediterranean regions. Marine Pollution Bulletin, 64(5): 926–931.

Naturvårdsverket (2007) Status, potential och kvalitetskrav för sjöar, vattendrag, kustvatten och vatten i övergångszon. En handbok om hur kvalitetskrav i ytvattenförekomster kan bestämmas och följas upp. Handbok 2007:4. Stockholm: Naturvårdsverket.

NFS 2008:1 (2008) Naturvårdsverkets föreskrifter och allmänna råd om klassificering och miljökvalitetsnormer avseende ytvatten. Naturvårdsverkets författningssamling NFS 2008:1. ISSN 1403–8234.

Neto J.M, Orfanidis S., Dencheva K., Mascaró Vidal O., Muxika I., Borja A. (2012) Report/manuscript on benthic macroflora indicators for transitional waters, including classification boundaries, definition of reference conditions and uncertainty. WISER deliverable 4.2-4. http://www.wiser.eu/download/D4.2-4.pdf

Noges P., van de Bund W., Cardoso A.C., Solimini A.G., and Heiskanen A.S. (2009) Assessment of the ecological status of European surface waters: a work in progress. Hydrobiologia, 633: 197–211.

Quinn G.P., Keough M.J. (2002) Experimental design and data analysis for biologists.. Cambridge, UK: Cambridge University Press.

Satterthwaite F.E. (1946) An approximate distribution of estimates of variance components. Biometrics Bulletin, 2(6): 110–114.

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

Schneider D.C. (2001) The rise of the concept of scale in ecology. BioScience, 51(7): 545–554.

SEPA (2010) Status, potential and quality requirements for lakes, water courses, coastal and transitional waters: A handbook on how quality requirements in bodies of surface water can be determined and monitored. 2007:4. Stockholm: Swedish Environmental Protection Agency.

SFS 2004:660 (2004) Förordning (2004:660) om förvaltning av kvaliteten på vattenmiljön.

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

Thackeray S., Nõges P., Dunbar M., Dudley B., Skjelbred B., Morabito G., Carvalho L., Phillips G., Mischke U. (2012) Uncertainty in Lake Phytoplankton Metrics. WISER deliverable 3.1-3. http://www.wiser.eu/download/D3.1-3.pdf

Underwood A.J. (1992) Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world. Journal of Experimental Marine Biology and Ecology, 161: 145–178.

Wiens J.A. (1989) Spatial scaling in ecology. Functional Ecology, 3(4): 385–397.

Annex A: Statistical terminology

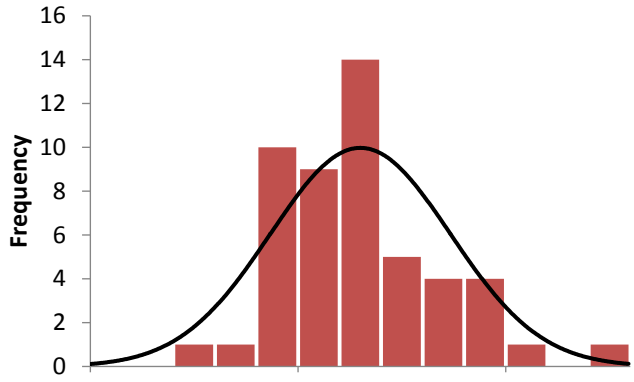

FIGURE A.1
Illustration of difference between observations and model (see text for further explanations).

In statistics, it is important to distinguish between data and the model used to describe the data. Therefore the terminology used to separate model and estimated parameters differs. In figure A.1, the histogram represents the sampling distribution, whereas the solid line represents the modelled distribution, assuming the data to be normally distributed.

TABLE A1
Statistical terminology and the distinction between observations (estimates) and models (statistical populations).

Observations	Model
x_1, x_2, x_3, ….. , x_n are observed values, e.g. x_1 = 10.5 and so forth The average is $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. It is also called the sample mean. The sample standard deviation is: $$s = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$	X is a stochastic variable that can be described by a distribution, e.g. the normal distribution N(μ; σ^2). The mean of the distribution is μ and is normally unknown. It is estimated ($\hat{\mu}$) by the sample mean. The standard deviation of the distribution is σ and is normally unknown. It is estimated by the sample standard deviation (*s*).
The standard error is the standard deviation of the sampling distribution of a statistic, e.g. the standard error of the mean ($SEM = \sqrt{\frac{s^2}{n}}$, the estimated standard deviation of $\hat{\mu}$) or the standard error of a regression slope. Thus, standard error refers to the standard deviation of estimated parameters used to characterize a distribution.	-

Calculation of precision and confidence

The precision of an indicator is a measure of uncertainty and equals the half-width of the confidence interval, i.e. for a symmetrical distribution (e.g. normal) it is the distance from the mean value to the lower or upper confidence limit.

If the indicator is normally distributed or is estimated with a large number of df, i.e. ≥30, the precision can be found as:

$$CI\% = \sqrt{V[\bar{y}]} \times z_{1-\alpha/2}.$$

If it is estimated with few samples, the confidence interval is estimated using the *t*-distribution:

$$CI\% = \sqrt{V[\bar{y}]} \times t_{1-\alpha/2,df},$$

where α is the desired level of type 1 error (and $1-\alpha/2$ is the desired confidence in the estimate). The WFD does not give any recommendations as to appropriate levels of type 1 error.

The confidence in classification 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).

Reporting the confidence of a classification.

Once the probabilities of the different classes are calculated, there are various options for classification. The CIS guideline no. 7 outlines three approaches for classification:

1) The *fail-safe approach* is a precautionary approach for the environment. It generally assumes a worse ecological status (null hypothesis) unless a better ecological status can be documented with sufficient confidence. Thus, for testing the G–M boundary, the null hypothesis is that the status is moderate or less, and sufficient confidence is needed to

reject this hypothesis to conclude that the status is actually better than moderate. This is the environmentally friendly option, in which the burden of proof rests with the polluter.

2) The *benefit-of-doubt approach* is a precautionary approach for the polluters. It generally assumes a better ecological status (null hypothesis) unless a worse ecological status can be documented with sufficient confidence. Thus, for testing the G–M boundary, the null hypothesis is that the status is good or better, and sufficient confidence is needed to reject this hypothesis to conclude that the status is actually worse than good. This is the polluter-friendly option, in which the burden of proof rests with the environment.

3) The *face-value approach* does not consider confidence or the uncertainty of an indicator. The value of the indicator strictly gives the status class, disregarding whether or not the indicator is well-determined. Thus, if the indicator value is above the G–M boundary, the status is good (or high), and if the indicator value is below the G–M boundary, the status is moderate (or below). Since the uncertainty of the indicator is disregarded, the burden of proof is shared equally between the polluter and the environment.

It is argued in the CIS guideline no. 7 that if the true value of an indicator is close to the G–M boundary, increased monitoring is needed to achieve better indicator precision and thus more confidence in the classification. However, if the true ecological status is moderate or below, there may not be an incentive to increase monitoring efforts with the benefit-of-doubt or face-value approaches, because a high uncertainty would increase the probability of actually achieving a better than moderate status.

The CIS guideline no. 13 recommends that confidence be reported for: 1) the status class reported, 2) the status worse than reported, and 3) the status class better than reported. Confidence reporting using a confidence level of 95% is illustrated for the three classification approaches above using an indicator with a distribution given in the figure A.2.

1. The status according to the *fail-safe approach* is moderate, because the status class cannot be classified as better than moderate with 95% confidence (Figure A.2). Thus, the confidence reporting should state that the ecological status is moderate with 24.2% confidence and better than moderate with 75.8% confidence.

2. The status according to the *benefit-of-doubt approach* is high, because the status class cannot be classified as lower than high with 95% confidence. Thus, the confidence reporting should state that the ecological status is high with 8.2% confidence and less than high with 91.2% confidence (Figure A.2).

3. The status according to the *face-value approach* is good, because the mean of the indicator is in the interval for good status. Thus, the confidence reporting should state that the ecological status is good with 67.6% confidence, less than good with 24.2% confidence, and better than good with 8.2% confidence (Figure A.2).

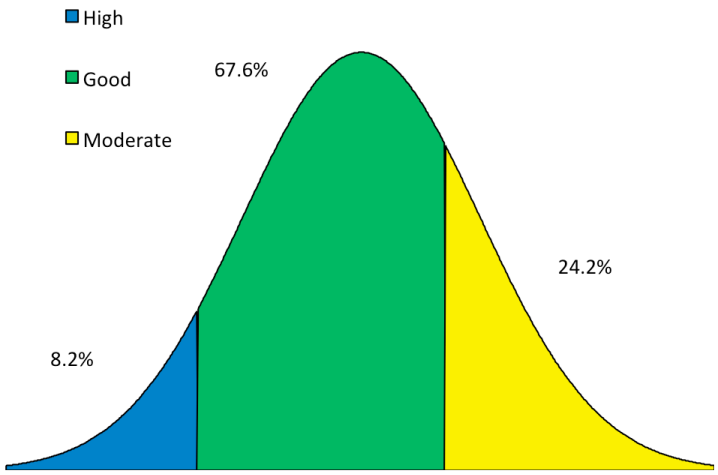

FIGURE A.2
Figure illustrating confidence in classification and classification approaches. Probabilities for three different status classes are given.

Annex B: Review of indicators, sampling requirements, and uncertainty procedures for Swedish WFD indicators

	Indicator characteristics			Sampling kriteria (prescribed by BG or standard)			
	Unit of indicator	**Responds to**	**Sampling method**	**Depth**	**Substrate**	**Time of sampling**	**Other criteria**
Lakes							
Phytoplankton							
Total biomass	microg/l	Nutrients	Water sampler (Ruttner)	In epilimnion	not relevant	July-Aug	Samples taken of epilimnion witin 200 radius.
Proportion cyanobacteria	prop. biomasscyb	Nutrients	Water sampler (Ruttner)	In epilimnion	not relevant	July-Aug	Samples taken of epilimnion witin 200 radius.
TPI (trofic index)	Index based on biomasses and indictaor values	Nutrients	Water sampler (Ruttner)	In epilimnion	not relevant	July-Aug	Samples taken of epilimnion witin 200 radius.
Species richness	# species	Acidity	Water sampler (Ruttner)	In epilimnion	not relevant	July-Aug	Samples taken of epilimnion witin 200 radius.
Chlorophyll	microg/l	Nutrients	Water sampler (Ruttner)	In epilimnion	not relevant	July-Aug	Samples taken of epilimnion witin 200 radius.
Macrophytes							
Trofic index (TMI)	Index based on precence of species and indicator values	Nutrients	Visual, quadrates, rake	Transects over the whole depth	all	Late summer	Transects are placed "subjektivt optimalt" to maximise number of species
Macrofauna							
ASPT	Index based on precence of families and indicator values	Ecological quality	Grabs (profundal) and net (hâv) in littoral	profundal (deep) and littoral (<1 m)	accumulation bottom	Autumn	within a radius of 100 or 50 m, little slope
BQI	Index based on proportions of certain midge species and indicator values	Nutrients	Grabs (profundal) and net (hâv) in littoral	profundal (deep) and littoral (<1 m)	accumulation bottom	Autumn	within a radius of 100 or 50 m, little slope
MILA	Multimetric index based on six indices	Acidity	Grabs (profundal) and net (hâv) in littoral	profundal (deep) and littoral (<1 m)	accumulation bottom	Autumn	within a radius of 100 or 50 m, little slope
Fish							
EQR8	Multimetric index based on eight indices (abundance of species, multiple regression)	Ecological quality	Fishing net, "nordiskt översiktsnät"	Stratified according to depth.	all substrates	July-August	whole lake, lake need to fulfill environmental criteria occurrence
Streams							
Diatoms							
IPS	Index based on relative abundance, indicator value and sensitivity of taxa	Nutrients and organic enrichment	Field: brush, Lab: microscope	Whole deth range except very shallow	stones 10-25 cm (or vegetation)	Late summer/autumn	running water, accessible, 5 pooled samples within 10 m
ACID (supporting variable)		Acidity	Field: brush, Lab: microscope	Whole deth range except very shallow	stones 10-25 cm (or vegetation)	Late summer/autumn	running water, accessible, 5 pooled samples within 10 m
%PT (supporting variable)		Organic enrichment	Field: brush, Lab: microscope	Whole deth range except very shallow	stones 10-25 cm (or vegetation)	Late summer/autumn	running water, accessible, 5 pooled samples within 10 m
TDI (supporting variable)		Nutrients	Field: brush, Lab: microscope	Whole deth range except very shallow	stones 10-25 cm (or vegetation)	Late summer/autumn	running water, accessible, 5 pooled samples within 10 m
Macrofauna							
ASPT	Index based on precence (?) of families and indicator values	Ecological quality	Net ("sparkprov"), 0.5 mm sieve	<1 m depth	hard bottom (stones?), homogenous	Autumn	running water (10 cm/s), within 10 m stretch, not dried out, 50 m of homogenous conditions upstream, 100 m from lake
DJ-index	Multimetric index based on five indices	Nutrients	Net ("sparkprov"), 0.5 mm sieve	<1 m depth	hard bottom (stones?), homogenous	Autumn	running water (10 cm/s), within 10 m stretch, not dried out, 50 m of homogenous conditions upstream, 100 m from lake
MISA	Multimetric index based on six indices	Acidity	Net ("sparkprov"), 0.5 mm sieve	<1 m depth	hard bottom (stones?), homogenous	Autumn	running water (10 cm/s), within 10 m stretch, not dried out, 50 m of homogenous conditions upstream, 100 m from lake
Fish							
VIX	Multimetric index based on six indices (abundance of species, multiple regression)	Ecological quality	Electrofishing	<1 m depth	stones, gravel, possibly various types	August-october	running water, 20-70 cm/s
VIXsm (additional index)	Multimetric index based on four indices (abundance of species, multiple regression)	Acidity/morphology	Electrofishing	<1 m depth	stones, gravel, possibly various types	August-october	running water, 20-70 cm/s
VIXh (additional index)	Multimetric index based on four indices (abundance of species, multiple regression)	Hydrology	Electrofishing	<1 m depth	stones, gravel, possibly various types	August-october	running water, 20-70 cm/s
Coastal waters							
Phytoplankton							
Biovolyme	mm3/l	Nutrients and eutrophication	hose or water sampler	0-10 or 0.5 m	not relevant	June-august	Station representative to water-body
Chlorophyll a	microg/l	Nutrients and eutrophication	hose or water sampler	0-10 or 0.5 m	not relevant	June-august	Station representative to water-body
Macrophytes							
MSMDI	Index based on lower depth limits of selected (≥3) taxa	Nutrients, Eutrophication and visibility	Visual, diving	0-20 m	Hard or soft substratum	July-september	Salinity within type-specific interval
Macrofauna							
BQIm	Index based on proportions of species, sensitivity, richness and abundance	Eutrophication	Grab (van Veen or Smith-McIntyre)	>5 m	accumulation bottoms	May-june	

Sampling design (prescribed by BG or standard)					Uncertainty in assessment procedure		
	# replicates within sites	# sites	# times of sampling per year	# years of sampling	Measure of uncertainty	Classification approach	Uncertainty in reference values
Lakes							
Phytoplankton							
Total biomass	1 (pooled)	1-5	1	>3	SD	Face-value	not reported
Proportion cyanobacteria	1 (pooled)	1-5	1	>3	SD	Face-value	not reported
TPI (trofic index)	1 (pooled)	1-5	1	>3	SD	Face-value	not reported
Species richness	1 (pooled)	1-5	1	>3	SD	Face-value	not reported
Chlorophyll	1 (pooled)	1-5	1	>3	SD	Face-value	not reported
Macrophytes							
Trofic index (TMI)	1	>8	1	-	Deviation in EQ-value=0.05	Face-value	not reported
Macrofauna							
ASPT	5	1	1	-	SD or method	Face-value	not reported
BQI	5	1	1	-	SD or method	Face-value	not reported
MILA	5	1	1	-	SD or method	Face-value	not reported
Fish							
EQR8	1 (pooled)	>4 strongly dependent on lake size	1	1-6	SD or method	Face-value	not reported
Streams							
Diatoms							
IPS	1 (≥5 pooled)	≥1	1	>1	SD or method	Face-value	not reported
ACID (supporting variable)	1 (≥5 pooled)	≥1	1	>1	SD or method	not relevent	not relevent
%PT (supporting variable)	1 (≥5 pooled)	≥1	1	>1	SD or method	not relevent	not relevent
TDI (supporting variable)	1 (≥5 pooled)	≥1	1	>1	SD or method	not relevent	not relevent
Macrofauna							
ASPT	5	1	1	-	SD or method	Face-value	not reported
DJ-index	5	1	1	-	SD or method	Face-value	not reported
MISA	5	1	1	-	SD or method	Face-value	not reported
Fish							
VIX	1 (pooled)	5-30 depending on catchment size	1	-	SD or expected SD based on multiple regression	Face-value	not reported
VIXsm (additional index)	1 (pooled)	5-30 depending on catchment size	1	-	SD or expected SD based on multiple regression	Face-value	not reported
VIXh (additional index)	1 (pooled)	5-30 depending on catchment size	1	-	SD or expected SD based on multiple regression	Face-value	not reported
Coastal waters							
Phytoplankton							
Biovolyme	1	1	3-5	3	not reported	Face-value	not reported
Chlorophyll a	1	1	3-5	3	not reported	Face-value	not reported
Macrophytes							
MSMDI	1	≥3 required	1	-	SD	Face-value	not reported
Macrofauna							
BQIm	1 Bothnian bay, Baltic proper; 4 Halland, 2-4 Bohuslän	≥5 required	1	-	Bootstrap confidence interval	Fail-safe	not reported

Uncertainty of biological indicators for the WFD in Swedish water bodies: current procedures and a proposed framework for the future

Monitoring and status assessment of the environment is always associated with uncertainties. Therefore the Water Framework Directive (WFD) defines two aspects of uncertainty, precision and confidence, that need to be estimated and reported by the member states. In this report we review the basic requirements from the Directive and the routines defined in the Swedish assessment criteria. We conclude that there are substantial differences in the way uncertainty is assessed among biological quality elements and that the overall uncertainty of estimates for whole six-year WFD cycles is not addressed by any of the quality elements. We propose a general framework which can be used to harmonise assessments of uncertainty across quality elements and to reduce uncertainty by a more efficient use of existing data, optimisation of monitoring programs and by accounting for environmental factors that may explain some of the random variability. These possibilities will be further explored in coming work within WATERS.

WATERS is coordinated by:

Havsmiljöinstitutet
Swedish Institute for the Marine Environment

WATERS is financed by:

NATUR VÅRDS VERKET
SWEDISH ENVIRONMENTAL PROTECTION AGENCY