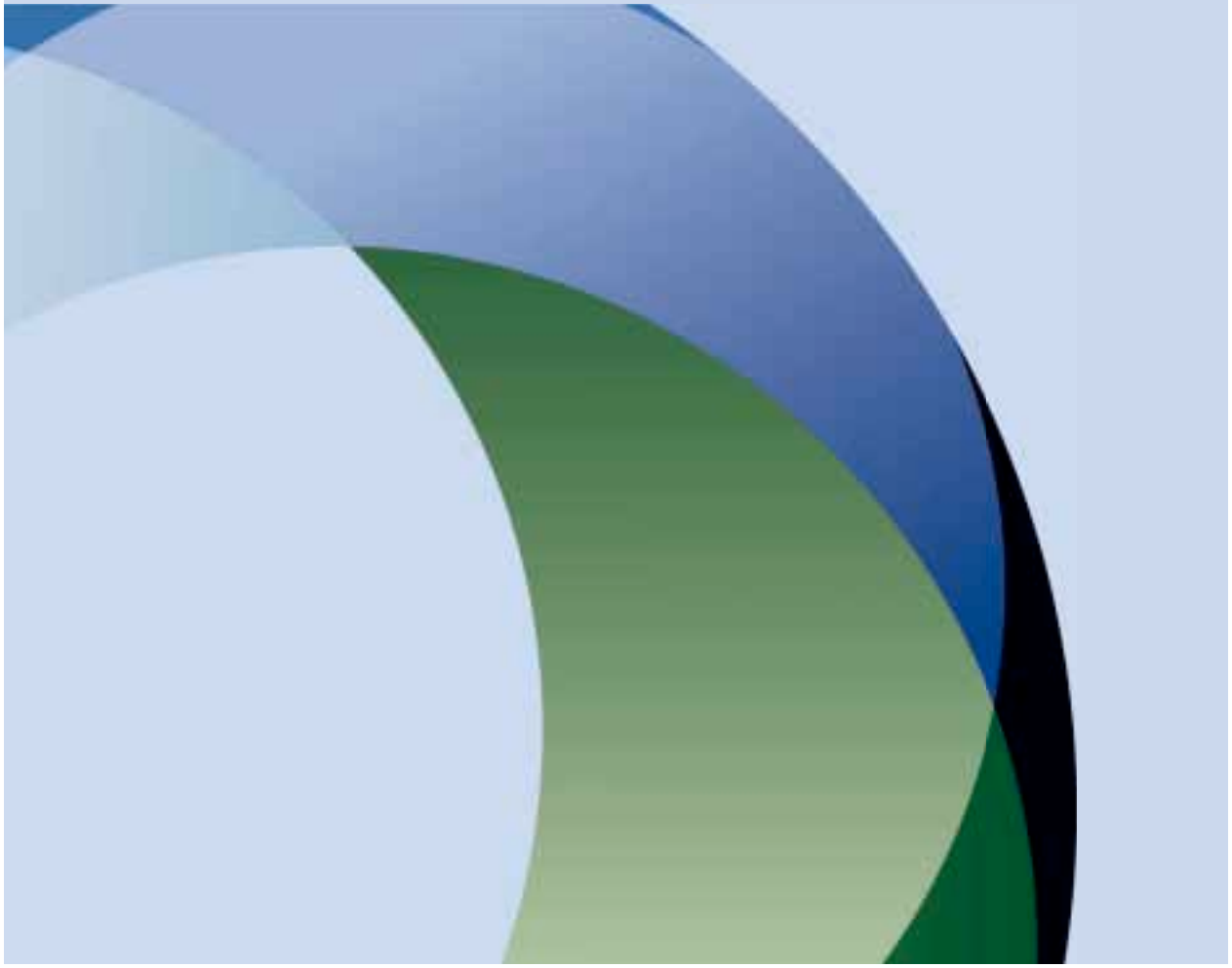
CLIMATE AND POLLUTION AGENCY

Statistical analysis of Norwegian long-term national and regional monitoring data

Estimation of variance components and detectable trends

Effects of changes in monitoring design on precision of time trend assessments for contaminants in biota

TA
2939
2012



NIVA
Norwegian Institute for Water Research

Norwegian Institute for Water Research
– an institute in the Environmental Research Alliance of Norway

REPORT

Main Office
Gaustadalléen 21
NO-0349 Oslo, Norway
Phone (47) 22 18 51 00
Telefax (47) 22 18 52 00
Internet: www.niva.no

Regional Office, Sørlandet
Jon Lilletuns vei 3
NO-4879 Grimstad, Norway
Phone (47) 22 18 51 00
Telefax (47) 37 04 45 13

Regional Office, Østlandet
Sandvikaveien 59
NO-2312 Ottestad, Norway
Phone (47) 22 18 51 00
Telefax (47) 62 57 66 53

Regional Office, Vestlandet
Thormøhlens gate 53 D
NO-5006 Bergen Norway
Phone (47) 22 18 51 00
Telefax (47) 55 31 22 14

Regional Office Central
Pirsenteret, Havnegata 9
P.O.Box 1266
NO-7462 Trondheim
Phone (47) 22 18 51 00
Telefax (47) 73 54 63 87

Title	Report No..	Date
Effects of changes in monitoring design on precision of time trend assessments for contaminants in biota *Statistical analysis of Norwegian long-term national and regional monitoring data* *Estimation of variance components and detectable trends*	NIVA 6336-2012 Klif TA-2939/2012 Project No. 11481	25.6.2012 Pages Price 77
Author(s) Birger Bjerkeng	Topic group Marine ecology Geographical area Norwegian coastal waters	Distribution Open Printed NIVA

Client(s)	Client ref.
Climate and Pollution Agency/Klima- og forurensningsdirektoratet (Klif), Norway.	Thomas Hartnik

Abstract
Data from the Norwegian CEMP program has been used to analyse how the precision of trend assessments will be affected by changes in the monitoring program. Monitoring with two or three year intervals instead of annually will increase the minimum trend that can be detected with specified power and significance level by 50 to 100 % compared to yearly monitoring, based on trend detection with the current CEMP procedure. It cannot be generally recommended to decrease the monitoring frequency but it may be considered for stations where it is established that concentrations stays well below levels of any concern, and without any signs of upward trend over a number of years. Decreasing the sample size for cod liver from 25 to 20 fish would increase the minimum detectable trend with only 2-7 % (e.g. from 10 % to well within 11 % yearly change). Reduction to 15 fish would increase the minimum detectable trend by 3 to 22 %, less than 10 % for most stations and parameters. Analysing 25 fish as 5 batch samples of 5 fish each will give even less reduction of trend detection ability than reduction to 20 fish analysed individually. Even considering the uncertainty of these estimates, the general conclusion is that reductions in the cod liver program as described here will not reduce trend detection ability to an extent that it has practical significance. For areas where 3 or 4 repeated or distributed mussel sampling is done each year, a reduction to a single mussel sample per year may lead to a considerable reduction in trend detection ability (probably up to about 50 % increase in minimum detectable trend, possibly 80-90 %). A more cautious reduction to fewer samples, but still more than one, could probably be implemented without a large effect on the ability to detect trends.

4 keywords, Norwegian	4 keywords, English
1. Miljøgifter	1. Contaminants
2. Biota	2. Biota
3. Marin	3. Marine
4. Trendanalyse	4. Trend analysis

Birger Bjerkeng
Project Manager

Kai Sørensen
Research Manager

Brit Lisa Skjelkvåle
Research Director

ISBN 978-82-577-6071-7

Preface

This report has been written on behalf of the Climate and Pollution Agency (Klif) in Norway. The goal of the project was to study how changes in the monitoring of hazardous substances in biota may affect the ability to detect trends, and the present report addresses three specific issues; the effect of reduced monitoring frequency, the effect of changing sample size for monitoring of concentrations in cod liver, and the effect of reducing number of mussel samples per year in locations where current monitoring include repeated or distributed sampling within a local area, also partly with replicate samples each time or site.

NIVA, Oslo 25. June 2012

Birger Bjerkeng

Table of Contents

Summary **1**

1. Introduction **4**

2. Effect of multiyear monitoring intervals – CEMP procedure with subsampling **5**

2.1 Method description 5

2.1.1 Modified CEMP procedure for yearly data 5

2.1.2 Subsampling procedure for assessing effect of reduced monitoring frequency 6

2.2 Selecting data for analysis 7

2.3 Results 7

2.4 Choice of optimal trend test 11

3. Effects of reduced sample size or analysis of batch samples for cod liver data **13**

3.1 Statistical model – method of analysis 13

3.2 Selecting data for the analysis 15

3.3 Results of variance analysis 18

4. Repeated sampling within year and area – analysis of mussel data for Grenland and Sørfjord. **24**

4.1 Statistical model - method of analysis 24

4.2 Selection of data for analysis 26

4.2.1 Data from the Grenland area 26

4.2.2 Mussel data from Sørfjord, Hardanger 28

4.3 Results 30

4.3.1 Grenland dioxin data in Grenland mussels 30

4.3.2 PCB, DDT, HCB and Mercury in Grenland mussels 33

4.3.3 Metals Cd, Cu, Pb and Zn in Grenland Mussels 34

4.3.4 Metals, PCB and pp'- DDE in mussels from Sørfjord 36

4.4 Conclusions for analysis of mussel data 42

5. References **43**

6. Explanation of acronyms and abbreviations **44**

Appendix A. Technical description of trend assessment procedure for Norwegian CEMP data 45
A.1 Aggregating data into yearly medians 45
A.2 Assessing trend for yearly data 45
A.3 Assessing trend for monitoring each second or third year. 46
A.4 Smoother fit by weighted least squares regression (LOWESS) 47
A.4.1 Setting regression weights 47
A.4.2 Estimating LOWESS parameters and fitting trend values 47
A.5 Assessing trend from smoother function 48

Appendix B. Linear regression trend test 50

Appendix C. Significance and power – the ability to detect trends 51
C.1 General description 51
C.2 Application to linear trend test 53

Appendix D. Estimating the effect of design changes on between-year variance from analysis of variance (ANOVA) with two nested, random random factors. 54
D.1 Expressions for mean square terms 54
D.2 Estimating variance components 55
D.3 Estimating mean squares for a new sampling design 56
D.4 Estimating relative change of between-year variance 58
D.4.1 Nested sampling within year (more than one sample, some with replicates). 58
D.4.2 One main sample per year (divided in replicate subsamples) 59

Appendix E. Statistical properties of batch samples when individual concentrations are log-normally distributed. 61

Appendix F. Results from CEMP procedure with subsampling to simulate 2 or 3 year monitoring intervals 64

Appendix G. Time series plots of blue mussel data from Sørfjord 72

Summary

Data from the Norwegian part of OSPAR's CEMP[1] program has been used to analyse how the precision of trend assessments will be affected by changes in the monitoring program for hazardous substances in biota. The study focuses on **three issues**:

1. The effect of monitoring with 2 or 3 years intervals instead of yearly
2. The effect of changing monitoring of cod liver from 25 individual livers per year and station to 20 or 15 individual samples, or 5 batch samples of 5 individuals each
3. The gain of repeated or distributed mussel sampling each year compared to taking a single sample.

All analyses are done on log-transformed data, using natural logarithms. For data series with measurements below quantification or detection limit, either the whole data series is excluded from analysis, or all data for years with such observations are excluded, depending on the type of analysis.

The first issue (monitoring with 2 or 3 years intervals instead of yearly) has been studied by running the Norwegian CEMP trend assessment procedure on subsets of data corresponding to monitoring each 2nd or 3rd year, running over all possible starting points. The procedure uses yearly medians. According to the CEMP trend assessment procedure, linear regression is used for time series with 5 or 6 data points; for series with 7 or more data points local weighted regression with a 7 year window was used. The results for 2 and 3 year intervals are compared with the results from the complete yearly data set for each time series. Comparisons are done both for the minimum trend that could be detected with specified certainty, and the actual trend estimates.

The results show that the minimum trend that can be detected with specified power and significance level will typically increase by 50 to 100 % when monitoring frequency is decreased. That means that it takes longer time to detect trends or that time trends have to be fairly large to be detected. For cod data, a large number of time series give minimum detectable trends (with significance level 0.1 and power of 90 %) in the range from 0.05 to 0.1 on natural log scale (meaning a 5 to 10 % change in concentration within one year). Monitoring with 2 or 3 years intervals would give detectable trends from 0.06 to 0.2 instead for these time series (6 to 20 % change per year). For a majority of the existing mussel time series, the detectable trend for yearly monitoring is in the range 0.05-0.2 (5-22% change per year). Monitoring each 2nd or 3rd year would increase the minimum detectable trend to10-50 % change per year.

For many time series the trend estimates based on monitoring each 2nd or 3rd year are quite different from fairly clear trends based on yearly data, and the trends with 2 or 3 years monitoring interval often shows strong random variation depending on starting points, e.g. monitoring even or odd years. Therefore it cannot be recommended generally to decrease the monitoring frequency in cases where possible trends are of concern, but it may be considered for stations where established time series show concentrations well below levels of any concern, and without any upward trend over a number of years. For trend analysis using smoothing (local weighted regression) on data with 2 year intervals, the results are only marginally better if the local regression window for the smoother is expanded from 7 years to 9 years. For monitoring

[1] The **C**o-ordinated **E**nvironmental **M**onitoring programme of the **Os**lo **Par**is Commission.
(*Note that* Chapter 6 contains an explanation of abbreviations and acronyms occurring in the document).

interval of 3 years, the results are substantially improved if the regression window is expanded to 13 years, making them nearly comparable to 2 year interval. A comparison of trend detection by smoother and log-linear regression indicates that revisions to the trend detection procedure might be appropriate.

The second issue (changing monitoring of cod liver) has been studied by analysing long cod liver time series with approximately 25 fish per year. Only series without missing years and without any observations below quantification limit are used. For these analyses the smooth trends fitted by the CEMP procedure have been removed from the data, leaving the irregular noise component of between-year variation. These de-trended data are subjected to Analysis of Variance (ANOVA) to estimate variance terms and calculate the effect of changing the number of replicates. It can be concluded that reducing the number of replicates per sampling location from 25 to 20 fish per year has only a marginal effect on the trend detection ability, increasing the minimum detectable trend under given conditions by only 2-7 %, while a reduction to 15 fish would increase the detectable trend by 3 to 22 % (less than 10 % for most stations and parameters). These increases show a reduced ability to detect trends when reducing the number of replicates, but the effect is generally small or moderate.

Analysing 25 fish in 5 batch samples of 5 fish each will decrease the trend detection ability less than reducing to 20 fish analysed individually, but individual liver analysis might give better protection against outliers and allow for adjustments for biological characteristics. One clear example of outliers is the occurrence in 2002 of very high PCB concentrations in four out of 25 fish from Sørfjord station 53B. It was established that it was related to a special event of removal of paint from an old building the year before (Ruus et al. 2006). If the fish had been analysed in only a few batch samples it would not have been possible to detect and exclude this disturbance as one can with individual analyses.

The 70 % confidence intervals[1] for the relative change of detectable trend generally extend from 0.65 to 1.4 times the estimated changes listed above (e.g. an interval from 6.5 to 14 % around an estimated increase of 10 % in detectable trend). Even with this uncertainty, the general conclusion is that reducing the number of replicates in the cod liver program as described here will not reduce trend detection ability so much that it has practical significance.

The third issue (reducing number of yearly samples for mussel monitoring) has been studied by analysing subsets of mussel data in the Norwegian CEMP program from the Grenland region southwest of Oslo, and from Sørfjord in Hardanger, in both cases supplemented by data from local or regional monitoring programs. The subsets consist of data from years and localities where mussel sampling has been repeated within the same season within a 1-2 month interval, or where samples are taken from different sites at the same location. The residual between-year variance for yearly averages in time trend analysis will depend both on the number of main samples (different sampling sites and/or dates) and on how the shells collected in each main sample are distributed on different subsamples for chemical analysis. The data are analysed in a General linear model (GLM) with two levels Year, and Main sample nested in year. For Sørfjord the data sets from 4 stations have been analysed together, with Year nested in Station, to estimate

[1] A confidence interval is an uncertainty range between a lower and an upper limit, determined in such a way that it will with a prescribed probability (confidence level) include the true value to be estimated, provided that the assumptions that are made about the form of underlying distributions are valid (e.g. that observations are normally distributed). Note that the confidence level is not the probability that the true value is between the specific limits; it is the probability of finding limits so that the true value is included in the range. The probability applies to repeated derivations of confidence limits for new data sets; it does not apply to the numbers determined for a specific instance.

within-year variance as an average over all stations, while allowing for differences in time trend between stations. Where it is needed, the between-year variance estimated by the GLM model has been analysed further to separate the irregular short-term between-year variation from the effect of medium- and long-term time trends and large sudden changes in levels because of major discharge events. It is the irregular short-term between-year variation that is used as "noise" variance component to estimate capability for detecting future trends in a situation dominated by gradual change without any major events.

Monitoring of dioxins, PCBs, DDT with metabolites, HCB and mercury (Hg) in mussels from Langesundsfjorden in the Grenland area has been done with 3 or 4 main samples per year, on average divided into 2 subsamples for analysis. If the monitoring is reduced to one single mussel sample per year, the between-year standard deviation is estimated to increase with 35-50 % for dioxins, 45-65 % for PCBs, p,p'-DDE (DDT metabolite, also referenced by its ICES code DDEPP), HCB and Hg. If it is reduced to 2 main samples, each divided in 2 or three subsamples, the estimated increase is kept within 12 % for dioxins and within 15 % for PCBs, p,p'-DDE, HCB and Hg. 70 % confidence intervals extend from 0.5 to 1.5 times these estimated changes.

Long time series for metals cadmium (Cd), copper (Cu), lead (Pb) and zinc (Zn) exist only for one station in the Grenland region with one main sample analysed in three subsamples (with data from the last two years from two other stations). If the monitoring is reduced to one analysed sample per year, it is estimated that the between-year standard deviation will increase with 15-30 %. With reduction from 3 to 2 subsamples the increase is estimated to be within 8 %. The two-sided 70 % confidence limits extend from 0.65 to 1.45 times these estimated changes.

For mussels from Sørfjorden, Hardanger, the current monitoring at some stations includes two main samples per year, with 3 and 1 subsample, respectively. Reduction to one analysed sample per year is estimated to increase between-year standard deviation by 25 - 50 %, with the largest effect for DDEPP, less for metals and PCB. The 70 % confidence intervals extend from 0.8 to 1.2 times the estimated changes.

The conclusion is that reducing to a single mussel sample per year may lead to a considerable reduction in trend detection ability. A more cautious reduction, to fewer, but still more than one sample, could probably be implemented without a large effect on the ability to detect trends.

1. Introduction

The goal of the project is to analyse how the precision of trend assessments will be affected by changes in the monitoring program for hazardous substances in biota, specifically for the Norwegian CEMP program. The Norwegian Climate and Pollution Authority (Klif) has raised three questions:

1. Many sampling locations in the Norwegian CEMP program have yearly monitoring of fish and blue mussels. How will precision of trend estimations change if monitoring is done with 2 or 3 years intervals instead of annually? Is there any difference between impacted stations and stations in areas with background concentrations?
2. In the current CEMP program 25 individual cod livers are analysed per sampling occasion for each station. How will the power and precision of trend assessments change by instead analysing
    a. 20 or 15 individual samples?
    b. 5 batch samples of 5 individuals each?
3. In the Grenland region and in Sørfjorden in Hardanger, sampling of blue mussels has been repeated with two months interval during autumn, and at more than one site in the same local area, partly also with replicate subsamples from each time or site. Does this monitoring strategy give a better trend assessment power than for one sample per year?

The questions are treated separately in the three following chapters. Each chapter describes the analysis of one main question, with methods, selection of data and results presented together. A final chapter summarizes the conclusions. Technical details are described in Appendices.

Chapter 6 contains an explanation of abbreviations and acronyms occurring in the document.

2. Effect of multiyear monitoring intervals – CEMP procedure with subsampling

2.1 Method description

The effect of increasing the monitoring interval to 2 or 3 years is assessed empirically by applying a customized version of the CEMP assessment procedure to all suitable time series, first using all data, and then for subsets of data which correspond to monitoring each second year or each third year. For two-year monitoring intervals, each time series is analysed twice, using data only from even and odd years, respectively. For three year monitoring intervals, the analysis is repeated three times, using available data from years $j*3+k$, where j is any integer, and k is repeat number (=1, 2, 3). For each monitoring interval, the repetitions with different starting points will analyse different, non-overlapping data sets and give results that are unrelated to each other.

An empirically based analysis of the consequences of reduced monitoring frequency is performed by summarizing how results vary between different subsets of data and how precision of the trend assessment is changed compared to using the full set of data.

The procedure can in principle be done on all data series in the CEMP database, with the amount of results varying between time series. For the present purpose, the analysis is restricted to time series with data from all the last 11 years (2000-2010), and results are summarized only for series where the yearly sample medians within this period are well-defined values, independent of observations below analytical limit[1].

2.1.1 Modified CEMP procedure for yearly data

The CEMP assessment procedure (used in Green et al. 2011) is applied separately to each data series (each combination of station, species, tissue and contaminant). The data are first aggregated into a simple time series with one value per year, extracting the yearly median where there is more than one value per year. The log-transformed time series is then analysed for time trend if there is enough data. For series with data for 5 or 6 years, linear regression is used; for longer series a smoother curve based on 7-year local regression window (LOWESS) is fitted to the data. This means that for each year, the fitted value is normally based on weighted regression on data from that year and 3 years before and after, with necessary shifts in the window position near the end of the series. If there are gaps in the series, the regression will be based on fewer points; the time span of the regression window is kept unchanged. The time trend is tested by comparing the fitted value for the last year with the value 10 years before (or as long before as possible if the time series is shorter). Details of the procedure described above are given in Appendix A.

Since the interest here is on future monitoring, some supplemental results are added to the original CEMP procedure.

[1] Analytical limit is used here to refer to the upper limit reported by laboratories for observations where the measurement is too uncertain for a definite quantification. It can be a detection limit or a quantification limit, depending on laboratory practice; also called reporting limit. For such observations, any assignment of a definite, assumed value in the range from 0 to the limit (e.g. half the reported limit) is arbitrary, and for analysis of log-transformed values the arbitrary choice can have a large impact on variance component estimates. If the yearly sample median is independent of any such assumption, it is a well-defined value (see Appendix A.1 for a more detailed discussion).

For series of 7 or more years, a linear trend (on log-scale) is fitted for the last 10 years, or as long as the data permits if less than 10 years, regardless of the length of the data series, and a separate trend test is done on this regression, independent of the smoother. This gives a set of results that may be better for investigating the effect of varying monitoring interval, as described in the following section.

For both the smoother and the separate recent 10-year linear trend, as an addition to testing the actual trend, the results also include estimation of the smallest trend that will be detected with 90 % power in a two-sided test with 10 % significance level[1], using the CEMP procedure.

2.1.2 Subsampling procedure for assessing effect of reduced monitoring frequency

The repeated analysis with data subsets is performed in a customized version of the automated procedure that performs the statistical trend analysis in the CEMP procedure. The procedure is implemented through a set of queries and Visual Basic code in a system of Microsoft ACCESS databases. The results are stored in Microsoft ACCESS tables for further summarizing and analysis.

If the subset analyses are done with unchanged LOWESS regression window, only at most three data points will be included in each local regression, even when subsampling from complete data series without missing years. This leads to considerably higher uncertainty of estimates and trends than for 1 year monitoring interval.

As an alternative, the local regression window and time scale for trend test can be expanded to compensate for the reduced frequency. The reasoning behind this could be that a reduced monitoring frequency would be selected in locations where changes are expected to take place slowly, and where it is acceptable with a longer monitoring period being needed to detect trends. The regression window can be expanded by requiring that at least 5 points are included as basis for each local regression. This leads to local regression windows of 9 years for monitoring each second year, and 13 years for monitoring each third year. The expanded regression window means that the smooth trend curve has higher stiffness, but it will improve the statistical basis behind each fitted value (Trade-off between bias and precision).

The analyses for 2 or 3 year monitoring interval is done both with regression window kept fixed to 7 years (time span of 6 years), including at most 3 years in each local regression, and with local regression windows expanded by requiring 5 points in each regression.

The effect of reducing monitoring frequency is analysed by comparing estimated detectable trend for the full dataset of each series with the results for subsampling of the same series. The results are then summarized by presenting graphs and statistics on the ratios between detectable trends for subsampling and the full dataset.

A theoretical description of the statistical properties of the smoother and the effect one can expect by increasing the monitoring interval, including selection of regression window and time scale for the trend test, are found in Appendix A.

[1] The significance level is the accepted risk of drawing a false conclusion when there is actually no trend. The power is the probability of detecting a real trend. The power will increase with the size of the actual trend and with the selected significance level. If the significance level is changed from 10 % to 5 %, the required minimum trend will increase by 10 to 20 %, depending of the degrees of freedom (Table 12 in Appendix C.1).

2.2 Selecting data for analysis

For the empirical analysis of the effect of multiyear monitoring interval on the CEMP trend fitting procedure, time series are selected from the Norwegian CEMP database (Green et al. 2011). Only time series with yearly data continuing until 2010 are used. Only time series with at least 11 years of data are included.

Yearly data series continuing up to 2010 mostly exist for blue mussels and cod, with more limited data sets for other species. Table 1 shows the number of stations for each species, and lists the basis (dry, wet) used for statistical analysis in the Norwegian CEMP report on Hazardous substances (Green et al. 2011) . Data are converted to these bases before analysis, and data in other bases are excluded if there is not sufficient information for the specific sample to convert to preferred basis.

Table 1. Overview per species of number of stations with suitable data series for empirical assessment of multiyear monitoring interval with the CEMP procedure

Species code	English name, Norwegian name (*Latin name*)	Station count	Preferred basis
MYTI EDU	blue mussel, blåskjell (*Mytilus edulis*)	46	Wet for dioxins Dry for other parameters
NUCE LAP	dogwhelk, purpursnegl (*Nucella lapillus*)	7	Dry (Data only for tributyltin (TBTIN))
GADU MOR	atlantic cod, torsk (*Gadus morhua*)	8	Wet
LEPI WHI	megrim, glassvar (*Lepidorhombus whiffiagonis*)	2	
LIMA LIM	dab, sandflyndre (*Limanda limanda*)	2	
PLAT FLE	flounder, skrubbe (*Platichthys flesus*)	1	
PLEU PLA	plaice, rødspette (*Pleuronectes platessa*)	1	

2.3 Results

The results are summarised by comparing, per station and parameter, the estimated detectable trend for yearly data with the average of detectable trends estimated for each combination of monitoring interval and window size (3 or 5 regression points required). For 2 year monitoring interval this is the average of results for even and odd years and for 3 year monitoring interval it is the average of 3 repeated analyses with different start points. In some cases the subsampling means that fit method changes from smoother to linear trend. Figure 1 summarises results for cod liver and for mussels for the alternatives with expanded local regression window. The figure compares average detectable trend for 2 or 3 year monitoring across the possible starting points with the detectable trend estimated from yearly data; the individual results for each monitoring interval will vary much more compared to the detectable trend. Detailed results for each station are shown in graphs in Appendix F.

Generally, with monitoring each second year, the minimum trend that can be detected within a specified time frame and with specified power increases by 50 to 100 % compared to monitoring each year. The results are marginally better if the local regression window is expanded to include 5 data points in each local regression than for a 7 year smoother.

Monitoring only each third year would for many of the time series result in far worse possibility of detecting trends if the local regression window is kept at 7 years, in many cases the minimum detectable trend increases ten-fold, compare to monitoring each year. By increasing the local regression window to 13 years, so that 5 points are included in each regression, the empirical results for existing data series are comparable to monitoring each second year, with minimum detectable trend often about twice as large as for monitoring each year.

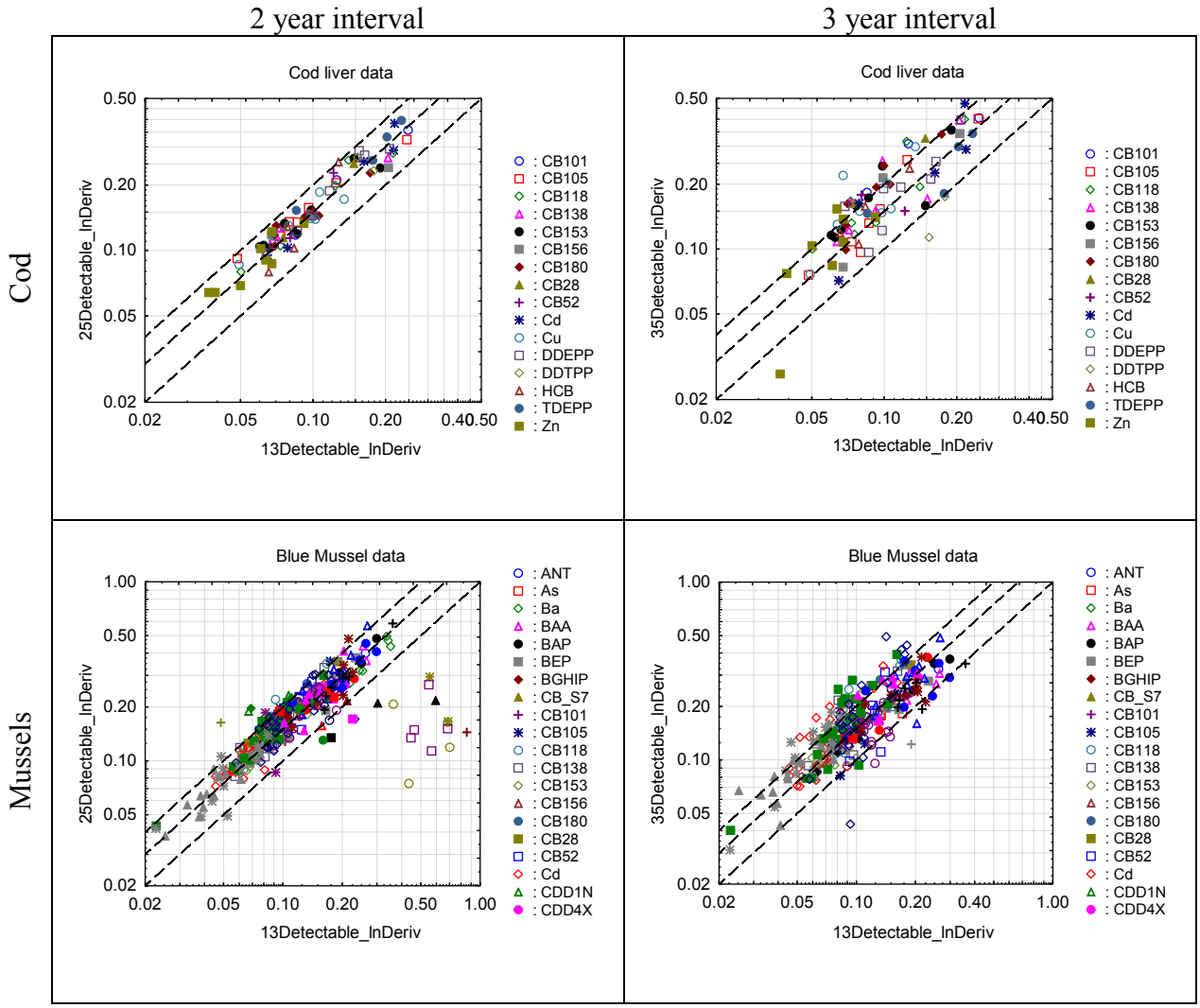


Figure 1. Relationships between detectable trend (natural log scale) for yearly data (horizontal axes) and for 2 and 3 year interval (vertical axes) for cod liver (top) and Blue Mussels (Bottom). The plots on the left show average results across different starting points for 2 year interval with 9 year local regression window (5 points required) on the vertical axes. The plots on the right show average results for 3 year interval and 13 year local regression window. Straight lines are drawn at ratio 1:1, 1.5:1 and 2:1

It may seem like a paradox that figure 1 shows that for some time series 2 year interval give much better trend detectability (lower minimum detectable trend) than yearly monitoring, and the 3 year interval shows better result than for 2 year monitoring for a few time series with upward trends . This happens because the procedure sometimes shifts from a smoother trend test to linear regression for increased monitoring interval because of fewer data points. In addition such results will occur occasionally because the subsampling might exclude precisely strong outliers from the data series.

Another way of showing the effect of reduced monitoring frequency is to compare trend estimates for different sampling frequency for each time series. Figure 2 shows one such comparison. Trends based on yearly data are plotted on the horizontal axis while trends based on two or three years monitoring intervals are plotted on the vertical axis for each possible choice of starting point (even or odd years for 2 year interval; years n*3+j, j=0,1,2 for 3 year intervals). Thus, each time series is represented by 2+3 points lying along a vertical line in the plot, Only time series where the trend from yearly data is significant with p<0.02 are included. A closer look on the data shows that these trends are typically estimated with 15 to 30 % relative standard error. For series with a downward trends of between -0.1 and -0.05 on natural log scale (10 to 5 % reduction per year), the estimated trends from monitoring each second or third year vary from 0 to -10 % between series, and there is also large variation for each series depending on which years are monitored. For the few time series with a clearly significant upward trend, monitoring with 2-year intervals would lead to a false downward trend in more than 30 % of the cases.

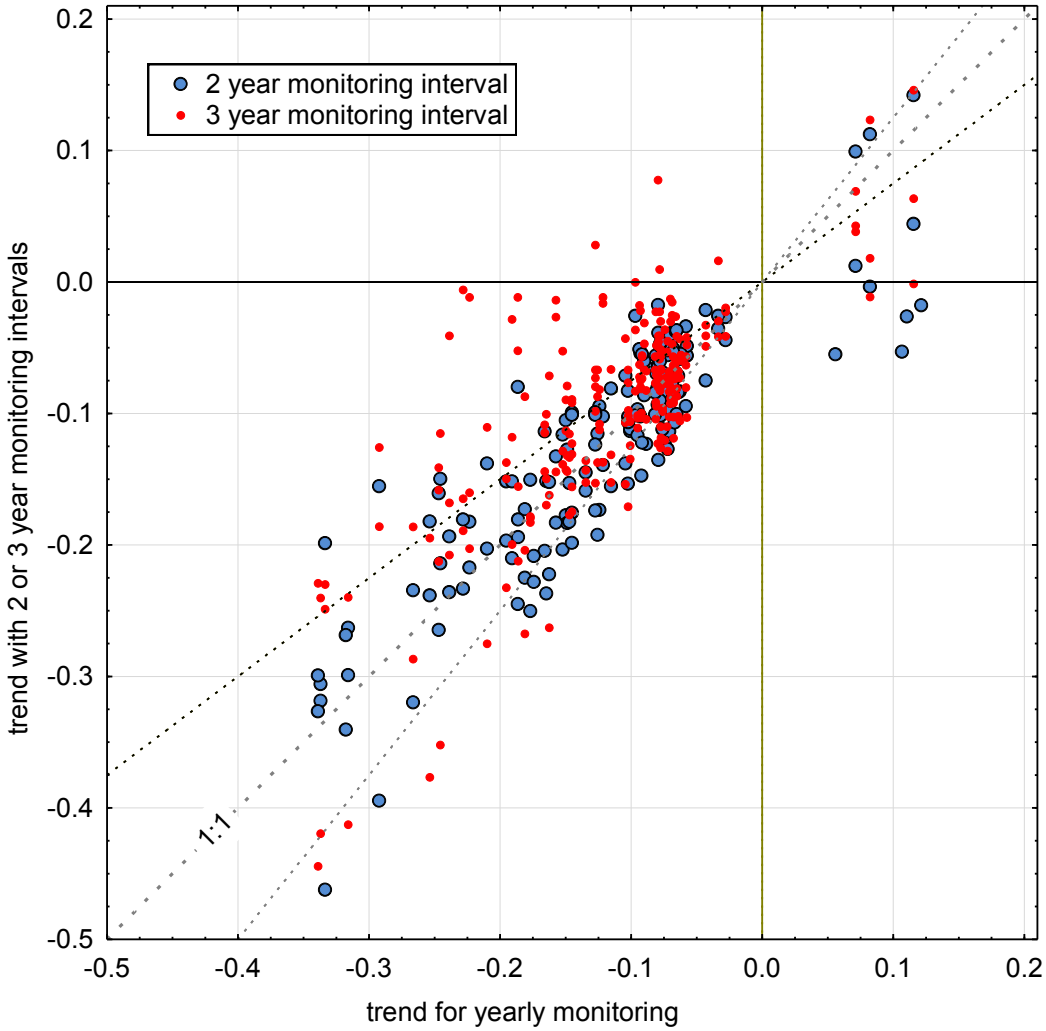


Figure 2. Comparison of estimated trend on natural log-transformed values for different monitoring intervals. The horizontal axis shows trend estimated with yearly data, and the vertical axis shows trends when only data each second or third year is available. The plot includes only time series where the trend from yearly data is significance with p<0.02. Results for each time series is shown in a vertical band of points, with each point representing a specific subsampling scenario (interval and starting point). Only the multiyear schemes with expanded local regression windows are shown. The thin dotted lines indicate ±25 % deviation from a 1:1 relationship.

For some series there have been fluctuations lasting 4-5 years (e.g. Hg in Cod at station 30B, shown in Figure 3). For such series, monitoring each second or third year would have meant that the apparent pattern becomes more random, depending on how monitoring years coincide with the peaks in the data. For this specific time series, all the alternative starting points for 3 year monitoring give the impression of a more general increase, without the multiyear fluctuation pattern seen in the yearly data. The figure also shows that the confidence interval is very different depending on exactly which years are monitored, which means that assessment of trend becomes more random.

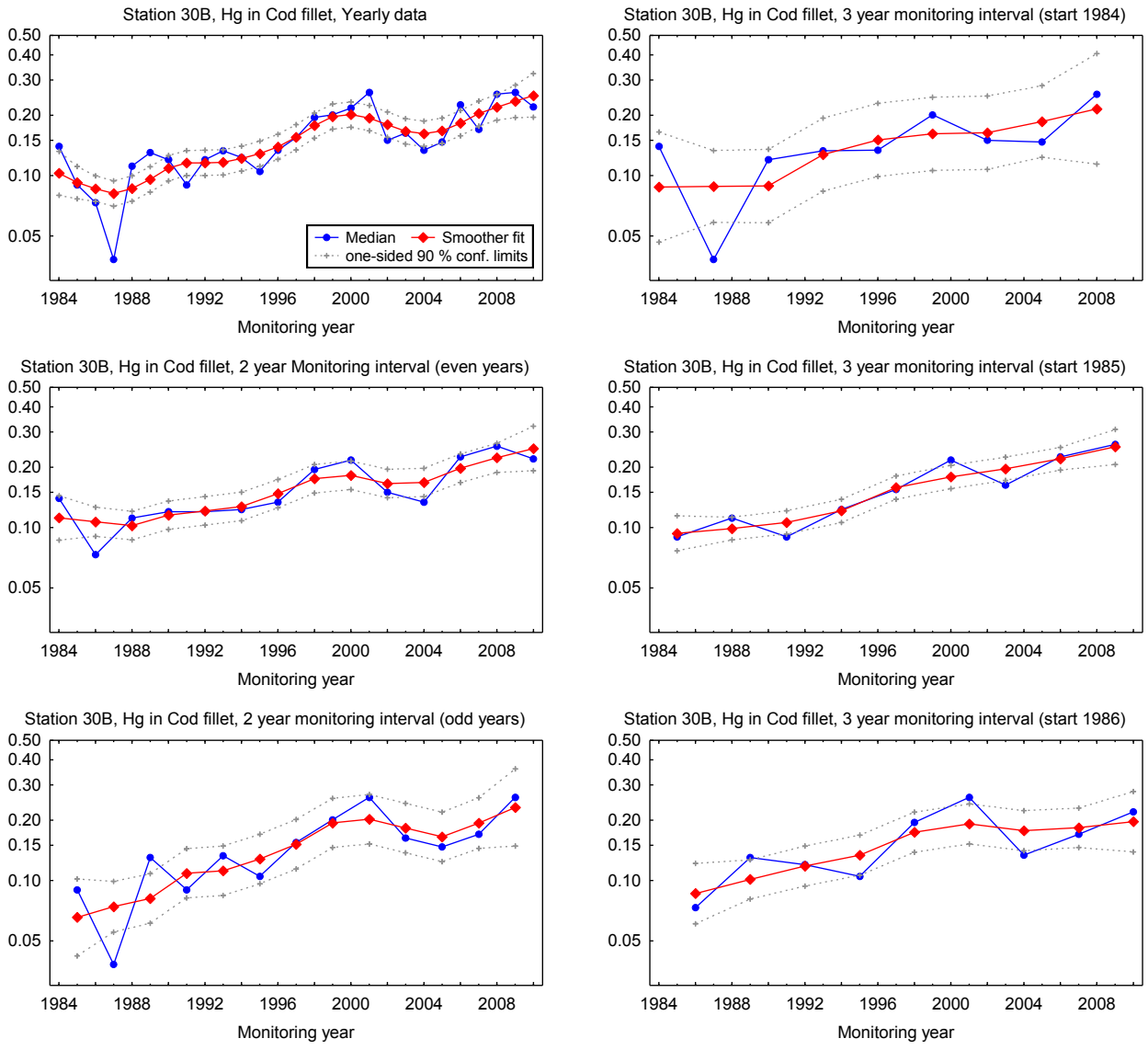


Figure 3. Data for mercury in cod liver, inner Oslofjord (Station 30B) as assessed with different sampling scenarios. For 2 or 3 year monitoring interval, the local regression window has been expanded to 9 and 13 years, respectively

Conclusions:

Monitoring with 2 or 3 year intervals will typically increase the minimum detectable trend for specified power and significance level by 50 to 100 %, compared to yearly monitoring.

For cod data, a large number of time series give minimum detectable trends[1] in the range from 0.05 to 0.1 on natural log scale (5 or 10 % change per year) with yearly monitoring. Monitoring with 2 or 3 years intervals would give detectable trends from 0.06 to 0.2.

For a majority of the existing mussel time series, the detectable trend is in the range from 0.05 to 0.2. With monitoring each 2nd or 3rd year this would instead be 0.1 to 0.4 (10 to 50 % change per year).

For many time series the trend estimates from monitoring at 2 or 3 year intervals deviate strongly from fairly clear trends estimated from yearly data, and the trends with 2 or 3 monitoring interval often show strong random variation depending on starting points, e.g. monitoring even or odd years.

For smoother trend testing based on data with 2 year intervals, the results are marginally better if the local regression window for the smoother is expanded from 7 years to 9 years. For monitoring interval 3 years, the results are much worse than with 2 year interval if the regression window is kept at 7 years; if the window is expanded to 13 years, the results are more comparable to 2 year interval for many time series.

It cannot be recommended to decrease monitoring frequency for time series where time trend detection is a concern, but it may be considered for stations where it is established that concentrations stays well below level of any concern, without any signs of upward trend over a number of years.

2.4 Choice of optimal trend test

In the current CEMP procedure, the trend for long data series is tested by the difference (contrast) between smoother fit for the last year and 10 year before. The adapted smooth function has a fairly large number of degrees of freedom, and each smoother fit is based predominantly on just a few years of the series. The pattern of the time series in the years close to the middle of the tested period does not affect the results much, so the test is more or less the same whether there is a gradual change or whether the values show large nonlinear fluctuations in between the two tested years. If there is a gradual change over time, for instance in the form of a log-linear regression, it would seem reasonable to consider that as confirmation that the difference between the endpoint years represents a real trend, compared to series where the value shows a non-linear, perhaps non-monotonic variation in intervening years.

Figure 4 shows the relation between significance levels for the smoother time trend test and the log-linear regression for series where there is no indication of a significant nonlinear deviation from log-linear regression, that is, where the p value for a nonlinear component in the regression is larger than 0.2. The plot shows that when the log-linear regression is significant at $p<0.05$ for such series, the smoother trend test is much weaker, often with p values 100 to 1000 times higher than the log-linear regression test. By using the log-linear result instead of the smoother trend

[1] With significance level 0.1 and power of 90 %. it was chosen to use a balanced risk of assessment error both ways. Reducing the significance level to the customary level of 0.05 would increase the minmum detecable trend by 10 to 20 %.

test if the nonlinear component is not significant, the ability to detect trends may increase considerably.

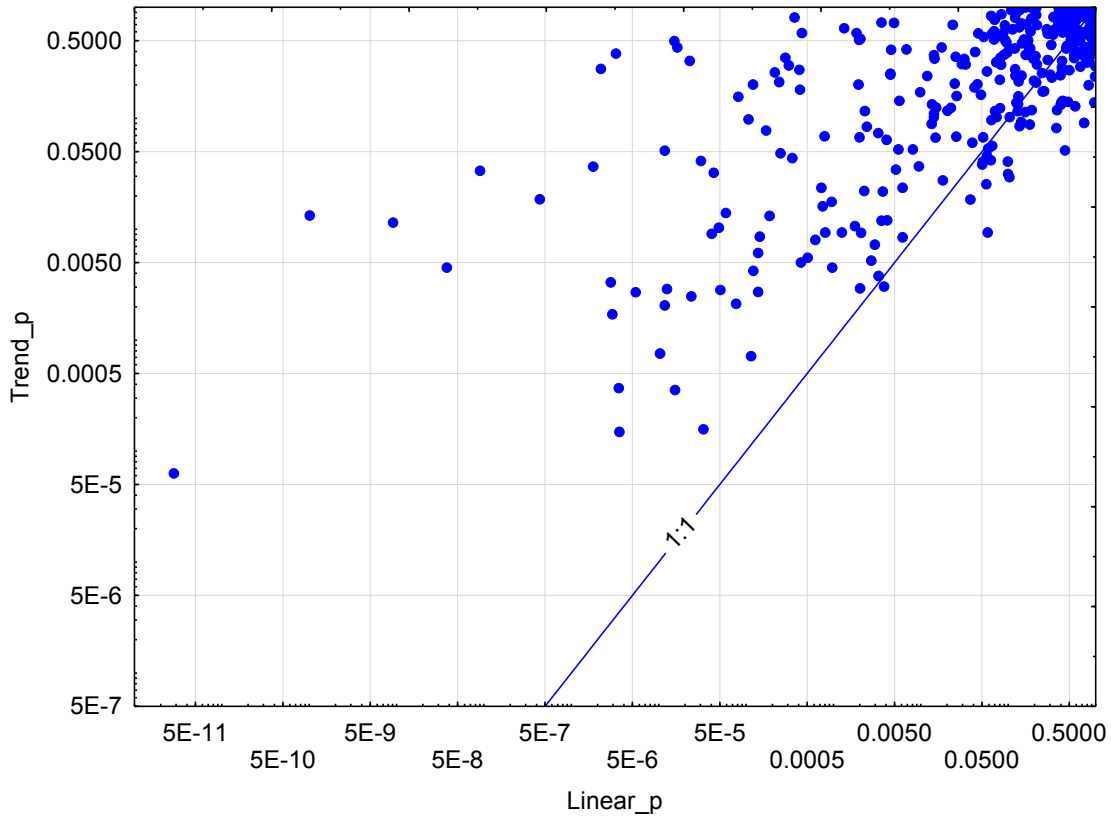


Figure 4. Relation between significance level (p values) for the smoother trend test and the linear regression trend test with smoother residual estimate. The plot only includes time series where the non-linear component is not significant (Nonlinear p>0.2). It is seen that when the log-linear regression is significant at p<0.05 for such series (points to the left of 0.05 on the horisontal axis), the smoother trend test (p value on vertical axis) is much weaker, often with p values 100 to 1000 times higher than for the log-linear regression test.

For future monitoring, with less occurrence of large variability due to on-going discharges, and more dominance of slow changes due to diffuse large-scale impact, a modification of the trend assessment procedures in line with should be considered. One possibility might be a step-wise procedure where a set of different types of trends are tested (e.g. log-linear, LOWESS smoother fit, sigmoid curve etc.). It should be kept in mind that the significance levels should be adjusted for this: testing different models increases the total probability of finding spurious trends above the assigned probability (significance level) of each test.

In the CEMP procedure, log-transformed yearly medians are used as averages. Arithmetic means of log-transformed individual values (geometric mean on linear scale) might be just as good or better. The mean estimates might possibly be made robust by excluding a certain fraction of extreme values.

3. Effects of reduced sample size or analysis of batch samples for cod liver data

A major part of current monitoring program consists of analysing about 25 individual cod livers from a number of stations each year. The time trend assessment for each station is based on the time series of yearly averages from these data. It is of interest to study how the ability to detect time trends will change if the number of individuals per year is reduced to 20 or 15 fish analysed individually, or alternatively, if the 25 livers are combined into 5 samples consisting of 5 livers each.

3.1 Statistical model – method of analysis

The log-transformed concentrations of a time series can be described by the additive statistical model

$$y_{t,s} = f(t) + \alpha_t + \varepsilon_{s(t)} \quad (1)$$

The total variation is a linear combination of a trend function $f(t)$ of time t and irregular variation between and within years:

α_t = between-year fluctuation of yearly population average around the smooth trend. These fluctuations are common to all individual samples, and have the same effect on the yearly averages independently of the number of individuals each year.

$\varepsilon_{s(t)}$ = within-year deviation of analysed samples $s=1,...,S$ (with index nested within year t) from the population average in year t. This deviation includes both real variation between samples and chemical analysis error. For analysis of individual livers, the index s denotes individual fish; for combination into batch samples, s will be an index on analysed subsamples. In both cases, S is the number of analysed samples.

Transformed back into the linear concentration scale, the model becomes multiplicative $\{\log(a) + \log(b) \Rightarrow a \cdot b\}$, describing concentrations as a trend function multiplied by two factors, the first varying irregularly between years, the other varying between samples within each year.

Analysis of variance (ANOVA) with year as random factor is used to estimate these two sources of variation, as between-year and within-year variance[1] components. These estimates are then used to assess how between-year variance changes with number of individual samples per year or combining them into batch samples for analysis. The statistical analysis is done using arithmetic means of log-transformed values as yearly averages. How this can be applied to yearly medians used in the CEMP procedure is discussed in connection with the presentation of the results.

The variation of yearly sample means $\bar{y}_t$ around the trend will be a combination of the between-year fluctuations α_t (real variation over time) and the residual effect of within-year variation averaged over subsamples (sampling error). If α_t and $\varepsilon_{s(t)}$ have standard deviations σ_T and σ_S respectively, and the number S of samples is the same each year, the total between-year variance relative to the trend is:

$$Var(\bar{y}_t - f(t)) = \sigma_T^2 + \frac{\sigma_S^2}{S} \quad (2)$$

[1] Variance is the average squared deviation from the mean, or the square of the standard deviation.

For a given length and frequency of a time series, the power to detect trends is a function of this between-year variance. The minimum trend that can be detected with a certain power when testing with a chosen significance level is proportional to the between-year standard deviation, which is the square root of the between-year variance. Details are given in Appendix C.

If individuals are analysed separately, $\sigma_S^2 = \sigma_i^2 + \sigma_a^2$ where σ_i^2 is the variance of the "true" individual concentrations as deviations from the true population average and σ_a^2 is the variance of chemical analysis error. The total variance of the yearly average is

$$Var(\bar{y}_t - f(t)) = \sigma_T^2 + \frac{\sigma_a^2}{S} + \frac{\sigma_i^2}{S} \quad (3)$$

If *S* individuals are combined into a smaller number *B* of batch samples each containing *N=S/B* individuals, the variance of yearly averages is:

$$Var(\bar{y}_t - f(t)) = \sigma_T^2 + \frac{N\sigma_B^2}{S} \quad (4)$$

where σ_B is the between-batch standard deviation, which will be lower than σ_S, but larger than $\sigma_S/\sqrt{N}$. The between-year variance for averages of *B* batch samples per year will therefore be larger than for averages of *S* individually analysed samples. Partly, this is because the effect of the statistical error of chemical analysis is larger for a single batch sample analysis than for the calculated average of *N* individual analyses. Additionally, individual contributions have a skewed distribution of concentrations on the linear scale, typically close to a lognormal distribution. When they are combined into batch samples, they do not contribute equally to the between-batch variance. Details are given in Appendix E. Using the results in the appendix, it can be estimated that σ_B is related to σ_i^2, σ_a^2 and *N* by the function

$$\sigma_B^2 = \sigma_a^2 + \frac{\sigma_i^2}{N^{1/(1+\sigma_i^2/7.7)}} \quad (5)$$

The between-year variance for batch samples then becomes

$$Var(\bar{y}_t - f(t)) = \sigma_T^2 + \frac{\sigma_a^2}{B} + N^{q/(1+q)} \cdot \frac{\sigma_i^2}{S} \qquad \text{where} \quad q = \sigma_i^2/7.7 \quad (6)$$

The ANOVA for estimating variance components is done on log-transformed and de-trended data for cod livers that have been chemically analysed individually. The de-trending is done by removing the smooth trend that was fitted by the 7-year CEMP LOWESS smoother for yearly data (Chapter 2.1.1), leaving only the irregular part of the between-year variation as residual deviations from the smooth trend in addition to the within-year variation between individuals. The ANOVA estimates the total between-year variance $V(y - f(t))$ and the within-year individual variance σ_S^2 directly, and calculates the between-year variance component σ_T^2 from equation (2). With these estimates, the effect of varying number of fish or combining fish in batch samples can be calculated with the relevant formulas given above.

The calculations can also be based directly on the mean square estimates from the ANOVA analysis. The numerical results are the same as when the variance is derived from the extracted components of variance as in equation (2), but the formulas based on the mean squares provide a basis for assessing confidence limits on the results. Appendix D describes this in detail for the case of individually analysed livers. This also includes how to estimate the uncertainty of estimated changes, in the form of confidence intervals for estimated absolute values and relative changes of between-year standard deviations. The intervals are estimated with a chosen confidence level, which is the probability of getting an interval that contains the true value, if assumptions about the form of distributions of random deviations are fulfilled. For the intervals estimated in this report, a 70 % confidence limit has been chosen, with equal error probability in

both directions. This means there is a 15 % probability that the lower limit will be above the true value, and a 15 % probability that the upper limit will be below the true value.

It should be noted that the smoother trend is fitted to yearly medians, while the variance analysis model is valid for yearly arithmetic means of log-transformed concentrations. The effect of within-year variance on yearly arithmetic mean, which is described by the equations above, may not be directly applicable to the residual effect on yearly medians. How this effect should be corrected to apply to median values is investigated empirically by numerical simulations, drawing repeated random samples of normally distributed test data with equal expectation and standard deviation. The simulation was done for sample sizes 5, 15, 20, 30, 50 and 100. For each sample size, the mean and median for each of 10,000 samples were calculated, and then the residual between-sample variance for means and medians. The final result is the ratio between variance of medians and variance of means for each sample size.

3.2 Selecting data for the analysis

The data are all from the CEMP database (Green et al. 2011). Values flagged as suspect values in the database are excluded. Of the remaining data, only series without missing years and without any observations below analytical limit are used. This is to ensure that such incomplete quantification does not affect results to a large degree, neither for between-year variance nor within-year variance[1]. The parameters with suitable time series for cod according to that criterion are Hg in fillet, and in liver CB101, CB105, CB118, CB138, CB153, CB156, CB180, the DDT metabolite p,p'-DDE and metals Cd, Cu and Zn. Table 2 shows number of years with data for each combination of station and parameter which has an accepted time series for the analysis in this chapter. Of the PCBs, CB153 covers the largest number of stations, but CB138 also has fairly good coverage. For some of the other PCBs, only a few of the stations have available data series according to the selection criterion.

Table 2. Total number years with data in cod time series used in the statistical analysis of within-year variation

	Parameters											
Stations	CB101	CB105	CB118	CB138	CB153	CB156	CB180	p,p'-DDE	Cd	Cu	Zn	Hg
10B			17	17	17			17	17	17	17	17
15B			17	17	17		17	17		17	17	
23B				17	17			17		17	17	17
30BA*	17	17	17	17	17	17	17			17		
36B	17	17	17	17	17		17			17	17	
53B			16	16	16		16	16		16	16	16
67B			17	17	17		17	17		17	17	
98BA*					17			17	17	17	17	17

*: Station 30 BA and 98BA are group names in the CEMP database, mainly including data from station 30B and 98B, respectively. The group names pools data from nearby locations that can be considered as sites within the same station area.

[1] Note that this is a stricter requirement than for the analysis in Chapter 2, where it was only required that yearly medians should be independent of any assumptions for such values. Even with this stricter requirement, the reduced numerical precision for low values close to analytical limit (e.g. values like 0.1 or 0.2), might still influence the result, but should not to have a large impact for series where no values are actually below the limit.

Figure 5 to Figure 8 shows de-trended log-transformed data as yearly arithmetic means and individual values for CB153, DDEPP and Cu in liver and Hg in fillet. For most years there are about 25 individuals. The variation between yearly means in the figures is due to the combined effect of two variance terms in the variance model as described above.

The de-trending procedure is fairly successful, and leaves data without long-term variations. There is considerable difference between stations concerning data variability; station 10B has relatively small log-scale variation both between years and between individuals within year, while in particular 53B shows high variability compared to the other stations.

Some of the time series plots show more or less clear outliers, indicating that it might be an advantage to use robust statistics to estimate time trends, also using robust estimates of variance. One particularly clear case of outliers is found in the time series for, CB153 at Station 53B (Figure 5), where four individuals in 2002 had much higher PCB levels than the other fish from the same years. The levels were 20 to 50 times higher than the largest concentration among the remaining 21 fish, and 80 to 200 higher than the average in the other 21 fish. By analysis of differences in PCB profiles it was established that the high concentrations in the four fish was related to removal of paint and plaster from an old power station the year before (Ruus et al. 2006). This had also affected mussel concentrations in nearby stations (see section 4.3.4).

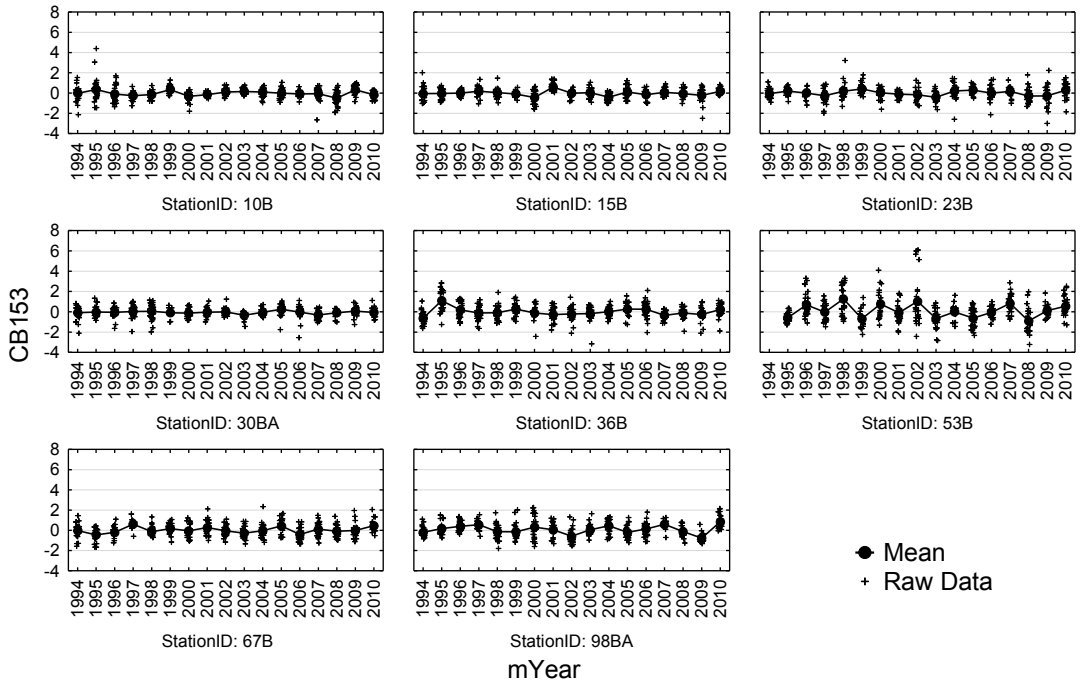


Figure 5. De-trended log-transformed data for selected time series of CB153 in cod liver. Vertical axis is on natural log scale. Yearly means are arithmetic averages of log-transformed values as deviations from fitted smooth trend.

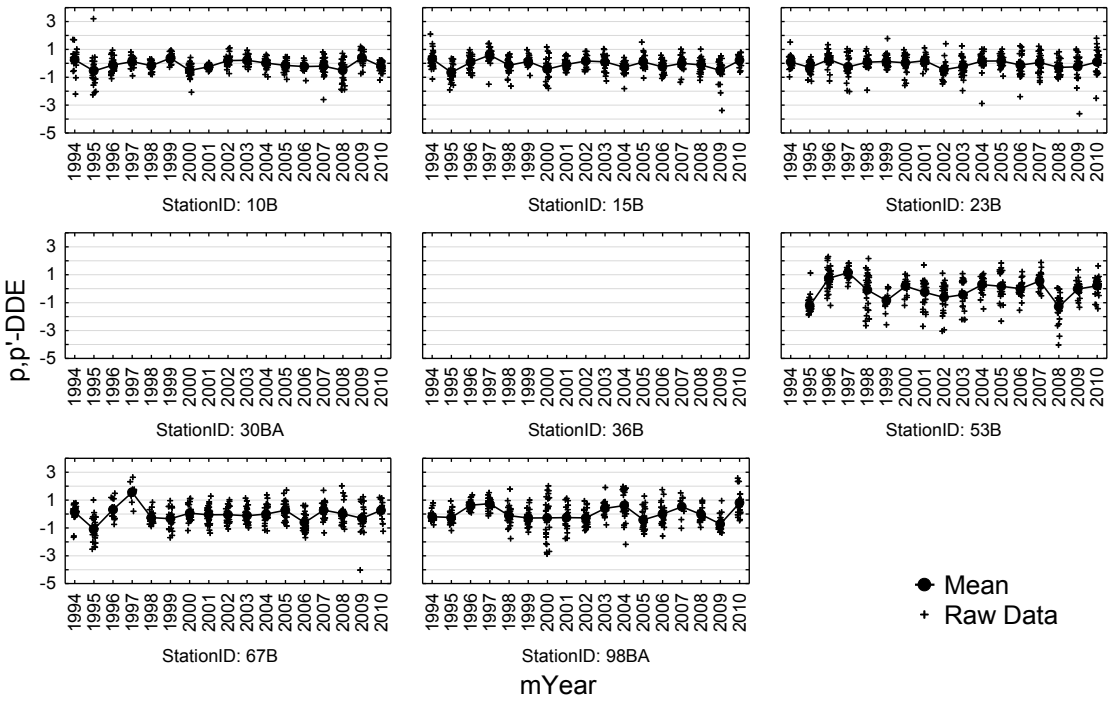


Figure 6. De-trended log-transformed data for selected time series of p,p'-DDE in cod liver. Vertical axis is on natural log scale. Yearly means are arithmetic averages of log-transformed values as deviations from fitted smooth trend.

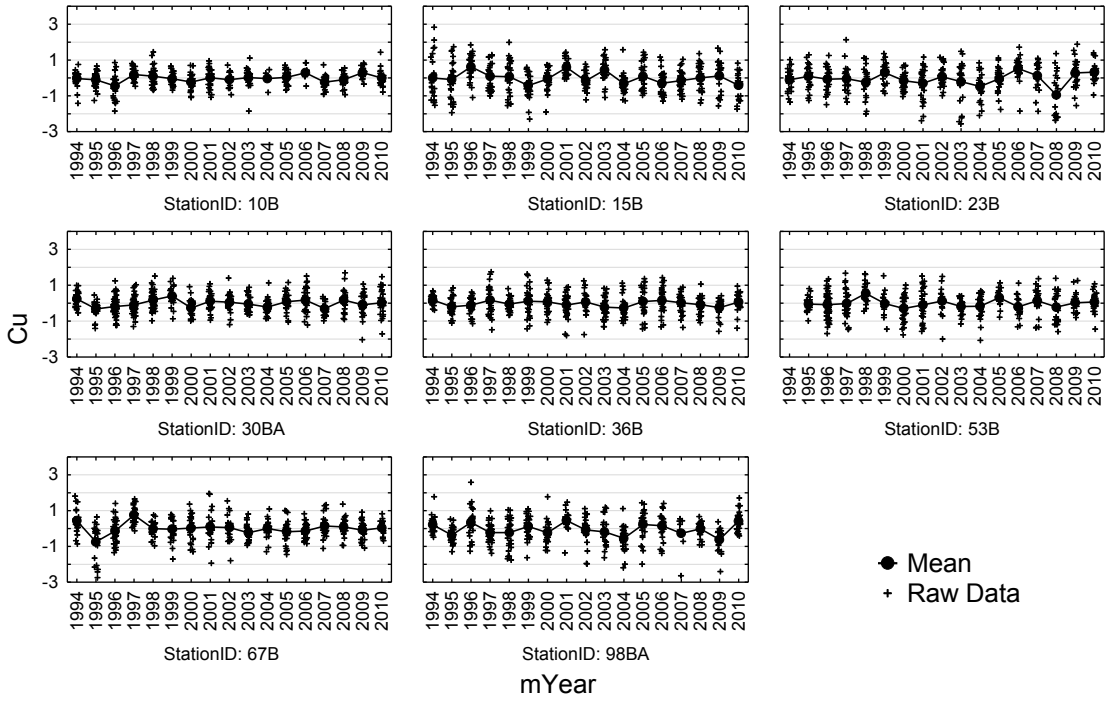


Figure 7. De-trended log-transformed data for selected time series of Cu in cod liver. Vertical axis is on natural log scale. Yearly means are arithmetic averages of log-transformed values as deviations from fitted smooth trend.

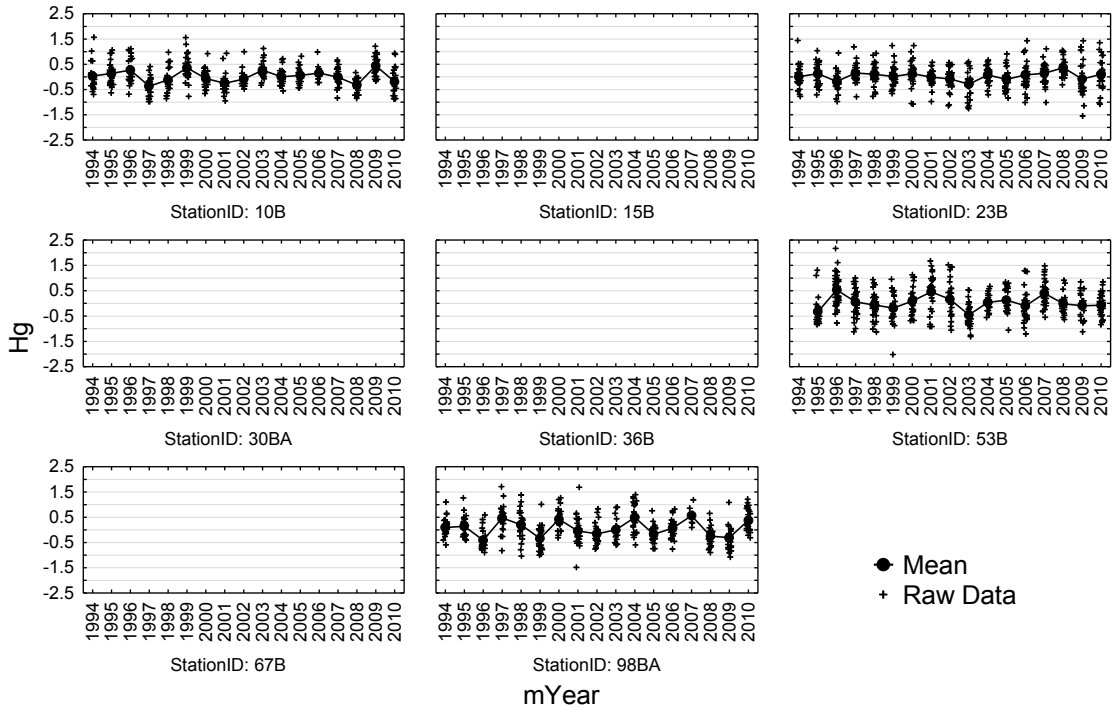


Figure 8. De-trended log-transformed data for selected time series of Hg in cod fillet. Vertical axis is on natural log scale. Yearly means are arithmetic averages of log-transformed values as deviations from fitted smooth trend.

3.3 Results of variance analysis

Variance analysis of the de-trended data is done separately for each station and parameter, with year as random factor. The parameters selected for analysis are CB138, CB153, P,P'-DDE, Cu and Hg. Table 3 shows the resulting estimates for variance components.

Table 3. Estimated variance components for monitoring of individual cod livers, per station and parameter.

Referring to description in Chapter 3.1, page 13 :

Between Years are estimates of variance component σ_T^2,

Within Years are estimates of variance component $\sigma_S^2 = \sigma_i^2 + \sigma_a^2$

		CB138	CB153	P,P'-DDE	Cd	Cu	Hg
10B	**Between years**	0.0394	0.0486	0.0812	0.0207	0.0199	0.0488
	Within years	0.3802	0.3439	0.3169	0.3990	0.2310	0.1722
15B	**Between years**	0.0439	0.0495	0.0844		0.0750	
	Within years	0.2471	0.2088	0.3169		0.6047	
23B	**Between years**	0.0320	0.0446	0.0305		0.0932	0.0133
	Within years	0.4256	0.4211	0.4502		0.5741	0.2357
30BA	**Between years**	0.0052	0.0070			0.0315	
	Within years	0.2633	0.2587			0.2918	
36B	**Between years**	0.1344	0.1043			0.0072	
	Within years	0.5809	0.5138			0.3836	
53B	**Between years**	0.5070	0.4327	0.4278		0.0294	0.0696
	Within years	1.5095	1.2750	0.7292		0.4660	0.3074
67B	**Between years**	0.0598	0.0626	0.1911		0.0769	
	Within years	0.4261	0.4120	0.5045		0.3945	
98BA	**Between years**		0.1643	0.1908	0.4956	0.0859	0.0854
	Within years		0.4853	0.5695	0.7446	0.4747	0.2039

As noted in chapter 3.1, the between-year variances estimated by ordinary variance analysis are valid for variance of arithmetic population means around smooth trends, while the trend in the CEMP procedure is assessed by using yearly medians. This is not quite consistent, but there is no reason to assume that the true population median varies more between years than the true mean, so we assume that the between-year variance of the population mean can also be applied to the true median.

Numerical simulation with randomly created samples of lognormal distributions shows the contribution from within-sample variance between individual observations on between-sample variance is about 40 to 50% larger for sample medians than for sample means[1]. On the other hand, the actual data have outliers, which do not influence the median in the same way as the arithmetic mean. For the present analysis, estimates for within-year variance components Table 3 have been increased by a factor 1.45 when used as contributions to total between-year variance for medians with varying number of fish or collection in batch subsamples for analysis. The correction for batch samples uses the within-sample variance from Table 3. For the batch analysis variance it has been assumed that the chemical analysis is accurate to 10 %, which means that variance σ_a^2 due to chemical analysis error on natural log scale is 0.01

The variance components in Table 3 are combined as described by equations (2) and (6) in Chapter 3.1 for different alternatives of sampling and analysis, to estimate between-year standard deviations for fluctuations around the trend for natural logarithms of the concentrations. Table 4 lists the results. The values listed are the same as those achieved by using the mean squares (*MS*) directly as described in Appendix D.3 Median estimates (with probability 50 % that the estimate will exceed the true value) will be about 3 % larger.

As an example of how the values in this table are created, consider the results for CB138, station 10B. The between-year variance is calculated for *S* individual fish as

$$Var\left(\bar{y}_t - f(t)\right) = 0.0394 + 1.45\frac{0.3802}{S}$$

giving variance 0.061, 0.067 and 0.076 for 25, 20 and 15 fish, respectively. The between-year standard deviations listed in Table 4 are the square root of the variances: 0.248, 0.258 and 0.276, respectively. For *B* batch samples of *N* fish each the between-year variance for the same parameter and station is

$$Var\left(\bar{y}_t - f(t)\right) = 0.0394 + \frac{0.01}{B} + N^{0.0459}\frac{1.45 \cdot (0.3802 - 0.01)}{B \cdot N}$$

with the exponent on *N* given by $q = (0.3802 - 0.01)/7.7 = 0.0481$;
$q/(1+q) = 0.0459$ this results in variance 0.065 and standard deviation 0.256.

The uncertainty of the estimates in Table 4 can be expressed as 70 % confidence intervals[2] according to Appendix D.3 . The confidence interval extends from about 0.85 to 1.3 times the estimated between-year standard deviations for 25 fish, and from about 0.85 to 1.25 times the

[1] The simulations were done for sample sizes 5,15,20,30,50 and 100, and with 10.000 samples created for each sample size.

[2] The confidence level is the probability of getting a confidence interval that contains the true value, if assumptions about the form of distributions of random deviations are fulfilled. For the intervals indicated here, there is a 15 % probability that the lower limit will be above the true value, and a 15 % probability that the upper limit will be below the true value. A confidence level of 70 % has been used because it is approximately the confidence levels for an interval extending ± one standard error around a normally distributed statistic.

estimated values for 15 fish. As an example, the estimated between-year standard deviation 0.248 for CB138 at 10B with 25 fish per year has a 70 % confidence interval going from 0.213 to 0.313.

Table 4. Between-year standard deviation of natural logarithms of concentration around trend for different alternatives for sampling and analysis of cod liver

Station	Sampling and analysis	CB138	CB153	DDEPP	Cd	Cu	Hg
10B	25	0.248	0.262	0.316	0.209	0.182	0.243
	20	0.259	0.271	0.323	0.223	0.191	0.248
	15	0.276	0.286	0.334	0.243	0.205	0.256
	5x5	0.256	0.269	0.321	0.219	0.190	0.248
15B	25	0.241	0.248	0.321		0.332	
	20	0.249	0.254	0.328		0.345	
	15	0.260	0.264	0.339		0.365	
	5x5	0.248	0.254	0.326		0.342	
23B	25	0.238	0.263	0.238		0.356	0.164
	20	0.251	0.274	0.251		0.367	0.174
	15	0.270	0.292	0.272		0.386	0.190
	5x5	0.247	0.271	0.248		0.364	0.173
30BA	25	0.143	0.148			0.220	
	20	0.156	0.160			0.230	
	15	0.175	0.179			0.244	
	5x5	0.153	0.158			0.228	
36B	25	0.410	0.366			0.172	
	20	0.420	0.376			0.187	
	15	0.437	0.392			0.210	
	5x5	0.418	0.374			0.183	
53B	25	0.771	0.712	0.686		0.238	0.296
	20	0.785	0.725	0.693		0.251	0.303
	15	0.808	0.746	0.706		0.273	0.315
	5x5	0.789	0.727	0.692		0.248	0.301
67B	25	0.291	0.294	0.469		0.316	
	20	0.301	0.304	0.477		0.325	
	15	0.318	0.320	0.490		0.339	
	5x5	0.298	0.301	0.475		0.322	
98BA	25		0.439	0.473	0.734	0.337	0.312
	20		0.447	0.482	0.741	0.347	0.316
	15		0.460	0.496	0.753	0.363	0.324
	5x5		0.444	0.480	0.740	0.344	0.316

The between-year total standard deviations in Table 4 for 25 fish per year are mostly in the range 0.15 to 0.3, but up to 0.7-0.8 for PCBs and p,p'-DDE at station 53B. From the results in Appendix C.2 we see that if the data fits a log-linear regression over 10 years, a trend 0.41 times the between-year standard deviations can be detected with 90 % power and significance level 0.05. This can be applied to the standard deviations in Table 4.

The minimum values of trends that are detectable under these conditions vary from 0.06 for PCBs at Station 30BA[1] to 0.3 for PCBs at 53B. The minimum detectable trends are estimated with the same relative uncertainty as the between-year standard deviations (within -14 to +30 % for 25 fish, -13 to +26 % for 15 fish), since there is a direct proportionality (Appendix A.5 and Appendix B). For trend tests based on differences between smoother fits up to 10 years apart, the required minimum detectable difference, when converted to average change per year, will be somewhat larger than the value for log-linear regression over 10 years.

Table 5 shows relative change in total between-year standard deviation on log scale (square root of variance) when changing from 25 fish per year to 20 or 15 individuals, and when analysing 25 fish in 5 batch samples instead of individually. The calculated changes in Table 5 are based on the between-year standard deviations in in Table 4 and the calculation is done as described in Appendix D.4 .

Table 5. Estimated relative increase of between-year standard deviation of natural logarithms of cod liver concentrations around trend for reduced number of fish per year or analysis in batch samples (5x5).

Station	Sampling and analysis	% increase compared to 25 fish per year						
		CB138	CB153	DDEPP	Cd	Cu	Hg	Average
10B	20	4.4 %	3.6 %	2.3 %	6.4 %	4.9 %	2.1 %	3.9 %
	15	11.3 %	9.2 %	5.9 %	16.2 %	12.6 %	5.5 %	10.1 %
	5x5	3.2 %	2.7 %	1.7 %	4.6 %	4.3 %	2.2 %	3.1 %
15B	20	3.0 %	2.4 %	2.2 %		3.9 %		2.9 %
	15	7.8 %	6.3 %	5.8 %		10.1 %		7.5 %
	5x5	2.5 %	2.3 %	1.7 %		2.9 %		2.4 %
23B	20	5.3 %	4.3 %	5.6 %		3.2 %	6.1 %	4.9 %
	15	13.6 %	11.1 %	14.3 %		8.4 %	15.6 %	12.6 %
	5x5	3.8 %	3.1 %	4.0 %		2.4 %	5.3 %	3.7 %
30BA	20	8.9 %	8.2 %			4.3 %		7.1 %
	15	22.4 %	20.6 %			11.0 %		18.0 %
	5x5	7.4 %	6.8 %			3.3 %		5.8 %
36B	20	2.5 %	2.7 %			9.0 %		4.7 %
	15	6.4 %	7.1 %			22.6 %		12.1 %
	5x5	1.8 %	2.0 %			6.6 %		3.5 %
53B	20	1.8 %	1.8 %	1.1 %		5.8 %	2.5 %	2.6 %
	15	4.8 %	4.7 %	2.9 %		14.8 %	6.5 %	6.8 %
	5x5	2.4 %	2.1 %	0.9 %		4.2 %	1.9 %	2.3 %
67B	20	3.6 %	3.4 %	1.6 %		2.8 %		2.9 %
	15	9.3 %	8.8 %	4.3 %		7.3 %		7.4 %
	5x5	2.6 %	2.4 %	1.2 %		2.0 %		2.1 %
98BA	20		1.8 %	1.8 %	1.0 %	3.0 %	1.5 %	1.8 %
	15		4.7 %	4.8 %	2.6 %	7.7 %	4.0 %	4.8 %
	5x5		1.3 %	1.4 %	0.8 %	2.2 %	1.4 %	1.4 %

According to these estimates, reduction from 25 to 20 individual livers per year would increase the between-year standard deviation with 2-7 %. Reducing to 15 fish per year would increase the between-year standard deviation with 3-23 %.

[1] CEMP station group name, ref Table 2.

The minimum size of change per year that could be detected with given power and significance would increase proportionally with the between-year standard deviation, for the same length and frequency of a time series.

The estimated effect of reducing sampling and analysis effort varies between different parameters and stations. For station 67B and 98B the effects are fairly small for all parameters. The largest relative changes occur for PCBs at station 30BA and Cu at station 36B.

For PCBs at station 30BA the minimum detectable trend for a linear regression test would increase from about 0.06 to 0.072 (by about 20 %) if fish sample size were reduced from 25 to 15 fish. At station 53B it would increase with 5 % from 0.3 to about 0.315. As another example, for Cu at station 36B the minimum detectable trend would increase from about 0.07 to 0.086. All these changes refer to differences on natural log scale and are approximately equal to relative change per year in concentrations.

Sampling 25 individuals, but analysing only 5 batch samples of 5 fish each, is estimated to increase the standard deviation by 1.4 to 5.8 %, a little better than 20 individual fish. Compared to the alternatives with reduced number of individuals analysed, the batch sampling option means a much larger saving in analysis costs.

However, individual analyses has the advantage that individual outliers may be excluded more efficiently by using yearly median, as is currently done in the CEMP procedure, or a robust arithmetic mean that excludes the outliers. Chemical analyses of individuals instead of a few batch samples also give better possibilities for correcting for biological characteristics, such as fat content, which may increase the ability to detect trends. Such factors might mean that batch sample analysis somewhat less advantageous than estimated here. The outliers for PCBs (CB153) at station 53B in 2002 is an example of a time series where this is apparent (see page 16); without analyses of individual fish, these four fish would dominate completely the average concentration in batch samples they were included in, and this would reduce the possibility of detecting or excluding outliers. With individual analyses the outliers are easy to identify, and their effect may be eliminated from trend analysis by using robust statistics.

The uncertainty of the estimated changes relative to the present data set has been analysed as outlined in Appendix D.4.2. The results are not presented in detail, but they show that the estimated relative changes of between-year standard deviations have 70 % confidence intervals typically extending from 0.65 to 1.4 times the estimates in Table 5. So for instance the 11.4 % increase of between-year standard deviation that is estimated for CB138 at station 10B has a 70 % confidence interval from 7.3 to 15.4 %. The relative uncertainty is about the same for all stations and parameters.

Conclusion:

The current monitoring of 25 individual cod livers per year on each station is estimated to make it possible to detect log-linear trends in the range 6 to 35 % change per year (0.06 to 0.3 on natural log scale) with power 90 % when testing over 10 years with significance level 0.05. The detectable trend varies between stations and parameters and is for the majority of time series below 15 %. Decreasing the sample size to 20 fish would increase the minimum detectable trend with only 2-7 %, which means that a minimum detectable trend of 10% change per year with yearly sampling would increase to a maximum of 11% change per year when decreasing the sample size from 25 to 20 fish. A reduction to 15 fish would increase the minimum detectable trend by 3 to 22 %, less than 10 % for most stations and parameters.

Analysing 5 batch samples of 5 fish each would give even less reduction of trend detection ability than reduction to 20 fish. However, individual analyses has the advantage that outliers may be excluded more efficiently by robust statistics, and may also give better possibilities for correcting for biological characteristics, such as fat content.

The estimated relative changes are not precise assessments, but have a statistical uncertainty, expressed as 70 % confidence intervals. The confidence intervals for between-year standard deviation and required trend for detection are within 0.85 to 1.3 times the estimated values, and for relative change of detectable trend on reduction of number of fish the confidence intervals generally extend from 0.65 to 1.4 times the estimates.

4. Repeated sampling within year and area – analysis of mussel data for Grenland and Sørfjord.

The effect of repeated or distributed sampling of blue mussels within station and year is assessed by statistical analysis of subsets of the mussel data in the Norwegian CEMP program, supplemented by data from local or regional monitoring programs. The subsets consist of data from years and localities where mussel sampling has been repeated within the same season within a 1-2 month interval, or where samples are taken from different sites at the same location. There are typically 2 to 4 such main samples each year. Some or all of the main samples are divided into 2-3 subsamples for chemical analysis.

The variance of aggregated yearly averages around the time trend will depend both on the real between-year fluctuations and on the within-year sampling error, which in turn depends on the number of main samples and on how they are divided into subsamples for chemical analysis. The goal of the analysis is to estimate how the between-year variance around long-term trends of the yearly averages will change if the current monitoring program is reduced to include fewer samples per year, including the option of just one analysed sample per year.

4.1 Statistical model - method of analysis

The log-transformed concentrations for each time series are described statistically by the additive model

$$y_{t,s,r} = f(t) + \alpha_t + \beta_{s(t)} + \varepsilon_{r(t,s)} \tag{7}$$

with a trend function $f(t)$ of time t (year) and an irregular variation composed of:

α_t = between-year fluctuations of yearly population averages around the trend, different from year to year, but common to all samples from the same year.

$\beta_{s(t)}$ = within-year variation between true sampling populations for different sampling sites or sampling times (main sample s=1,...,S) as deviations from the overall yearly population average. This term represents spatial patchiness or short-term irregular fluctuations in time.

$\varepsilon_{r(t,s)}$ = within-year deviation of subsamples or replicates (r=1,...,R, with index nested within year t and main sample s) from the population average of the main sampling population. This deviation includes both real variation between subsamples and chemical analysis error.

As in chapter 3, the variation of yearly sample means $\bar{y}_t$ around the trend will be a combination of the real between-year variation and the residual effect of within-year variation averaged over subsamples, but now the within-year variance structure has two levels, main samples and sub-samples. If α_t , $\beta_{m(t)}$ and $\varepsilon_{s(t,m)}$ have standard deviations σ_T , σ_S and σ respectively, and the program is regular and balanced with S main samples and R replicate subsamples per main sample, the total between-year variance relative to the trend is:

$$Var(\bar{y}_t - f(t)) = \sigma_T^2 + \frac{\sigma_S^2}{S} + \frac{\sigma^2}{S \cdot R} \tag{8}$$

For the actual monitoring programs this is not the case, and the equation is then replaced by a more general form which can be written

$$Var(\bar{y}_t - f(t)) = \sigma_T^2 + \frac{k_2}{k_1}\sigma_S^2 + \frac{1}{k_1}\sigma^2 \tag{9}$$

The residual between-year variance given by these equations determines the power to detect trends for a given length and frequency of the aggregate time series as described in Appendix C.

The data are analysed with an ANOVA model with factors *Year* and *Main Sample*, with the last factor nested within *Year*. The subsamples are replicates within main sample. Both factors are defined as random, to get variance component estimates for each source of variation.

Monitoring stations are included in different ways in different data sets.

- In the Grenland data, stations are geographically close to each other, and in the statistical analysis considered to represent different sampling sites for the same location. Consequently, station is included as main sample, with multiple samples within station as sub-samples. The between-year variation estimated for the analysis is for the average over stations.
- In data from Sørfjord, stations are further apart, and have partly different time variation patterns. Some of the stations have a monitoring structure of main samples, some of which are analysed in a small number of replicate samples. The main samples here are the different monitoring programs: national CEMP and regional monitoring, and with replicate samples in particular for the CEMP monitoring. In the analysis of mussel data from Sørfjord, station is included as an additional factor to the design presented above. The analysis is done by nesting year within station, so time in reality means interaction Time*Station. It is the same as analysing each station separately, but pooling the within-year variance for all stations. The irregular between-year variance for a time-series is estimated by a special post-processing procedure as described below.

The analysis is done with the GLM model of Statistica version 10. The coefficients k_1 and k_2 are given from the analysis as the Expected Mean Square Coefficients. The mean square for between-year variation from the ANOVA analysis will be a multiple k_1 of the variance in equation (9):

$$EMS_A = k_1\sigma_T^2 + k_2\sigma_S^2 + \sigma^2 \qquad (10)$$

In equation (9) it is indicated that the long-term trend is removed before estimating variance terms. If this is not done the between-year variance estimate from the ANOVA model includes both trend and irregular fluctuations, and cannot be used directly for assessing the relative importance of within-year variance for trend detection ability. In the cod data analysed in Chapter 3 the trend was removed by de-trending data with the CEMP smoother fit procedure, and the de-trended data was analysed.

In the mussel data, we cannot rely on using the CEMP smoother fits to remove trend. Partly, we are combining data from different CEMP stations located close to each other, seen as sites within a location, or as different main samples. We also include data from regional programs in addition to CEMP data. Finally, there are some major and sudden events in the data; sharp drops or increases from one year to another and in some instances short-term but very large peaks, which must be assumed to be part of the signals one wants to detect, but would turn out as part of noise in the 7-year smoother of the CEMP procedure.

Extracting the noise component of between-year variance is therefore done by post-processing of the ANOVA results instead of doing it before the analysis.

First, a smoother is fitted by unweighted local regression (LOESS) with a 3-year window, and differences between yearly means and the fitted smoother values are adjusted to have expected variance equal to residual variance of data in a series with a linear trend. This simply means calculating adjusted deviations

$$\delta_t = \left(2y_t - y_{t-1} - y_{t+1}\right)/\sqrt{6} = 3\left(y_t - \bar{y}_{t-1,t,t+1}\right)/\sqrt{6} \qquad (11)$$

For data series where the *y* values follow a linear regression $y_t = a + b \cdot t + \varepsilon_i$ (including no trend where *b*=0) the set of deviations δ will have the same expected variance as the ε deviations of the linear regression, and be approximately normally distributed when the analysis is done on log-transformed values. For data with nonlinear trends and large sudden changes of levels or trend directions, the set of δ values will probably have variance much closer to the irregular variation of random character than the total between-year variance.

Any sudden changes or breakpoints in the time series that give large δ values can be identified as outliers and excluded. For this purpose, the deviations δ are analysed with robust statistics. No deviation is calculated for first and last year or for years just before or after gaps in the series, as indicated by the formula; only years *t* with dtaa for all three years *t*-1, *t* and *t*+1 are used. The set of δ values are ordered and assigned corresponding fractile values for the standard normal distribution. Linear regression of the δ values against the standard normal fractiles will estimate the between-year standard deviation as the regression slope. If there are outliers, the regression can be restricted to use only the central part of the ordered set of deviations (say, within a range ± 1 for the fractiles of the standard normal distribution), so that the estimated standard deviation is not influenced by outliers. If this regression is not forced through zero, it will also remove the effect of any constant nonlinearity in the time trend of the original data on the residual standard deviation estimate. The square of the between-year standard deviation from the regression is an estimate for $Var(\bar{y}_t - f(t))$ defined above. It can be used to calculate corrected Mean Squares for the Year effect and variance components for estimating how design changes will affect ability to detect trends, using the expected mean square coefficients from the ANOVA, as described in detail in Appendix D. As for cod livers, confidence intervals are calculated using the formulas in the Appendix, 70 % probability of getting intervals that contain the true value; 15 % probability that the lower limit will be above the true value, and a 15 % probability that the upper limit will be below the true value.

The analysis of statistical uncertainty of arithmetic means (of log-transformed values) is not directly applicable to uncertainty of the yearly medians which are used in the CEMP procedure. Numerical simulation with random draws of a large number of small samples from a normal distributions shows that the standard deviation of the median does not decrease with increasing sample size as effectively as for the arithmetic mean; the analysis on arithmetic means might overestimate the effect of reducing from 3-4 samples to a single sample per year if the trend test is done on yearly medians. Occurrence of outliers might modify this, since the use of medians protects better against outliers than an arithmetic mean.

4.2 Selection of data for analysis

4.2.1 Data from the Grenland area

Blue mussel data from this area have been collected since 1983. The data selected for the analysis her are from four CEMP stations close to each other in Langesundsfjord, shown in the map in figure 9: 71A Bjørkøya (Risøyodden), I711 Steinholmen, I712 Gjemesholmen and I713 Strømtangen. For details, see monitoring yearly reports for 2010 (Green et al. 2011and Ruus et al. 2011a) which also contain references to previous reports.

Bjørkøy/Risodden has the longest time series, with monitoring starting in 1983. Monitoring at CEMP stations Steinholmen and Gjemesholmen started in 1995. Steinholmen has not been sampled since 2002, while monitoring at Strømtangen started that year. The monitoring covers metals, pesticides, organotin, dioxins, PCBs and other organochlorines.What contaminants are included varies between stations and over time. Gjemesholmen is also included in the regional monitoring of dioxins in mussels (under the name Croftholmen); this monitoring started in 1989.

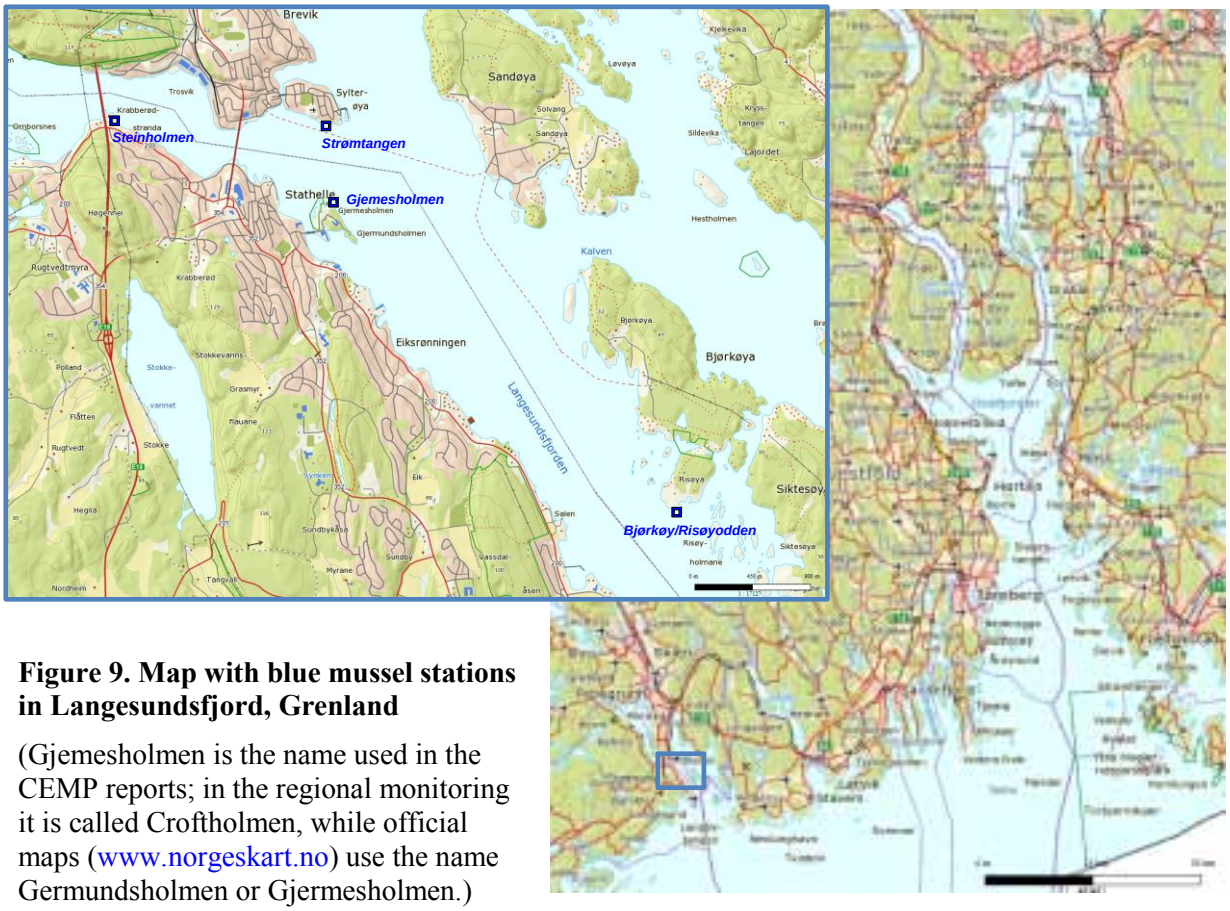


Figure 9. Map with blue mussel stations in Langesundsfjord, Grenland

(Gjemesholmen is the name used in the CEMP reports; in the regional monitoring it is called Croftholmen, while official maps (www.norgeskart.no) use the name Germundsholmen or Gjermesholmen.)

Dioxins

For dioxins from Grenland, only data from 2002-2010 have been included in the statistical analysis here. The reason is that before 2002 discharges were still going on; variability in time and space from that time may be influenced by fluctuations in discharges and are considered not to be representative of the present and future variability. In addition, it is mostly since 2002 that more than one mussel sample have been analysed regularly from some of the stations. For the selected dioxin data set, observations reported as < analytical limits occur in just a few instances, and has very little impact on sums of toxic equivalents.

There are CEMP data from Strømtangen and Gjemesholmen and Bjørkøy/Risøyodden. From Strømtangen only one sample is analysed for dioxins per year; from the other two stations there are two subsamples analysed each year. From Gjemesholmen (=Croftholmen) there are additional data from the regional monitoring; with one sample per year, most years collected 1-4 weeks later in the autumn.

Mercury, Organochlorines, pesticides

In order to get a complete dataset for ANOVA models, and to get estimates of variance that are representative of the current monitoring program, only data from 2002-2010 are used, and the data from Steinholmen in 2002 are not included.

The components suitable for statistical analysis are Hg, CB118, CB138, CB153, HCB and p,p'-DDE (ICES code DDEPP). Other contaminants have a large fraction of observations below reporting limit.

HCB is included in the statistical analysis, although there is one observation from 2003 reported as below analytical limit for station 71A. After conversion to dryweight concentration, the values

are 0.327, <0.32 and 0.512 µg/kg dw. It has been chosen to include the two well-defined values and exclude the observation <analytical limit, to get a complete dataset.

Metals Cd, Cu, Pb and Zn

For cadmium, copper, lead and zinc, station 71A Risøya/Risøyodden has long time series with 3 subsamples per year; the other stations only have data from 2009 and 2010. Analysis of data over all three stations from just the last two years would give estimates of station*year interaction, but with so few degrees of freedom that the result would be very unreliable. The analysis is therefore done only for station 71A.

4.2.2 Mussel data from Sørfjord, Hardanger

Mussel data from Sørfjord has been collected through the national CEMP program, supplemented by a specific monitoring program. The stations in the two programs are shown in maps in Figure 10, and Table 6 summarises the station identification used in the two programs. For details, see Green et al. (2011) and Ruus et al. (2011b)

Table 6. List of mussel stations in Sørfjord, Hardanger, with correspondence between station codes used in the regional and national monitoring programs.

In the Regional monitoring (Ruus et al. 2011b)	In the CEMP database (Green et al. 2011)	Identification used in statistical analysis	Distance from Odda (km)
B1	51A	51AA	2
B2	52A	52A	3
B3	B3x		6
B4	B4		10
Måge			15
B6	56A	56A	18
B7	57A	57AA	37
Utne			40

The locations with multiple samples over a number of years are 51AA (=51A), 52A, 56A and 57AA(=57A), and the statistical analysis is restricted to these four stations. Station 51AA and 57AA are group names in the CEMP database. The group names pools data from nearby locations that can be considered as sites within the same station area.

From 1996 at station 51AA, and from 1994 at the other three stations, the CEMP program has included chemical analysis of three replicate samples in September each year, while the regional monitoring has collected one additional sample in November. The data are analysed as coming from two main samples each year (CEMP and regional program), with 3 subsamples and only one sample, respectively. In the CEMP database, the main samples are identified with different Sequence Numbers (Seqno), while the replicates are identified with sub number (Subno) 1, 2 and 3.

Additional locations (56A1-56A5, 57A1, 57A2) were sampled in 1999 and 2002. For the analysis here, data from 56A1 is included as part of the 56A data, and 57A1 with the 57AA data, as the locations are quite close to the respective main stations. Data from the other additional locations in 1999 and 2002 are not used, as they are considered too far away to represent the same stations. In 1999 there was also an additional main sample from a 51A, analysed in three replicate samples. In 2007 the regional program included three replicate samples from stations B1 (=51A) and B2 (=52A).

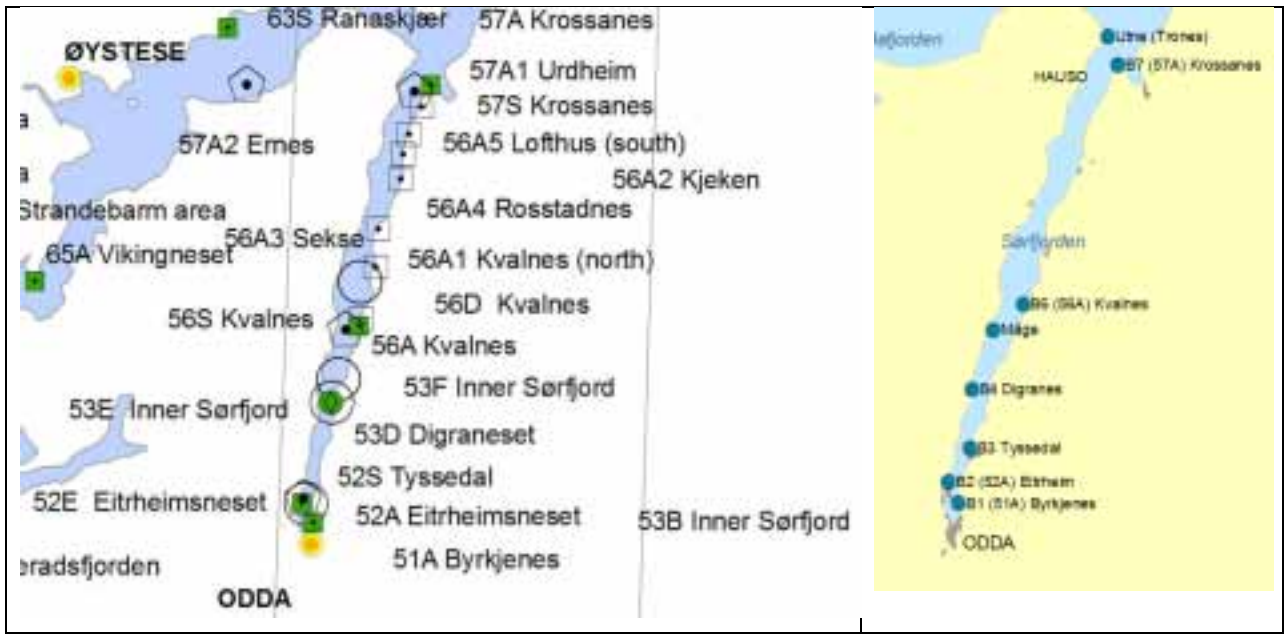


Figure 10. Excerpts from Station maps in relevant monitoring reports (Green at al. 2011, Ruus at al. 2001b)

In this dataset, a number of parameters are not suited for statistical analysis, either because levels during the last decade are low compared to the analytical limit (HCB, OCS, QCB, γ-HCH, α-HCH), or because there is too little data (As, Cr).

The parameters that may be suitable for analysing the gain of multiple sampling per year are metals (Cd, Cu, Hg, Pb and Zn), some of the PCB components (CB118, CB138 and CB153), pp'-DDD and pp'-DDE. Of these parameters, the most interesting ones may be those where classification above class I occurs, they are pp'-DDE, Pb, Cd, Hg. The recent levels of PCBs are all in class I, but PCB is still included in the analysis since the results for the between-year and within-year variance structure can be of interest generally.

Even for these contaminants, some values have to be excluded from the data set used in the statistical analysis, to avoid results that are influenced by the occurrence of observations reported as below analytical limit.

- All PCB values before 1994 are excluded, because most of the values have very low numerical precision in the database (only one significant digit in the range 0.1 to 0.3, i.e. with numerical error of up to 20-50 %), and some observations also are below analytical limit.
- All values are removed for any combination of station, year and component where one or more observations are below analytical limit. After a preliminary selection of contaminants, this is the case for
    - 1999: pp'- DDD data from station 51A
    - 2002: CB118 and CB153 from 56A.
    - 2003: CB118 from 56A
    - 2006: CB153 from 56A
    - 2010: CB118 from 56A, 57A and Utne, CB138 from 57A.

The selection of data for statistical analysis is further based on whether there is sufficient data to separate variance between years, between main samples and between subsamples (or replicates). Years with less than three values are excluded, considered separately for each station and analysed parameter.

The final selection of data is as follows:

- All 5 metals are included. The analysis is based on data from 1996-2010 at station 51A, from 1989 to 2010 except 1993 for 52A (only a single sample in 1989), and for 1987-2010 for 56A and 57A. For copper and zinc, station 51AA is excluded from the analysis, since there are too many years with only one value from the regional monitoring and no values for the CEMP samples.
- For PCB, the component CB138 has the best coverage, and is selected among the PCB components for statistical analysis. The selected data with at least 3 samples from each station and year are 1996-2010 at station 51A, 1994-2010 except 1996 at station 52A and 56A, and 1994-2009 except 1996 at station 57A.
- For pp'- DDE the years with more than two samples are 1996-2010 for station 51A, 1994-2010 for 52A, and 1992-2010 for 56A and 57A.

4.3 Results

4.3.1 Grenland dioxin data in Grenland mussels

Figure 11 shows data included in the statistical analysis, as separate time series for different stations and sampling programs and with both single values and yearly arithmetic means of log-transformed values.

As mentioned in chapter 4.2.1, most of the data are from the National CEMP monitoring, but there is one supplementing set of observations at Croftholmen from the regional Grenland monitoring program, usually sampled 1-4 weeks later than the national program. The plots show three different measures of dioxin: Toxic Equivalent (TE), the sum of 12 congeners that account for about 90 % of the total sum (all congeners without observations <analytical limit), and the single component 2378-TCDF which represent on average about 14 % of TE and 22 % of the total concentration in these data. The TE plot indicates visually slightly increasing values in the last three years. The unweighted sum has larger variations between time series than TE or 2378-TCDF.

From CEMP stations Croftholmen and Bjørkøy/Risøyodden there are data for two subsamples per year. The figures show that the variation between the two subsamples is usually small compared to the variation between stations and over years. The exception is the very low value from Croftholmen in 2005, which stands out as a clear outlier. It seems reasonable to exclude this value from the statistical analysis as untypical and possibly erroneous.

The Croftholmen data from the regional monitoring differs from the CEMP data from the same location (Gjemesholmen), and the difference is generally much larger than the difference between the two CEMP subsamples, and also with systematically lower values, if the low outlier in the CEMP data from 2005 is ignored. It is therefore reasonable to treat the regional monitoring as a separate time series, and not pool the CEMP and regional data as subsamples from the same station.

The data are analysed in an ANOVA model with Year and Station as main factors. Year has 9 levels (2002-2010) and Station has the 4 levels shown in Figure 11; the two sampling series from Croftholmen are treated as separate stations. Year is defined as a random factor, to get variance component estimates for variation between years, between different main samples (stations) within year and for subsamples within each main sample. Using this model means that we combine the data series into a common average time trend in the form of yearly means across stations, and looks at the effects of reduced number of stations and number of replicates in assessing an average trend.

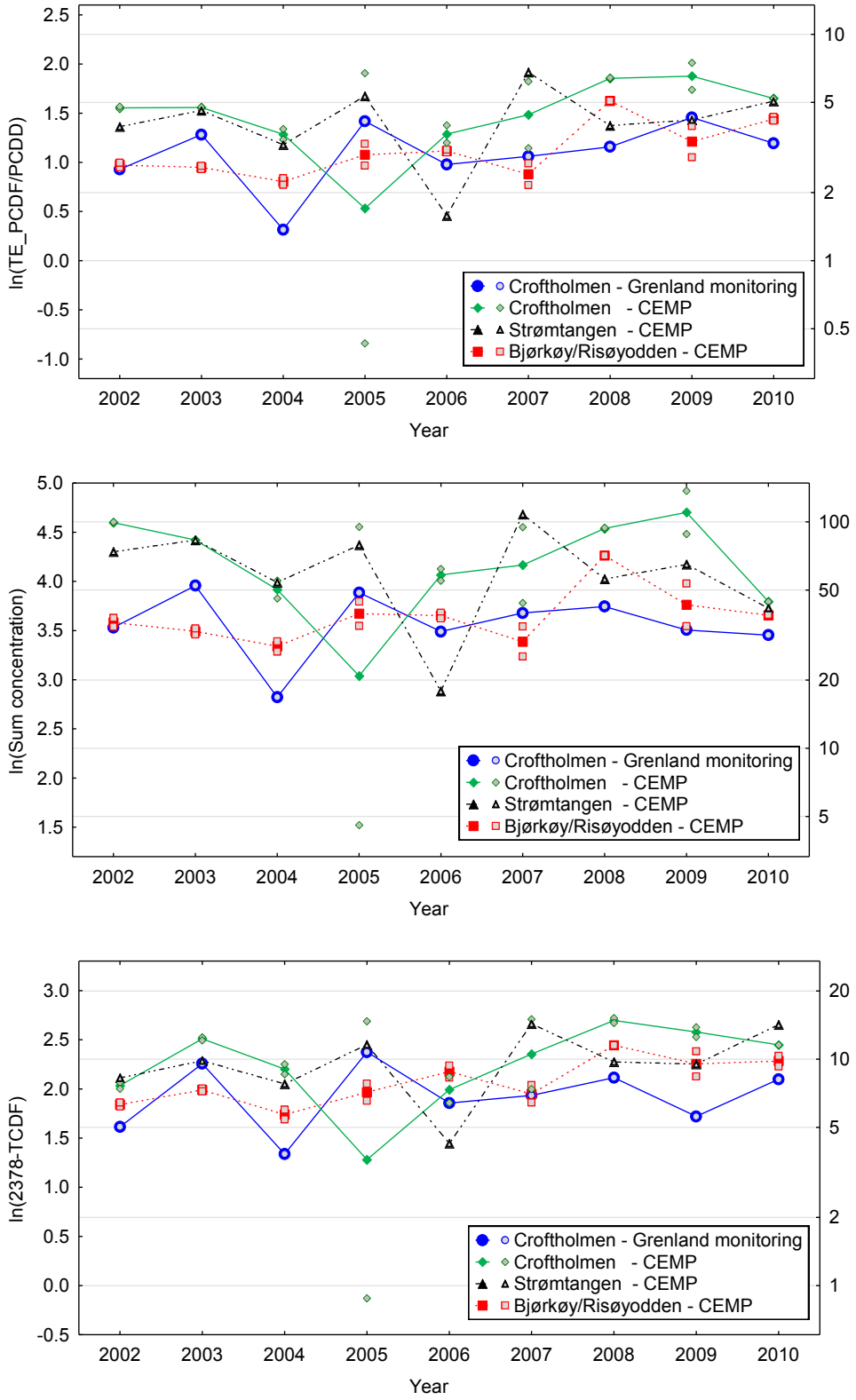


Figure 11. Time series of dioxin concentration, as toxic equivalent (TE_PCDF/PCDD), as sum over the 12 congeners) and as 2378-TCDF The right axis show the raw concentrations and the left axes the natural log transformed values.

The results are shown in Table 7 for the toxic equivalent (TE), for the sum of the 12 congeners and for the single congener 2378-TCDF. The current data design ("*Actual*" in the table) corresponds to 4 main samples and between 1 and 2 subsamples on average per main sample. As

one must expect, one sample from each of 3 main samples (different times or sites) are better than 3 replicates from one main sample.

In these data there are no indications of between-year variance except for short-term noise. The post-processing of between-year variation described in Chapter 4.1 (with 3-year smoother and robust regression on the results) does not give lower between-year variance than the direct estimates from the GLM model. Consequently, the variance components from the GLM analysis are used directly to estimate the effect of varying number of stations or subsamples.

The sum concentration has larger variance within years than either TE or the selected single congener 2378-TCDF, but about the same variance component between years, so the estimated gain of increased number of main samples or subsamples is a little larger.

The uncertainty of the relative changes of the between-year standard deviation (rightmost three columns of the table) has been calculated as confidence limits as described in Appendix D.4.1. Approximate 70 % confidence intervals for the change caused by a reduction to a single sample (S=R=1) extends from about 0.5 to 1.5 times the relative changes shown in the table. As an example, for the Sum of 12 congeners the table shows that the between-year standard deviation is estimated to increase by 48 % if the monitoring is changed from the current progam ("*Actual"*) to one single sample (S=1, R=1). The 70 % confidence interval for this value increases from 26 to 76 % of the between-year standard deviation. For three replicates from one main sample the upward confidence limit is at 1.73 to 1.8 times the estimate; for the estimate 33 % change for Sum of 12 congeners, the confidence limits are 17 and 58 %.

Table 7. Between-year residual standard deviation for dioxin in blue mussels from Langesundsfjord, Grenland estimated for different yearly numbers of main samples and subsamples.

	Variance component estimates			Number (S) of main samples	Expected between-year standard deviation (square root of variance)					
	Between years	Within-year			Absolute value			change relative to actual estimate 2002-2010		
		Between main samples	Between sub-samples		*Number R of replicate samples per main sample*					
					1	*2*	*3*	*1*	*2*	*3*
$TE_{PCDF/PCDD}$	0.0424	0.0362	0.0262	1	0.324	0.303	0.295	36 %	27 %	24 %
				2	0.271	0.259	0.255	14 %	9 %	7 %
				3	0.251	0.242	0.239	6 %	2 %	1 %
				4	0.241	0.234	0.231	1 %	-2 %	-3 %
				Actual	***0.238***					
Sum of 12 congeners	0.0421	0.0607	0.0401	1	0.378	0.351	0.341	48 %	37 %	33 %
				2	0.304	0.287	0.281	19 %	12 %	10 %
				3	0.275	0.263	0.258	8 %	3 %	1 %
				4	0.259	0.250	0.246	2 %	-2 %	-4 %
				Actual	***0.255***					
2378-TCDF	0.0388	0.0309	0.0257	1	0.309	0.287	0.280	36 %	27 %	23 %
				2	0.259	0.246	0.242	14 %	8 %	7 %
				3	0.240	0.231	0.228	6 %	2 %	0 %
				4	0.230	0.223	0.221	1 %	-2 %	-3 %
				Actual	***0.227***					

4.3.2 PCB, DDT, HCB and Mercury in Grenland mussels

Data from 2002-2010 are considered for analysis. Figure 12 shows the data as separate time series for different stations with both single values and yearly arithmetic means of log-transformed values. All these data are from the National CEMP program. There are from 2 to 4 subsamples for all stations and years for these parameters, in most cases three subsamples or replicates. Data from Station 71A has a different time patterns than the other stations during the first three years, with markedly lower values for PCBs, pp'- DDE and HCB in 2004, for HCB also for 2002 and 2003. Because of this, it has been chosen to analyse only data from 2005-2010.

The data are analysed in an ANOVA model with Year and Station as main factors, in the same way as the dioxin in Chapter 4.3.1, with year as a random factor. The ANOVA results are used as described for the two-level model in Appendix D, with Year corresponding to level A and interaction Year*Station corresponding to nested level B, and the subsamples as replicates.

For HCB all stations have a strong peak in 2005; apparently an outlier compared to the pattern of between-year variation in later years. To avoid exaggerating the irregular between-year variance, the results of the ANOVA for HCB have been post-processed by calculating adjusted deviations from a 3-point local regression and using that to estimate a corrected between-year variance as described in Chapter 4.1. For the other parameter the ANOVA results are used directly; post-processing has not been found to have a considerable effect.

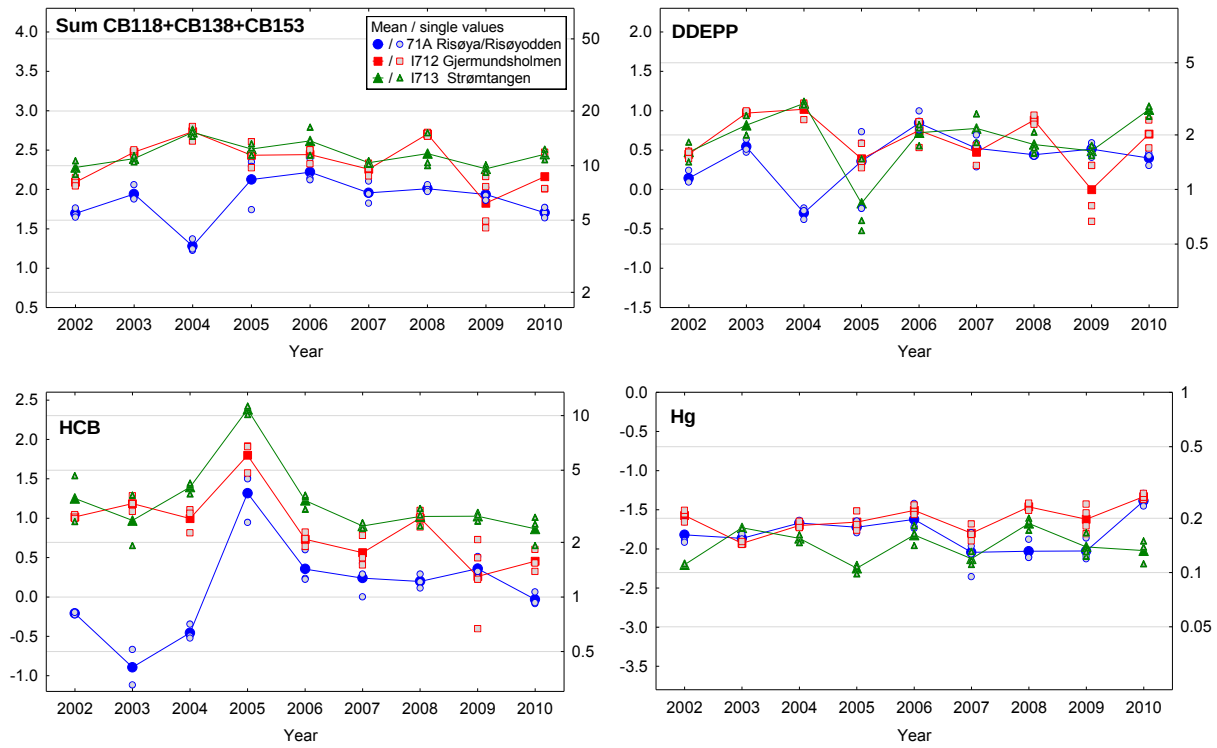


Figure 12. Data for PCBs, pp'- DDE (ICES code DDEPP), HCB and Hg from mussels in Langesundsfjord. Right axes show concentrations on dry weight basis on log scale; the left axes show the corresponding natural logarithms.

The results are shown in Table 8. Reducing from 3 replicates to a single sample at each station will most likely increase the between-year standard deviation by about 10 %; the 70 % confidence intervals run from approximately 0.4 to 1.8 times the estimated relative change; for instance for pp'- DDE from 5 to 19 % relative change.

Reducing to 2 stations with 2 or 3 replicates each has an effect of about the same size as 3 stations with one sample each.

If the program is reduced to a single sample per year (S=R=1), the between-year standard deviation is estimated to increase by 45-63 %, with 70 % confidence intervals from about 0.4 to 1.6 times the estimated relative change; for pp'- DDE the upper confidence limit give an increase of 100 %. The corresponding upper confidence limit for the between-year standard deviation is 0.44, which would mean that the trend would have to be about 15-20 % per year to be detected with 90 % power after 10 years when testing with 5 % significance level (Appendix C.2)

Table 8. Between-year residual standard deviation for PCB (sum of three components), DDT (represented by metabolite p,p'-DDE), HCB and Hg in blue mussels from Langesundsfjord, Grenland estimated for different yearly numbers of main samples and subsamples.

The analysis is based only on data from 2005-2010.

	Variance component estimates				Expected between-year standard deviation (square root of variance)					
	Between years	Between main samples	Between sub-samples	Number (S) of main samples	Absolute value			relative change from actual estimates 2005-2010		
					Number R of replicates per main sample					
					1	*2*	*3*	*1*	*2*	*3*
Sum	0.0211	0.0163	0.0290	1	0.257	0.228	0.217	50 %	32 %	26 %
CB118				2	0.209	0.191	0.184	21 %	11 %	7 %
+CB138				3	0.190	0.177	0.172	10 %	3 %	0 %
+CB153				***Actual***	***0.172***					
p,p'-DDE	0.0266	0.0556	0.0541	1	0.369	0.331	0.317	63 %	46 %	40 %
				2	0.285	0.261	0.252	26 %	15 %	11 %
				3	0.251	0.233	0.226	11 %	3 %	0 %
				Actual	***0.226***					
HCB	0.0340	0.0168	0.0418	1	0.304	0.268	0.254	45 %	27 %	21 %
	(between-year variance adjusted by post-processing of ANOVA results)			2	0.252	0.230	0.222	20 %	9 %	6 %
				3	0.231	0.216	0.210	10 %	3 %	0 %
				Actual	***0.210***					
Hg	0.0130	0.0288	0.0171	1	0.243	0.224	0.218	55 %	43 %	39 %
				2	0.190	0.178	0.174	21 %	14 %	11 %
				3	0.168	0.160	0.157	8 %	2 %	0 %
				Actual	***0.156***					

4.3.3 Metals Cd, Cu, Pb and Zn in Grenland Mussels

For these metals, there are only data from the last two years for stations I712 and I713, while there are long time series from station71A Risøya/Risøyodden, beginning in 1983. Analysis of data just from the last two years would give estimates of station*year interaction for I712 and I713, but with so few degrees of freedom that the result would be very unreliable. The analysis is therefore done only for station 71A, which means that it will estimate the effect of varying number of yearly samples within one station, but not the effect of having more than one station for estimating trend.

For the analysis here, data from 1992-2010 are used. Figure 13 shows the data, both single observations and yearly means. There are considerable long-term variations over time, both compared to the variation between subsamples each year and the average short-term between-year variation of yearly means. There are also some large, sudden changes that should be seen as part of the signal to be detected and not of the noise in data around trends. Thus, for these data, it is necessary to correct the between-year variance from the ANOVA results by post-processing the least-square yearly means as described above in Chapter 4.1. The results shown use the within-year variance from the ANOVA combined with the between-year variance component from the post-processing, and calculates between-year variance and Mean Squares based on the adjusted variance component for assessing of confidence limits for estimated changes due to changed sampling. The degrees of freedom for the ANOVA results are used in this assessment.

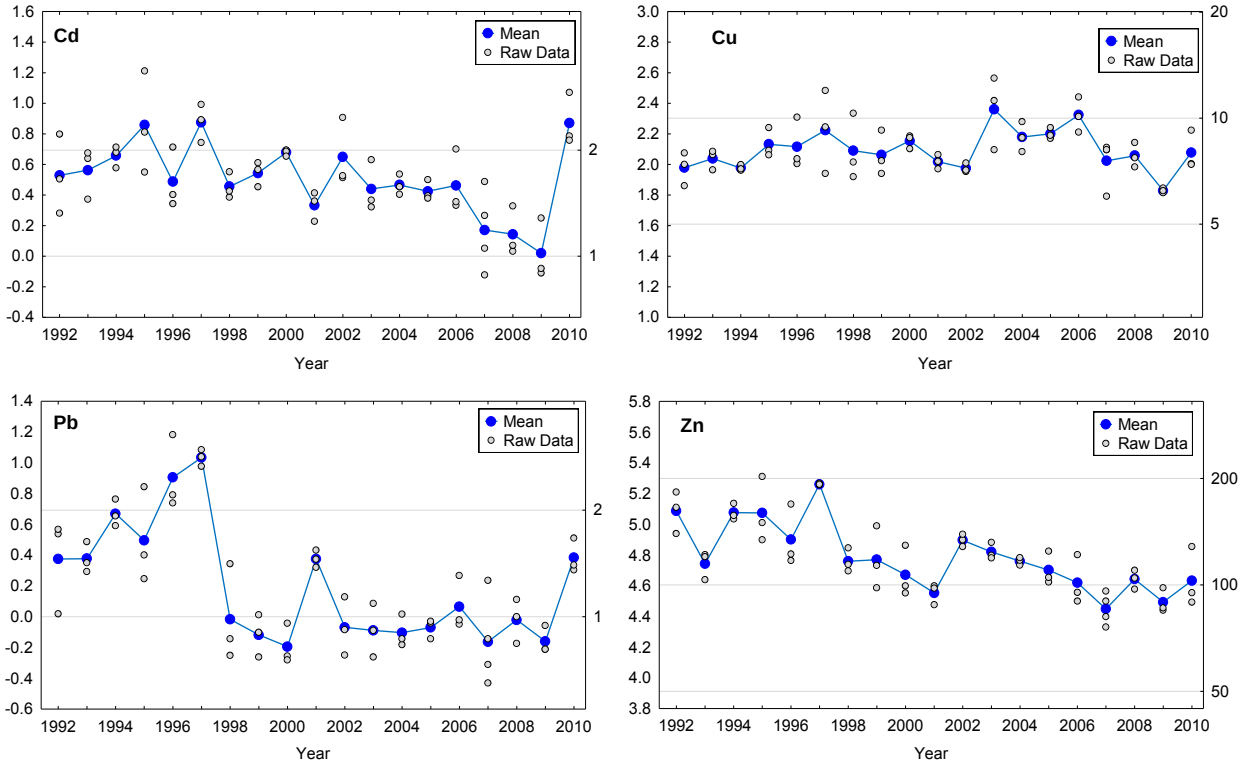


Figure 13. Metal concentrations in mussel from Grenland, station 71A Risøya/Risøyodden from 1992-2010. Right axes show concentrations on dry weight basis on log scale; left axes show corresponding natural logarithms.

The results are shown in Table 9. They indicate that by reducing monitoring to only one sample of the same size as each of the three samples that are currently analysed, the between-year standard deviation would increase by 15 to 30 %. The two-sided 70 % confidence intervals around the estimates extend from 0.65 to 1.45 times the estimated relative change shown in the table. For instance for lead, it is estimated that the relative increase would be about 18 %. This estimate has an uncertainty, which can be described by a 70 % confidence interval with lower limit 12 % and upper limit 26 %. For copper, with an estimated 31 % relative change, the 70 % confidence interval has lower limit 20 % and upper limit 43 %.

The between-year standard deviation for yearly averages of natural log-transformed concentrations for the current program is in the range from 0.12 to 0.2. This means that after 10 years of monitoring with the current program one should be able to detect log-linear trends of 5 to 10 % change per year with 90 % power and 5 % significance (Appendix C.2). With reduction to one sample per year the detectable trend would most probably increase to 7 -12 % per year, and possibly to 10-17 % per year if the upper confidence limit of the estimated change of the between-year standard deviation is considered.

Note that only the effect of changed number of replicates for a single station is estimated here. Since 2009 the metals have been monitored at three stations in Langesundsfjord, with three replicates each per year. If it is considered to reduce this expanded program, one should also take into account the effect of reduced number of sampling sites for estimating an average trend over all stations, as for dioxins and other contaminants in preceding sections. The effects of reduced number of stations can then probably be comparable to the results in chapter 4.3.2.

Table 9. Between-year residual standard deviation for metals in blue mussels from Langesundsfjord, Grenland, estimated for different yearly numbers of main samples and subsamples.

	Variance component estimates			Expected between-year standard deviation (square root of variance)					
				Absolute value			relative change from actual estimate 1992-		
	Between years	Between sub-samples		*Number R of replicate samples per year*					
				1	*2*	*3*	*1*	*2*	*3*
Cd	0.0337	0.0317		0.2559	0.2227	0.2105	21.8 %	6.0 %	0.2 %
			Actual	***0.2101***					
Cu	0.0107	0.0175		0.1680	0.1395	0.1286	31.0 %	8.8 %	0.3 %
			Actual	***0.1282***					
Pb	0.0475	0.0354		0.2879	0.2553	0.2435	18.4 %	5.0 %	0.2 %
			Actual	***0.2431***					
Zn	0.0279	0.0155		0.2083	0.1888	0.1818	14.7 %	4.0 %	0.1 %
			Actual	***0.1816***					

4.3.4 Metals, PCB and pp'- DDE in mussels from Sørfjord

The data series are plotted in a series of graphs in Appendix G. Data from the different stations follow in general much of the same time pattern, but there are also important differences between stations with regard to peaks and other large changes. For instance, copper (Cu) at station 52A has a sharp isolated peak in the CEMP monitoring in 1991, which is not present at the other stations (page 73). Another example is Hg, where station 51AA and 56A has high concentrations in 1999 and 2000 compared to the years just before and after; this feature is not so clear at 52A and 57AA (page 74). For PCBs all stations have large peaks in 2001, but of different sizes. By analysis of PCB profiles, it was established that this peak was caused by removal of old paint and plaster from an old power station close to the shore. This was part of a renovation of the building which had been designated a national historical monument the year before (Ruus et al. 2006).

There are also differences between stations in temporal coverage of data, as was also described in chapter 4.2.2.

In order to handle these differences, but still get overall estimates of within-year variance terms, data from all stations are analysed together in an ANOVA model with *Station* as main fixed factor and *Year* as a random factor nested within station. *Main Sample* (SeqNo) is a random factor nested within Station*Year (independent random numbering of samples from year to year). The subsamples are replicates within main sample, normally 3 for the CEMP Main Sample, and 1 for the regional monitoring. The analysis is done with the GLM module in Statistica version 10.

When the data are analysed in this design, with year nested in station, the yearly means are estimated independently for each station, without any between-year component as average over all stations. This allows for different stations having different time coverage as described in chapter 4.2.2, and gives yearly means that adapt to the different time variation patterns at for different stations. The within-year variance between main samples and subsamples are calculated as pooled estimates based on data from all stations.

Table 10 shows the results from the ANOVA analysis for each parameter, with degrees of freedom (df) and Mean Squares (MS) for the effects Year (nested in Station) and seqno (main sample, nested in Year*Station) and for residual error (between subsamples from the same main sample. The right side of the table lists the Expected Mean Square Coefficients for the data set. These coefficients define the relation between the variance components and the ANOVA Mean Squares as described in chapter 4.1; also see Appendix D for explanation and details. The table also shows the Between-year Standard deviation; this is the Mean Square (MS) of the factor Year divided by the coefficient k_1.

Table 10. Data for mussels from Sørfjord: ANOVA results with correction of between-year variance by post-processing.

All results are for natural logarithms of concentrations.

Parameter	Degrees of freedom and Mean Square for different factors in the nested ANOVA model.			Between-year standard deviation	Expected Mean Square (EMS) Coefficients		
		df	Mean square		k1	k2	k3
Cd	Year(Station) including trend	80	3.144	0.9692	3.347	1.754	
	seqno(Station*Year)	82	0.263				1.691
	Error	192	0.104				
	Year(Station) excluding trend		0.4712	0.3752			
Cu	Year(Station) including trend	66	0.268	0.2837	3.333	1.748	
	seqno(Station*Year)	67	0.085				1.692
	Error	157	0.033				
	Year(Station) excluding trend		0.1360	0.202			
Hg	Year(Station) including trend	80	1.492	0.6678	3.347	1.754	
	seqno(Station*Year)	82	0.240				1.691
	Error	192	0.061				
	Year(Station) excluding trend		0.3300	0.314			
Pb	Year(Station) including trend	80	1.499	0.6692	3.347	1.754	
	seqno(Station*Year)	82	0.355				1.691
	Error	192	0.139				
	Year(Station) excluding trend		0.5817	0.4169			
Zn	Year(Station) including trend	66	1.433	0.6542	3.348	1.763	
	seqno(Station*Year)	67	0.092				1.692
	Error	158	0.035				
	Year(Station) excluding trend		0.2364	0.2657			
p,p'-DDE	Year(Station) including trend	66	0.479	0.3827	3.272	1.712	
	seqno(Station*Year)	69	0.422				1.648
	Error	155	0.034				
	Year(Station) excluding trend		0.3198	0.3128			
CB138	Year(Station) including trend	57	1.379	0.6400	3.367	1.793	
	seqno(Station*Year)	59	0.141				1.704
	Error	139	0.028				
	Year(Station) excluding trend		0.3295	0.3128			

The ANOVA analysis is done on the original data, without removal of long-term trends. As discussed in chapter 4.1, this means that the between-year Mean square and Standard deviation from the ANOVA, shown in the shaded first row for each parameter, contains long-term time trends as well as peaks and irregular fluctuations, and in additions differences between stations. In order to estimate the irregular between-year variance within station, without the effect of trend and sudden major changes in the data, and without effect of differences between stations, the post-processing procedure described in chapter 4.1 is applied to the ANOVA results.

The first step is to calculate deviations between Year*Station means and 3-point LOESS smoother fits through the time sequence of means. This is done separately for each time series, that is, for each parameter and station. Figure 14 shows cadmium data from station 52A as an example. The differences between yearly means and the smoother values as they appear in the plot are multiplied by a factor $(3/\sqrt{6})$ to get the adjusted deviations δ described in equation (11)

on page 25. The adjusted deviations has a variance that should represent irregular fluctuations between years, but may possibly have some large outliers due to peaks or sudden changes over time. The identification of such outliers is done in with deviations from all stations combined, as explained below. Figure 14 shows which years for station 52A which are then excluded as outliers.

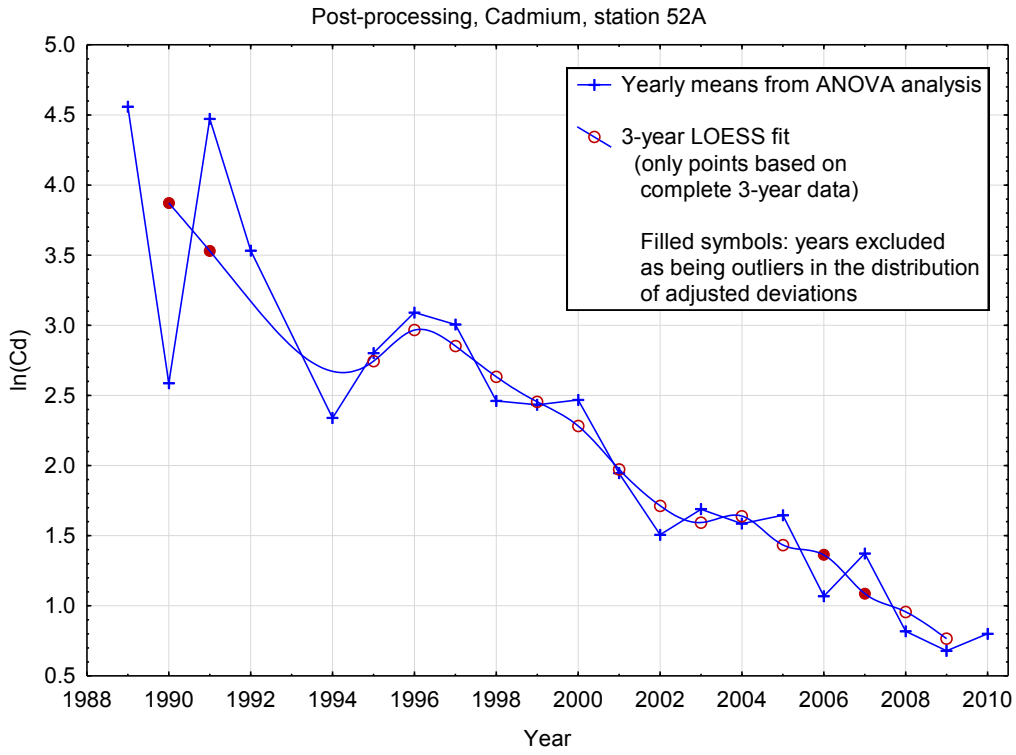


Figure 14. Least-square yearly means for cadmium at station 52A, with fitted 3-point unweighted local regression (LOESS) smoother.

The second step of the post-processing attempts to estimate the between-year irregular variance after having removed the outliers. For each parameter, the adjusted deviations from all stations are pooled into one set, ordered and assigned corresponding standardised normal fractiles. From this set of value pairs, the standard deviation for the irregular, between-year variation can be estimated as explained in chapter 4.1. Outliers are excluded after visual inspection. Generally, this is done by ignoring data with standard normal fractile outside the interval ±1, and calculate the regression for the central part of the distribution of residuals, which generally follows a normal distribution. Figure 15 illustrates this process for cadmium, with the tails excluded as outliers marked as open rectangles.

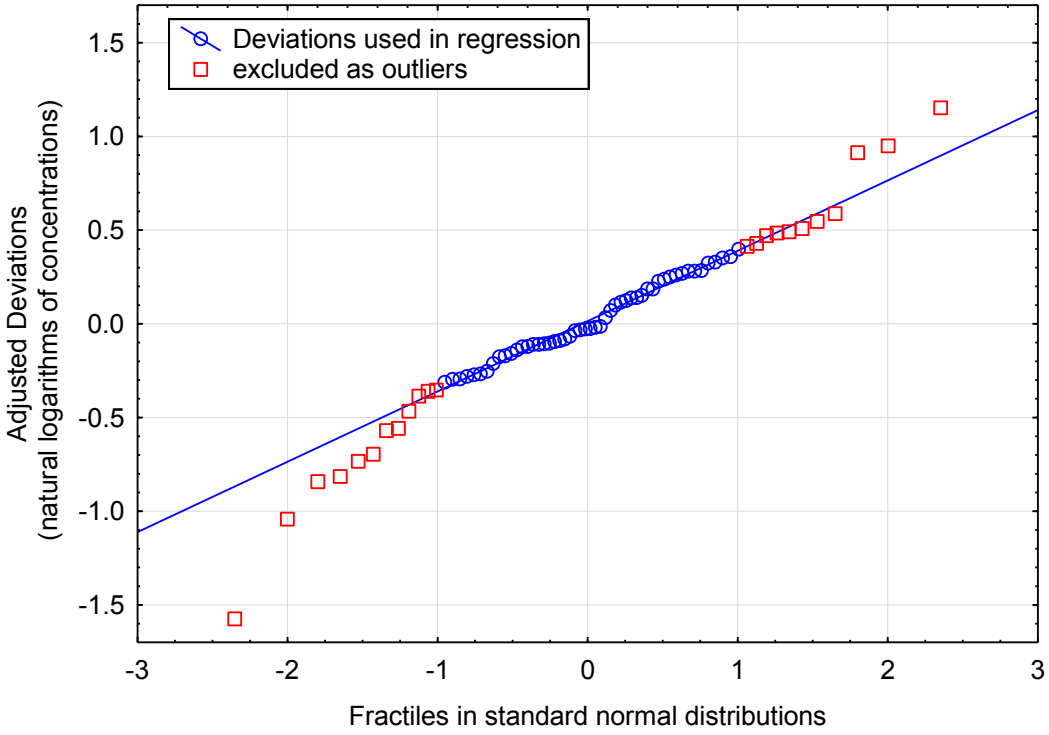


Figure 15. Adjusted deviations for cadmium (log-transformed) in Sørfjord, pooled for all stations, ordered and assigned fractiles of standard normal distribution. The tails of the distributions deviate from normal distribution and are excluded from the linear regression.

The standard deviation of irregular, ordinary between-year variation is estimated by the slopes of the regressions. The results for the different parameters are shown in the last row for each parameter (labelled *Year(Station) excluding trend*) in Table 10. For cadmium the value derived from the data in Figure 15 is 0.3752. The corresponding Mean square is calculated by multiplying the square of the between-year standard deviation with the coefficient k_1. For cadmium the calculated Mean square is $0.4712 = 3.347 \cdot 0.3752^2$. This value is used instead of the Mean Square for Year(Station) from the ANOVA (3.144 for cadmium) when calculating the effect of changing monitoring design according to Appendix D.

Table 11 shows the results of the calculations of expected changes in between-year standard deviations if the monitoring program is changed, based on the results in table 10. The left side of table 11 lists the variance components for Year, Main sample and Replicate. The within-year variance components are calculated from the ANOVA results, while the between-year variance component is found by combining these with the Mean Square from the post-processing of yearly means. For all parameters except CB138 this gives a negative between-year variance component, which means that the analysis does not make it possible to detect any real between-year variance except the residual effect of within-year variance. Variance terms can never be really negative; the negative estimates are due to uncertainty, or statistical error. This may occur in a regular ANOVA, but the modification here by the post-processing makes it more likely to happen. However, the effect of changing monitoring program can still be estimated. Both the estimated effect and confidence limits for the effects are best calculated using the Mean Squares directly instead of the variance components , as shown in Appendix D. The right half of Table 12 shows this, in the same way as in previous sections in this chapter. This part of the table shows the result of changing the monitoring design, both as expected between-year standard deviation and as relative change compared to the value for the present data set (Shown in bold as ***Actual***).

The results indicate that by reducing monitoring to a single sample per year, one can expect to increase between-year standard deviation by 25 - 50 %, with the largest effect for pp'- DDE. The 70 % confidence intervals for these estimated changes runs from approximately a factor 0.8 to 1.2 times the estimate, so for instance from 41 to 62 % for pp'- DDE and from 18 to 29 % for CB138. The estimates for different parameters cannot be considered independent, since they are derived from almost the same set of samples, so a simple analysis of whether the differences are significant is not meaningful.

Table 11. Between-year residual standard deviation for selected contaminants in blue mussels from Sørfjord, Hardanger, estimated for different yearly numbers of main samples and subsamples.

	Variance component estimates			Number (S) of main samples	Expected between-year standard deviation (square root of variance)					
	Betw. years	Within year			Absolute value			Change relative to actual estimate 2002-2010		
		Between main samples	Between repl. samples		*Number (R) of replicate samples per main sample*					
					1	*2*	*3*	*1*	*2*	*3*
Cd	-0.0280	0.0941	0.1037	1	0.508	0.454	0.435	35.5 %	21.1 %	15.9 %
				2	0.399	0.365	0.353	6.4 %	-2.6 %	-5.8 %
				Actual	***0.375***					
Cu	-0.0137	0.0307	0.0331	1	0.280	0.249	0.238	24.2 %	10.3 %	5.3 %
				2	0.216	0.196	0.189	-4.3 %	-13.2 %	-16.3 %
				Actual	***0.226***					
Hg	-0.0454	0.1059	0.0609	1	0.438	0.402	0.389	39.4 %	27.9 %	23.8 %
				2	0.329	0.305	0.297	4.8 %	-2.8 %	-5.5 %
				Actual	***0.314***					
Pb	-0.0537	0.1278	0.1389	1	0.576	0.512	0.489	38.2 %	22.9 %	17.4 %
				2	0.446	0.405	0.390	6.9 %	-2.9 %	-6.4 %
				Actual	***0.417***					
Zn	0.0118	0.0336	0.0350	1	0.333	0.306	0.296	25.4 %	15.1 %	11.5 %
				2	0.277	0.261	0.255	4.3 %	-1.9 %	-4.0 %
				Actual	***0.266***					
p,p'-DDE	-0.1250	0.2355	0.0338	1	0.492	0.474	0.468	51.0 %	45.6 %	43.8 %
				2	0.327	0.314	0.310	0.5 %	-3.5 %	-4.9 %
				Actual	***0.326***					
CB138	0.0145	0.0667	0.0275	1	0.385	0.367	0.361	23.1 %	17.3 %	15.3 %
				2	0.318	0.307	0.303	1.7 %	-1.8 %	-3.0 %
				Actual	***0.313***					

4.4 Conclusions for analysis of mussel data

Conclusions:

In general the data sets from Langesundsfjord in the Grenland region has residual between-year variation corresponding to 3 or 4 main samples per year, on average divided into 2-3 subsamples for analysis. Reducing the sampling to one single mussel sample per year for Langesundsfjorden, Grenland, is estimated to increase the residual between-year standard deviation with 35-40 % for dioxins, with 45-65 % for PCBs, p,p'-DDE, HCB and Hg. For reduction to two main samples, each divided in two or three subsamples, is estimated to keep the increase within 12 % for dioxins and within 15 % for the other parameters. 70 % confidence intervals extend from about 0.5 to 1.5 times the estimated relative changes for dioxins, and from 0.4 to 1.6 times the estimates for the other parameters.

The Grenland metal data cover only one mussel station, reducing from three subsamples to one single sample is estimated to increase between-year standard with 15-30 %. Reducing from 3 to 2 subsamples is estimated to increase the between-year standard deviation within 8%. The two-sided 70 % confidence limits extend from 0.65 to 1.45 times these estimates.

For the Sørfjord data, results indicate that for stations where the current monitoring program consists of two replicate samples from the CEMP monitoring and one sample from the regional monitoring, a reduction to a single sample per year, would increase the between-year standard deviation by 25 - 50 %. The 70 % confidence intervals extend from 0.8 to 1.2 times these estimated increases.

Reduction from the current programs to a single sample does not seem to be advisable, since it may mean a substantial reduction of trend detection ability, but a more cautious reduction to fewer main samples, or fewer subsamples per main sample, may take place without significant reduction of time trend detectability.

5. References

Bjerkeng, B, 2006: Statistical assessment of monitoring of dioxins in organisms, Quantification of uncertainty and assessment of statistical assessment power – basis for planning of monitoring programs NIVA report 5123, ISBN 82-577-4833-1, 110 pages (In Norwegian).

Fryer, R.J. and M. D. Nicholson, M.D., 1999: Using smoothers for comprehensive assessments of contaminant time series in marine biota. ICES Journal of Marine Science, 56: 779–790. 1999

Milliken, G.A. and Johnson, D.E., 1992: Analysis of Messy Data Volume I: Designed Experiments. Chapman & Hall.

Green, N.W., Schøyen, M., Øxnevad, S., Ruus, A., Høgåsen, T., Beylich, B. Håvardstun, J., Gudmundson Rogne, Å. K. and Tveiten, L., 2011: Hazardous substances in fjords and coastal waters – 2010. Levels, trends and effects. Long-term monitoring of environmental quality in Norwegian coastal waters. Klif monitoring report 1111/2011, TA-no. 2862-2011, NIVA-report 6239-2011, ISBN 978-82-577-5974-2. 252 pages.

Ruus, A., Bakke, T, Bjerkeng, B. and Knutsen, H., 2011a: Monitoring of contaminants in fish and shellfish from Grenlandsfjordene 2010 (Overvåkning av miljøgifter i fisk og skalldyr fra Grenlandsfjordene 2010). Klif monitoring report 1104/2011, TA-no. 2836-2011, NIVA-report 6203-2011, ISBN 978-82-577-5938-4. 65 pages. *In Norwegian, with English summary.*

Ruus, A., Skei, J., Lundmark Daae, K., Green, N.W. and Schøyen, M., 2011b: Monitoring of environmental quality in the Sørfjord 2010. (Overvåkning av miljøforholdene i Sørfjorden 2010). Klif monitoring report 1103/2011, TA-no. 2824-2011, NIVA-report 6199-2011, ISBN 978-82-577-5938-4. 65 pages. *In Norwegian, with English summary.*

Ruus, A., Green, N. W., Maage, A. and Skei, J., 2006: PCB-containing paint and plaster caused extreme PCB-concentrations in biota from the Sørfjord (Western Norway) — A case study. Marine Pollution Bulletin 52 (2006), 100-103

6. Explanation of acronyms and abbreviations

Chemical elements and chemical substances	
2378-TCDF	2,3,7,8-tetrachloro dibenzofuran
α-HCH	alpha-hexachlorocyclohexane, by-product in production of Lindane
γ-HCH	gamma- hexachlorocyclohexane, Lindane, pesticide
As	arsenic
CB101	PCB congener: 2,2',4,5,5'-pentachlorobiphenyl
CB105	PCB congener: 2,3,3',4,4'-pentachlorobiphenyl
CB118	PCB congener: 2,3',4,4',5-pentachlorobiphenyl
CB138	PCB congener: 2,2',3,4,4',5'-hexachlorobiphenyl
CB153	PCB congener: 2,2',4,4',5,5'-hexachlorobiphenyl
CB156	PCB congener: 2,3,3',4,4',5-hexachlorobiphenyl
CB180	PCB congener: 2,2',3,4,4',5,5'-heptachlorobiphenyl
Cd	cadmium
Cr	chromium
Cu	copper
DDEPP	ICES code for p,p'-DDE
DDT	dichlorodiphenyltrichloroethane, insecticide
HCB	hexachlorobenzene
Hg	mercury
ICES	International Council for exploration of the Seas
OCS	octachlorostyrene
p,p'-DDD	DDT metabolite: p,p'-Dichlorodiphenyl dichloroethane
p,p'-DDE	DDT metabolite, p,p'-Dichlorodiphenyl dichloroethylene
Pb	lead
PCB	polychlorinated biphenyls
PCDD	polychlorinated dibenzodioxins
PCDF	polychlorinated dibenzofurans
QCB	pentachlorobenzene
TBTIN	tributyltin
TDEPP	ICES code for p,p'-DDD
TE	Toxic Equivalent, in this document used for TE_PCDD/PCDF which is calculated as weighted sum of dioxin and dibenzofuran components
Zn	zinc

Organisations etc.	
CEMP	Co-ordinated Environmental Monitoring programme of OSPAR
ICES	International Council for Exploration of the Sea
Klif	Norwegian Climate and Pollution Agency ("Klima- og forurensningsdirektoratet")
OSPAR	The Oslo Paris Commission

Statistical terms	
ANOVA	Analysis of variance
GLM	General Linear Model
LOWESS	Local weighted regression
LOESS	Local regression, here used about unweighted regression
OSPAR	The Oslo Paris Commission

Appendix A. Technical description of trend assessment procedure for Norwegian CEMP data

The description below describes the customized version of the CEMP procedure used for the present analysis, noting differences from the regular CEMP procedure used for the National CEMP reports on hazardous substances (Green et al. 2011). In particular, the customized version treats observations < analytical limits different from the regular procedure.

A.1 Aggregating data into yearly medians

For each time series, the data are first aggregated into yearly median[1] values. For years where all measurements give definite values above analytical limit, extraction of a well-defined median value is straightforward. For years where one or more measurements are below an analytical limit, lower and upper bounds for the median are extracted. The lower bound for the median is found by assuming observations <analytical limit to be = 0, while the upper bound is found by using the analytical limit for these observations. The result may still be a definite median value, with lower bound equal to the upper bound, depending on how many observations below limit there are, and how the analytical limits are distributed compared to definite values in the sample. For instance, the sorted sample

<0.4, <0.5, 0.8, 0.9, 1.2, 1.4, 2.6

will have a well-defined median of 0.9. If more than 50 % of observations are below limit, the median will necessarily be specified as a low-high range, with 0 as lower bound. Otherwise, if the median is affected by observations <analytical limit, the lower bound may also be a positive value.

In the original CEMP procedure, observations below analytical limit are included as definite values set to half the analytical limit. This always gives a definite median value, but the statistics may partly be a result of variations in analytical limit over time, and not reflect real variability, and the effect on trend estimates may go in both directions compare to what the result would have been with lower analytical limits.

In the customized CEMP procedure used for the present report, the upper bound is used in such cases. However, results are presented only for time series where all median values are well-defined, independently of any further assumptions about the observations below analytical limit. The method described above for extracting median may help to include time series with a moderate number of <limit observations, which would be excluded by a simpler but stricter requirement that none of the observations should be <limit.

A.2 Assessing trend for yearly data

In the CEMP procedure used for the National Reports, trend is analysed by fitting a regression curve to log-transformed medians versus year.

- For series with data for less than 5 years, no trend is estimated.
- For time series with data for 5 or 6 years, only linear regression is fitted and used in the trend test (Appendix B).
- For time series with at least 7 years of data, a nonlinear smoother curve is fitted by local weighted regression (LOWESS) (details in Appendix section A.4 and A.5). Linear

[1] For an odd number of observations, the sample median is the value in the middle of the sequence of observations ordered by size; for an even number of observations the mean of the two middle observations.

regression is also tested, and in addition the significance of nonlinear deviations from the linear fit, but the trend is assessed by comparing smoother fits for last year with the fit up to 10 years earlier.

The smoother uses a moving regression window of 7 years. For a complete series with data for n consecutive years, the smoother estimates at each target year k=3,…,n-2 are based on the data values from the 7 surrounding years (j = k-3, k-2, …, k+3). For k=1 or 2 the regression is based on data for the first 7 years and for k=n-1and n the last 7 years are used.

Wherever a time series has missing years, the local regression window can be expanded where necessary to contain a minimum number of data points for each local regression. The expansion is done in such a way that the window time scale is minimized, to keep the window centred around the target point as well as possible. In the present report, for analysis of yearly data it is required that at least 3 points be included in each local regression. The CEMP procedure as run for the National Comments report has no such requirement, which is in line with the OSPAR procedure, but the emphasis in the present report is on time series with data for all years, where expansion of the window does not play a role for yearly data.

For the smoother regression on yearly data, the time trend is tested by comparing the fitted smoother value for the last year of the series with the fitted value 10 years before, or for the first year of series if the time span of the series is shorter than 10 years. The contrast is tested at a 5 % two-sided significance level. In addition to the smoother, a linear regression is also fitted for time series with more than 7 years of data, and a separate test for nonlinearity (Lack-of-fit for linear regression) is conducted. If the time series has only 5 or 6 years with data, the linear trend is used for assessment.

The standard deviation of yearly medians around the time trend is estimated as part of the smoother fit and used to estimate residual statistical errors in the trend assessment[1]. The standard deviation of yearly medians around the trend is assumed to be homogeneous across years[2]. However, they may be due both to irregular short-term between-year fluctuations around the fitted trend and to differences between the actual trend and the fitted linear or smooth regression. For instance, there are examples of sharp increases from one year to another, followed by a gradual reduction towards lower levels. In some cases, such an event should properly be considered as part of the signal to be detected, and not as part of the statistical uncertainty. Such lack of fit for either smoother or regression means that the ability to detect trends for future monitoring may be underestimated, not taking into account the possibility of models that fit the data better.

A.3 Assessing trend for monitoring each second or third year.

The procedure described above is adapted to yearly monitoring. If the monitoring interval is systematically increased to 2 or 3 years, a 7 year window becomes a bit short, even if the data series is complete, since it means that only at most three data points will be included in each local regression, as discussed in Chapter 2.1.2. Therefore, an alternative with expanded regression window is also used. The expansion is implemented by requiring that at least 5 points are included as basis for each local regression. This leads to local regression windows of 9 years

[1] In the CEMP procedure, the residual standard deviation of data around the smoother fit is used also to test the significance of the linear regression and the nonlinear component.

[2] If the number of data behind each median is very variable between years, this ought to be taken into account when weighting data points in the regression, but the current procedure presupposes approximately equal number of values all years for each series.

for monitoring each second year, and 13 years for monitoring each third year. The expanded regression window means that the smooth trend curve has higher stiffness, but it will improve the statistical basis behind each fitted value (Trade-off between bias and precision).

The analyses for 2 or 3 year monitoring interval has been done both with regression window kept fixed to 7 years (time span of 6 years), including at most 3 years in each local regression, and with local regression windows expanded by requiring 5 points in each regression. The detailed description below is generally valid for all monitoring frequencies and regression windows.

A.4 Smoother fit by weighted least squares regression (LOWESS)

In the smoother fitting process for a data series $x_k, y_k, (k = 1, \dots, n)$, the corresponding fitted values $\hat{y}_k$ $(k = 1, \dots, n)$ are calculated from the local regressions on data points in the vicinity of each target point

$$\hat{y}_k(x) = a_k + x \cdot b_k \tag{12}$$

Each parameter set a_k, b_k is estimated by local weighted regression (LOWESS) on a subset of data points within a local regression window around the target point t_k. For each regression *k,* the data points, now indexed *i*=1,…,*n*, are assigned weights w_{ik} according to distance from t_k.

A.4.1 Setting regression weights

The weights are set by a tricube function of the distance from the target point to each data point *i* within the regression window:

$$w_{ik} = 1 - \left(\frac{|x_i - x_k|}{\Delta T_i}\right)^3 \tag{13}$$

The regression window for each target point is the minimal best centred window that is of specified minimum width and has a required minimum number of data points. The time scale ΔT_i in the weight function is set so that all weights are positive:

$$\Delta T_i = \max_k(|x_i - x_k|) + 1 \tag{14}$$

If there are gaps in the time series, so that the specified window does not include the required minimum number of points, the window is expanded until the required number is reached. The expansion is done so as to minimize the time scale (ΔT_i) for each target point.

For a long time series with data for all consecutive years, the inner target points x_i (3 years or more from both ends of the series) have a 7-year centred window ($(k = i - 3, \cdots, i + 3)$ and a time scale ΔT= 4 years, and the non-zero weights (scaled to have sum 1.0) are:

k =	$i \pm 3$	$i \pm 2$	$i \pm 1$	i
w_{ik} =	0.042	0.145	0.206	0.2164

For target points closer to the end of the series, the window is shifted as necessary to keep it within the time series.

A.4.2 Estimating LOWESS parameters and fitting trend values

The parameters a_k, b_k for each local regression are then estimated by minimising the weighted sum of squared differences between data and fitted regression over all included data points:

$$\min\left(L_w = \sum_i (y_i - \hat{y}_k(x_i))^2 w_{ik} = \sum_i (y_i - a_k - x_i b_k)^2 w_{ik} \right) \tag{15}$$

The summation can be considered to include only the included data points for each regression, or to include all data, with w_{ik} =0 for data outside the different regression windows. The solution is found by setting partial derivatives of L_w on a_k, b_k to zero:

$$\sum_i (y_i - a_k - x_i b_k) w_{ik} = 0 \quad ; \quad \sum_i x_i (y_i - a_{k0} - x_i b_k) w_{ik} = 0 \tag{16}$$

If the x scale is assumed to be adjusted so that x_i is centred on the weighted average $\left(\sum_i x_i w_{ik} = 0\right)$, and the summation index for terms independent of y_i is changed from i to j, the result is

$$a_k = \sum_i y_i w_{ik} \Big/ \sum_j w_{jk} \quad ; \quad b_k = \sum_i x_i y_i w_{ik} \Big/ \sum_j x_j^2 w_{ik} \tag{17}$$

The fitted value $\hat{y}_k$ at t_k can be expressed as a weighted average over the observed y values:

$$\hat{y}_k = \sum_i y_i s_{ik} \tag{18}$$

with weights

$$s_{ik} = w_{ik} \left(\frac{1}{\sum_j w_{jk}} + \frac{x_i x_k}{\sum_j x_j^2 w_{jk}} \right) \tag{19}$$

For the unadjusted t values this changes to:

$$s_{ik} = w_{ik} \left(\frac{1}{\sum_j w_{jk}} + \frac{(x_i - \bar{x}_k)(x_k - \bar{x}_k)}{\sum_j (x_j - \bar{x}_k)^2 w_{jk}} \right) \quad \text{where } \bar{x}_k = \sum_j x_j w_{jk} \Big/ \sum_j w_{jk} \tag{20}$$

If the target position x_k is equal to the weighted average $\bar{x}_k$ of the data points, the weights s_{ik} are simply w_{ik} scaled to have sum 1 over index i.

In the following, the summation terms are taken as going over all data points for all k, with w_{ik} and s_{ik} set to zero if the data point i is outside the local regression window for target point k. The formulas can then be expressed in a compact matrix notation, see Fryer and Nicholson 1999.

A.5 Assessing trend from smoother function

The smoother trend assessment in the CEMP procedure consists in comparing the fitted value for the last year (n) and a year (m) up to 10 years before and testing whether the difference is larger than expected as a random result of residual errors of data points around the trend. The estimated difference is given by

$$\hat{D} = \hat{y}_n - \hat{y}_m = \sum_{i=1}^{n} (s_{in} - s_{im}) y_i \tag{21}$$

The residual variance of data point fluctuations around the trend is estimated by

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\nu} \tag{22}$$

where

$$\nu = \sum_{i=1}^{n} [(1 - s_{ii})^2 + \sum_{k \neq i} s_{ik}^2] \tag{23}$$

is the approximate degrees of freedom for the variance estimate.

These fluctuations around the smooth trend lead to an uncertainty of the difference $\widehat{D}$ as an estimate of an assumed true difference between the two years. The uncertainty is defined by the variance of $\widehat{D}$:

$$V(\widehat{D}) = \widehat{\sigma^2} \cdot \sum_{i=1}^{n} (s_{in} - s_{im})^2 \quad (24)$$

A two-sided test with significance level **α** will conclude that there is a real difference if the estimated difference meets the condition

$$|\delta| > t(0)_{1-\alpha/2,\nu} \quad (25)$$

where $t(0)_{p,\nu}$ is the upper p fractile of the central t distribution with ν degrees of freedom and δ is given by:

$$\delta = \frac{D}{\sqrt{V(\widehat{D})}} = \frac{\sum_{i=1}^{n} (s_{in} - s_{im}) \cdot y_i}{\sqrt{\widehat{\sigma^2} \cdot \sum_{i=1}^{n} (s_{in} - s_{im})^2}} \quad (26)$$

Appendix B. Linear regression trend test

The linear regression can be seen as a special case of the weighted smoother regression, with all data points included for all target points and infinite time scale, so that weights $w_{ik} = \frac{1}{n}$ for all *i,k*. The slope coefficient *b* in the linear regression model $y_i = a + b \cdot x_i + \varepsilon_i$ is estimated by:

$$\hat{b} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{27}$$

where $\bar{x}, \bar{y}$ are averages over all data, The estimate of *b* has error variance:

$$V(\hat{b}) = \sigma^2 \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{28}$$

The residual variance σ^2 for the fluctuations ε_i of data around the regression has ν=*n*-2 degrees of freedom and is estimated by

$$s^2 = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y} - \hat{b}(x_i - \bar{x})\right)^2}{n-2} \tag{29}$$

A two-sided test with significance level **α** will conclude that there is a real difference if the estimated difference meets the condition

$$\left|\hat{b}\right| > \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} t(0)_{(1-\alpha/2),n-2} \tag{30}$$

Where $t(0)_{p,\nu}$ is the upper *p* fractile of the central t distribution.

For a contiguous data series with one value for data each of *n* years ($x_i = x_0 + i - 1$ for $i = 1, \ldots, n$):

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} i\right)^2 = \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} = \frac{(n-1)n(n+1)}{12} \tag{31}$$

For a series with *h* years between values ($x_i = x_0 + (i-1)h$ for $i = 1, \ldots, n$) covering a time span $T = (n-1)h + 1$ years:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} h^2(i - \bar{i})^2 = h^2 \frac{(n-1)n(n+1)}{12} \tag{32}$$

which can be written in terms of *T* and *h* as

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{(T-1)(T-1+h)(T-1+2h)}{12h} \tag{33}$$

The trend test condition then becomes:

$$|\delta| \geq t(0)_{1-\alpha/2,\nu} \tag{34}$$

where

$$\delta = \frac{b}{s}\sqrt{\frac{(T-1)(T-1+h)(T-1+2h)}{12h}} \tag{35}$$

Appendix C. Significance and power – the ability to detect trends

C.1 General description

Equations (25) in Appendix A.5 and (34) in Appendix B give the trend test conditions for the smoother and for a linear regression. Both tests can be written

$$|\delta| \geq t(0)_{1-\alpha/2,\nu} \tag{36}$$

where δ is a test value calculated from the data.

The tests are constructed to give a chosen level of protection against falsely rejecting the null hypothesis of no trend when in fact it is true; the significance level is the chosen acceptable risk of making such an error. Both for the smoother and the linear regression, the test is based on the assumption that if there is no trend, the test statistic follows the central t distribution, which is symmetric with the peak at zero.

The probability of detecting an actual trend is the probability that the test statistic will exceed the criteria given by the equations. If there is a trend, the test statistic will follow a noncentral t distribution, which is skewed and has a non-zero peak. Thus, the probability of detecting the trend is the probability that a value drawn at random from a noncentral *t* distribution exceeds the test value defined by the central t distribution. If the test statistic really follows a noncentral t distribution with noncentrality parameter δ, the probability of rejecting the null hypothesis is given by the equation

$$t(|\delta|)_{\beta,\nu} = t(0)_{1-\alpha/2,\nu} \tag{37}$$

Where $1-\beta$ is the probability of rejecting the null hypothesis (β being the risk of falsely accepting the null hypothesis); $t(\delta)_{\beta,\nu}$ is the β fractile of the noncentral t distribution. i.e. the value such that there is a probability β that a randomly drawn value is lower.

An approximate condition can be obtained by using only the central *t* distribution, which is more available in statistical computational tools:

$$|\delta| = t(0)_{1-\beta,\nu} + t(0)_{1-\alpha/2,\nu} \tag{38}$$

If the degrees of freedom ν is >30, the results are practically identical. For fewer degrees of freedom the approximation deviates from the theoretically correct solution. Bjerkeng (2006) has found that for $\nu > 2$, $\alpha = 0.05$ and with β in the range 0.1 to 0.2, the approximate formula overestimates required δ by 2-4 %, which is in practice unimportant. In this report, estimated ability to detect trends is based on the central t approximation.

Table 12 shows critical values of δ for different combinations of degrees of freedom, significance level and power. The table can be used to check how changes in the amount of available data, differences in residual standard deviation or size of trend will affect the available power.

Table 12. Critical values of test value (noncentrality parameter) for two—sided t test as function of degrees of freedom, significance level and power

Significance level α	df	power (1-β)						
		0.95	**0.9**	**0.85**	**0.8**	**0.7**	**0.6**	**0.5**
0.05	4	4.91	4.31	3.97	3.72	3.35	3.05	2.78
	5	4.59	4.05	3.73	3.49	3.13	2.84	2.57
	6	4.39	3.89	3.58	3.35	3.00	2.71	2.45
	7	4.26	3.78	3.48	3.26	2.91	2.63	2.36
	8	4.17	3.70	3.41	3.19	2.85	2.57	2.31
	10	4.04	3.60	3.32	3.11	2.77	2.49	2.23
	12	3.96	3.54	3.26	3.05	2.72	2.44	2.18
	16	3.87	3.46	3.19	2.98	2.65	2.38	2.12
	20	3.81	3.41	3.15	2.95	2.62	2.34	2.09
	25	3.77	3.38	3.12	2.92	2.59	2.32	2.06
	50	3.68	3.31	3.06	2.86	2.54	2.26	2.01
0.1	4	4.26	3.67	3.32	3.07	2.70	2.40	2.13
	5	4.03	3.49	3.17	2.93	2.57	2.28	2.02
	6	3.89	3.38	3.08	2.85	2.50	2.21	1.94
	8	3.72	3.26	2.97	2.75	2.41	2.12	1.86
	7	3.79	3.31	3.01	2.79	2.44	2.16	1.89
	10	3.62	3.18	2.91	2.69	2.35	2.07	1.81
	12	3.56	3.14	2.87	2.65	2.32	2.04	1.78
	16	3.49	3.08	2.82	2.61	2.28	2.00	1.75
	20	3.45	3.05	2.79	2.58	2.26	1.98	1.72
	25	3.42	3.02	2.77	2.56	2.24	1.96	1.71
	50	3.35	2.97	2.72	2.52	2.20	1.93	1.68

An example:

For a dataset with df=10 degrees of freedom, if we test with α=0.05 a trend that gives δ=3.60 will be detected with power = 0.9. The (given the residual variance in the data and the time span of data involved in calculating the test value). (If testing many such data sets, the trend wold be detected in 90 % of the cases). If the degrees of freedom is reduced to 4 because of less data, but other factors are unchanged, the power for detecting the same trend will be about 0.8.

If the test is a linear regression test on yearly data, and the residual standard deviation increases by 25 % due to for instance fewer observations behind each yearly average, but the data time series is otherwise the same, δ will be reduced by 20 %, from 3.60 to 2.70. The power to detect the same trend will then be reduced from 90 % to a little below 70 %.

C.2 Application to linear trend test

For a linear regression test, the trend detection ability is defined by the equation

$$\frac{|b|}{s} = \left(\frac{12h}{(T-1)(T-1+h)(T-1+2h)}\right)^{1/2} \cdot \left(t(0)_{1-\beta,\nu} + t(0)_{1-\alpha/2,\nu}\right) \qquad (39)$$

The equation defined the relation between the smallest absolute size of a trend b that will be detected with power $1-\beta$ when testing with significance level α, and the time series covers T years with h years between succeeding values and residual standard deviation s for data points around the linear trend, and otherwise fulfilling the assumptions in least squares linear regression. The right hand side of the equation is simply the ratio between detectable trend and residual standard deviation of the time series.

Table 13 shows this ratio for different combinations of time series length and frequency and different requirements for significance level and power.

With yearly data over 10 years, the ratio varies from 0.41 to 0.30, depending on the choice of α and β. With $\alpha = 0.05$, a trend of 5 % per year (about 0.05 per year in natural logarithms) can be detected with power $1-\beta = 0.9$ if the residual standard deviation on natural log scale is ≤0.12 (since 0.41·0.12 = 0.0492). With 2 years monitoring interval, the time series would have to cover 3 more years to achieve as good detection ability (it would then be somewhat better than 10 years of yearly data). With 3 year intervals, 16 years would be needed (which would give better ability than 10 years of yearly data).

Yearly data over 20 years allows detection of a 5 % change per year with the same significance and power if the residual standard deviation is within about 0.3 on natural log scale. The same trend detection ability is reached after 25 years of monitoring each second year, and 28 years of monitoring each 3rd year.

Table 13. Trend detection ability for linear regression as ratio between detectable trend and residual standard deviation a function of data span and monitoring frequency.

Monitoring time span T (years)	Monitoring interval h (years)	Number of data points n=(T-1)/h+1	0.05	0.1	0.1	Significance level α
			0.9	0.9	0.8	Trend detection power $1-\beta$
10	1	10	**0.41**	**0.36**	**0.30**	**Ratio between detectable trend *b* and residual standard deviation *s***
15	1	15	**0.21**	**0.19**	**0.16**	
20	1	20	**0.13**	**0.12**	**0.10**	
11	2	6	**0.52**	**0.44**	**0.37**	
13	2	7	**0.38**	**0.33**	**0.28**	
19	2	10	**0.20**	**0.18**	**0.15**	
25	2	13	**0.13**	**0.12**	**0.10**	
10	3	4	**0.92**	**0.72**	**0.59**	
13	3	5	**0.51**	**0.42**	**0.35**	
16	3	6	**0.34**	**0.29**	**0.24**	
22	3	8	**0.20**	**0.17**	**0.15**	
28	3	10	**0.14**	**0.12**	**0.10**	

Appendix D. Estimating the effect of design changes on between-year variance from analysis of variance (ANOVA) with two nested, random random factors.

In this appendix, the ANOVA model is analysed with the purpose of using ANOVA results from an existing dataset to assess between-year variation for other sampling designs than the one that has been used to collect the data. The discussion is based on a nested model, with a factor A that is thought to represent year, as a random factor since we are interested in the irregular between-year variation, and a factor B nested within A to represent the main samples (sites, repeated visits) each year; each main sample (B level) can be divided in a number of subsamples or replicates for chemical analysis.

The analysis leads to formulas for estimates of the absolute size of between-year variation for new designs and for the relative change of between-year variation when the design is changed, and also for giving confidence limits for these estimates.

The results can also be used for data without a nested design, with a number of independent replicate samples per year that are not grouped in main samples. This is described explicitly below.

The formulas for estimated fractiles (e.g. median and confidence limits) rely on the assumption that deviations from expectation values in the ANOVA model are normally distributed (for log-transformed values).

D.1 Expressions for mean square terms

Consider an ANOVA model with two random levels: one main factor A and another factor B nested in A. The expected mean squares in the ANOVA table for such a model is defined as functions of the variance[1] components for A, for B within A and for residual variance (Error) for replicates within B(A).

If the data are complete and balanced, with N different levels of factor A, S levels of factor B for each level of A, and R observations for each A*B combination, the Expected Mean Squares (EMS) for different levels are mutually independent and defined as:

Factor	Expected mean squares	Degrees of freedom
A	$EMS_A = S \cdot R \cdot \sigma_A^2 + R\sigma_B^2 + \sigma^2$	N-1
B (within A)	$EMS_B = R\sigma_B^2 + \sigma^2$	$N(S$-1)
Error (within B)	$EMS_r = \sigma^2$	$NS(R$-1)

Here σ_A^2 is the variance of "true" population averages across different levels of A, σ_B^2 is the average variance for "true" values of the different B levels within each A level, and σ^2 is the average variance for different observations within a combination A*B. The variance terms are not known; the ANOVA table lists mean squares (*MS*) based on the analysed data as estimates of the *EMS* statistics, and the *MS* values can be used to estimate the variances.

[1] Variance of a stochastic variable is the expected mean square of deviations from expected mean. The standard deviation is the square root of the variance.

If the dataset is unbalanced, with different number of *B* levels for each A level, or different number of observations for different combinations of A and B, the coefficients on the variance terms are different from the ones listed above, and the estimated expected mean squares will not always be independent. As a more general form of the table above, which also covers unbalanced designs, we can write:

Factor	Expected mean squares	Degrees of freedom
A	$EMS_A = k_1\sigma_A^2 + k_2\sigma_B^2 + \sigma^2$	ν_A
B (within A)	$EMS_B = k_3\sigma_B^2 + \sigma^2$	ν_B
Error (within B)	$EMS_{error} = \sigma^2$	ν_r

The ANOVA table will list the *MS* estimates and the effective degrees of freedom for the different levels, and available ANOVA results also include the coefficients k_i in a table of Expected Mean Square Coefficient (Statistica v.10, GLM module).

D.2 Estimating variance components

The equations can be solved to give expressions for the variance terms σ^2 as functions of the *EMS* terms, and by using the mean square estimates *MS* from the ANOVA table instead of the unknown *EMS* we get estimates s^2 for the variance components σ^2:

$$s_A^2 = \hat{\sigma}_A^2 = \frac{MS_A - \frac{k_2}{k_3}MS_B - \left(1 - \frac{k_2}{k_3}\right)MS_{error}}{k_1} \tag{40}$$

$$s_B^2 = \hat{\sigma}_B^2 = \frac{MS_B - MS_{error}}{k_3} \tag{41}$$

The uncertainty of the estimated variance components compared to the unknown "true" values can be calculated using the Satterthwaite approximation (Milliken and Johnson 1992). This approximation says that if we have a series of independent mean square estimates MS_i^2 i=1,…,N with degrees of freedom ν_i, and combine them into a weighted sum:

$$Q = \sum_{i=1}^{N} k_i MS_i^2 \tag{42}$$

then

$$\nu_Q = \frac{Q^2}{\sum_{i=1}^{N}\left(k_i MS_i^2\right)^2 \nu_i^{-1}} \tag{43}$$

is the effective degrees of freedom for an approximately Chi-square distributed stochastic variable $\nu_Q Q/E(Q)$, where $E(Q)$ is the expectation value of Q:

$$E(Q) = \sum_{i=1}^{N} k_i EMS_i^2 \tag{44}$$

If all k_i are positive, ν_Q will be larger than any of the individual ν_i. If some k_i are positive and some are negative, the degrees of freedom may be lower, and in some cases even less than 1.

The Satterthwaite approximation can be used for linear combinations of any independent quadratic forms s_i^2 that except for scaling factors are Chi-square distributed, for instance variance estimates with expected values σ_i^2:

$$Q = \sum_{i=1}^{N} k_i s_i^2 \qquad \nu_Q = \frac{Q^2}{\sum_{i=1}^{N} \left(k_i s_i^2\right)^2 \nu_i^{-1}} \qquad E(Q) = \sum_{i=1}^{N} k_i \sigma_i^2 \tag{45}$$

Applied to the variance component estimates from the ANOVA model, we get that the stochastic variables $\nu_{sA} s_A^2 / \sigma_A^2$ and $\nu_{sB} s_B^2 / \sigma_B^2$ are Chi-square distributed with ν_{sA} and ν_{sB} degrees of freedom calculated as:

$$\nu_{sA} = \frac{\left(k_1 s_A^2\right)^2}{\left(MS_A\right)^2 / \nu_A + \left(\frac{k_2}{k_3} MS_B\right)^2 \Big/ \nu_B + \left(\left(1 - \frac{k_2}{k_3}\right) MS_{error}\right)^2 \Big/ \nu_{error}} \tag{46}$$

$$\nu_{sB} = \frac{\left(k_3 s_B^2\right)^2}{\left(MS_B\right)^2 / \nu_B - \left(MS_{error}\right)^2 / \nu_{error}} \tag{47}$$

D.3 Estimating mean squares for a new sampling design

To see how much a change of sampling design will affect the ability to detect trends, we need to estimate absolute size or relative change of the expected variances (EMS^*) compared to the current design, and are not interested in the change of variance components per se. Expressed as functions of the unknown variance terms in the same way as for the current design that have given the available data, the new expected mean squares are:

$$EMS_A^* = g_1 \sigma_A^2 + g_2 \sigma_B^2 + \sigma_{error}^2 \tag{48}$$

$$EMS_B^* = g_3 \sigma_B^2 + \sigma_{error}^2 \tag{49}$$

where g_i are the *EMS* coefficients for the changed design. If the new design is complete and balanced, with S samples (levels of factor B) for each A level, and R replicates (subsamples) per sample, the coefficients are $g_1 = SR; \quad g_2 = g_3 = R$.

The expressions for EMS^* can be rewritten as functions of the original *EMS* for the current design or data set with coefficients k_i as described in section D.1 :

$$EMS_A^* = f_A EMS_A + f_B EMS_B + f_{error} EMS_{error} \tag{50}$$

$$\text{where} \quad f_A = \frac{g_1}{k_1} \qquad f_B = \frac{g_2 - k_2 f_A}{k_3} \qquad f_{error} = 1 - f_A - f_B \tag{51}$$

$$EMS_B^* = \frac{g_3}{k_3} EMS_B + \left(1 - \frac{g_3}{k_3}\right) EMS_{error} \tag{52}$$

All the *EMS* terms are unknown, but we can estimate the EMS^* values for the new design from the original *MS* estimates for the current design as:

$$MS_A^* = f_A MS_A + f_B MS_B + f_{error} MS_{error} \tag{53}$$

$$MS_B^* = \frac{g_3}{k_3} MS_B + \left(1 - \frac{g_3}{k_3}\right) MS_{error} \tag{54}$$

These estimates are stochastic variables with uncertainty, just as the *MS* terms they are calculated from. Since the original *MS* terms are at least approximately mutually independent if the data set is not too unbalanced, the Satterthwaite approximation can be used to estimate approximate degrees of freedom and probability distribution for the calculated mean squares for the new design. We have that

$$\nu_{xA} \frac{MS_A^*}{EMS_A^*} = \nu_{xA} \frac{f_A MA_A + f_B MS_B + f_{error} MS_{error}}{f_A EMA_A + f_B EMS_B + f_{error} EMS_{error}} \tag{55}$$

is approximately Chi-square distributed with ν_{xA} degrees of freedom calculated as

$$\nu_{xA} = \frac{\left(f_A MS_a + f_B MS_B + f_{error} MS_{error}\right)^2}{\left(f_A MS_A\right)^2 / \nu_A + \left(f_B MS_B\right)^2 / \nu_B + \left(f_{error} MS_{error}\right)^2 / \nu_{error}} \tag{56}$$

while

$$\nu_{xB} \frac{MS_B^*}{EMS_B^*} = \nu_{xB} \frac{\frac{g_3}{k_3} MS_B + \left(1 - \frac{g_3}{k_3}\right) MS_{error}}{\frac{g_3}{k_3} EMS_B + \left(1 - \frac{g_3}{k_3}\right) EMS_{error}} = \frac{MS_B + \left(\frac{k_3}{g_3} - 1\right) MS_{error}}{EMS_B + \left(\frac{k_3}{g_3} - 1\right) EMS_{error}} \tag{57}$$

is approximately Chi-square distributed with ν_{xB} degrees of freedom calculated as

$$\nu_{xB} = \frac{\left(MS_B + \left(\frac{k_3}{g_3} - 1\right) MS_{error}\right)^2}{\frac{\left(MS_B\right)^2}{\nu_B} + \frac{1}{\nu_{error}} \left(\left(\frac{k_3}{g_3} - 1\right) MS_{error}\right)^2} \tag{58}$$

This gives approximate confidence limits for the expected mean squares of the new design, which can be converted into confidence limits for the standard deviations (*StdDev*) of variation across levels of A, and across B levels within A:

$$\sqrt{\frac{\nu_{xA} MS_A^*}{g_1 \chi^2\left(\nu_{xA}\right)_{1-\alpha/2}}} \le \left(StdDev_A^* = \sqrt{\frac{EMS_A^*}{g_1}}\right) \le \sqrt{\frac{\nu_{xA} MS_A^*}{g_1 \chi^2\left(\nu_{xA}\right)_{\alpha/2}}} \tag{59}$$

$$\sqrt{\frac{\nu_{xB} MS_B^*}{g_3 \chi^2\left(\nu_{xB}\right)_{1-\alpha/2}}} \le \left(StdDev_B^* = \sqrt{\frac{EMS_B^*}{g_3}}\right) \le \sqrt{\frac{\nu_{xB} MS_B^*}{g_3 \chi^2\left(\nu_{xB}\right)_{\alpha/2}}} \tag{60}$$

where $\chi^2(\nu)_p$ is the *p* fractile of the Chi-square distribution with ν degrees of freedom.

Inserting the *MS* estimates in the formulas for *StdDev* as function of *EMS* corresponds to setting the Chi-square variables at their expectation values, equal to the degrees of freedom. This gives estimate:

$$StdDev_A^* \cong \sqrt{\frac{EMS_A^*}{g_1}} \tag{61}$$

$$StdDev_B^* \cong \sqrt{\frac{EMS_B^*}{g_3}} \tag{62}$$

Another option, more logical as a central estimate between the bounds set by the confidence limits, is to set the Chi-square variables equal to their median values:

$$StdDev_A^* \cong \sqrt{\frac{\nu_{xA} MS_A^*}{g_1 \chi^2(\nu_{xA})_{0.5}}} \tag{63}$$

$$StdDev_B^* \cong \sqrt{\frac{\nu_{xB} MS_B^*}{g_3 \chi^2(\nu_{xB})_{0.5}}} \tag{64}$$

D.4 Estimating relative change of between-year variance

In the context of this report, it is of special interest to assess the uncertainty of the estimated relative change of the between-year variance when the design is changed, as this determines the ability to detect trends. The subsequent sections show how this can be done.

D.4.1 Nested sampling within year (more than one sample, some with replicates).

If factor *A* is the irregular between-year variation, as in most of the mussel data sets analysed in Chapter 4, the ratio of the between-year standard deviations for the new design over the current design is given by:

$$\frac{StdDev_A^*}{StdDev_A} = \sqrt{\frac{EMS_A^*/g_1}{EMS_A/k_1}} = \sqrt{1 + \frac{f_B EMS_B + f_{error} EMS_{error}}{f_A EMS_A}} \tag{65}$$

This is the factor by which the residual standard deviation in time trend testing will change if the design is changed from current practice to a new design as described above. Using the Satterthwaite approximation, we have that the statistic

$$X_1 = \nu_x \frac{f_B MS_B + f_{error} MS_{error}}{f_B EMS_B + f_{error} EMS_{error}} \tag{66}$$

is approximately Chi-square distributed with degrees of freedom

$$\nu_x = \frac{(f_B MS_B + f_{error} MS_{error})^2}{(f_B MS_B)^2/\nu_B + (f_{error} MS_{error})^2/\nu_{error}} \tag{67}$$

We also have that

$$X_2 = \frac{\nu_A MS_A}{EMS_A} \tag{68}$$

is Chi-square distributed with ν_A degrees of freedom. These equations can be combined into:

$$\frac{f_B MS_B + f_{error} MS_{error}}{MS_A} = \frac{f_B EMS_B + f_{error} EMS_{error}}{EMS_A} \left[\frac{X_1/\nu_X}{X_2/\nu_A} \right] \tag{69}$$

where the bracketed factor of the second term on the right side is approximately *F* distributed, with degrees of freedom (v_x, v_A). The equation expresses the estimated mean square ratio as a function of the true, unknown ratio, combined with a stochastic, *F* distributed term. From this equation we can define confidence limits for the true mean squares ratio as function of the estimated ratio:

$$\frac{\left(\frac{f_B MS_B + f_{error} MS_{error}}{MS_A}\right)}{F(\nu_X, \nu_A)_{1-\alpha/2}} \leq \left(\frac{f_B EMS_B + f_{error} EMS_{error}}{EMS_A}\right) \leq \frac{\left(\frac{f_B MS_B + f_{error} MS_{error}}{MS_A}\right)}{F(\nu_X, \nu_A)_{\alpha/2}} \tag{70}$$

where $F(\nu_X, \nu_A)_p$ is the p fractile of the *F* distribution. When this is combined with the definition of the standard deviation ratio defined above, we get the following confidence limits:

$$\sqrt{1 + \frac{f_B MS_B + f_{error} MS_{error}}{f_A MS_A \cdot F(\nu_X, \nu_A)_{1-\alpha/2}}} \leq \frac{StdDev_A^*}{StdDev_A} \leq \sqrt{1 + \frac{f_B MS_B + f_{error} MS_{error}}{f_A MS_A \cdot F(\nu_X, \nu_A)_{\alpha/2}}} \tag{71}$$

Inserting the *MS* estimates instead of the unknown *EMS* terms in the formulas for the *StdDev* ratio corresponds to setting the *F* variables at their expectation values, equal to 1:

$$\frac{StdDev_A^*}{StdDev_A} \cong \sqrt{1 + \frac{f_B MS_B + f_{error} MS_{error}}{f_A MS_A}} \tag{72}$$

This does not exactly correspond to the expectation value of the *StdDev* ratio, but gives reasonable estimates for the relative change of between-year standard deviation.

An alternative which is more consistent as a central estimate between the confidence limits is to set the *F* variable equal to the median value from the *F* distribution:

$$\frac{StdDev_A^*}{StdDev_A} \equiv \sqrt{1 + \frac{f_B MS_B + f_{error} MS_{error}}{f_A MS_A \cdot F(\nu_X, \nu_A)_{0.5}}} \tag{73}$$

D.4.2 One main sample per year (divided in replicate subsamples)

For the cod liver data analysed in Chapter 3, and in general for time series based on one sample per year, but divided into replicates, there is only one level. In that case the equations for level *B* can apply to the between-year irregular variation (noise), and the equations for level *A* are not used[1]. The relative change of the between-year standard deviation is now given by

$$\frac{StdDev_B^*}{StdDev_B} = \sqrt{\frac{EMS_B^*/g_3}{EMS_B/k_3}} = \sqrt{1 + \left(\frac{k_3}{g_3} - 1\right)\frac{EMS_{error}}{EMS_B}} \tag{74}$$

[1] Alternatively, one may use the equations for level *A*, but set coefficients k_i, g_i for i=2 and 3 equal to 1 (no replicates), which leads to the equivalent result.

In the identity

$$\frac{MS_{error}}{MS_B} = \frac{EMS_{error}}{EMS_B}\left[\frac{MS_{error}/EMS_{error}}{MS_B/EMS_B}\right] \tag{75}$$

the bracketed term is the ratio of two Chi-square variables each divided by its degrees of freedom, and therefore F distributed with degrees of freedom (v_{error}, v_B). This gives confidence limits for the standard deviation ratio:

$$\sqrt{1+\frac{(k_3/g_3-1)MS_{error}}{MS_B\cdot F(v_{error},v_B)_{1-\alpha/2}}} \le \frac{StdDev_B^*}{StdDev_B} \le \sqrt{1+\frac{(k_3/g_3-1)MS_{error}}{MS_B\cdot F(v_{error},v_B)_{\alpha/2}}} \tag{76}$$

As in the previous section, inserting the *MS* estimates instead of the unknown *EMS* terms in the formulas for the *StdDev* ratio to get an estimate corresponds to setting the *F* variables at their expectation values, equal to 1:

$$\frac{StdDev_B^*}{StdDev_B} \cong \sqrt{1+\frac{(k_3/g_3-1)MS_{error}}{MS_B}} \tag{77}$$

As in section D.4.1, an alternative is to set the *F* variable at its median value:

$$\frac{StdDev_B^*}{StdDev_B} \cong \sqrt{1+\frac{(k_3/g_3-1)MS_{error}}{MS_B\cdot F(v_{error},v_B)_{0.5}}} \tag{78}$$

Appendix E. Statistical properties of batch samples when individual concentrations are log-normally distributed.

When analysing contaminants in biota, batch samples composed of a number of individuals are commonly used to get a better estimate of average conditions for a given analytical cost. There may be a large variation in contaminant concentrations between single individuals at a certain time and place, and by using batch samples these variations will tend to be averaged out.

If concentrations in individual specimens are symmetrically distributed with variance σ_i^2, batch samples of N randomly selected individuals will have variance σ_i^2/N around the true population average. The expected (population) median will then be equal to the expected arithmetic mean, and will be the same for individuals and for batch samples.

For hazardous substances in biota, concentrations often show large variation between individuals, with the highest concentrations several times higher than the expected median, while the spread downward spread is limited by zero as an absolute lower bound. The distribution is then skewed, with a much longer tail towards high concentrations. For such data the log-transformed concentrations are often much more symmetrically distributed, and may be well approximated by a normal distribution; the concentrations are then log-normally distributed.

The concentration in a batch sample will still be an arithmetic average of concentrations of the individuals included in the sample, but with a skewed distribution with a long tail towards high values it may be determined almost completely by one or two individuals with very high concentrations. The expected arithmetic average is still the same for batch samples and for individuals. However, the expected median will now be lower than the arithmetic mean, and not the same for batch samples as for individuals.

With a skewed distribution of concentrations in the individuals, the residual variation of batch samples will not decrease as effectively with increasing number of included individuals as for a symmetrical distribution; this is the case both for arithmetic mean and median of batch sample values. The median value of batch samples may have smaller expected relative variance than the arithmetic mean.

The batch samples will also have a skewed distribution, although to a smaller degree than the individual concentrations. Even if the individual concentrations are exactly log-normally distributed, the batch samples will not be log-normally distributed. However, log-transformation of the batch sample values may still be give deviations from fitted trends or averages that are more symmetrically distributed and more independent of expected values, with correspondingly better properties for standard hypothesis testing and confidence limit estimation.

In order to assess what can be achieved by using batch samples, it is necessary to know, or at least estimate, how the variance of batch samples depends on number of individuals contributing to the sample. A numerical simulation has been done to answer this question. For different combinations of number of individuals per batch sample (N) and the variance (σ_i^2) of log(C) for individuals, 10 000 sets of N single values has been drawn from a lognormal distribution with average 0 for log(C) (median for C=1). The arithmetic mean of each set is calculated, to correspond to the analytical value for a batch sample. For the resulting 10 000 batch sample values, the arithmetic mean on linear concentration scale as well as average μ_b and variance σ_b^2 for log-transformed values are calculated for each combination of N and σ_i^2 .

By empirical nonlinear function fit it has been found that the variance of log-transformed batch sample concentration can be described as a function of the number of individuals in the sample and the variance for log-transformed individual concentrations:

$$\sigma_b^2 = \frac{\sigma_i^2}{N^{1/\left(1+\sigma_i^2/7.7\right)}} \tag{79}$$

This formula can be used directly to estimate the batch sample variance if the variance for individual concentrations have been estimated. Inversely, if the batch sample variance σ_b^2 has been estimated for known N, the formula can be used iteratively to estimate σ_i^2, for instance by:

$$\sigma_{(k+1)_i}^{\ 2} = \frac{1}{2}\left[\sigma_{(k)_i}^{\ 2} + \sigma_b^2 N^{1/\left(1+\sigma_{(k)_i}^{\ 2}/7.7\right)}\right] \tag{80}$$

With B batch samples based on different number of individuals N_b and with residuals s_b from some model fit, the iteration can be:

$$\sigma_{i\,(k+1)}^2 = \frac{1}{2}\left[\sigma_{i\,(k)}^2 + \frac{\sum_{b=1}^{B} s_b^2 N_b^{\,1/\left(1+\sigma_{i\,(k)}^2/7.7\right)}}{B}\right] \tag{81}$$

The iterated value of σ_i^2 can then be used to estimate σ_b^2 for other values of N.

Compared to the statistics for symmetrical distributions, an 'effective' number of individuals can be defined from the ratio between variances on log scale:

$$N_{eff} = \frac{\sigma_i^2}{\sigma_b^2} = N^{1/\left(1+\sigma_i^2/7.7\right)} \tag{82}$$

For instance, a variance 0.1 for batch samples of 50 individuals could be explained as the averaged effect of individual variance 2.14. If this was on natural log scale it would mean that about 70 % of individuals are within an interval from 0.1 to 10 times the median), and the averaging effect would be the same as for 20 individuals from a symmetric distribution. If batch samples of 50 individuals had variance 0.03 it would correspond to an individual variance of about 1.0. If this was on natural log scale if would mean that about 70 % of individuals were within a range from 0.37 to 2.7 times the median, and the averaging effect would be as for 32 individuals from a symmetric distribution.

Figure 16 shows the match between numerical simulation results and values calculated by equation (79), for a wide range of different σ_i^2, N.

In the discussion above it has been assumed that the batch sample consists of equal amounts of material from each individual. If this is not the case, the weight variation between individual contributions to the sample must also be taken into account. This may be the case for species like blue mussels, where the whole soft body of each individual is used in the sample. The results of using batch samples will then depend on whether contaminant concentrations are correlated with amount of material used.

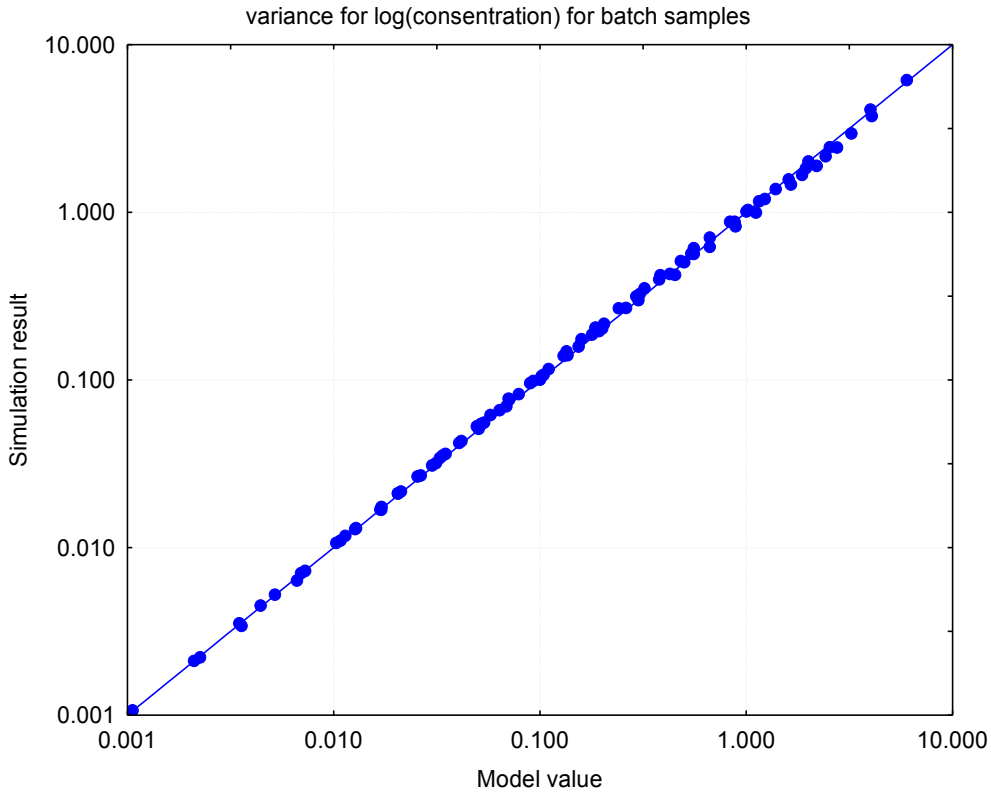


Figure 16. Relationship between numerical simulation results and model values calculated by equation (79), for a wide range of different σ_i^2, *N*.

Appendix F. Results from CEMP procedure with subsampling to simulate 2 or 3 year monitoring intervals

The graphs below shows the relation between detectable trends per station and parameter assessed by the customized CEMP procedure when using data from each year and when using data only for each second or each third year.

The detectable trend is estimated from the residual variance, degrees of freedom for the residual variance and the contrast coefficients for calculating trend from either the smoother or from a linear trend.

The horizontal axis of each plot shows results when using all data, i.e. with yearly monitoring.

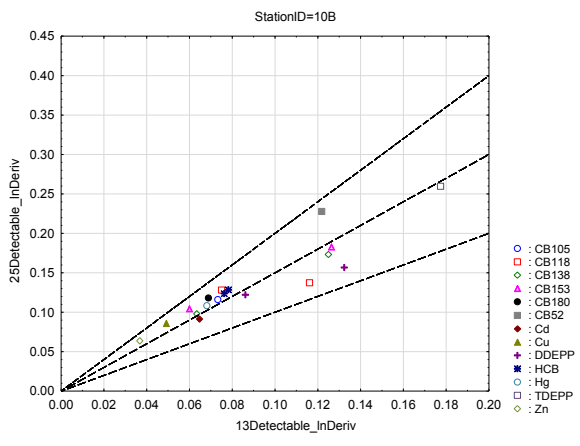
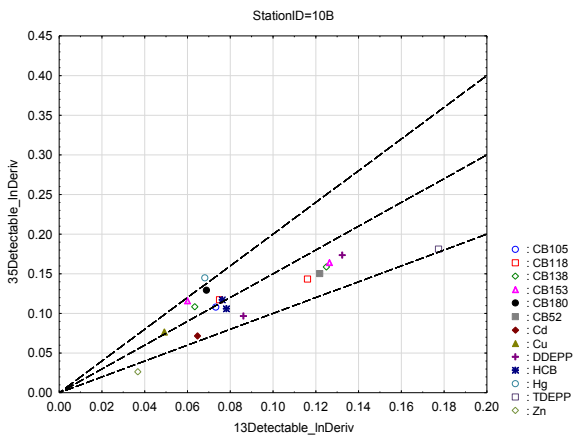
Each pair of plots shows results for one station, with the parameters that were included in the analysis according to selection criteria described in XXX.

- The plots on the left shows detectable trend for 2 year monitoring interval along the vertical axis as average over the two independent subsets for even and odd years.
- The plots on the right shows detectable trend for 3 year monitoring interval along the vertical axis as average over the three independent subsets.

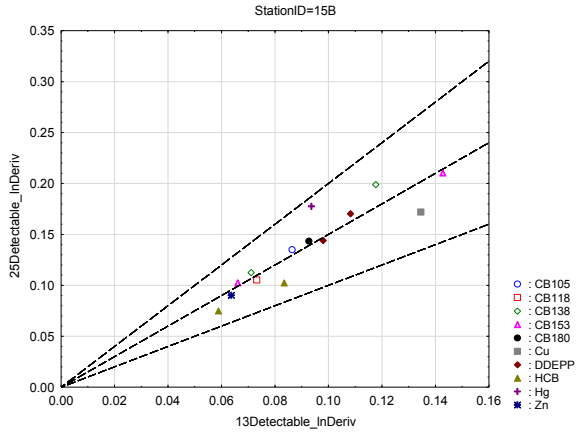
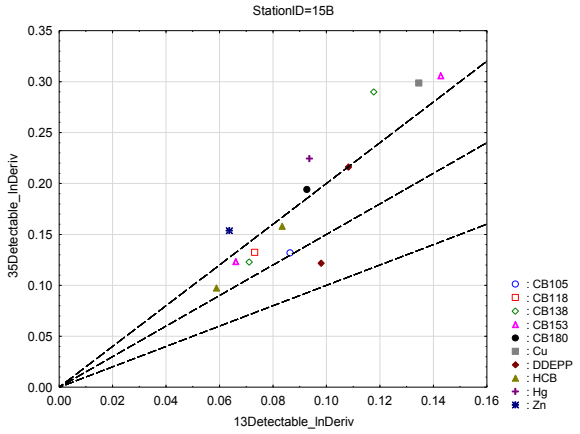
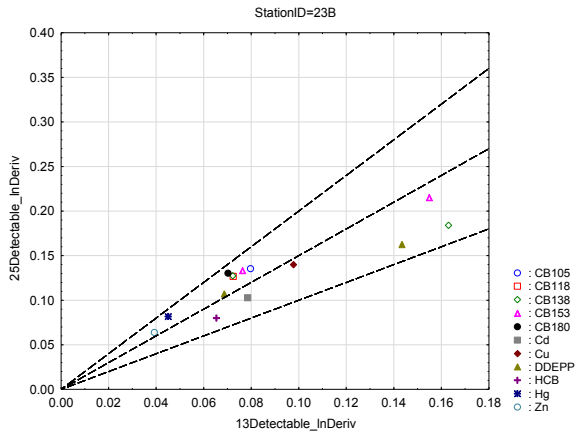
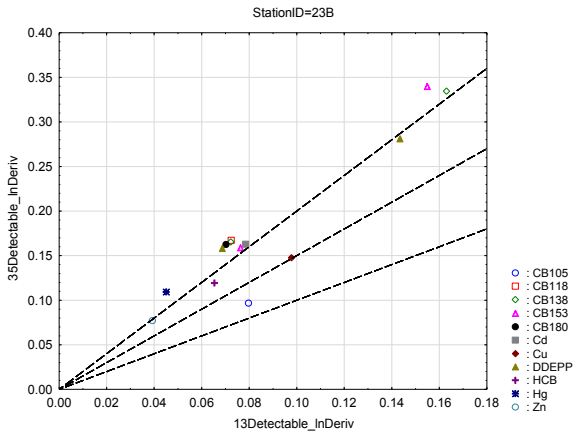
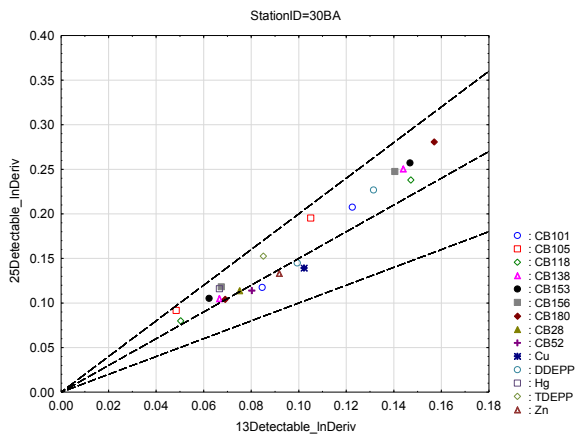
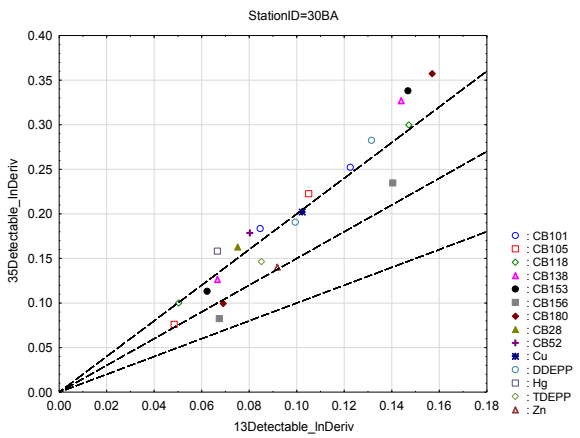
In both cases, the alternative with expanded local regression window is used.

The lines show ratios 1:1, 1.5:1 and 2:1.

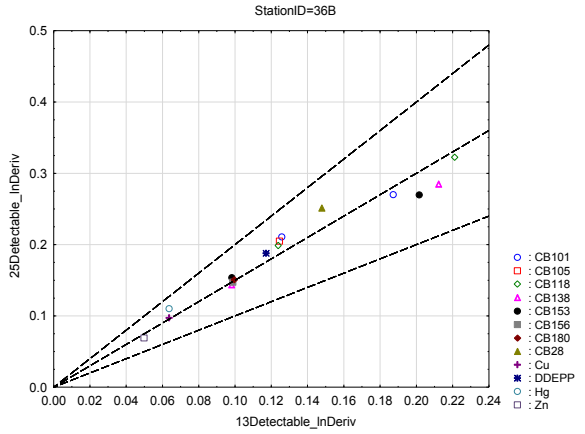
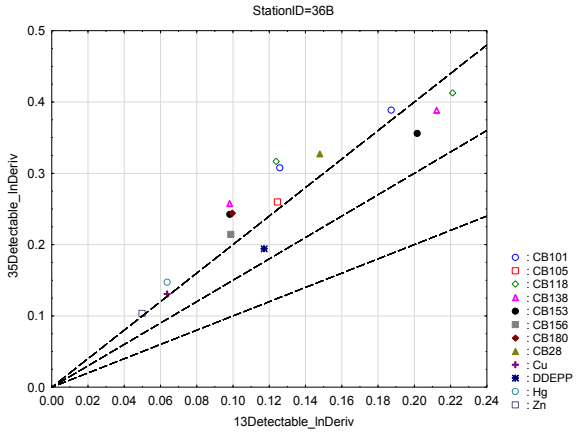
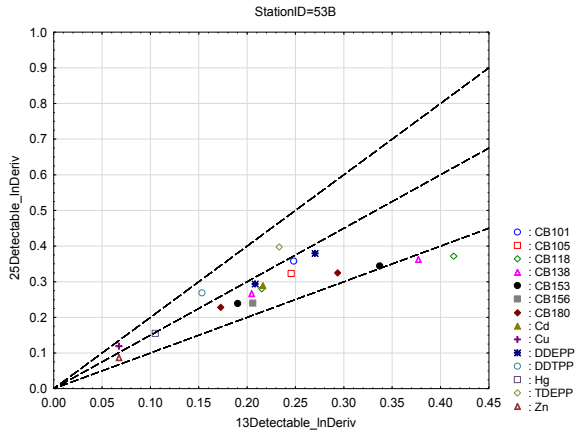
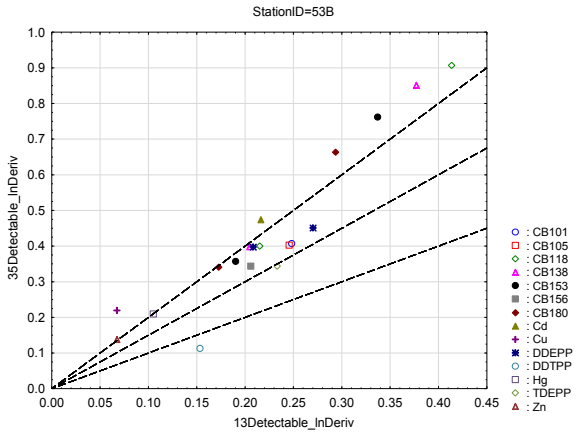
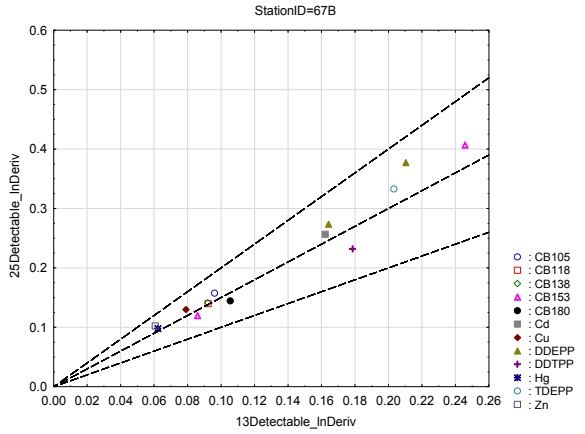
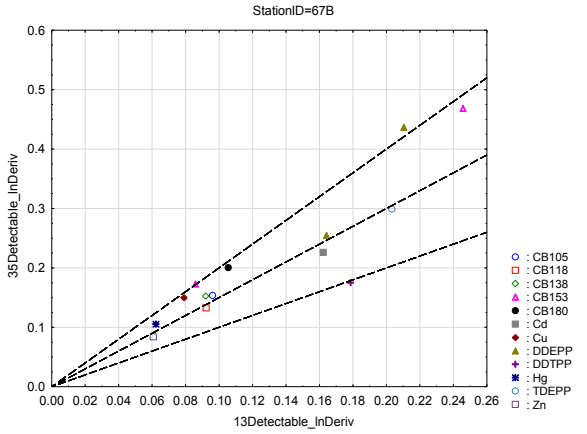
Relations between detectable trend (natural log scale) for yearly data (13Detectable_lnDeriv, horizontal axis) and for 2 and 3 year interval (vertical axis). *Straight lines show ratio 1:1, 1.5:1 and 2:1*

2 year interval, 9 year local regression window (25Detectable_lnDeriv)	3 year interval, 13 year local regression window (35Detectable_lnDeriv)
	

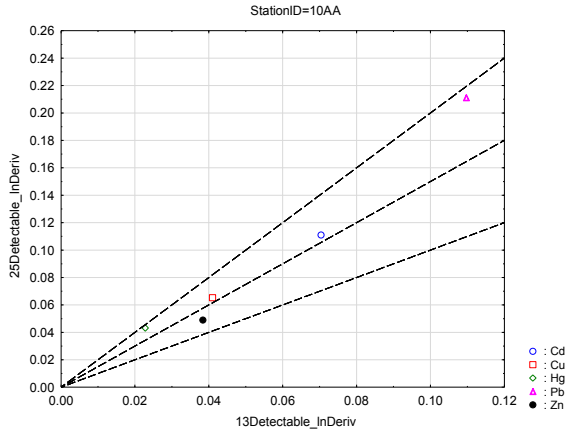
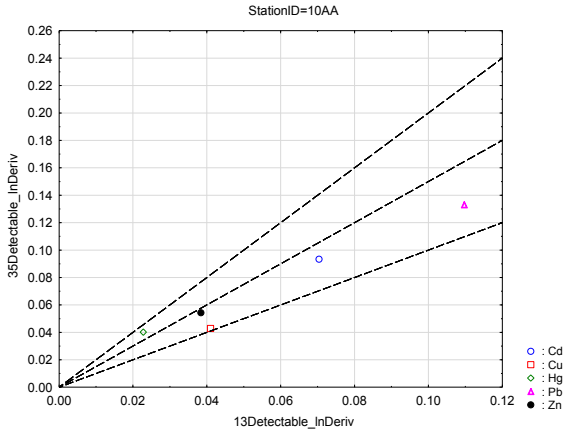
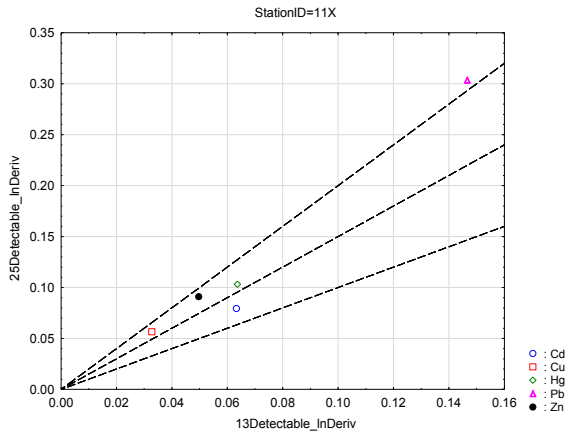
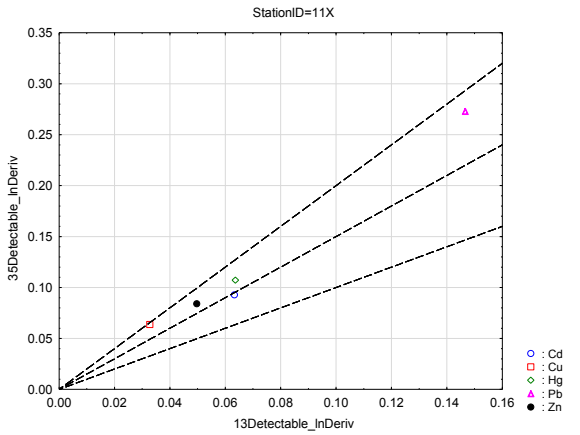
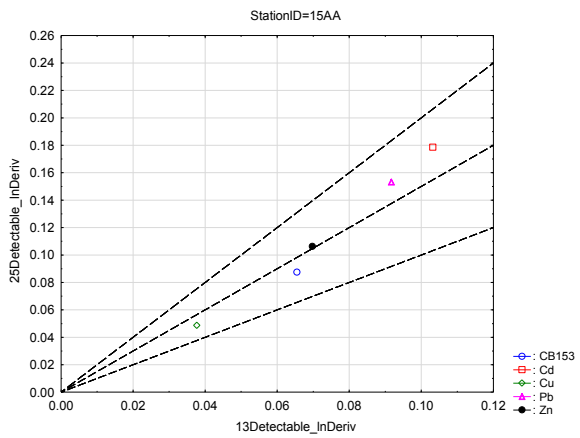
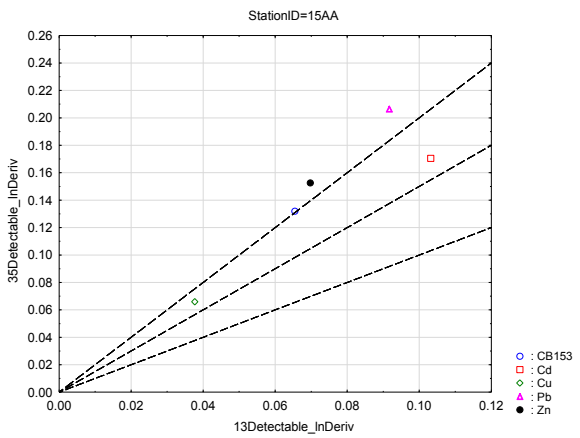
Relations between detectable trend (natural log scale) for yearly data (13Detectable_lnDeriv, horizontal axis) and for 2 and 3 year interval (vertical axis). *Straight lines show ratio 1:1, 1.5:1 and 2:1*

2 year interval, 9 year local regression window (25Detectable_lnDeriv)	3 year interval, 13 year local regression window (35Detectable_lnDeriv)
	
	
	

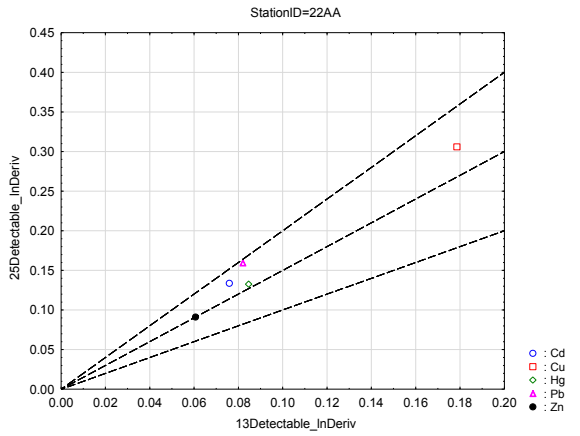
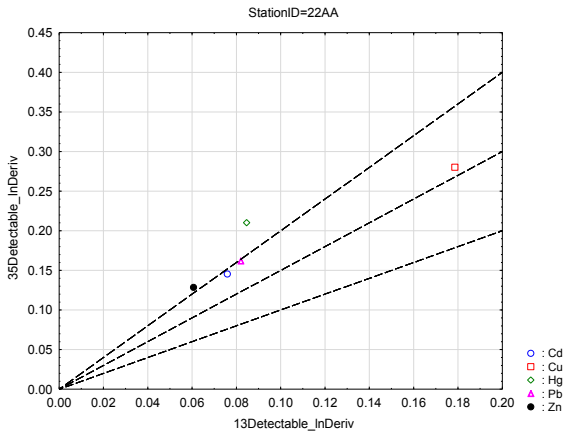
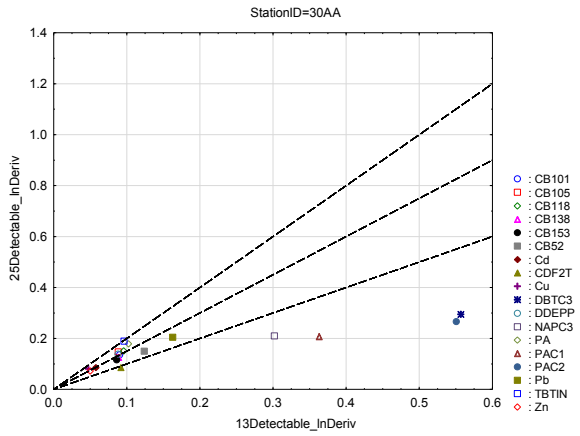
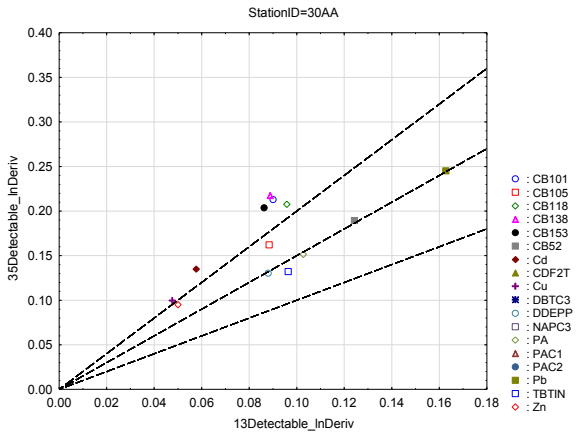
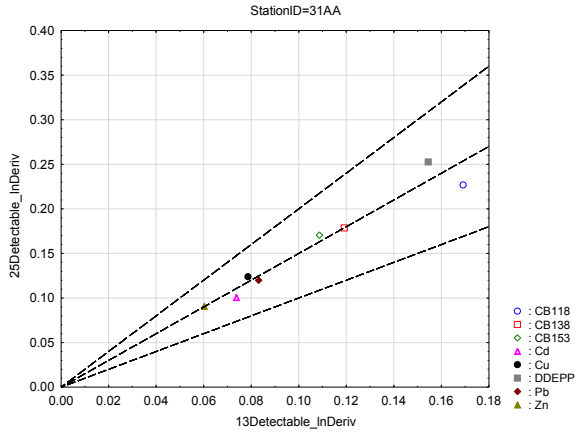
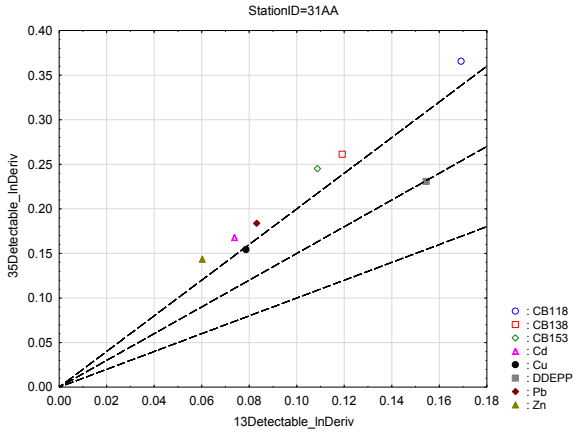
Relations between detectable trend (natural log scale) for yearly data (13Detectable_lnDeriv, horizontal axis) and for 2 and 3 year interval (vertical axis). *Straight lines show ratio 1:1, 1.5:1 and 2:1*

2 year interval, 9 year local regression window (25Detectable_lnDeriv)	3 year interval, 13 year local regression window (35Detectable_lnDeriv)
 cAS	
	
	

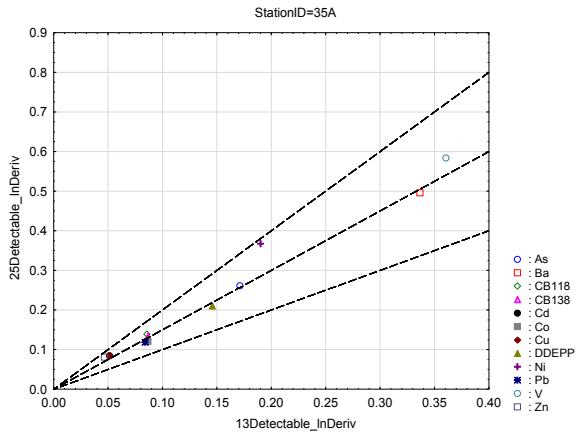
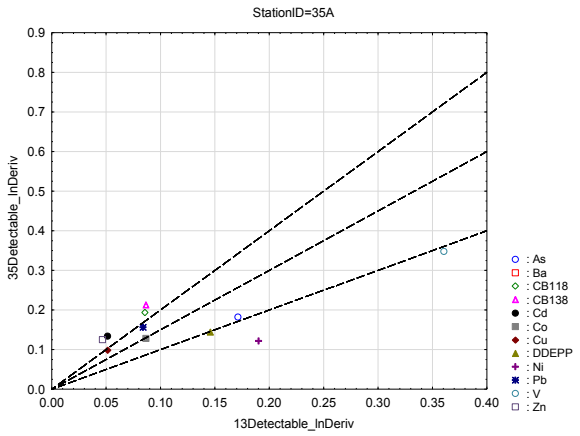
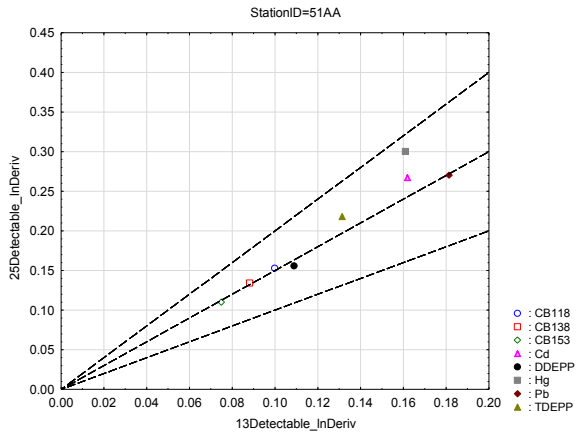
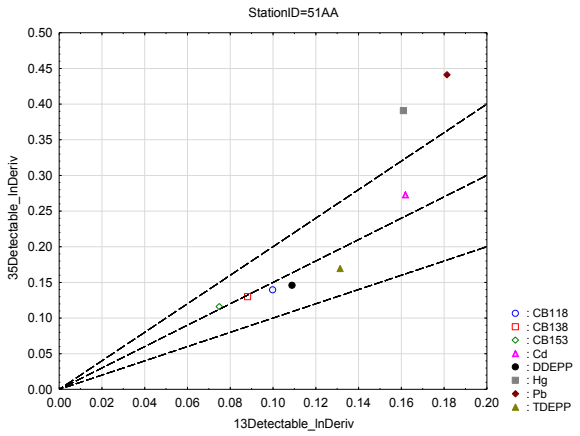
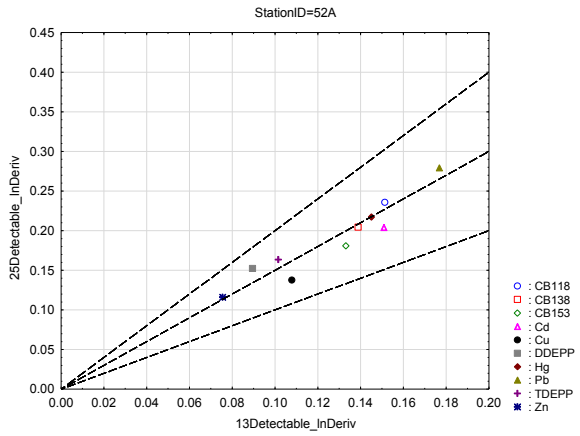
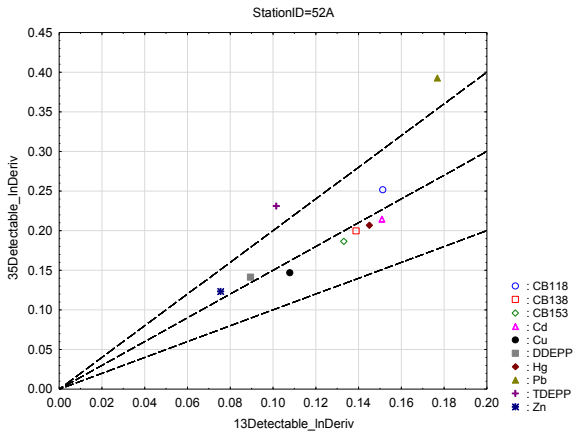
Blue Mussel stations:Relations between detectable trend (natural log scale) for yearly data (13Detectable_lnDeriv) and for 2 and 3 year interval. *Straight lines show ratio 1:1, 1.5:1 and 2:1*

2 year interval, 9 year local regression window (25Detectable_lnDeriv)	3 year interval, 13 yearlocal regression window (13Detectable_lnDeriv)
	
	
	

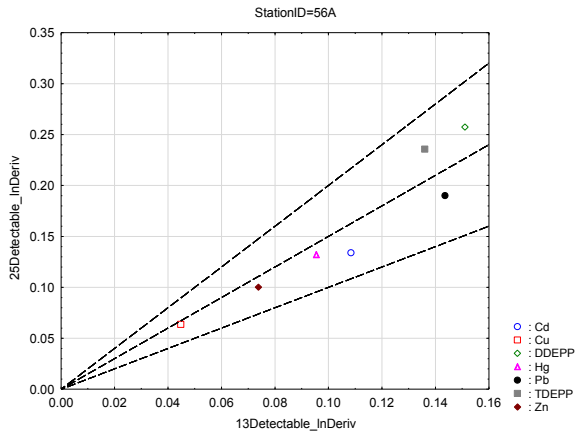
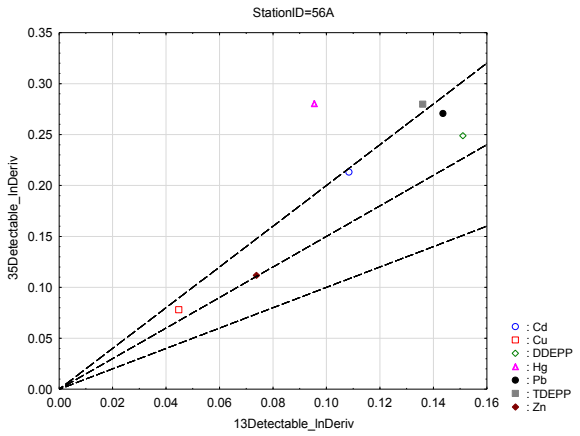
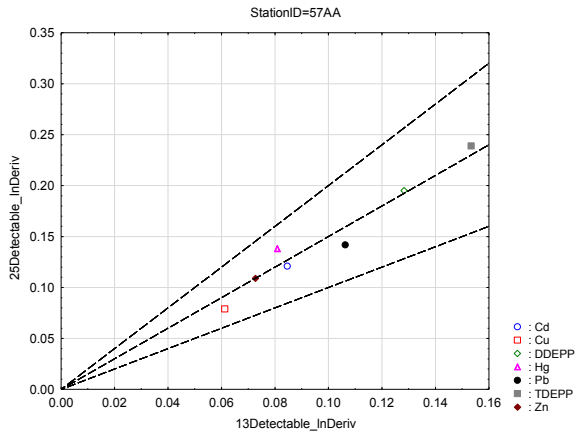
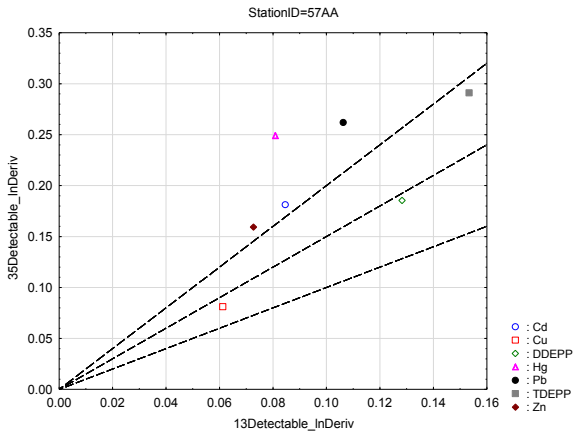
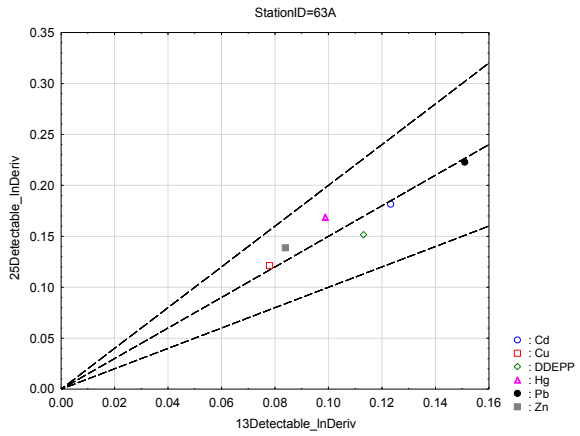
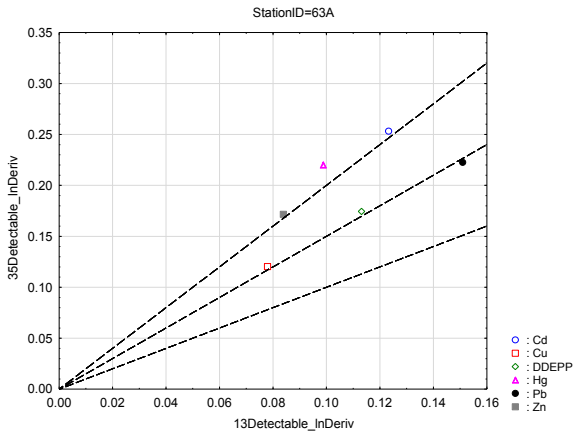
Blue Mussel stations:Relations between detectable trend (natural log scale) for yearly data (13Detectable_lnDeriv) and for 2 and 3 year interval. *Straight lines show ratio 1:1, 1.5:1 and 2:1*

2 year interval, 9 year local regression window (25Detectable_lnDeriv)	3 year interval, 13 yearlocal regression window (13Detectable_lnDeriv)
	
	
	

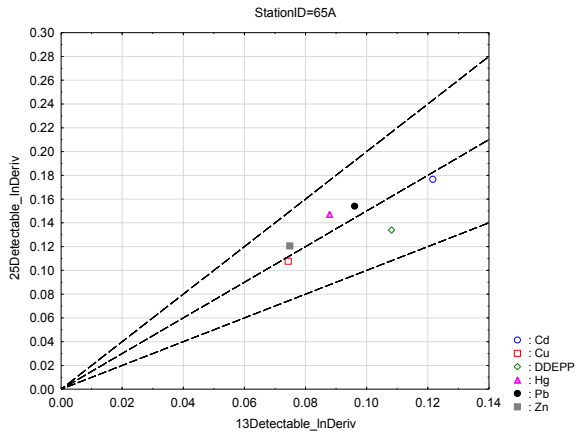
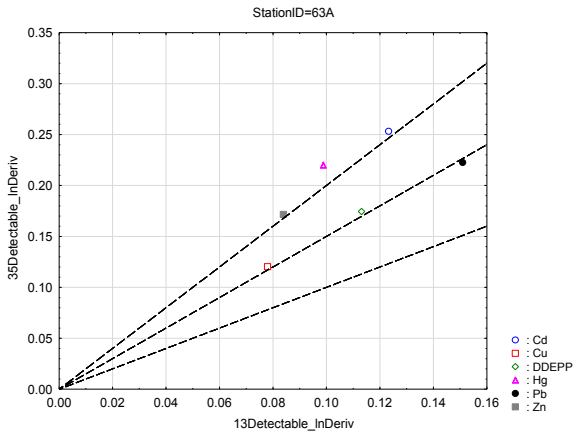
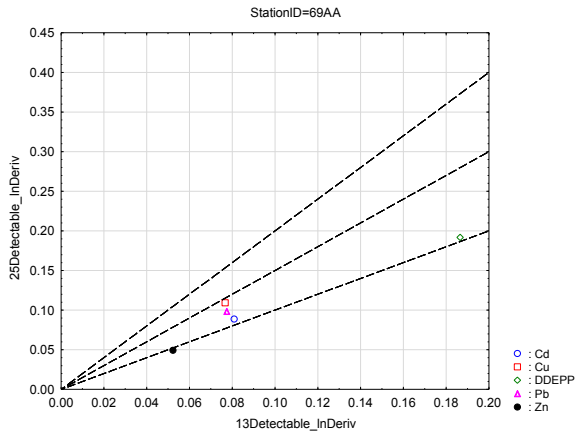
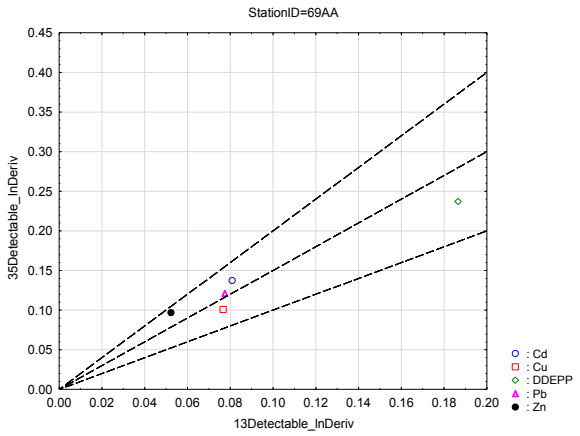
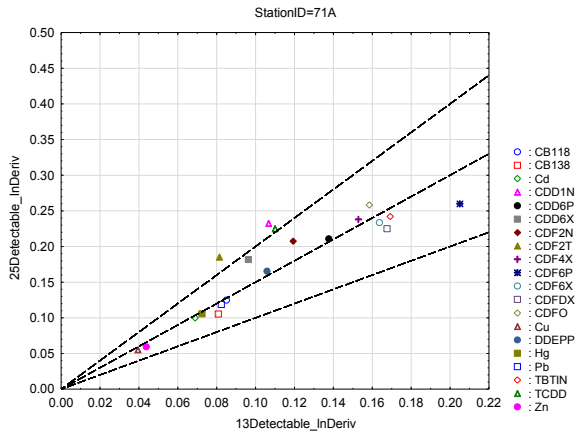
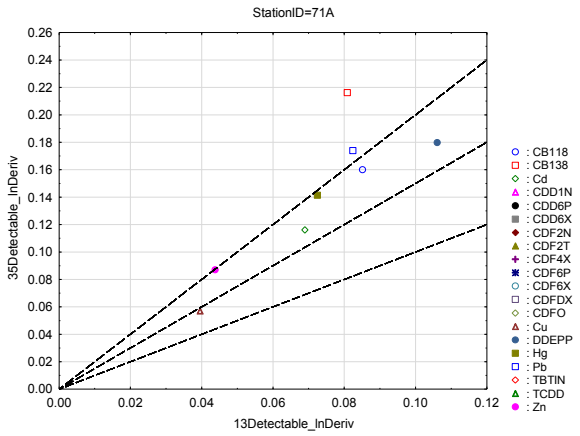
Blue Mussel stations:Relations between detectable trend (natural log scale) for yearly data (13Detectable_lnDeriv) and for 2 and 3 year interval. *Straight lines show ratio 1:1, 1.5:1 and 2:1*

2 year interval, 9 year local regression window (25Detectable_lnDeriv)	3 year interval, 13 yearlocal regression window (13Detectable_lnDeriv)
	
	
	

Blue Mussel stations:Relations between detectable trend (natural log scale) for yearly data (13Detectable_lnDeriv) and for 2 and 3 year interval. *Straight lines show ratio 1:1, 1.5:1 and 2:1*

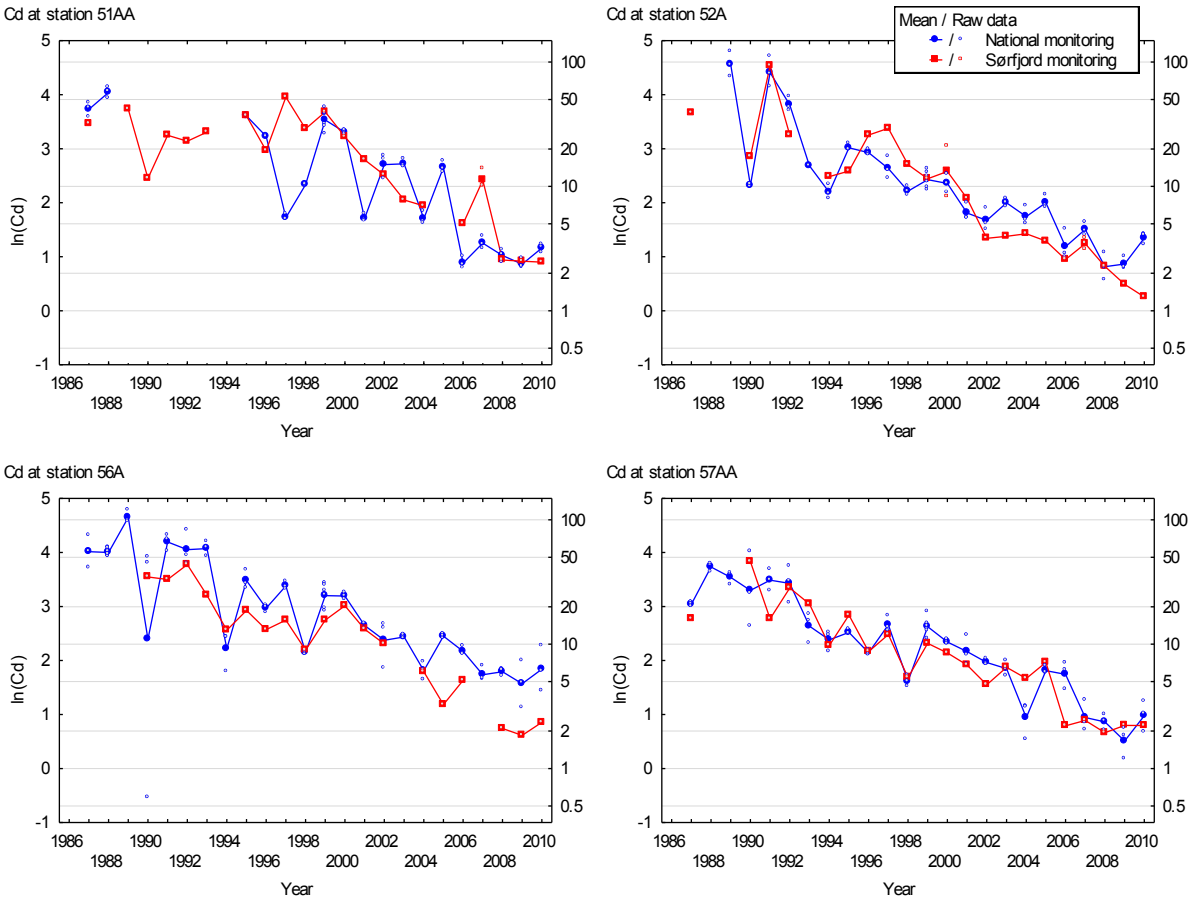
2 year interval, 9 year local regression window (25Detectable_lnDeriv)	3 year interval, 13 yearlocal regression window (13Detectable_lnDeriv)
	
	
	

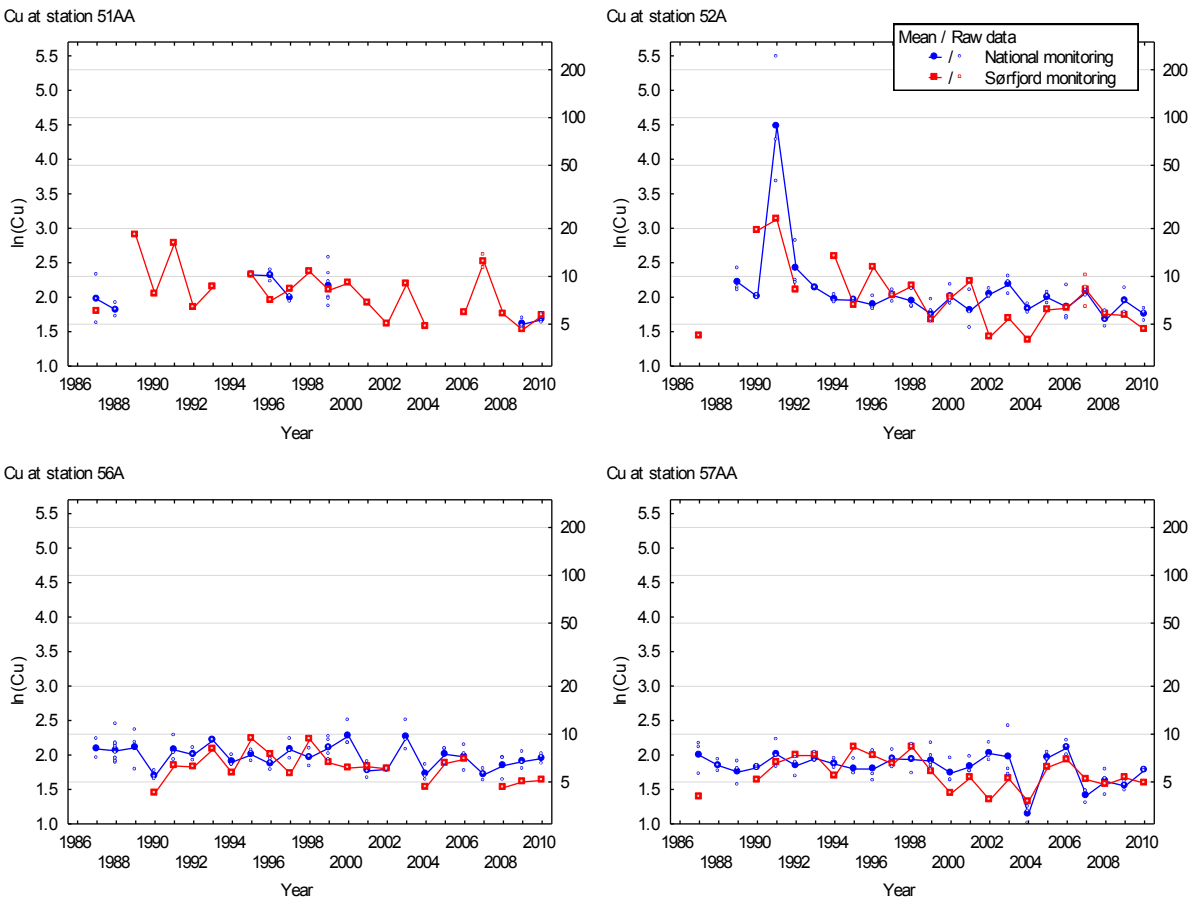
Blue Mussel stations:Relations between detectable trend (natural log scale) for yearly data (13Detectable_lnDeriv) and for 2 and 3 year interval. *Straight lines show ratio 1:1, 1.5:1 and 2:1*

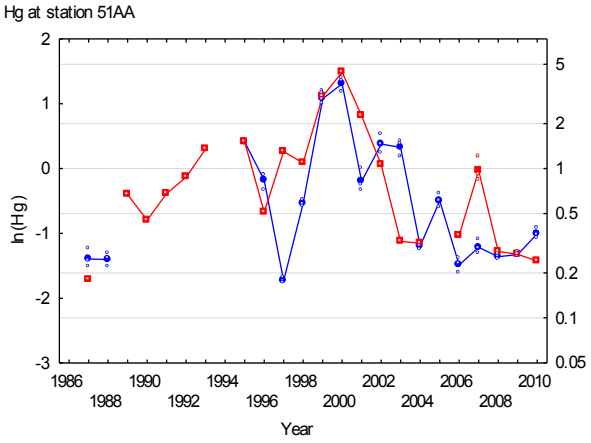
2 year interval, 9 year local regression window (25Detectable_lnDeriv)	3 year interval, 13 yearlocal regression window (13Detectable_lnDeriv)
	
	
	

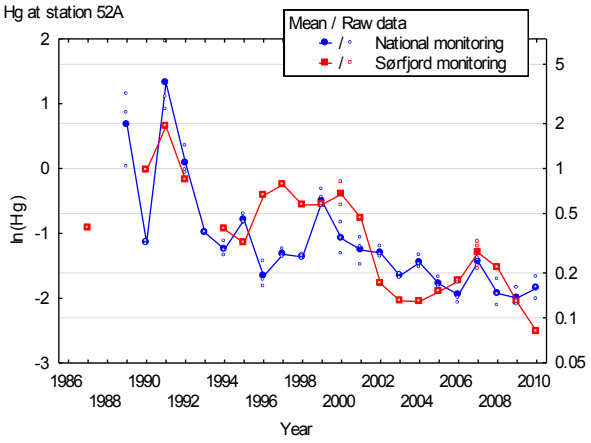
Appendix G. Time series plots of blue mussel data from Sørfjord

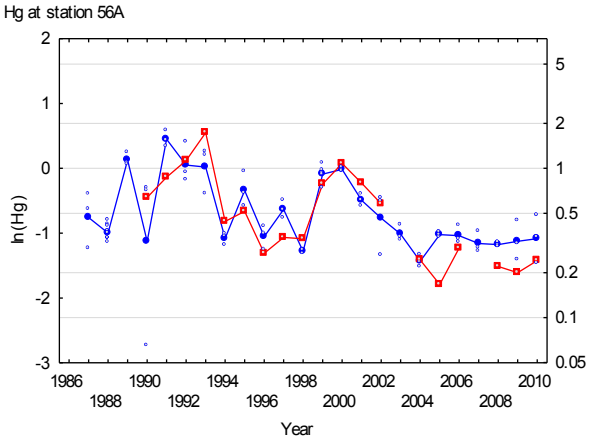
The plots below show the blue mussel time series from Sørfjord selected for the statistical analysis in Chapter 4. Each plot shows data for one station and one parameter, both yearly medians and raw data. Data series from the national CEMP program and the local Sørfjord monitoring are shown as separate time series in the same plot. The right axes show concentrations, for metals as µg/g dry weight, for the others as µg/kg wet weight. The left vertical axes show the corresponding natural logarithms used in the statistical analysis. Only data from the period 1990-2010 are used in the analyses and shown in the plots.

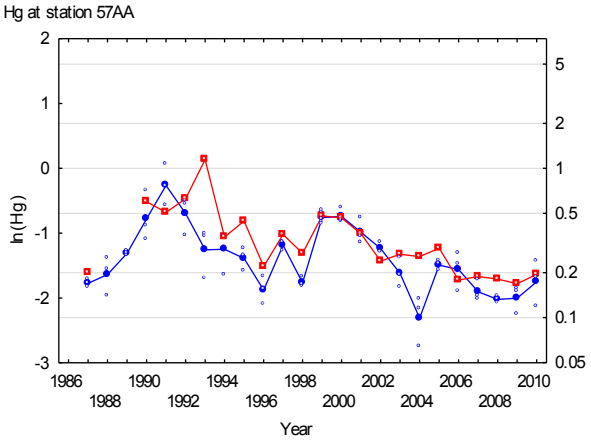

Cd at station 51AA
Cd at station 52A
Mean / Raw data
National monitoring
Sørfjord monitoring
Cd at station 56A
Cd at station 57AA
ln(Cd)
Year

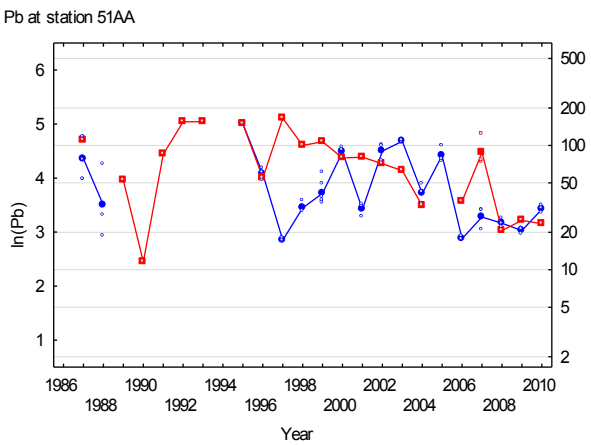

Cu at station 51AA
Cu at station 52A
Mean / Raw data
National monitoring
Sørfjord monitoring
Cu at station 56A
Cu at station 57AA
ln(Cu)
Year

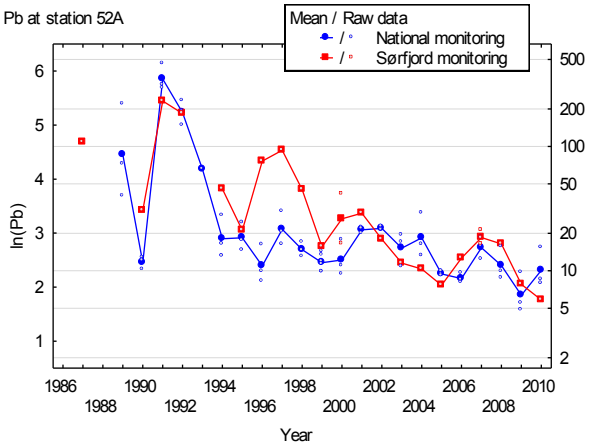

Hg at station 51AA
ln(Hg)
Year

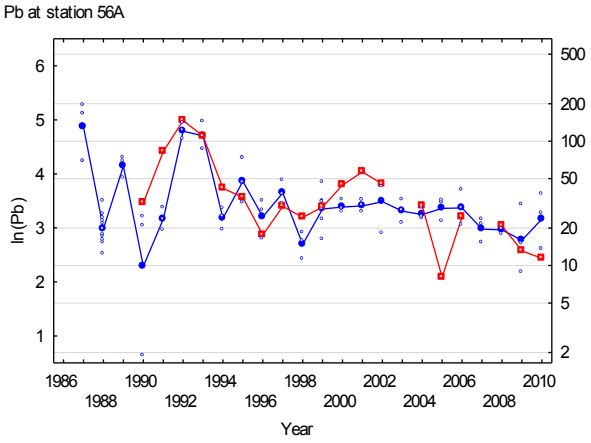

Hg at station 52A
Mean / Raw data
National monitoring
Sørfjord monitoring
ln(Hg)
Year

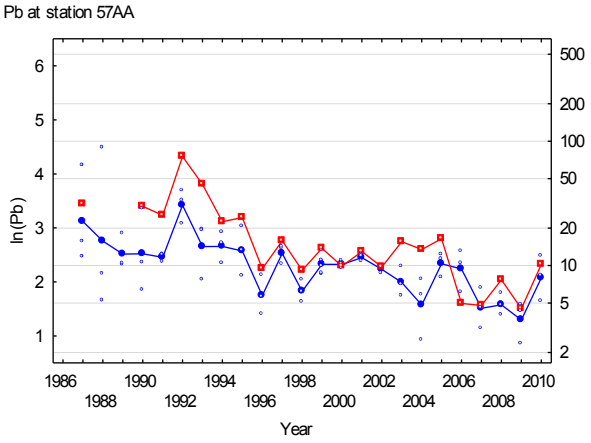

Hg at station 56A
ln(Hg)
Year

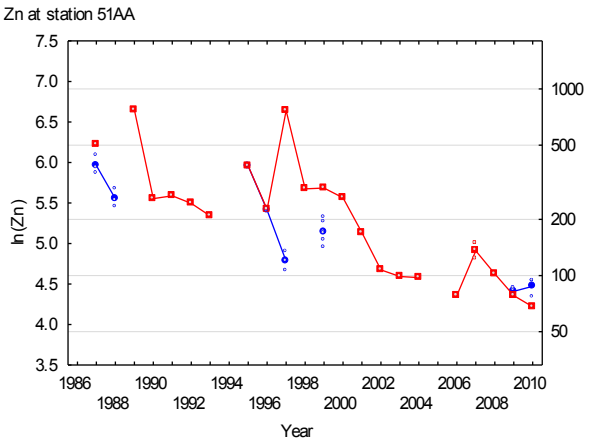

Hg at station 57AA
ln(Hg)
Year

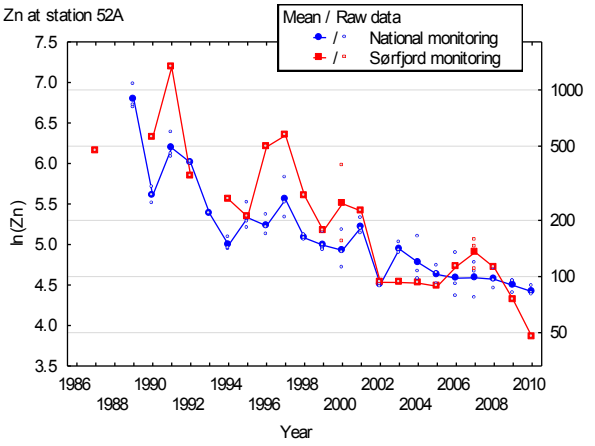

Pb at station 51AA
ln(Pb)
Year

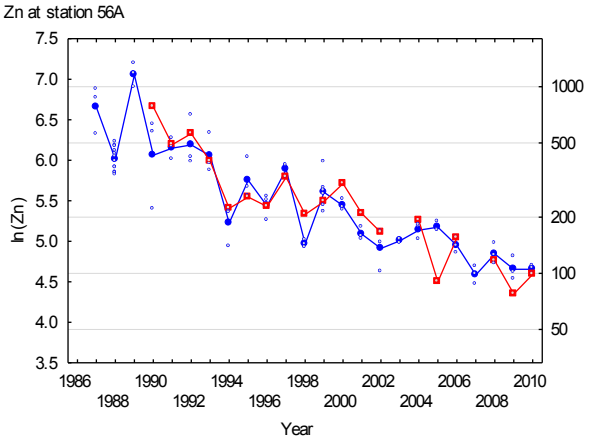

Pb at station 52A
Mean / Raw data
National monitoring
Sørfjord monitoring
ln(Pb)
Year

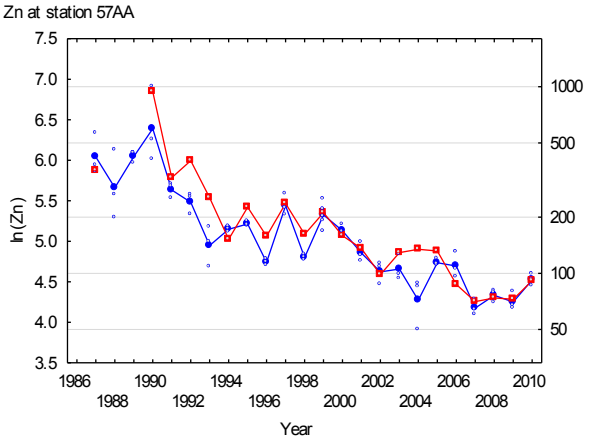

Pb at station 56A
ln(Pb)
Year

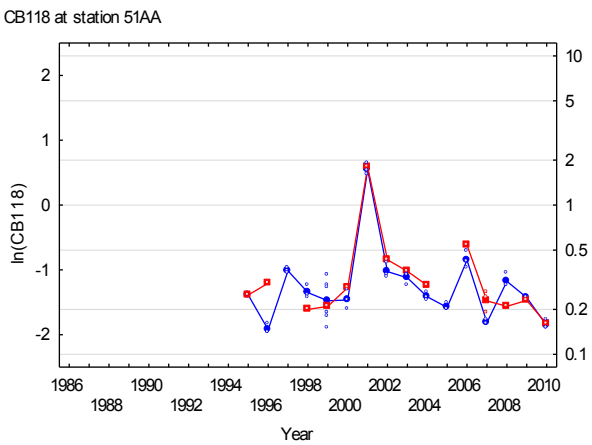

Pb at station 57AA
ln(Pb)
Year

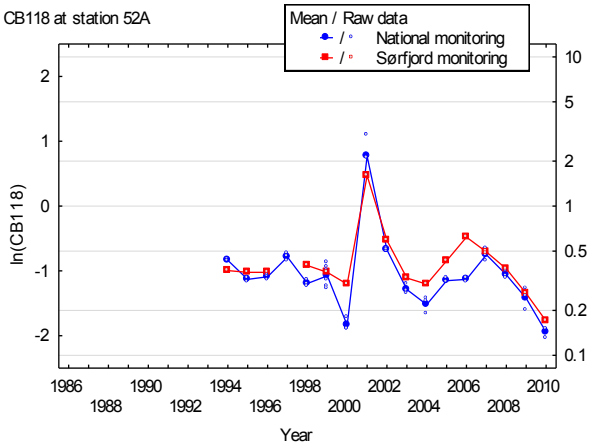

Zn at station 51AA
ln(Zn)
Year

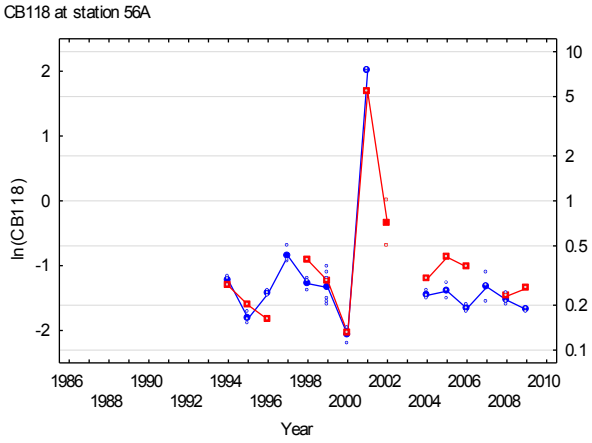

Zn at station 52A
Mean / Raw data
National monitoring
Sørfjord monitoring
ln(Zn)
Year

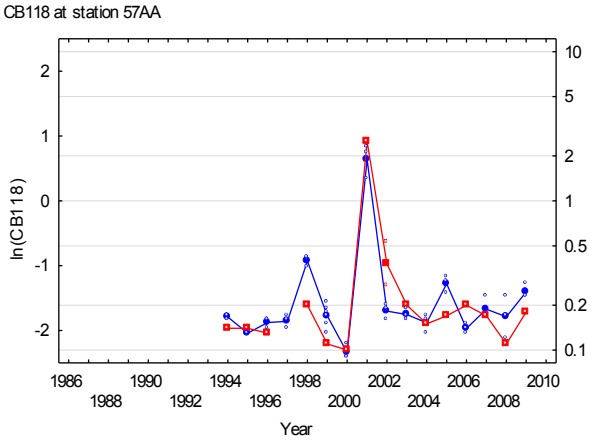

Zn at station 56A
ln(Zn)
Year

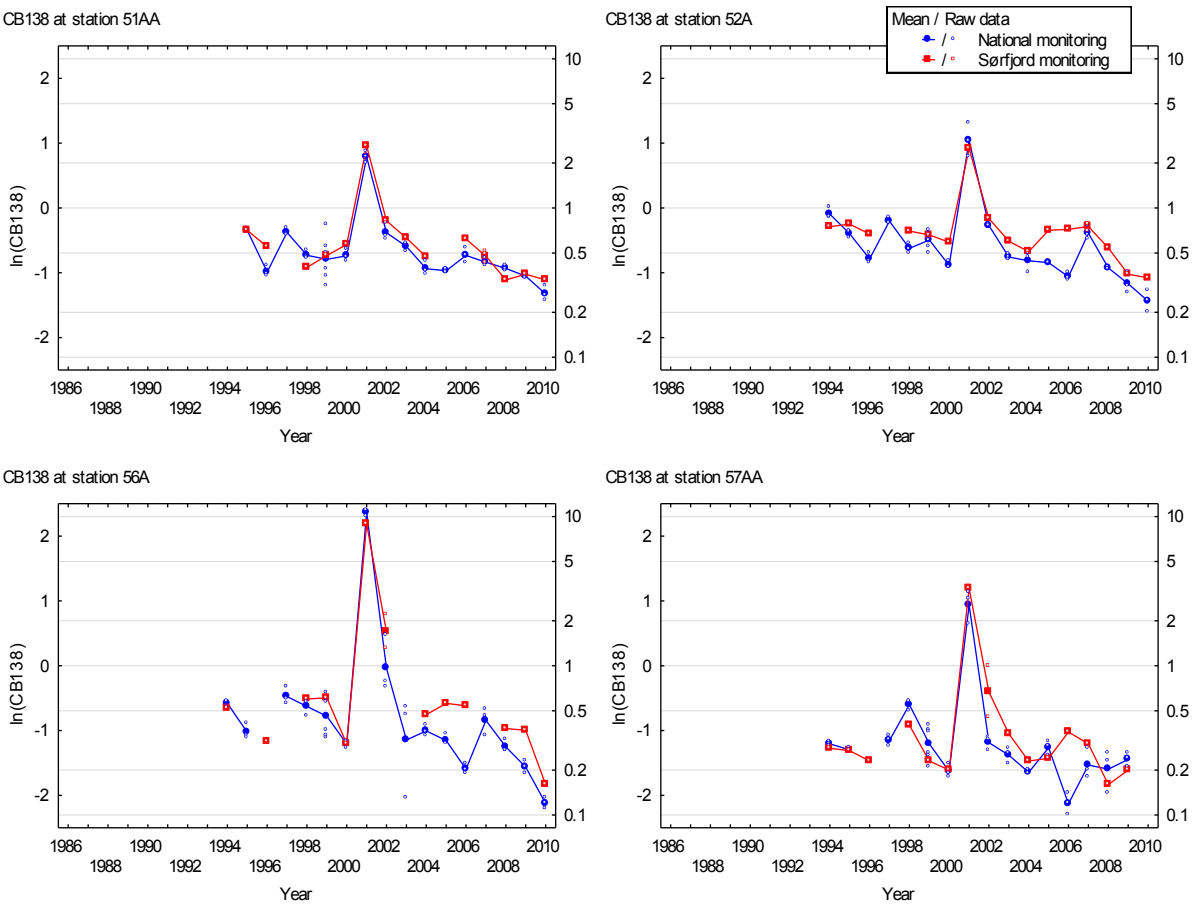

Zn at station 57AA
ln(Zn)
Year

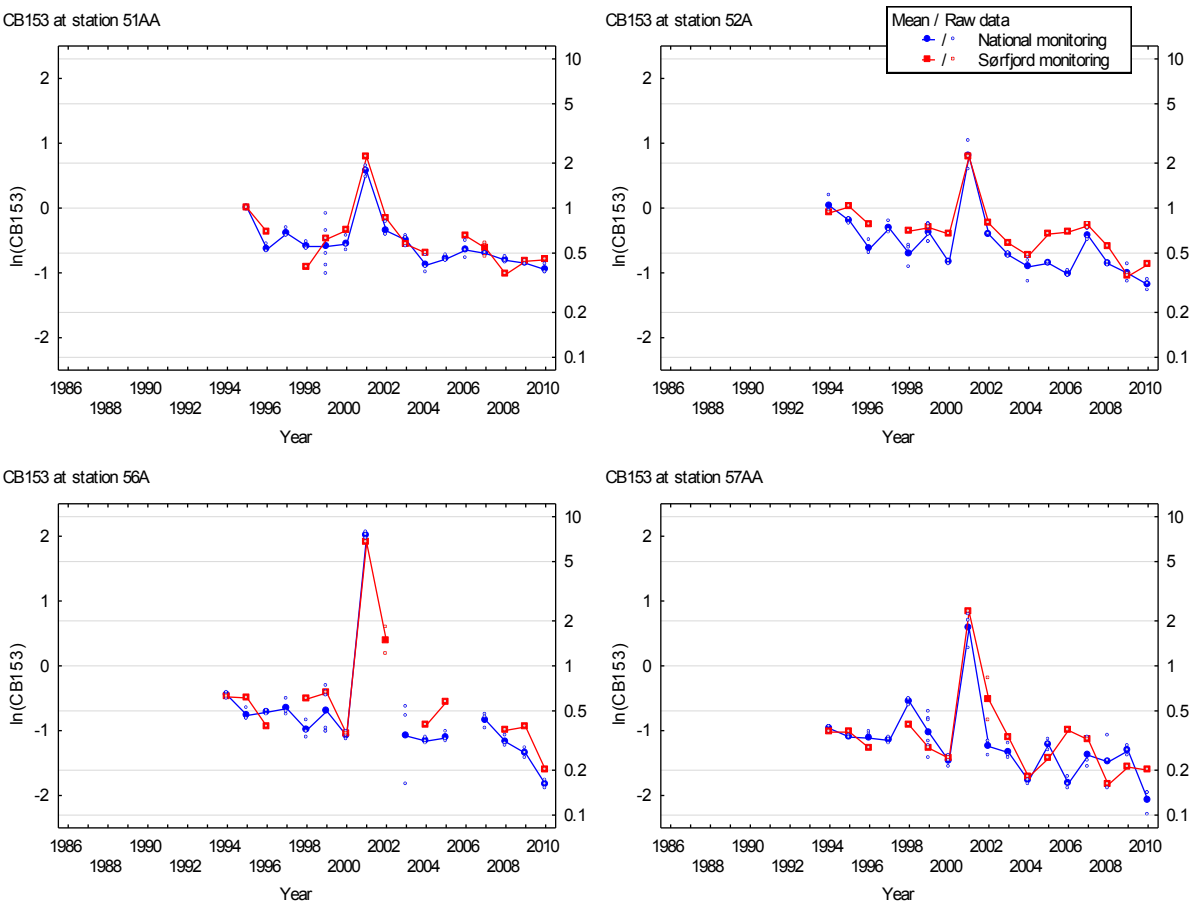

CB118 at station 51AA
ln(CB118)
Year

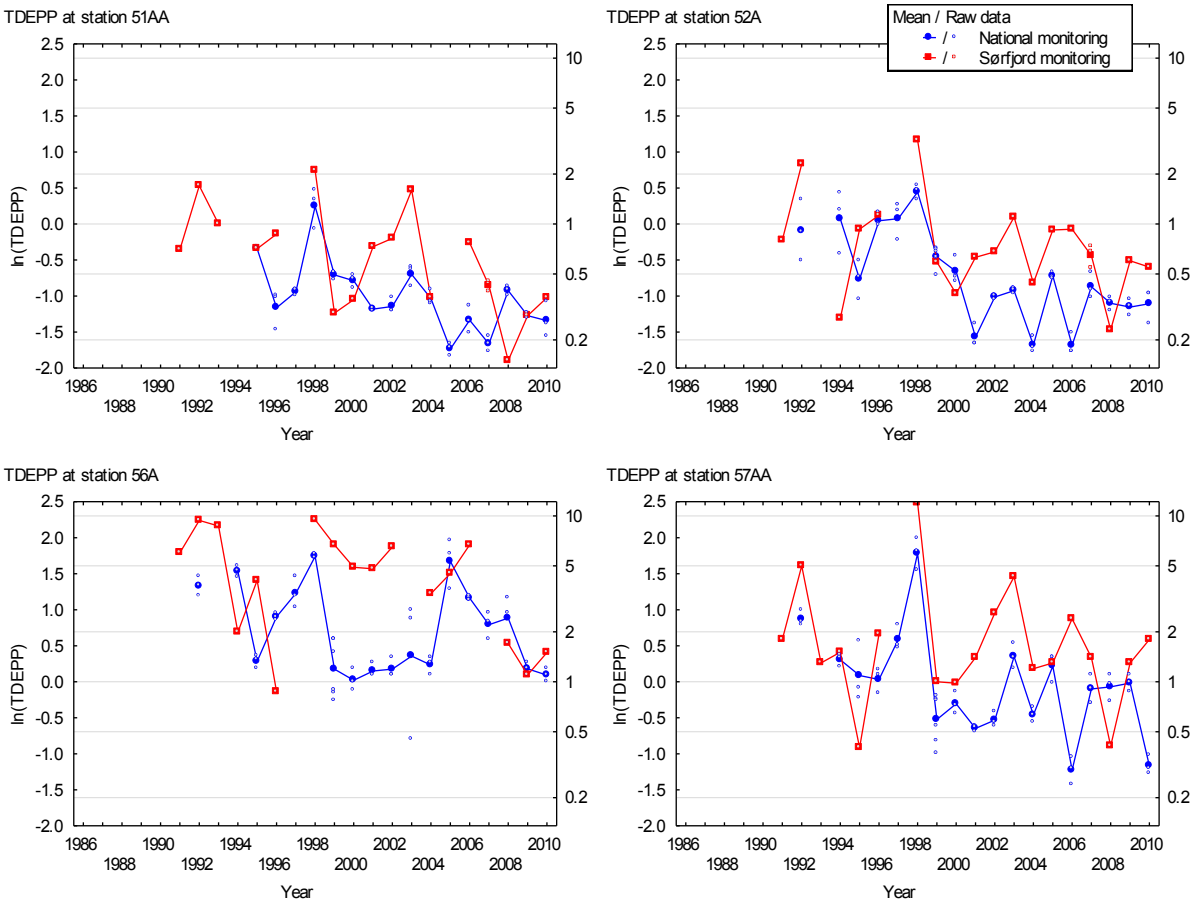

CB118 at station 52A
Mean / Raw data
National monitoring
Sørfjord monitoring
ln(CB118)
Year

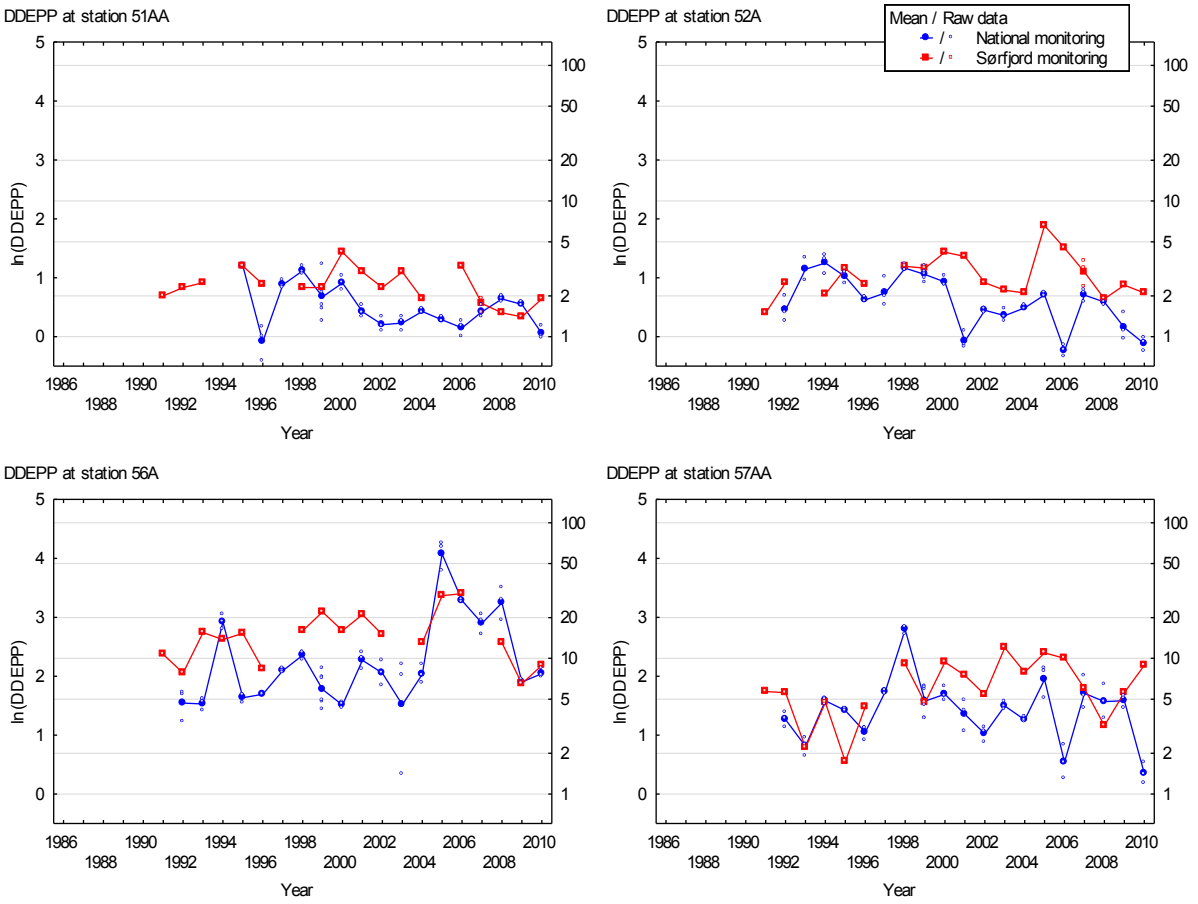

CB118 at station 56A
ln(CB118)
Year


CB118 at station 57AA
ln(CB118)
Year


CB138 at station 51AA
CB138 at station 52A
Mean / Raw data
National monitoring
Sørfjord monitoring
CB138 at station 56A
CB138 at station 57AA
ln(CB138)
Year


CB153 at station 51AA
CB153 at station 52A
Mean / Raw data
National monitoring
Sørfjord monitoring
CB153 at station 56A
CB153 at station 57AA
ln(CB153)
Year


TDEPP at station 51AA
TDEPP at station 52A
Mean / Raw data
National monitoring
Sørfjord monitoring
TDEPP at station 56A
TDEPP at station 57AA
ln(TDEPP)
Year


DDEPP at station 51AA
DDEPP at station 52A
Mean / Raw data
National monitoring
Sørfjord monitoring
DDEPP at station 56A
DDEPP at station 57AA
ln(DDEPP)
Year

Climate and Pollution Agency
P.O. Box 8100 Dep
N-0032 Oslo
Tel: +47 22 57 34 00
Fax: +47 22 67 67 06
E-mail: postmottak@klif.no
Internet: www.klif.no/english

Street address: Strømsveien 96, Helsfyr, Oslo

Climate and Pollution Agency

The Climate and Pollution Agency reports to the Ministry of the Environment and has 325 employees, based mainly in Oslo. We implement government policy on pollution. We act as advisers guardians and stewards for the environment. Our most important fields of work include climate change, chemicals, marine and freshwater environment, waste management, air quality and noise. Our vision is a future without pollution.

We are working to

- reduce greenhouse gas emissions
- reduce the spread of hazardous substances harmful to health and the environment
- achieve integrated and ecosystem-based management of the marine and freshwater environment
- increase waste recovery and reduce emissions from waste
- reduce the harmful effects of air pollution and noise

TA-2939 /2012