Marine Habitat Committee **ICES CM 2000/E:05**
Ref.: ACME

REPORT OF THE

WORKING GROUP ON STATISTICAL ASPECTS OF ENVIRONMENTAL MONITORING

Nantes, France
27–31 March 2000

International Council for the Exploration of the Sea

Conseil International pour l'Exploration de la Mer

Palægade 2–4 DK–1261 Copenhagen K Denmark

TABLE OF CONTENTS

Section **Page**

TABLE OF CONTENTS

1 OPENING THE MEETING

The Chair, Steffen Uhlig, opened the meeting of the Working Group on Statistical Aspects of Environmental Monitoring (WGSAEM) at 10.00 hrs on 26 March 2000. He welcomed the participants and thanked Benoit Beliaeff for the smooth organisation of and arrangements for the meeting.

2 TERMS OF REFERENCE

ICES C.Res. 1986/2:25

The general terms of reference originally formulated for WGSAEM were to:

- develop statistical protocols for the determination of temporal and spatial trends in the concentration and distribution of contaminants in marine biota, sediments and sea water;
- analyse data for the elucidation of temporal and spatial trends in contaminants in marine biota, sediments and sea water;
- provide statistical advice with respect to other monitoring issues, as required;
- liaise with the Statistics Committee as appropriate.

ICES C.Res. 1999/2:E:05

Specific tasks for the 2000 WGSAEM meeting are to:

a) continue the development of trend detection methods in order to:

   i. consider further development and assessment of robust smoother methods and the development of appropriate techniques for revealing outlying data values [OSPAR 2000/2.1],
   ii. consider further development of statistical methods for adjustment of input loads [OSPAR 2000/2.2],
   iii. develop provisions for the use of monthly data in trend detection methods [OSPAR 2000/2.3];

b) review and report on results of investigations concerning the characteristics of sampling and analytical variability of biomarkers and biological endpoints, the design of effective sampling schemes relative to specified objectives, and concerning the development of appropriate management tools for integrating and interpreting biological effects;

c) continue to review statistical methods for assessing and designing monitoring programmes;

d) fully exploit the Voluntary International Contaminant Monitoring in Temporal Trends (VIC) data on contaminants in biota, in order to improve sampling strategies for the analysis of both temporal and spatial trends;

e) develop the work on spatial issues by requesting presentations of national monitoring programmes, and by using this material to make generalisations and specific recommendations, e.g., on the number of replicate samples of sediments or biota needed to characterise an area;

f) continue the development of sampling allocation strategies, especially the development of dynamic sampling strategies;

g) based on available data for a suite of biological, chemical, biomarker, and endpoint measurements, carry out the following tasks, to the extent possible:

   i. explore, on a univariate basis, the minimum difference in level between two stations that can be detected with 90 % power following the JAMP guidelines, for each chemical, biomarker, and biological endpoint variable,
   ii. investigate statistical methods for modelling the relationships between the biological, chemical, biomarker, and biological endpoint measurements,
   iii. begin the consideration of ways of combining a suite of biological, chemical, biomarker, and biological endpoint measurements into summary indices that may be suitable for management or statistical purposes.

3 ADOPTION OF THE AGENDA AND ORGANIZATION OF THE WORK

Several working papers were submitted for discussion. The Chair proposed that these to be allocated to and discussed under the appropriate agenda item. The group agreed, and the agenda was accepted. The agenda is attached as Annex 1. The list of participants is given in Annex 2. Additional requests for the work on trend detection methods for input data of OSPAR INPUT are given in Annex 3, and a letter regarding these requests is presented in Annex 4. The working papers are included as Annexes 5–7.

4 ICES C. RES. 1999/2:E:05 CONTINUE THE DEVELOPMENT OF TREND DETECTION METHODS

WGSAEM noted the OSPAR request to ICES for advice on trend assessment tools: to continue the development of trend detection methods in order to:

i) consider further development and assessment of robust smoother methods and the development of appropriate techniques for revealing outlying data values [OSPAR 2000/2.1],

ii) consider further development of statistical methods for adjustment of input loads [OSPAR 2000/2.2],

iii) develop provisions for the use of monthly data in trend detection methods [OSPAR 2000/2.3];

OSPAR had also agreed on further requests to be considered on a voluntary basis for advice on trend assessment tools, presented in Annex 3. They consisted of:

iv) providing comments on the procedure described in the HARP document,

v) the use of annual mean concentrations instead of annual loads,

vi) providing an example,

vii) reviewing the Trend-y-tector,

viii) cooperating in writing draft guidelines, to be prepared by the Netherlands.

On behalf of the Netherlands, Otto Swertz announced in a letter to WGSAEM (presented in Annex 4) that the Netherlands is intending to use Section 6.4 from the 1999 ACME report as a basis for draft guidelines, and he asks WGSAEM to comment on that document.

First, in Section 4.1, below, we will report some comments regarding the use and applicability of robust smoother methods. Secondly, in Section 4.2 the work on adjusted load is reviewed and considerations on the usefulness of concentration mean values for the assessment of reduction measures are presented. Section 4.3 consists of some first steps in order to obtain a better understanding of the potentials of analysing monthly data instead of aggregated annual indices. General comments on the draft guidelines and on the Trend-y-tector are given in Section 4.4. Due to lack of time the HARP document could not be discussed further.

4.1 Consider Further Development and Assessment of Robust Smoother Methods and the Development of Appropriate Techniques for Revealing Outlying Data Values [OSPAR 2000/2.1]

The decision of whether or not to use a robust smoother should depend on the purpose of the assessment, as discussed in 1999 ACME Section 6.4. Further, note that it is not useful to apply a robust smoother on aggregated data to control the influence of a single outlier within the year. If there are single outliers within a year, for example, they have to be detected and eliminated before producing an annual index. Alternatively, a robust statistic could be used for the annual index. A standard robust smoother can be found in Cleveland (1993).

It should be noted that the decision to use a robust smoother has to be made in view of the specific problem under study. In Section 6.7, below, an example is given for a trend detection approach (trends in fish disease prevalences), which deliberately avoids a robust approach against outliers, because the aim to immediately inform about alarming recent trends was given higher priority than the aim to make a statement unaffected by incidental outliers. The reason is that the intended display of recent trends should trigger a more detailed analysis and possibly further action, if recent trends are extremely upwards or downwards. The chance of being misled, if the most recent data was contaminated by outliers, is acknowledged.

A second example for employing concepts different from robustness is addressed in Section 5.1. Here, a cyclic series was studied, and the problem of potential high effects due to outliers at the ends of the series was resolved at least

partially by adding one copy of the series at the beginning of the series and another copy at the end. Then a smooth curve was fitted to the now threefold as long series, thus transforming the former end points to interior points, which do not have the particular susceptibility to outliers that end points have.

Reference

Cleveland, W.S. 1993. Visualising data. Hobart Press. Summit, New Jersey.

4.2 Consider Further Development of Statistical Methods for Adjustment of Input Loads [OSPAR 2000/2.2]

Steffen Uhlig presented a paper on flow adjustment, describing the procedure of adjustment, its interpretation and several adjustment models. For the Rhine River (Lobith station) and the Ems River (Herbrum), the adjusted loads according to these models were calculated for nitrate, ammonia, phosphorus, orthophosphate, lead, cadmium, and dissolved particles.

It turns out that there is no method that is optimal for every river and every substance. However, method L1 (see Annex 5) performs reasonably well in terms of smoothness of the annual adjusted load for the nutrients and partly also for heavy metals. Method L1 is based on a linear model for the concentration, with time and reciprocal flow as independent variables, together with a parametric seasonal component of second order. The calculation of the parameters is performed by local regression based on a running window of a given number of years. According to the model assumptions and the outcome of the statistical calculations, it can be concluded that adjustment according to method L1 performs reasonably well for nutrients and if the load flow relation is approximately linear.

Some reservations were expressed about using the "smoothness" in the time series of adjusted loads as the only criterion for choosing an adjustment method. Many of the adjustment methods used data from several adjacent years to construct each annual index, and this will induce correlations in the annual indices. It was unclear what impact this would have on statistical tests made using these indices. One possibility is that the correlations would increase the type I error, resulting in spurious significant trends.

A subgroup evolved to look at this issue. They considered three things:

1) the correlation structure induced by various adjustment methods, and its impact on inference,
2) the performance of some simpler adjustment methods,
3) the use of concentrations instead of annual loads.

These are discussed in Sections 4.2.1, 4.2.2, and 4.2.3, respectively.

4.2.1 Correlation structure induced by adjustment

The subgroup focused on adjustment method L1, given that this had performed reasonably well averaged across a range of substances (Annex 5), and given that it was amenable to some theoretical manipulation.

The flow data series from the Lobith/Rhine was used to calculate the covariance matrix of the annual (additive-) adjusted loads, assuming that there are 26 measurements each year, and assuming a window width of 7 years. It was assumed that only ten years of data were available. The covariance matrix depends on the flow regime, i.e., on which ten-year period in the whole time series was chosen. The covariance matrix below corresponds to the period 1986–1995, and shows some positive correlation in the annual indices in the middle of the period, and negative correlation at the ends.

29	–2.6	–2.0	0.40	–0.19	–0.10	–0.083	–0.099	0.025	0.064
-2.6	37	–5.8	0.35	0.33	–1.5	–1.7	–0.36	0.35	0.72
-2.0	–5.8	45	0.51	0.38	2.1	–1.8	–0.51	0.40	0.87
0.40	0.35	0.51	19	2.3	2.9	3.4	1.2	–0.71	–0.94
-0.19	0.33	0.38	2.3	19	2.6	3.1	2.4	–0.47	–1.3
-0.10	–1.5	2.1	2.9	2.6	18	2.5	2.3	2.1	–0.95
-0.083	–1.7	–1.8	3.4	3.1	2.5	21	1.6	0.30	–0.33
-0.099	–0.36	–0.51	1.2	2.4	2.3	1.6	22	0.34	–0.55
0.025	0.35	0.40	–0.71	–0.47	2.1	0.30	0.34	30	–4.1
0.064	0.72	0.87	–0.94	–1.3	–0.95	–0.33	–0.55	–4.1	37

A simulation showed that this resulted in a slight increase in the type I error of an F-test based on linear regression of the annual indices. Under a variety of flow regimes (ten-year snapshots from the Lobith/Rhine series), the type I error increased from 5 % to about 6 % at most. This is only a slight increase in type I error, so that for these data, and for the 7-year window, the annual adjusted loads behave almost like a stochastically independent time series. It can be concluded that under the given circumstances the Trend-y-tector can be applied to the annual adjusted loads as well as to the unadjusted loads.

Further investigation showed that if the window width was reduced, the induced correlations increased, and care should be taken in the choice of window width accordingly. There was no time to investigate a time series with only 12 measurements per year.

Note that with a seven-year window, the autocorrelations in the annual adjusted loads were mainly due to the autocorrelations in the annual mean flows. When flow data were simulated that were stochastically independent, the autocorrelation in the annual adjusted loads virtually disappeared. This would not necessarily be the case with smaller windows. Further note that the OSPAR load is highly autocorrelated due to the autocorrelation of the annual flow, causing some increase of the actual level of a trend test based on the OSPAR load as well.

4.2.2 Some simpler adjustment methods

Some alternative adjustment methods were considered that constructed the annual index using the data collected in that particular year. They had the great advantage of simplicity. Let (L_{ij}, q_{ij}), j=1,…,m, be the pairs of load and flow measurements in year i, let $\left(q_t^{(i)}\right), t = 1, \cdots, 365(366)$, denote the series of daily flows and $t_{i1}, t_{i2}, ..., t_{im}$ the sampling times in year i, i.e., $q_{ij} = q_{t_{ij}}^{(i)}$. Let Q be the mean flow over the entire time series. The annual indices were:

- Mean: the annual mean load $\overline{L}_i = \frac{1}{m}\sum_{j=1}^{m} q_{ij} c_{ij}$
- OSPAR: the annual load calculated according to the "OSPAR-formula": $L_{i,OSPAR} = \frac{\frac{1}{365}\sum_{t=1}^{365} q_t^{(i)}}{\frac{1}{m}\sum_{j=1}^{m} q_{ij}} \overline{L}_i$.
- Ratio: $\overline{L}_i \times Q / \left(\frac{1}{m}\sum_{j=1}^{m} q_{ij}\right)$, also known as method A0 in Annex 5 and 1A1 in the draft HARP Protocol.

- Linear: the model $L_{ij} = \alpha_i + \beta_i q_{ij} + \varepsilon_{ij}$ is fitted assuming gamma errors and identity link: the annual index is then the predicted load at Q, $\hat{\alpha}_i + \hat{\beta}_i Q$.

- GAM: the model $L_{ij} = s_i(q_{ij}) + \varepsilon_{ij}$ is fitted assuming gamma errors and identity link, where $s_i(.)$ denotes a year-specific smoothing spline on 2 degrees of freedom that allows for a non-linear load-flow relationship: the annual index is then the predicted load at Q.

These indices were fitted to four time series: nitrate and total P in the rivers Rhine (Lobith) and Ems (Herbrum). To assess their performance, the residual standard deviation of the annual indices was calculated by fitting a LOWESS smoother with a span of eight years (and corrected for differences in the mean level). For comparison, the residual standard deviation was also calculated for the annual index L1 of Annex 5, which uses information from adjacent years.

The results (in tonnes per year) are:

	Rhine/Lobith		**Ems/Herbrum**	
	Nitrate	**Total P**	**Nitrate**	**Total P**
Mean	38 085	3 938	4 347	267
OSPAR	37 412	3 544	3 586	226
Ratio	11 920	4 952	1 993	120
Linear	12 808	3 241	2 879	152
GAM	13 002	2 757	2 083	152
L1	9 076	2 545	1 300	115

In terms of smoothness, method L1 performs best throughout. The GAM index also performs reasonably across all four time series. However, note that calculating the GAM index was sometimes difficult for the Herbrum, where there were only 12 observations per year, due to convergence problems.

Although annual indices that only use within-year information are simple to calculate, they can still be autocorrelated if there are strong seasonal differences in the load flow relationship, and there are also autocorrelations in annual flows.

4.2.3 On the use of concentrations instead of annual loads

The subgroup noted that, in case of seasonal variation, yearly averaged concentrations instead of annual loads or annual adjusted loads may not reflect properly the effectiveness of reduction of inputs, not even in the long run.

In the following a hypothetical example is presented that is based on the monthly mean runoff at Herbrum/Ems. In this example, it is assumed that reduction measures were effective in reducing inputs by more than 40 %. This is obvious in the load (and would also be appearant in an adjusted load) with a reduction of 41.2 %, but due to a shift in the seasonal variation of the mean concentration there is no reduction of the averaged concentration at all.

In order to avoid the risk of obtaining such a result, the subgroup felt that the use of concentration mean values can hardly be recommended.

Month	Runoff Herbrum (Ems)	Concentration	Load	Concentration	Load
	Monthly mean	Before	Before	After	After
1	183.2	1.5	274.8	0.2	36.6
2	154.5	1.8	278.1	0.2	30.9
3	149.3	2.3	343.4	0.3	44.8
4	102.6	2.0	205.2	0.7	71.8
5	63.8	2.0	127.6	1.5	95.7
6	50.1	1.5	75.2	1.8	90.2
7	48.0	1.0	48.0	2.3	110.4
8	35.7	0.5	17.9	2.0	71.4
9	43.8	0.2	8.8	2.0	87.6
10	61.3	0.2	12.3	1.5	92.0
11	90.7	0.3	27.2	1.0	90.7
12	133.1	0.7	93.2	0.5	66.6
Mean	**93.0**	**1.2**	**126.0**	**1.2**	**74.1**

4.3 Develop Provisions for the Use of Monthly Data in Trend Detection Methods [OSPAR 2000/2.3]

The purpose of this section is, in light of an example, to present different approaches for linear trend detection in time series showing strong seasonality. Two different approaches are presented where season is modelled versus fitting a trend on some annual indices such as yearly means or medians. Section 4.3.1 contains a description of the data set, and the two approaches are described in Sections 4.3.2 and 4.3.3. Concluding remarks are presented in Section 4.3.4.

4.3.1 Data set

To illustrate the effect of monthly measured values instead of yearly values (this could be the mean or median values of the underlying monthly data) on the trend detection method, we used total phosphorus values (mg l^{-1}) measured at the location Eijsden in the river Meuse. This is a location near the Dutch-Belgian border. The monitoring programme started in 1988 on a biweekly frequency; in 1993 onwards the frequency was changed to nearly every week.

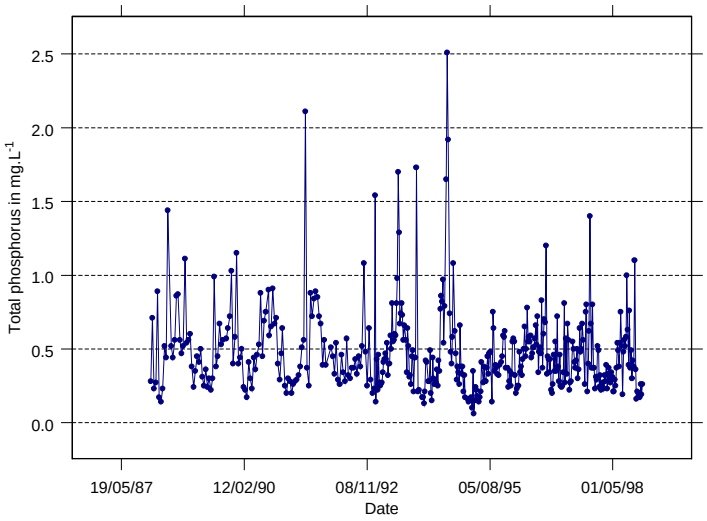


Figure 4.3.1. The time series of total phosphorus at Eijsden in the Meuse River.

4.3.2 Statistical analysis: First approach

The statistical analysis is presented step by step.

i) Raw data are pooled according to Julian day of sampling (years are ignored), in order to highlight the average shape of the intraannual (i.e., between-months) variability. The median of pooled values of each month is then estimated, and a first identification of outliers is performed at the same time (Figure 4.3.2, Box-and-Whisker plots, also used and explained in Section 6.1).

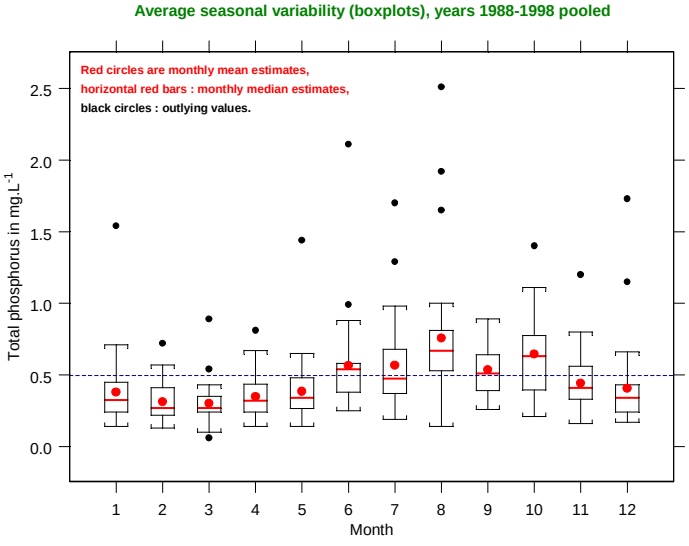


Figure 4.3.2. These boxplots provide a graphical illustration of the seasonal pattern, the main features of which are summer maximum values and a local minimum in September. The circles within the boxes are monthly mean estimates; the horizonatal bars within the boxes are monthly median estimates.

ii) The previous monthly median estimates are subtracted from the original data values, thus removing a robust estimate of the seasonal component of the time series (Figure 4.3.3).

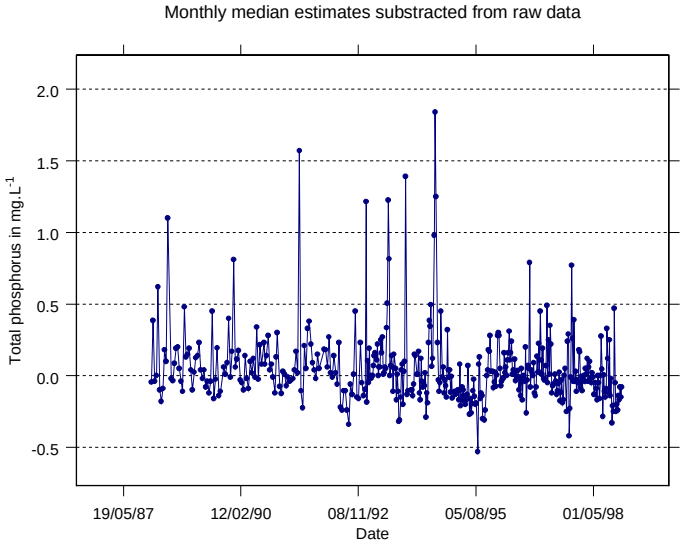


Figure 4.3.3. Plot of the result of subtracting the monthly mean estimates from the original data values.

NB 1: the existence of seasonality is *a priori* known, and no test is performed to assess its statistical significance.
NB 2: in the following, the variability of monthly terms estimates will be neglected; however, a comprehensive analysis should manage this uncertainty, especially by quantifying its effect on trend estimates (e.g., by resampling).

iii) A robust linear regression model (Least Trimmed Squares regression) is then fitted to the deseasonalized time series, in order to separate outliers from the bulk of observations (Figure 4.3.4), the latter exhibiting a linear trend "polluted" by serial correlation. The intention here is to identify clearly outlying observations which obviously will not adequately fit to a simple model of trend (here a linear trend).

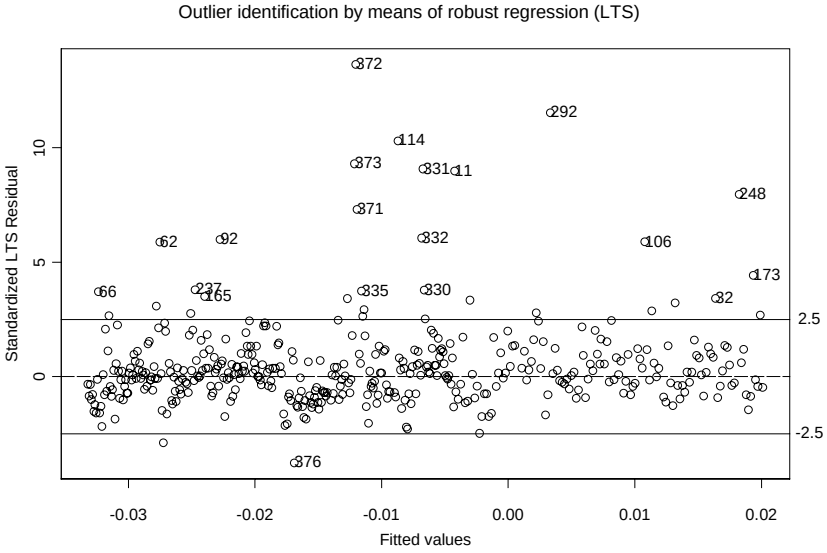


Figure 4.3.4. Plot of the result of fitting the Least Trimmed Squares regression model to the deseasonalized time series, to identify outliers.

Twenty troublesome observations are thus removed; this amounts to approximately 10 % of the total number of measurements. From a phenomenological point of view, it is likely that these "extreme" values are highly informative. They will not be taken into account here, but it is worth emphasising that explaining their occurrence should constitute a subject of study for biogeochemists.

From now on, attention will be paid to the "clean" deseasonalized time series only (Figure 4.3.5).

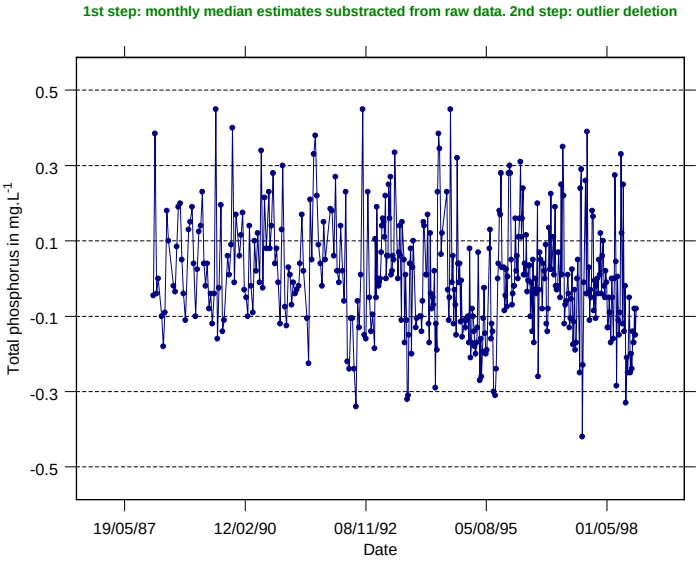


Figure 4.3.5. Plot if the "clean", deseasonalized time series of total phosphorus at Eijsden in the Meuse River.

iv) Before estimating the autocorrelation function (ACF), a regularization of the sampling time step is performed, because data have been collected at unequal time intervals. Some pooling over the shortest possible time scale is thus necessary, in order to build the average measurement values reproducing at best the data set we would have obtained by regular sampling. A time step of 2 weeks has been retained after trial and error; the corresponding ACF estimate is shown in Figure 4.3.6.

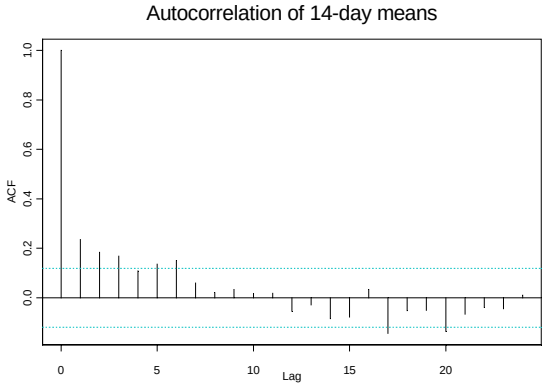


Figure 4.3.6. ACF estimated for the time series shown in Figure 4.3.5, after aggregation and averaging of observations over consecutive time periods of 14 days (avoiding the occurrence of "holes" in the series). Abscissa scale: lag in two-week units.

v) The ACF reveals a quite complex "memory" of the process, in accordance with the bumps observed on Figure 4.3.5. Nevertheless, for the sake of simplicity, and, first of all, in order to manage a limited set of unknown parameters, an AR-1 model is postulated, i.e.,

$$u_t = x_t - \rho x_{t-1} \quad , \quad v_t = y_t - \rho y_{t-1} \quad , \quad \omega_t = \varepsilon_t - \rho \varepsilon_{t-1} \quad [1]$$

where y, x, ε, and ρ are the response variable, the regressor, the error term, and the first order autocorrelation coefficient, respectively (the index 1 designates the time increment). The linear model whose error term ω is free from first order serial correlation can be written as:

$$v_t = \beta_0(1-\rho) + \beta_1 u_t + \omega_t \quad , \quad t = 1, ..., n \quad [2]$$

with n = 259 in the present case. Estimation of the parameter β_1 gives the rate of change of *y* with *x* [rate expressed in 1/(2 weeks)], i.e., the trend.

Results and discussion

Fitting model [1, 2] gives the following estimate b_1 of the slope β_1:

$b_1 = -4.10^{-4}$, $SE(b_1) = 10^{-4}$, both expressed in mg l^{-1} P (14-day period) $^{-1}$,

computed t-statistic = −3.072, with critical probability ≈ 2.10^{-3}.

It was further checked that the residuals in model [2] are no longer autocorrelated.

Roughly speaking, this result leads to conclude a decrease of total phosphorus concentration in water at an average rate of the order of 10^{-2} mg l^{-1} P year $^{-1}$.

It is worth pinpointing here some remarks:

- several uncertainties remain to be taken into account, for instance, the sampling noise associated with the seasonality, and also the autocorrelation coefficient estimates; it is not easy to guess the amplitude of their impact on the stability of the previous estimator;
- some steps of the analysis are highly non-linear, particularly the deletion of outlying values [step iii)];
- an attempt at a step function fit was achieved in order to catch a "jump" in the time series (see Figure 4.3.5, break in year 1995), but it did not lead to a substantial improvement;
- one of the major weakness of this analysis rests on the lack of "assimilation of information" from auxiliary variables such as flow, loads, etc.

4.3.3 Statistical analysis: Second approach

Statistical model

The second approach to analyse the phosphate data at the Eijsden station is based on a linear regression method, which accounts for the seasonality and autocorrelation in time. When time series data are used in regression analysis, the error term is often not independent through time. If the error term is autocorrelated, the efficiency of ordinary least-squares parameter estimates is adversely affected and standard error estimates are biased. The Durbin-Watson d-statistic can be used to test for the absence of first-order autocorrelation in OLS residuals. If an autocorrelation is detected, estimation methods which account for the autocorrelation give better estimates.

A small drawback of this way of modelling is that only ordered and equally spaced time series data could be used with no missing values. Therefore the data have been transformed to biweekly measurements. This model is defined in two parts: the structural and the error term. For the last term, an autoregressive process is assumed.

The model, which includes a linear trend and a seasonal component, is defined as:

$$y_t = \alpha + \beta_1 year + \beta_2 \sin\left(\frac{2\pi month}{12}\right) + \beta_3 \cos\left(\frac{2\pi month}{12}\right) + \nu_t \qquad [3]$$

where y_t is the measured total phosphate at time *t*,
α is the intercept,
β_1 is the coefficient of the global linear year trend,
β_2 is the coefficient of the sinus component,
β_3 is the coefficient of the cosinus component,

$year$ year value: 1988,…,1999

$month$ month value: 1,..12

v_t is the error for time *t*

The error is described as an autoregressive AR(p) process according to:

$$v_t = \varepsilon_t - \varphi_1 v_{t-1} - \cdots - \varphi_p v_{t-p}, \qquad [4]$$

where ε_t is Normally and independently distributed with a mean of 0 and a variance of σ^2.

As this model incorporates a systematic part and the autocorrelation process of the time series, it could be useful for forecasting.

Analysis of the data

In Figure 4.3.7, the results are presented for the biweekly time series. The residuals are analysed on the presence of autocorrelation. In the autocorrelation plot of the residuals of the least squares model (OLS), there appear two significant lags: 1 and 3. In the second figure, we have used model 1 which includes these autoregression coefficients and the autocorrelation has nearly vanished. It is nice to see that this model is capable of picking up some peaks (Figure 4.3.7). In Table 4.3.1, the coefficient β_1 and the standard error for the OLS model and the extended model are presented. It is clear that, because of the autocorrelated errors, the estimate of the standard error for β_1 is biased. For a further analysis it would be nice to include the characteristics of the flow of the river which could be used to pinpoint some peaks or level of the concentration.

In Figure 4.3.8, the results are presented of the trend analysis of the annual mean and the annual median This comparison should reflect the robustness of the median instead of the mean. It is clearly shown in Figure 4.3.7 that there are some peaks in the early years and also that the fluctuation is different during the years. Especially before 1995 the signal is much more fluctuating. In Figure 4.3.8 this fluctuation becomes clearer where the annual mean, annual median, and the maximum and minimum are shown. The median is less fluctuating than the mean (see also Figure 4.3.9).

We have chosen also a winter period, which is defined as the three-month period from December until February. The choice of the winter values is based on the fact that for this period the concentration of total phosphorus is not affected by biological activity, which partly caused the seasonal periodicity, and it would be a more stable period for making a comparison each year. It is however not the period with the highest value (see Figure 4.3.2), which rather may be also a good period to choose for detecting a trend in the data. The results of the regression on the annual values and the winter values are presented in Table 4.3.1.

It is interesting that, except for the median winter values, the trend coefficients are nearly the same with the same standard errors. The poor behaviour of the median could be caused by the high autocorrelation of the winter values.

Table 4.3.1. The results of the regression analysis on the biweekly values and the annual mean and median values and the mean and median winter values. The OLS-model refers to the structural part of model 1; OLS+AR(1,3) refers to the full model 1 with the AR-part at lag 1 and 3 without the cosinus part; the OLS-Mean is the linear trend model with the mean year values; the OLS-Median-year is the OLS linear trend model with median year values, and the OLS-Median-winter year is the same model with the median winter year values. Winter is defined as the three-month period from December until February.

Model	n	β_1 (mg l^{-1} $year^{-1}$)	Standard error	P(T<|t|) |H_0:$beta_1$=0)
OLS-model	262	–0.0147	0.0044	0.010
OLS+AR(1,3)	262	–0.0144	0.0066	0.030
OLS-Mean year	11	–0.0132	0.0064	0.068
OLS-Median year	11	–0.0117	0.0049	0.049
OLS-Mean winter year	11	–0.0118	0.0051	0.044
OLS-Median winter year	11	–0.0091	0.0065	0.193

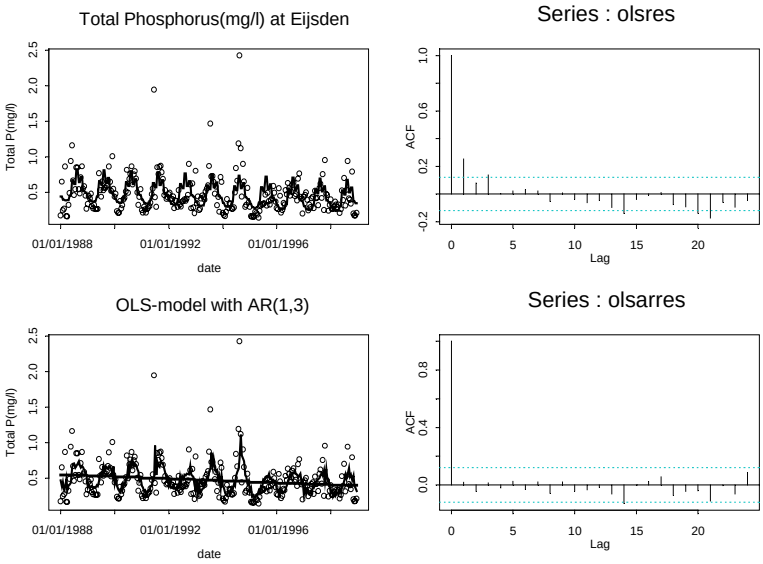


Figure 4.3.7. The biweekly time series for the concentration of total phosphorus (mg l^{-1}) at Eijsden. In the upper left corner the values are fitted with an ordinary least squares (OLS) model which incorporates the linear trend and the seasonality. In the upper right corner the residuals of this model are checked on autocorrelation. The dotted lines in this plot are the confidence limits of the autocorrelation with respect to H_0:ρ=0. In the left lower corner the results are shown for the OLS-model with the autocorrelated process included (AR(1,3)). In this figure also the linear trend line is included.

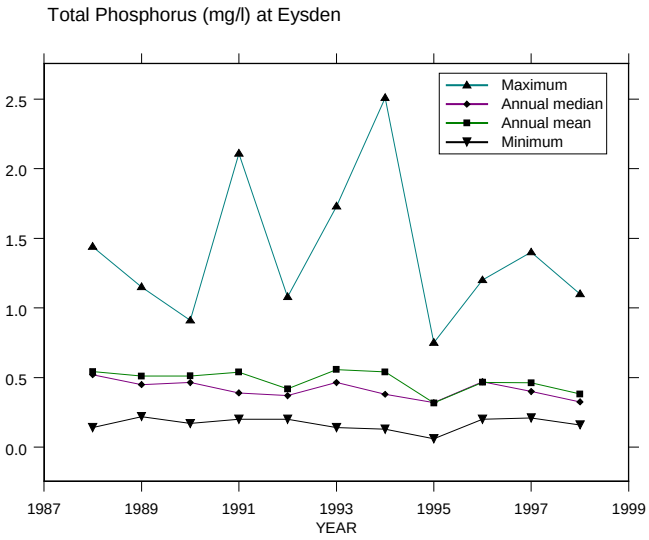


Figure 4.3.8. The annual mean, median, and the minimum and maximum values of total phosphorus (mg l^{-1}) at Eijsden.

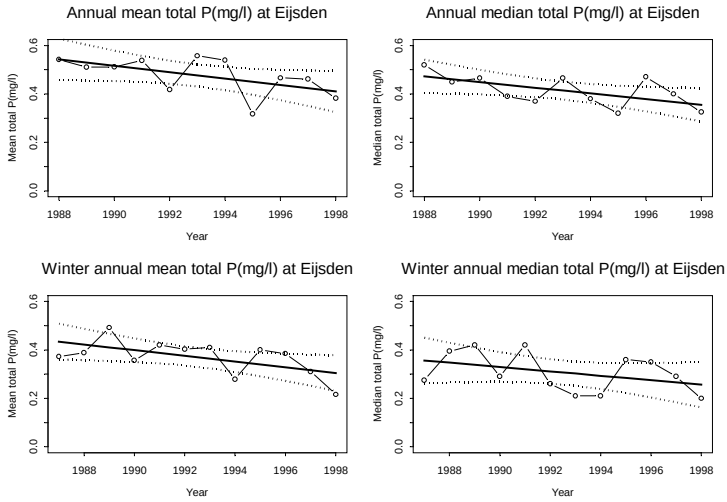


Figure 4.3.9. The linear trend lines of, respectively, the annual mean, annual median, winter annual mean and winter annual median of total phosphorus (mg l^{-1}) with their respective 95 % confidence limits.

4.3.4 Concluding remarks

The different approaches described above were aimed to detect a linear trend in a total phosphorus time series showing a strong seasonal pattern. They all attempted to model seasonal variations and autocorrelation.

1) The first approach, using 14-day pooling, led to a substantial reduction in the standard error of the slope estimate, presumably due to the removal of outlying observations. However, the slope standard error estimate is somehow optimistic as it does not take uncertainties on deseasonalization on the one hand, and outlier removal on the other hand into account.
2) Use of biweekly data instead of annual (winter) mean or annual (winter) median has the advantage of modelling all the data without having to choose some aggregating or choosing a special period, but the specification of the model is more complicated.
3) The use of annual mean or annual median looks promising for total phosphorus because no large difference in the trend coefficient compared with the biweekly trend estimation was present.
4) An annual median could be better used instead of an annual mean because of its robustness.

5) The presence of autocorrelation could inflate the robustness of the median (see the median winter year values) compared with the mean.

6) The winter values may not reflect the most important part of this variable to use for the trend estimation.

7) The use of monthly data instead of annual data in order to estimate trends over several years allows compensation for effects due to irregular seasonal variation which otherwise might bias a trend estimate based on annual data. In this case the use of a non-parametric approach might be required.

4.4 General Comments on Trend-y-tector and Draft Guidelines

WGSAEM discussed the ICES recommendation to SIME (SIME 00/5/4-E) that

"ICES considered Sections 6.2 and 6.4 [of the 1999 ACME report] *as a final work and hoped that they were clear enough for easy use by OSPAR, together with the material in Annex 1* [of the 1999 ACME report] *and the description of the Internet application Trend-y-tector. The software application Trend-y-tector could be considered as a tool to perform trend assessments according to 6.4, and it was currently being revised by the Netherlands to accord fully with the ICES advice in Section 6.4."*

WGSAEM noted that:

- Section 6.4 focused on the development of statistical methods that would be useful for large-scale assessments of many, many data sets, where the emphasis was on producing standard assessments and reports for each time series. For small numbers of data sets, more efficient methods might be possible, taking account of the specific features of each data set.

- Section 6.4 had been created as a general guide to the development of trend assessment software, in particular to suggest features that should be considered in the development of the Trend-y-tector software. It does not provide guidelines to making use of such software or the interpretation of results.

- Section 6.4 did not address assessment features that may arise in monitoring programmes other than the assessment of contaminant input data. For example, assessing the target of a decrease of 50 % in a ten-year period is specific to INPUT 2000.

- Appropriate guidelines for the use of the Trend-y-tector software should be developed.

5 JOINT MEETING OF WGSAEM AND WGBEC

WGSAEM met jointly with the Working Group on Biological Effects of Contaminants (WGBEC) "to review and report on results of investigations concerning the characteristics of sampling and analytical variability of biomarkers and biological endpoints, the design of effective sampling schemes relative to specified objectives, and concerning the development of appropriate management tools for integrating and interpreting biological effects".

Two data sets describing biomarker distribution in fish were presented; these emerged from the German "Stresstox" programme (Dr Kamman and Dr Wosniok) and from a Norwegian study (Dr Hylland and Dr Bjerkeng).

5.1 German "Stresstox" Dataset

The German programme used dab (*Limanda limanda*) from six North Sea (and one English Channel) sites and involved measurement of a suite of biomarkers (EROD, MT and others), chemical residues (conventional OCs, Zn, Cd and Cu plus others) together with biological factors such as age, size, sex, condition and reproductive status. All measurements were made at monthly intervals over a one-year period, using females from a limited size range. The thrust of this programme was to attempt to establish relationships between biomarkers and the presence of both contaminants and natural ecophysiological factors. All the biomarkers studied (EROD, MT, apoptosis, etc.) showed strong seasonal variation, but seasonal maxima and minima were not obviously correlated; that is, biomarkers followed different seasonal cycles, some of which could be related to reproductive cycles. Organochlorine residue concentrations and metal distributions also followed seasonal cycles, but these differed from those of the biomarkers with which they were expected to be correlated.

Details of the Data

Ulrike Kammann and Werner Wosniok presented an overview of a data set compiled in the framework of the "STRESSTOX" project funded by the German Ministry of Research. The data are to be used to investigate annual cycles of biomarkers and to investigate ways of describing the "health of an organism" using a suite of biomarkers. The data set includes contaminant, biomarker and fish disease measurements obtained from samples of dab liver from the same individual. The application of biomarkers in the field requires a good knowledge of factors which may influence induction. Without this information, differences of biomarker induction in organisms from various locations may lead to false interpretations. Female dab (n = 10) were collected monthly from February 1998 to January 1999 by bottom trawling. The sampling site was located northwest of Helgoland in the German Bight (North Sea). This site is only moderately contaminated compared with some inshore and estuarine areas of the North Sea. Data on bottom temperatures during the year were made available by the Federal Maritime and Hydrographic Agency in Germany. All liver samples were analysed by the University of Hamburg, Institute of Food Chemistry and the University of Mainz, Institute of Physiological Chemistry, as well as by the Federal Research Centre of Fisheries, Institute of Fisheries Ecology.

One objective of the OSPAR Joint Assessment and Monitoring Programme (JAMP) is the measurement of biological effects in fish once a year outside the spawning season, preferably simultaneously with measurements of relevant contaminants. This avoidance by JAMP of the spawning season had been thought to avoid the possibly confounding influence of elevated steroid hormone titres on such biomarker measurements as EROD induction. The present investigation suggests, however, that maxima in annual cycles exist beyond those that occur during the spawning season. If it is required that measurements be made outside of the spawning season during periods of low biomarker values, then a suitable period for each single biomarker can be taken from Table 5.1.1. Analogously, in order to use a battery of biomarkers simultaneously, a period containing only marker minima may be required.

However, Table 5.1.1 shows that a common "low" period for, e.g., EROD, MT, and HSP does not exist. The cycle of these variables can be assumed to change to some extent from year to year. Anthropogenic influences would induce additional variation in biomarker values. In order to derive a proper assessment of biomarkers, particularly to identify anthropogenic effects, the natural annual cycle variation of a biomarker must be accounted for. It should be noted that the conclusions about the existence of annual cycles in biomarker values are based upon the observation of a one-year cycle at a single location, and should therefore not be taken on their own as a sufficient reason to modify monitoring policy. Because of the explanations about why such cycles might exist and the parallels to some well-investigated biomarkers, it is assumed that such cycles may exist more generally, although the shapes of the cycles actually found cannot simply be carried over to other times or locations. A confirmation of this assumption requires the analysis of additional data, including information on varying residue concentrations, which may contribute to the observed seasonality in biomarker responses. However, the working hypothesis is that the majority of the seasonality is natural, and it is concluded that the observed annual cycles should be a major factor in the interpretation and evaluation of biomarkers.

Table 5.1.1. Maxima (dark) in annual cycles of biomarkers and related parameters measured in dab livers from the North Sea (February 1998 to January 1999). Data on water temperature were made available by the Federal Maritime and Hydrographic Agency of Germany. Temp. = temperature; lipid = lipid content of liver; MT = metallothionein; Apop. = apoptosis; SSF = DNA strand breaks**;** HSP = heat shock protein; GSI = gonadosomatic index; EROD = ethoxyresorufin-*O*-deethylase.

	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
Spawning												
Temp.												
Lipid												
Zinc												
MT-I												
MT-II												
Apop.												
SSF												
HSP												
GSI*												
EROD**												

* Saborowski, R. (1996), Zur Ökophysiologie der Kliesche, *Limanda limanda* (L.) Einfluß saisonaler Zyklen auf das hepatische Entgiftungssystem. Ph.D. Thesis, University of Hamburg, Department of Biology, 168pp. ** Vobach, M., Kellermann H.J., (1999) Entgiftungsstoffwechsel der Kliesche (*Limanda limanda*). *In* Jahresberichte 1998, Bundesforschungsanstalt für Fischerei, Hamburg.

For an analysis of relationships between biomarkers and contaminants, data on 151 dab (*Limanda limanda*, females) collected at 6 stations (North Sea: 4, Baltic Sea: 2) in December 1997 were available. The research question to be addressed here was whether there is a simple and low-cost approach to predict underlying contaminant concentrations from observed biomarker values. Biomarkers considered were MT-1, MT-2, apoptosis, HSP, EROD and DNA strand breaks; available contaminants data were for CB28, CB153, DDE, Cd and Zn in dab liver. All measurements were, as far as possible, made on the same individual. Age-length relationships were slightly different between stations, as only a restricted length class was sampled (20–24 cm), and not all stations could provide samples exclusively in that length class. Relations (second-order polynomial) between biomarkers and age were found for all biomarkers except MT-1 and MT-2. In order to avoid the confounding effects of age, and because the MT-1 data contain a few outliers, the following discussion will focus on MT-2.

The means from the stations with the highest and lowest MT-2 values are significantly different, so the complete set of MT-2 values comprises more than just random variation around a common mean. Hence, if causative factors are present among the variables under consideration, the detection of a relationship could be expected. Fitting a second-order polynomial, significant (α = 5 %) relationships between MT-2 and Cd, Zn, and CB28 were found. The relatively small coefficients of determination of 21 %, 18 %, and 16 % indicate, however, that additional conditions must be responsible for the MT-2 values. The confidence limits of the relationship between MT-2 and Zn are wide and do not allow a clear conclusion about underlying Zn values on the basis of MT-2 values (e.g., an MT-2 value of 3.5 leads to a predicted Zn value of 50 with a confidence interval of 35 to 60, which covers about half the range of Zn values).

Multivariate display and analysis approaches were demonstrated and discussed, but mainly for illustrative purposes, as for technical reasons not all required measurements could actually be made on each individual, which led to a severe loss in the number of cases available for multivariate analysis. Further elaboration in this direction was considered sensible only in connection with appropriate missing value replacement methods, which were beyond the scope of the present analysis. It was concluded, that (i) a simple approach as considered here was in this case not a promising method for relating biomarker changes to contaminants, (ii) other biomarkers and contaminants might show a clearer relationship, and (iii) investigations in areas with higher contaminant concentrations might provide a more readily understandable picture.

Reference

Lacorn, M., Piechotta, G., Simat, T.J., Kammann, U., Wosniok, W., Lang, T., Müller, W.E.G., Schröder, H.C. Jenke, H.S., and Steinhart, H. (2000) Annual cycles of apoptosis, DNA strand breaks, heat shock proteins, and metallothionein isoforms in dab (*Limanda limanda*): Influences of natural factors and consequences for biological effect monitoring. (submitted to Biomarkers)

5.2 Norwegian Data Set

The Norwegian programme focused on cod (*Gadus morhua*) from some industrialised fjords and reference sites, and involved sampling males and females once annually over a 2–3 year period. Again, a suite of biomarkers and chemical residues was examined along with "natural" variables. Some biomarkers were correlated with expected chemical causes (e.g., ALA-D and Pb) but correlations between others (e.g., EROD and OCs, or MT and metals) were less clear. However, statistical analyses of this data set were less complete than in the German study, and it would be premature to draw many conclusions yet.

Details of the Data

The results from preliminary analyses of Norwegian JAMP data were presented. The data comprise about 300 individuals where ALA-D, MT, EROD and PAH metabolites were measured together with physiology and contaminant concentrations, distributed over 6 stations and three JAMP monitoring years (1996–1998). For 1996 there were only data from two stations, and ALA-D was not measured. Three of the 6 stations are known to be more or less impacted by contaminants, whereas the other three sites were reference sites.

Unfortunately, some contaminants have a large proportion of cases with results less than the detection limit. The analysis has been focused on the contaminants without such problems.

The objectives of the study were: (1) to clarify relationships, if any, between biomarkers, physiology and contaminants, and (2) to identify explanatory variables for biomarker responses (EROD, MT, ALA-D).

(1) Relationships between biomarkers, physiology and contaminants.

PCA was performed on EROD, MT, physiology and contaminants. Data from 1996, 1997, and 1998 were used.

Factor 1 is the main general measure of contaminant content, all selected contaminants were positively correlated to Factor 1 (Figure 5.2.1). To a large extent this was determined by the CB congeners included. Biomarkers were not represented in Factor 1 or 2. Factor 3 appears mainly to be a relation between MT and the metals Zn and Cu (Figure 5.2.2).

Figure 5.2.1. Factors 1 and 2 in PCA with case-wise deletion.

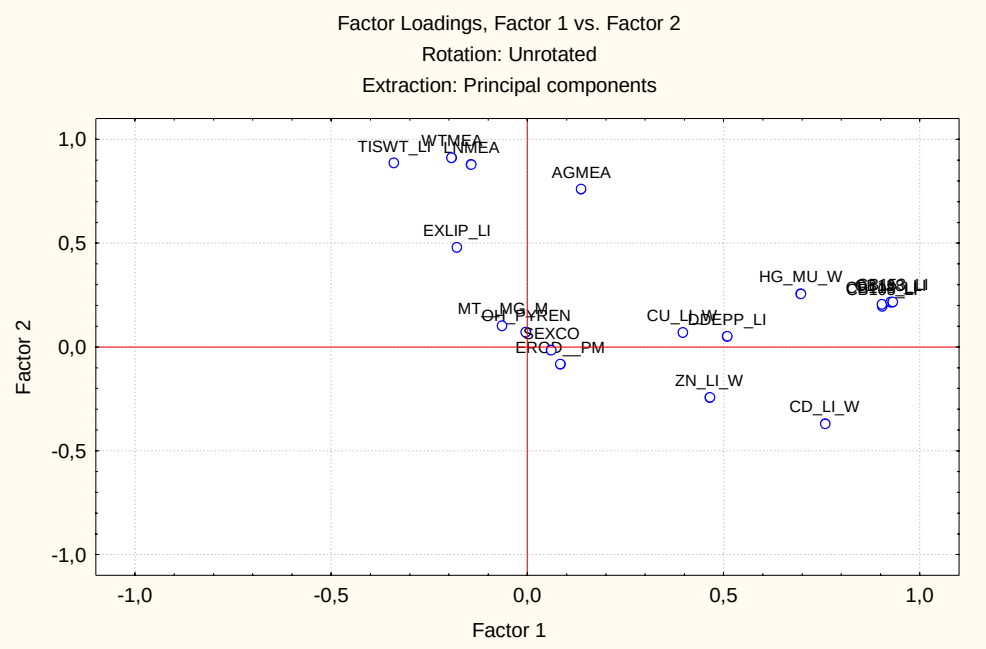
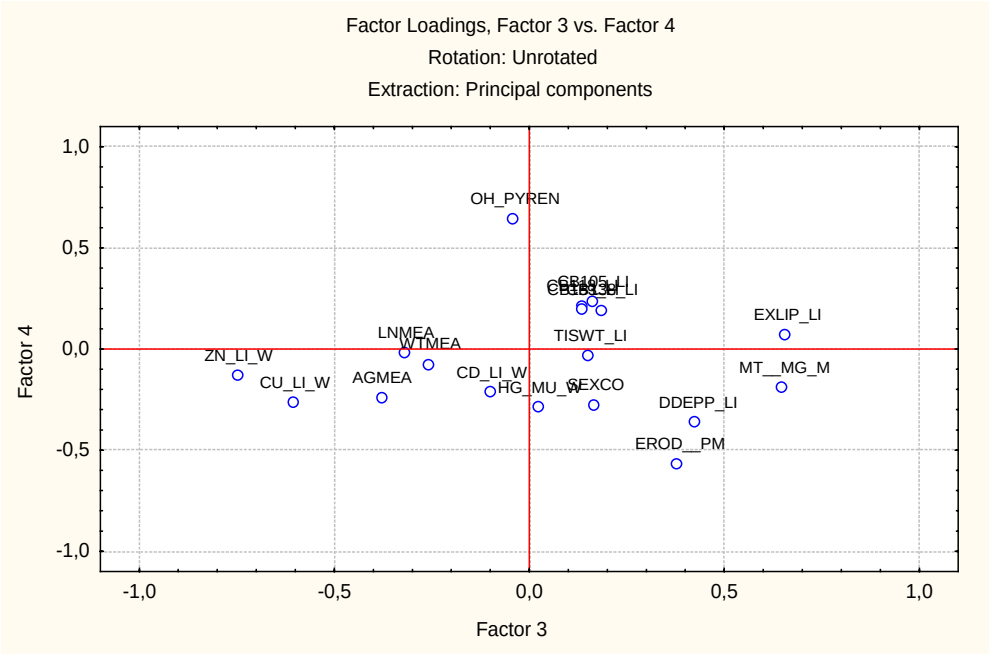


Figure 5.2.2. Factors 3 and 4 in PCA with case-wise deletion.



To identify any obvious differences between sites, EROD and MT responses of individual fish were analysed using PCA (Figures 5.2.3 and 5.2.4).

Figure 5.2.3. Factors 1 and 2 in PCA with individual fish from the six stations.

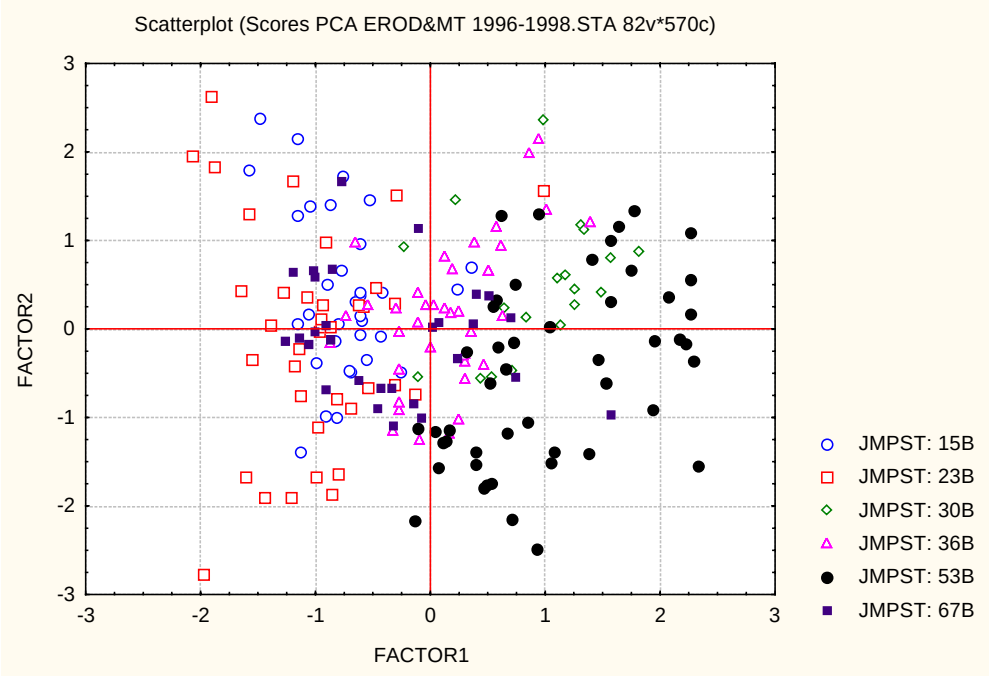
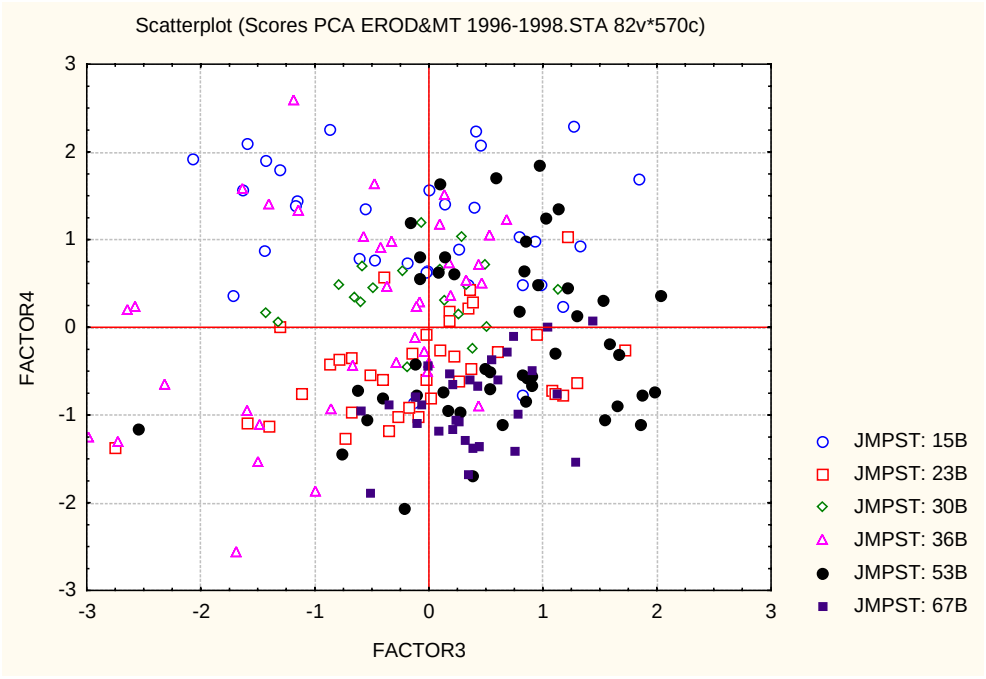


Figure 5.2.4. Factors 3 and 4 in PCA with individual fish from the six stations.



(2) Relationships between biomarkers, physiology and contaminants.

The biomarkers were analysed with GLM models for log-transformed variables, with and without station and/or year effects included.

Metallothionein (MT): There was a clearly significant negative relation between MT and zinc, with MT being proportional to the inverse square root of zinc concentrations, both with and without station*year included. However, Zn accounts for at most 15 % of the total variance of MT over the stations with data. There also appears to be a relation to fat content of the liver. For Cu, Cd in liver and Hg in muscle there are less clear relations with MT, with significance depending on what other terms are included in the model.

EROD: Without separation by station, EROD is significantly correlated to the contaminants OH-pyrene (in bile), Cd, *pp′*-DDE and Hg (in muscle), and also to physiology weight and liver fat condition measures. With station included as a factor, only Cd remains significantly correlated to EROD; for the other contaminants the correlation is mainly due to covariance of differences between stations with station differences of EROD.

ALA-D: Without station included as a factor, there is a significant positive correlation with Zn, and a negative correlation with Cd and Hg. Cadmium is the most important of these three contaminants, accounting for about 25 % of the total variance. When station is included as a factor, the relations are weaker, but still significant and in the same direction. Lead was not included in the analyses due to a large number of less-than values.

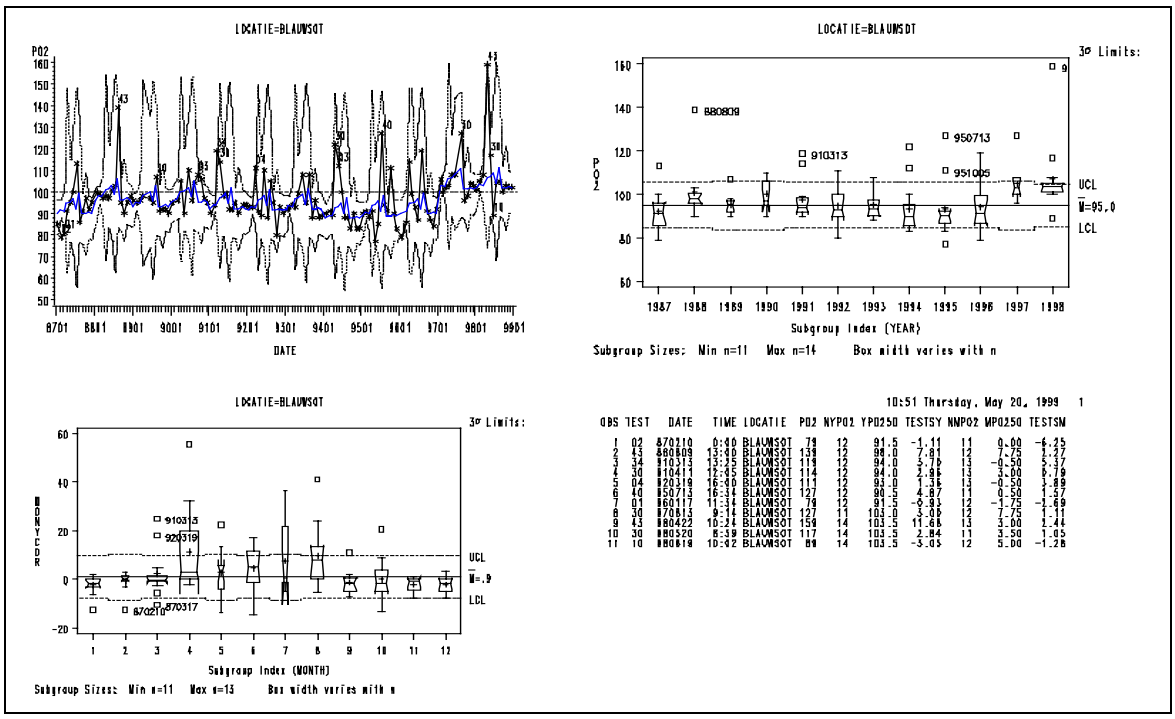
6 CONTINUE TO REVIEW STATISTICAL METHODS FOR ASSESSING AND DESIGNING MONITORING PROGRAMMES

6.1 Inspecting Monthly Time Series with the use of Box-and-Whisker Plots to Reveal Anomalies

R. Duin presented a simple and robust method that can be used to judge monthly time series on possible outliers or interesting phenomena. It is based on the terminology of the Box-and-Whisker plot which can be used to identify a value as an extreme or far extreme value. For a good judgement of an extreme, four figures are presented: (1) a box-plot for the values within each year; (2) a box-plot for the monthly values which are corrected for the annual median values; (3) the time series with a trend which is the annual median and the median residual monthly value and a roughly 99 % confidence interval, and (4) a table with the yearly and monthly extremes which for comparative purposes are standardised by the interquartile range. By presenting the extremes with respect to year and month, a better judgement could be given concerning whether these data are some sort of outlier or revealing something which could be interesting. In Figure 6.1.1 an example is given for a time series of the percentage oxygen in water for some location. In 1992, a monthly extreme for March is reported. For that time a phytoplankton bloom was started early in the year which

caused a high oxygen concentration. So be careful not to identify an extreme too quickly as an outlier, because it could be explained by other processes which are involved. It might be interesting also to look at the behaviour of the interquartile range, because it could be that the fluctuation of the signal can change during the years.

Figure 6.1.1. The template of four figures to judge a monthly time series. In the upper left corner, the original time series with the trend and the roughly 99 % confidence interval; in the upper right corner, a box-plot of the annual measurements; in the lower left corner, the box-plots for the median monthly value corrected for the annual median; in the lower right corner, a table with the extremes for the year and month.



6.2 Robust Trend Analysis based on a Penalized LS Approach

R. Duin presented a curve-fitting method which combines a least squares curve fitting with a penalty on the roughness of the curve and downweighting extreme values. It is used to analyse depth profiles which have much variation. The curve fitting is done by minimizing the following penalized least squares equation, which has two penalties in it:

$$Q = \sum_{i=1}^{n} w_i^{\,2}(z_i - t_i)^2 + \lambda \sum_{i=1}^{m} \left(\frac{\Delta t_i}{\Delta y_i}\right)^2 + \alpha^2 \sum_{i=1}^{m} (1 - w_i)^2 \ ,$$

the first part is the sum of the squared differences between the measured value (z_i) and the value of the smoothed curve (t_i), the second part is the penalty on the sum of the squared distance (Δy_i) between-points weighted roughness measure of the curve (the differences between the adjacent values Δt_i) and the third part is the penalty on the sum of the squared weights (w_i.). The striving is to give every value a weight 1, but when α is small this endeavour is not so difficult. High λ gives a smooth curve. By introducing a weight for every value the first part (the squared residuals) could be downweighted for influential values. Also this weight could be used for missing values. In that case the w_i =0 and the value of z_i is interpolated. Both parameters λ and α have to be given by the user, but for the use of the depth profiles a physical meaning can be given. An example of a depth profile is given in Figure 6.2.1.

Further development of this method for the analysis of data with, e.g., seasonality looks promising, and also including the second differences instead of the first differences for the roughness measure could be give better results.

Reference

Eilers, P.H.C., and Marx, B. 1996. Flexible smoothing with splines and penalties (with Discussion). Statistical Science, 11: 89–121.

Figure 6.2.1. A depth profile with the smoothed curve is given in the upper part of the figure. In the middle figure residuals are shown and the lower figure gives the individual weigths of the values of the series. Note that the influential values are having a nearly zero weight.

6.3 Mann Kendall Test of Trend with Missing Observations

One problem with the Mann Kendall test of trend is the presence of missing observations, that cause uneven spacings on the time axis. A common solution is simply to ignore the missing observations and treat the time series as if the observations were evenly spaced. However, Alvo and Cabilo (1995) described a modified version of the Mann Kendall test which incorporates the information about the spacing between the observations.

Writing $y_1, \ldots y_T$ for a series of k observations of which T–k of observations 2 …(T–1) are missing, the modified Mann Kendall statistic is

$$A_m = \sum_{i<j} a(i, j)$$

where

$$a(i,j) = \begin{cases} \text{sgn}[r(j) - r(i)] & \text{if } \delta_i \delta_j = 1 \\ \dfrac{2r(j)}{k+1} - 1 & \delta_i = 0, \ \ \delta_j = 1 \\ 1 - \dfrac{2r(i)}{k+1} & \delta_i = 1, \ \ \delta_j = 0 \\ 0 & \text{otherwise} \end{cases}$$

where

$$\delta_i = \begin{cases} 1, \text{ if there is an observation at time } i \\ 0, \text{ if not} \end{cases}$$

and $r(i)$ is the rank of observation i scored from 1 to k.

Critical values for A_m are given in Alvo and Cabilo (1993). Approximate large sample tests using the Normal distribution can be based on the variance formulae given by Cabilo and Tilley (1999).

The powers of the modified and un-modified Mann Kendall test are compared in Cabilo and Tilley (1999), together with the corresponding versions of the Spearman rank correlation test. These powers were computed by simulation for three monotonic trend scenarios:

Scenario 1 $y_i = c_1 i + \varepsilon_i$

Scenario 2 $y_i = c_2 i^2 + \varepsilon_i$

and

Scenario 3 $y_i = c_3 \sqrt{i} + \varepsilon_i$

where $\varepsilon_i \sim N(0,1)$ and the constant c is chosen so that the derivatives at observation $T/2$ is the same for the three scenarios. In Scenario 1, c_1 is obviously the slope of the (linear) trend. Figures 6.3.1, 6.3.2, and 6.3.3 show the three scenarios with $T = 10$ and $c_1 = 0.05$, together with the corresponding constant derivatives (of 0.05) at $T/2$.

The results of the simulations are extensive and somewhat complicated. As well as simulating three values of c_1, c_2 and c_3 (such that $c_1 = 0.05$, 0.1 and 0.2), for the modified Spearman and Mann Kendall tests, results were also derived for every possible arrangement of up to $T/2$ missing observations (in positions 2,…,$T-1$) for both $T = 10$ and $T = 12$.

When no observations were missing, the Spearman test was more powerful than the Mann Kendal test. However, with missing observations, all four methods were most powerful for some combinations of the trend, number of missing observations, and *T*.

Figure 6.3.4 provides a gross summary of the results presented by Cabilo and Tilley. It shows the percentage of times that the modified Mann Kendall test had the greatest power of the four methods for both $T = 10$ and $T = 12$. We see that with increasing numbers of missing observations, the power of the modified Mann Kendall test relative to the other three methods gradually increases, until it is most frequently the most powerful.

Figures 6.3.5, 6.3.6, and 6.3.7 show the powers of both versions of the Mann Kendall test for all three scenarios with a trend of 0.05 and $T = 10$. In this case, the modified Mann Kendall test is generally more powerful.

Note that some of the variation in power is due to variation in the true size of the test around the nominal 5 %. An improvement would have been, for the purpose of the power comparisons, to have maintained a fixed true size of the test by incorporating a random element in the test as in Fryer and Nicholson (1999).

Figure 6.3.1 Scenario 1

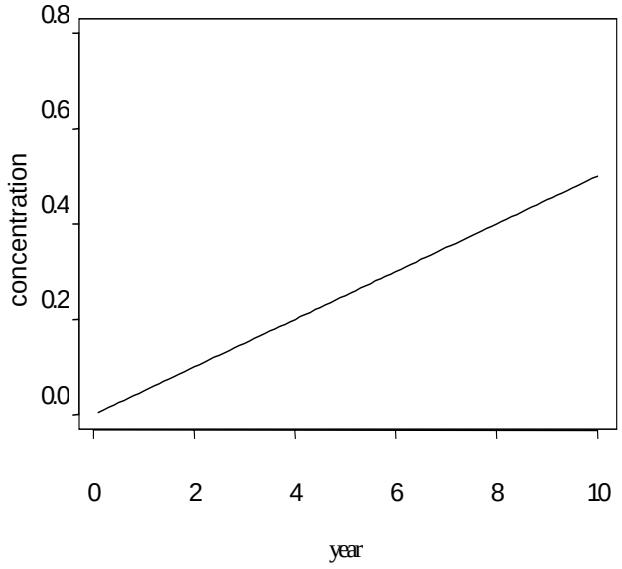


Figure 6.3.2 Scenario 2

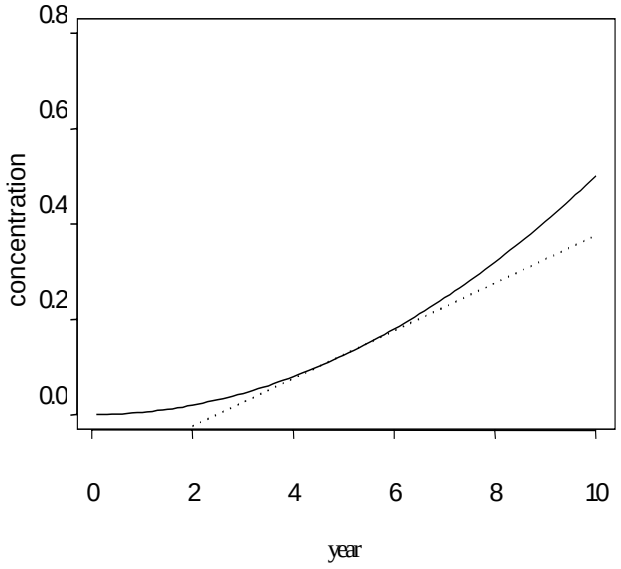


Figure 6.3.3 Scenario 3

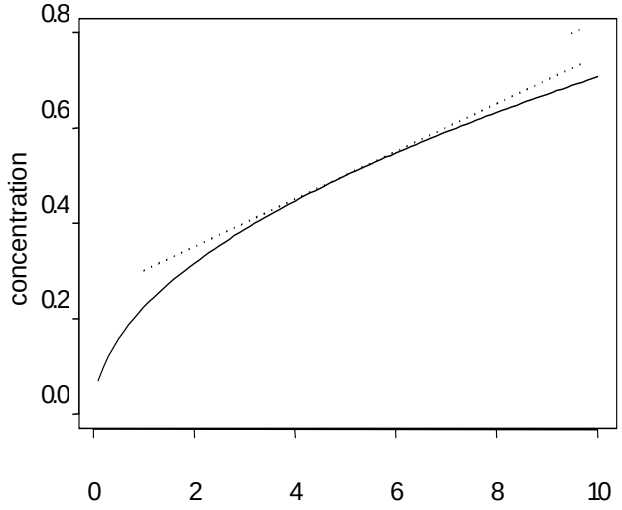


Figure 6.3.4 Overall Summary of Performance

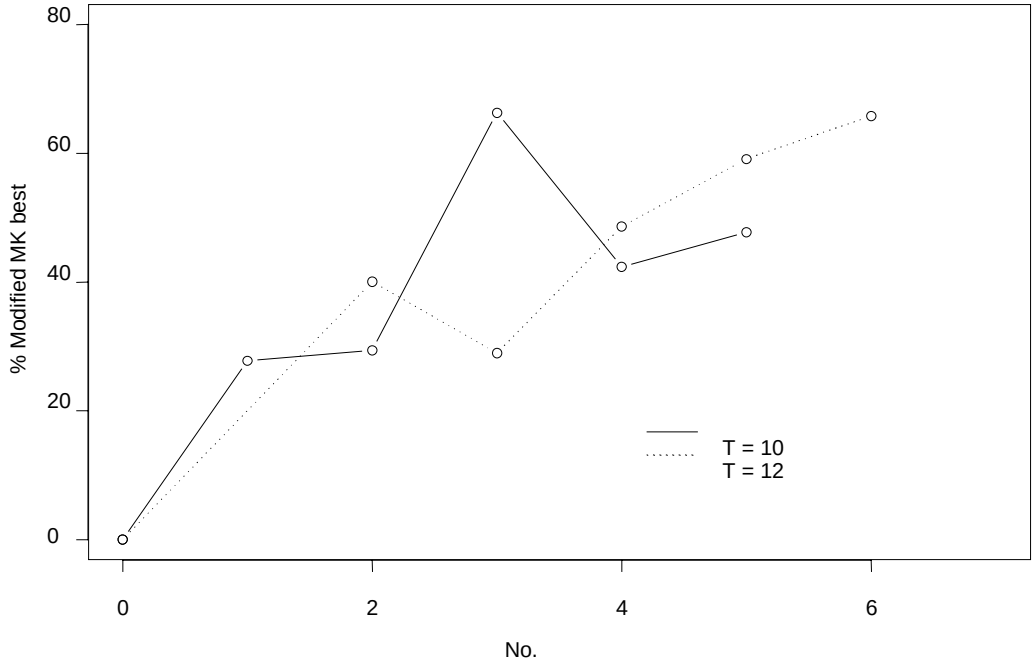


Figure 6.3.5

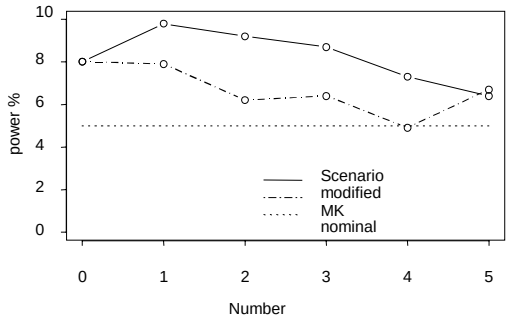


Figure 6.3.6

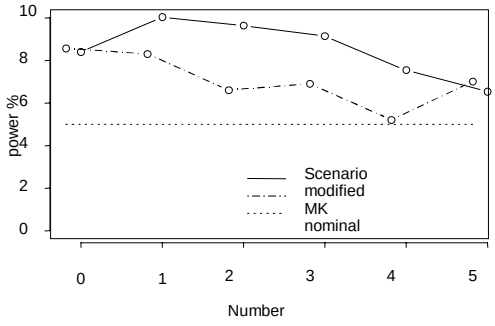
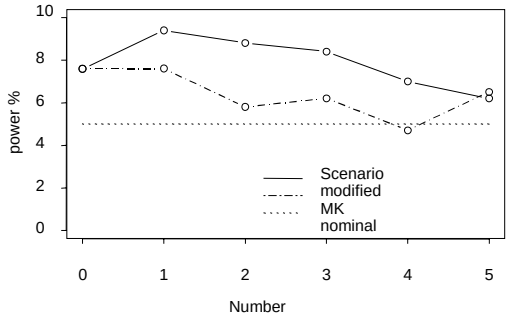


Figure 6.3.7



References

Alvo, M., and Cabilo, P. 1993. Tables of critical values of rank tests for trend when the data is incomplete. Technical Report No. 230. Laboratory for Research in Statistics and Probability, Carleton University and University of Ottawa, Canada.

Alvo, M., and Cabilo, P. 1995. Rank correlation methods for missing data. Canadian Journal of Statistics, 23: 345–358.

Cabilo, P., and Tilley, J. 1999. Power calculations for tests of trend with missing observations. Environmentrics, 10: 803–816.

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

6.4 Statistical Methods for the Analysis of the ICES Fish Disease Data Bank

The ICES Environmental Data Centre and other ICES Data Banks contain data collected in the framework of various monitoring programmes. As these programmes are not generally designed for pooling their data by time and location, the union of data from different programmes will often contain missing values for some parameters, time points and locations. In the holistic analysis of fish disease data as undertaken by the Working Group on Pathology and Diseases of Marine Organisms (WGPDMO) (cf. Report of the WGPDMO 2000 meeting, Section 7), the problem of missing values in time series of potentially explanatory variables was treated by the following procedure, which exploits the time series structure of the explanatory quantities:

1) fit a smooth estimate to the time series that contains missing values;
2) replace missing values by interpolation estimates from the smooth and record the local prediction error;
3) generate a random variation for each interpolated value (normally distributed with mean zero and standard deviation taken from the previously recorded prediction error);
4) generate a new set of response cases, using a binomial distribution with the observed number of cases and the empirical prevalence as parameters;
5) estimate the logistic parameters for the generated data set and record them;
6) repeat the preceding three steps sufficiently often (1000 times gave stable results);
7) calculate the empirical mean and quantiles of the estimated parameters and assign the corresponding p values to them.

This procedure leads to estimates for the logistic parameters, which account as well for the usual binomial variation as for the uncertainty that has been introduced by using interpolated instead of observed values. Comparing the estimates obtained with and without correction for using interpolated values led to two general observations: (i) p values for the test of the hypothesis “parameter value is zero” are increased when using the correction, which means that the hypothesis is rejected in fewer cases, and (ii) the estimated parameters move towards zero, which means that less relevance in the sense of explanatory power is assigned to the variable. Both observations are in line with what is to be expected, as the use of interpolated instead of observed values means introducing less information in the estimation process, hence the ability to detect relationships should decrease.

However, also after correcting for interpolated values, a considerable number of significant relationships were still found in univariate as well as in multivariate logistic relationships between fish disease prevalence and the variables under consideration. This finding shows that at least for the amount of missing values occurring in this analysis, the use of interpolated values does not completely destroy the chance to detect relationships. WGPDMO has thus recommended to continue the analysis of fish diseases with the intention of broadening the basis for the analysis by considering also other length classes/sexes than the currently used subset of female dab with lengths between 20 cm and 24 cm. It should, however, be noted that, where possible, the use of real data is the preferable alternative.

6.5 Superbeast

Annex 6 describes a simple model of the relationship between the concentration of a contaminant in an organism and

- the uptake and excretion rate of the organism,
- the ambient concentration.

A similar model was presented in WGSAEM (1993), and used to show how suitable uptake and excretion rates are important considerations in the choice of an appropriate monitoring organism. Here, the model was used to show how concentration-size relationships in the organism might evolve as a result of changes in the ambient concentration. In particular, steeper concentration-size relationships would be expected when ambient levels are decreasing. Some tentative support for the model was found in the literature.

The implications are that:

- assessment methodology for contaminant time series should be able to deal with evolving size-dependent trends,
- studying the form of size-dependent trends might help in the interpretation of changes in ambient concentration.

In his presentation on the French Mussel Watch programme (Section 8), Didier Claisse talked about uptake and excretion experiments that had been conducted on shellfish. WGSAEM noted that these experiments would provide estimates of uptake and excretion rates that could be used to further the development of the Superbeast model.

6.6 Standardized Statistical Analysis of Monitoring Data: the IFREMER Experience

Answering a request from the IFREMER coastal laboratories, a functional structure has been developed in IFREMER to allow coastal laboratories to perform standard statistical analyses using tools provided by the IFREMER “Operational Applications” Department in Nantes (France). This leads to the production of the bulletin of presentation of monitoring results, one for each laboratory. Data analysed are issued from the bacteriological (REMI), phytoplankton (REPHY), and contaminant (RNO) monitoring programmes.

Among other tools necessary to produce the bulletins, scripts S+ are provided to the laboratories. 1999 was the starting year and the people involved in making the document were trained for using S+ and the other tools, i.e., EXCEL Macros. Only graphics were produced in 1999. Trend analysis will be added in the 2000 bulletins (e.g., Figure 6.6.1). The seasonal Mann-Kendall test is used, and found to be very adequate, for montonic trend detection in bacteriological data. Polynomial adustment up to the third degree on yearly medians is used for trend detection of contaminants in biota time series.

Bulletins are available in PDF format at:

http://www.ifremer.fr/delao/surveillance/quadrige/produits/bulletins.html

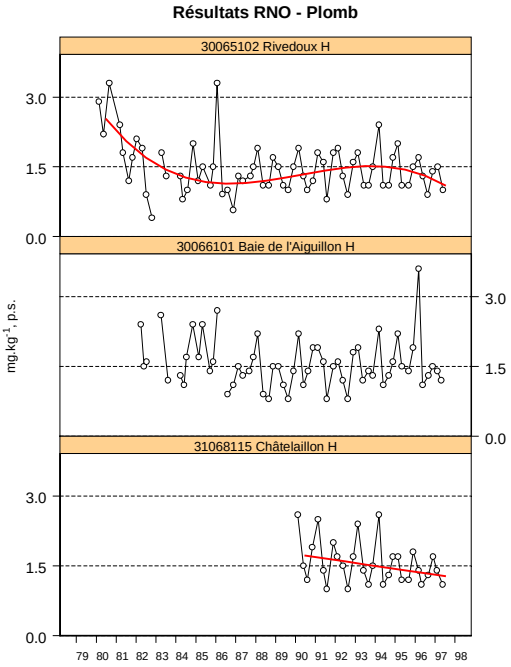


Figure 6.6.1. Lead results in oyster for La Rochelle coastal laboratory (three sampling points).

6.7 Statistical Methods to Summarize ICES Fish Disease Monitoring Data for Publication

W. Wosniok presented the analysis approach adopted by the WGPDMO during its 2000 meeting as a part of its effort to make information on the developments of fish disease prevalence accessible via the internet. This information is to summarize data that have been collected by ICES Member Countries in the framework of various monitoring programmes and submitted to the ICES Environmental Data Centre. The assessment of these data is performed with the intention to publish, in addition to a data inventory (number of data points and number of fish examined, per ICES statistical rectangle),

- maps with indicators for the character (up, down, constant) of recent trends for major diseases (lymphocystis, epidermal hyperplasia/papilloma, skin ulcerations), per ICES rectangle, and

- full plots of estimated disease prevalence trends for selected rectangles.

These displays are subject to the condition that a sufficient amount of data is available for the corresponding area. Only estimated prevalences are used for display. Raw prevalences were considered to be inappropriate because of their high variability and irregular temporal distribution, which would need too voluminous an interpretation aid. Besides that, estimated prevalences are the quantities actually employed for the trend assessment, and it was considered a crucial requirement to use the same material for assessment and presentation in order to avoid irritation of the reader. This makes it necessary to design a trend assessment that can be understood from the graphical display, hence the following procedure was developed.

Estimated prevalences are obtained from raw values by fitting a smooth curve to the latter, using a locally weighted second-order polynomial regression procedure. A recent trend is defined as a significant change in the estimated prevalence during a five-year period backwards from and including the reference year (actually 1997). A change is defined as “significant” in two different situations: (i) if the lower limit of the prevalence confidence interval for the reference year is above the minimum of the upper confidence bound of the trend within the assessment period (upward trend), and (2) if the upper limit of the prevalence confidence interval for the reference year is below the maximum of the lower confidence bound of the trend within the assessment period (downward trend). All other situations are labelled as “no trend”. It should be noted that, due to the smoothness of the estimated trend, the simultaneous occurrence of a minimum and a maximum within the assessment period is unlikely and has not yet been observed.

Confidence bounds are obtained by a bootstrapping procedure, in which for each observation on the time scale a binomial sample is generated, using the actual number of fish examined and the empirical prevalence from that time point as parameters. The collection of binomial samples for all time points with actual observations constitutes one bootstrap replicate. For each replicate, a smooth trend is estimated by locally weighted regression. This procedure is

repeated 1000 times, and for each estimated trend its Kolmogorov-Smirnov distance to the trend for the original observations is calculated. Replicate trends with distances larger than the empirical 95 % quantile of these Kolmogorov-Smirnov distances lie outside the 95 % confidence bound of the estimated trend. The graphical display shows the hull curves in which all replicate trends that lie inside the 95 % confidence bound are embedded. All calculations are performed separately for observations made between October to March and for observations made between April and September, as these two periods show clearly different trends.

This approach for trend assessment is not robust against outliers in the 5-year trend analysis period. The resulting danger of being erroneously alarmed by the indication of a trend induced by (an) outlier(s), is recognized, but assigned minor weight compared with the main goal that detection of a trend should initiate a more detailed analysis and possibly further action. The possible consequence of reacting on a falsely indicated trend is considered a less severe event than not reacting on a really existing trend.

7 FULLY EXPLOIT THE VOLUNTARY INTERNATIONAL CONTAMINANT MONITORING IN TEMPORAL TRENDS (VIC) DATA ON CONTAMINANTS IN BIOTA, IN ORDER TO IMPROVE SAMPLING STRATEGIES FOR THE ANALYSIS OF BOTH TEMPORAL AND SPATIAL TRENDS

Birger Bjerkeng reported on a paper (Annex 7) that he had presented to SIME that had exploited the data submitted by the Netherlands, and Norway to the VIC programme. This programme was designed to provide information about small-scale temporal and spatial variation, essential for providing estimates of components of variance that may improve the efficiency of the OSPAR monitoring programme.

From the Netherlands, data for flounder have been submitted, collected at different sites and at two times each year from the estuary of Westerschelde over the years 1996–1998, with both individual and pooled samples. From Norway there are data for cod and flounder collected at different sites and/or times in the Oslofjord, and in Sørfjord and Hardangerfjord on the west coast, with subsets covering both temporal and spatial small-scale variation. From Sweden there are data for herring from different locations in the southwestern part of the Baltic, the Kattegat and the Skagerrak. The within-year variation in these data are mainly spatial, and describe variation over larger distances in open waters.

The results of the analysis so far indicate that at least for some species and contaminants the within-year variance could be substantially reduced by sampling from more than one site or repeated sampling over time. Depending on cost components, this might result in improved efficiency per cost.

To demonstrate how these data might be used, Table 7.1 shows the estimated components of between-site, between-time, and between-individual-fish variances for log CB153 measured in cod livers from the Oslofjord. The spatial scale is 5–10 km within the total area of 20×50 km^2 of the Oslofjord. The temporal scale spans 2 weeks.

Table 7.2 provides a corresponding estimate of the between-year variance obtained using the full JAMP data for two stations (15B and 23B) with relatively small yearly fluctuations.

We can combine these estimates to give an estimate of the total between-year variance of an estimated annual mean log concentration, and evaluate the corresponding power, e.g., to detect a specific linear trend in a given number of years (c.f. Nicholson *et al.*, 1997). We can then explore the consequences of different allocations of sampling effort within a year.

Figure 7.1 gives plots of the powers to detect trends of 5 % and 10 % after 10 years assuming a 5 % significance level and 24 fish sampled per year. Fish are assumed to be collected in equal numbers from each visited site on each sampling occasion such that the total number of fish is constant. The power is plotted against the number of sites.

If spatial and temporal variation are generated by random fish movements, they may be measuring the same thing. In this case, only one of them should be included. The solid lines in each figure therefore show the change in power assuming that the between-time variance is zero, with the 24 fish sampled at 1 site, 2 sites, and so on to each fish coming from a different site.

The dotted lines assume that both between-site and between-time variance components are present. From top to bottom, the lines correspond to sampling at one time, 2 times, and so on up to 12 times.

The power corresponding to the current sampling guidelines of 24 (actually 25) fish sampled on one occasion from a single site falls somewhere between the point on the lowest dotted line at sites =1 (and times =1) and the solid line at sites=1. Clearly for this example, there is considerable opportunity for improvement.

WGSAEM noted that the data sets available for this kind of analysis are very patchy with respect to species, geographical and temporal scales and sampling design. Thus the results are fragmentary, and further, are based on low degrees of freedom. WGSAEM felt that it would be worthwhile to extend the database by locating and incorporating similar data sets.

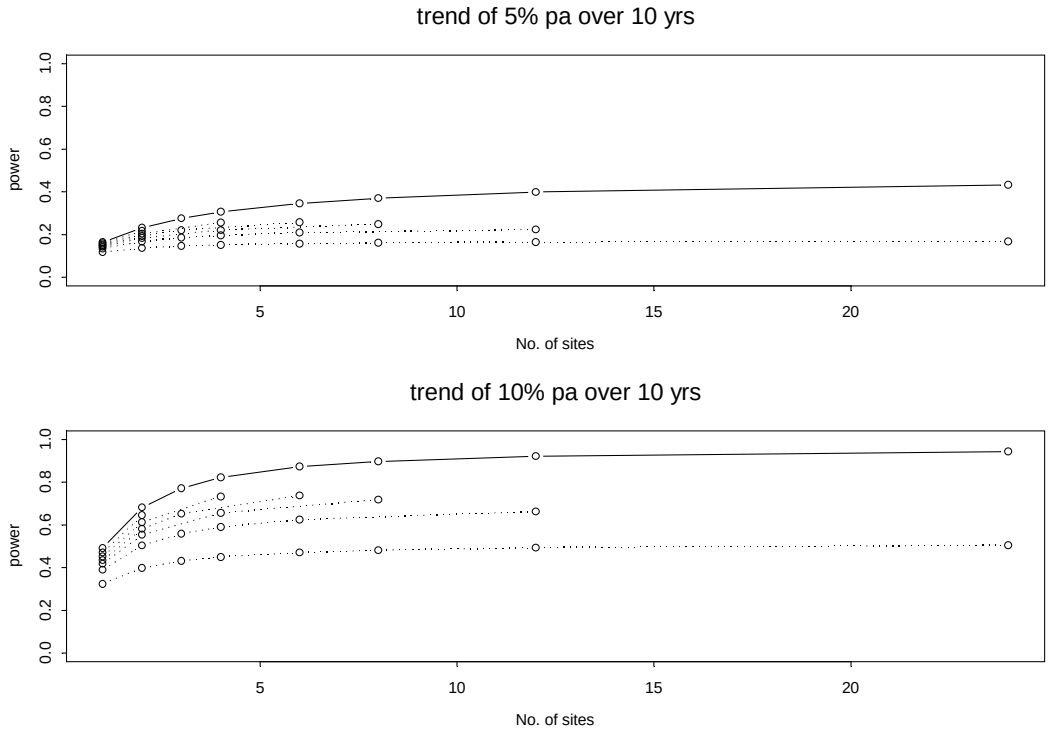
Table 7.1. Estimates of variance components for CB153 in cod livers derived from ANOVA of the VIC data for station 30B (location Maasene) in inner Oslofjord.

	CB153_LI
σ_t^2 between *Times* within year	0.114
σ_s^2 between *Sites* within year	0.125
σ_I^2 between *Individuals* within year	0.256
Total variance for individual fish within year	0.495

Table 7.2. Estimated between-year component of variance for CB153 in cod livers using the full JAMP data for two stations (15B and 23B) with relatively small yearly fluctuations.

	CB153_LI
Total between-year variance (20–25 fish samples)	0.055
Within-year variance for individual fish: 0.4–0.5	0.47
Within-year variance for mean over 22 fish	0.021
σ_y^2 Between-year variance component	0.034

Figure 7.1. Plots of power to detect trends of 5 % and 10 % after ten years assuming a 5 % significance level and 24 fish sampled each year.



8 PRESENTATIONS OF NATIONAL MONITORING PROGRAMMES, IN ORDER TO MAKE GENERALISATIONS AND SPECIFIC RECOMMENDATIONS, E.G., ON THE NUMBER OF REPLICATE SAMPLES OF SEDIMENTS OR BIOTA NEEDED TO CHARACTERISE AN AREA

8.1 The French Monitoring Programmes "Contaminants in Sediments and Contaminants in Biota"

Didier Claisse presented an overview of the French monitoring programmes "contaminants in sediments and contaminants in biota", stressing special issues of the monitoring strategies.

8.1.1 Chemicals in biota

The Mussel Watch-like programme was initiated in 1979. Eighty points are sampled quarterly for metals and organics along the French coast (Figure 8.1). Samples have been systematically kept freeze-dried since 1990 for further analyses. For example, individual CBs and PAHs can now be measured in samples collected at a time when only total CBs and PAHs were able to be measured.

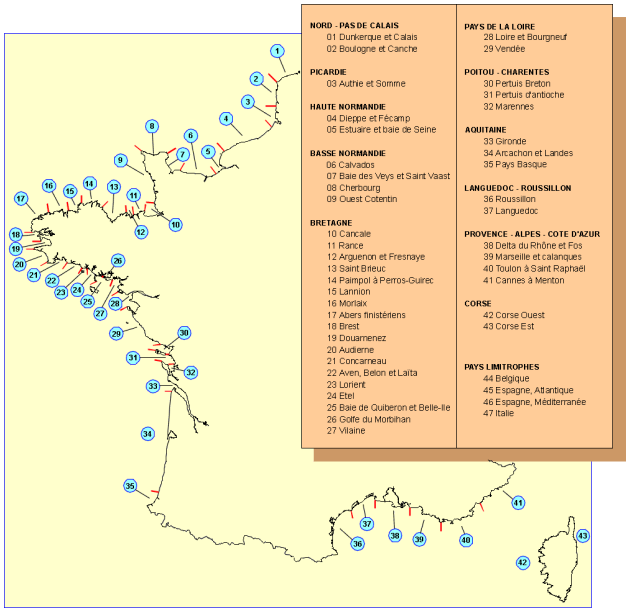


Figure 8.1. RNO monitoring sites.

8.1.2 Chemicals in sediments

Contaminants have also been analysed in sediments since 1979, but following the strategy explained below since 1993 only. The first cm of the superficial sediment layer is known to integrate several years of contamination. Therefore each site should not be sampled on a yearly basis. Each year only one site of the French coast (e.g., Figure 8.2), corresponding to 20–25 points, is sampled. Sediments are analysed for the same contaminants as for biota. Moreover, normalization parameters are measured, such as granulometry, organic carbon, carbonates, aluminium, iron, lithium and manganese.

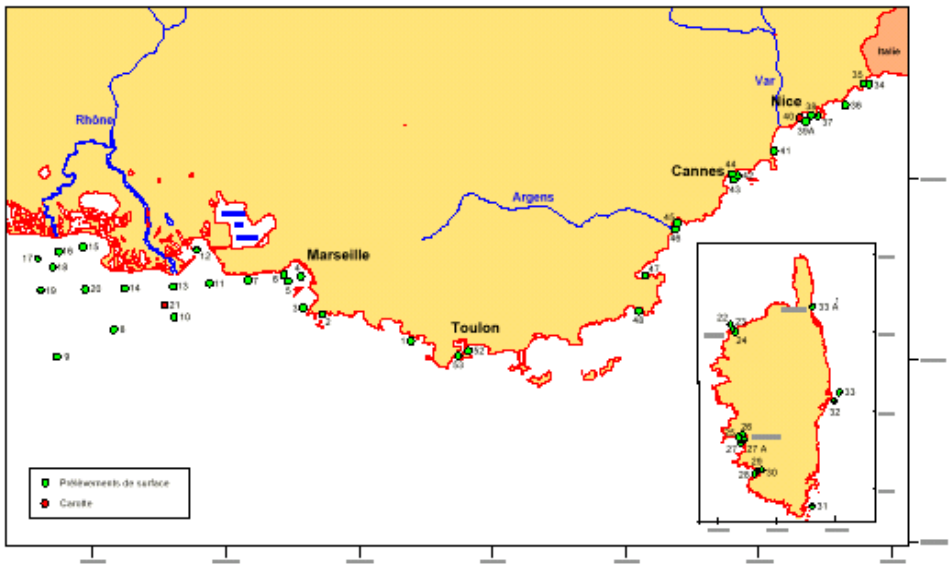


Figure 8.2. 1998 sediment survey in PACA (Provence and Côte d'Azur RNO monitoring sites).

8.2 The National Salt Water Monitoring Programme of the Netherlands

R. Duin presented the National Salt Water Monitoring Programme of the Netherlands. This programme consists of a chemical, a biological, and a physical component and is coordinated by the National Institute of Coastal and Marine Management/RIKZ.

The chemical monitoring started in 1966 and has two major objectives: temporal trend monitoring and compliance with national criteria. In 1995, an optimisation and modification of the programme took place with respect to these objectives. It consists of chemical monitoring in the water, the sediment, suspended matter, and biota and biological effects (fish disease) monitoring. The JAMP monitoring programme of the Netherlands is part of this national programme. The results are published yearly in the National Evaluation Report.

The biological monitoring started in 1990 and consists of benthic organisms, phytoplankton, zooplankton, water birds of coast and estuaries, seabirds and sea mammals. The main objective of this programme is to provide information on long-term developments. An evaluation of this programme is planned for 2000.

The physical monitoring deals with the monitoring of the bathymetry of the coastal part of the North Sea, the Wadden Sea and the Delta, the discharges, the waves on the sea, water height, and water temperature. An evaluation of this programme is planned for 2001.

There is an ongoing harmonisation of the three monitoring programmes.

9 CONTINUE THE DEVELOPMENT OF SAMPLING ALLOCATION STRATEGIES, ESPECIALLY THE DEVELOPMENT OF DYNAMIC SAMPLING STRATEGIES

Richard Duin presented some preliminary results for contaminants in sediments of the Dutch part of the North Sea. Since 1981, every five years about 60 locations are investigated for metals and organic contaminants. For the North Sea four areas are distinguished which are followed in time. The sampling design consists of some fixed stations (visited every sampling time) and new "randomly" chosen stations (see Table 9.1).

Table 9.1. The number of sites which are revisited with sampling of sediment for metals and organic contaminants in the North Sea.

Number of Sampling Times	Number of sites
4	9
3	18
2	34
1	145

Assessing trends with such data can be difficult, and WGSAEM discussed various approaches that could be taken. One approach which looked promising was that described by (our-old-friend-Bill) Warren (1994) who adopted a dynamic sampling strategy in which the sampling effort each year was allocated partly to a fixed subset of sites, and partly to a new, randomly chosen group of sites. Warren also discussed the conditions under which all-fixed, all-random and partial-replacement strategies would be superior.

Because of the complexity of the sampling design and lack of time, no direct answer for sampling allocation strategies could be given.

Reference

Warren, B. 1994. The potential of sampling with partial replacement for fisheries surveys. ICES Journal of Marine Science, 51: 315–324.

10 ANY OTHER BUSINESS

10.1 A Method to Analyse Seabird Data with both Kriging and the Application of the GLM

Richard Duin gave an overview of a method of mapping seabird densities on the North Sea. Seabird densities were estimated by a generalised linear modelling approach combined with geostatistics for the spatial interpolation.

10.2 Free, Downloadable Software for Computing the Power of Monitoring Programmes

A recent publication by Sheppard (1999) presented a discussion of the importance of assessing the power of monitoring programmes. This applies both to effective design, and to post-assessment evaluation when reporting non-significance. Many of the points were echoes of those made in previous WGSAEM reports.

The paper drew attention to a DOS programme (G*power) for computing power downloadable from www.psychologie.uni-trier.de:8000/projects/gpower.html. This website provides extensive guidelines, and a good discussion of the principles and theory of power assessment.

WGSAEM found G*power to be somewhat basic, focusing only on a t-test, one-way ANOVA and analysis of 2-way contingency tables. However, the corresponding non-central distributions might be useful for those without access to appropriate software.

Other software programmes that can be used to calculate the power of statistical hypothesis tests are listed in http://sustain.forestry.ubc.ca/cacb/power/index.html.

This website also contains a link to a paper by Thomas and Krebs (1997), which provides some evaluation of several power-evaluation programmes.

References

Sheppard, C.R.C. 1999. How large should my sample be? Some quick guides to sample size and the power of tests. Marine Pollution Bulletin, 38: 439–447.

Thomas, L., and Krebs, Ch. 1997. Bulletin of the Ecological Society of America.

11 ACTION LIST/PREFERENCES

Werner Wosniok	Replacement methods and multivariate methods for biological data
Rob Fryer	Dynamic sampling
Richard Duin	Analysis of sediment data
Birger Bjerkeng	Biomarker analysis and VIC data analysis
Benoit Beliaeff	Investigation of the methods for the analysis of quarterly data
Mike Nicholson	Sampling to control the risk of failing to detect hotspots
Steffen Uhlig	Investigation of trend detection methods for the analysis of monthly data

12 RECOMMENDATIONS

1) ICES should recommend to INPUT 2000 that a complete and fully documented version of the Trend-y-tector be evaluated, and also:

   a) for reference, the Trend-y-tector should be given version numbers;
   b) consideration should be given to providing more extensive additional analysis and reporting such as % variation explained and, in the case of the smoother, further analysis and reporting of the linear and non-linear components of the trend in the whole time series;
   c) consideration should be given to extending the methods to allow more sophisticated treatment of missing values;
   d) some post-hoc assessment of power is still a good idea;
   e) ICES may wish to advise INPUT 2000 to adopt some method for trial, and be prepared for the method and reporting format to evolve in response to their experience.

2) WGSAEM notes that at present no single method of calculating annual adjusted loads can be recommended for all river systems and all substances. However, when considering nutrients, if load is approximately linearly related to flow, method L1 will perform reasonably well and WGSAEM recommends that it be used on a trial basis.
3) ICES should note the small number of participants at WGSAEM 2000, and encourage national delegates to remember us in the future.
4) ICES should consider that a member of the ICES Secretariat should attend the WGSAEM meeting, in order to aid communications between WGSAEM, ICES, and OSPAR.
5) WGSAEM should meet for five days in March/April 2001 in Oslo.

13 CLOSING OF THE MEETING

There being no further business, the Chair closed the meeting at 17.00 hrs on Friday, 31 March 2000.

ANNEX 1: AGENDA

Item 1 Arrival, collecting and submitting papers

Item 2 Opening of the meeting

Item 3 Adoption of agenda and organisation of work

Item 4 Continue the development of trend detection methods in order to:

consider further development and assessment of robust smoother methods and the development of appropriate techniques for revealing outlying data values [OSPAR 2000/2.1],

consider further development of statistical methods for adjustment of input loads [OSPAR 2000/2.2],

develop provisions for the use of monthly data in trend detection methods [OSPAR 2000/2.3].

Item 5 Review and report on results of investigations concerning the characteristics of sampling and analytical variability of biomarkers and biological endpoints, the design of effective sampling schemes relative to specified objectives, and concerning the development of appropriate management tools for integrating and interpreting biological effects.

Item 6 Continue to review statistical methods for assessing and designing monitoring programmes.

Item 7 Fully exploit the Voluntary International Contaminant Monitoring in Temporal Trends (VIC) data on contaminants in biota, in order to improve sampling strategies for the analysis of both temporal and spatial trends.

Item 8 Presentations of national monitoring programmes, in order to make generalisations and specific recommendations, e.g., on the number of replicate samples of sediments or biota needed to characterise an area.

Item 9 Continue the development of sampling allocation strategies, especially the development of dynamic sampling strategies.

Item 10 Based on available data for a suite of biological, chemical, biomarker, and endpoint measurements, carry out the following tasks, to the extent possible:

explore, on a univariate basis, the minimum difference in level between two stations that can be detected with 90 % power following the JAMP guidelines, for each chemical, biomarker, and biological endpoint variable,

investigate statistical methods for modelling the relationships between the biological, chemical, biomarker, and biological endpoint measurements,

begin the consideration of ways of combining a suite of biological, chemical, biomarker, and biological endpoint measurements into summary indices that may be suitable for management or statistical purposes.

Item 11 Any other business

Item 12 Action list

Item 13 Recommendations

Item 14 Close the meeting

ANNEX 2: LIST OF PARTICIPANTS

Name	Address	Telephone	Fax	E-mail
Benoit Beliaeff	IFREMER DEL/AO BP 21105 44311 Nantes Cedex 3 France	+33 2 40374158	+33 2 40374073	Benoit.Beliaeff@ifremer.fr
Birger Bjerkeng	Norwegian Institute for Water Research P.O. Box 173 Kjelsaas N-0411 Oslo Norway	+47 2218 5151	+47 2218 5200	Birger.Bjerkeng@niva.no
Richard Duin	National Institute for Coastal and Marine Management P.O. Box 20907 2500 EX The Hague The Netherlands	+31 70 3114214	+31 70 3114231	R.N.M.Duin@rikz..rws.minvenw.nl
Rob Fryer	FRS Marine Laboratory Victoria Road Aberdeen AB11 9DB United Kingdom	+44 1224 295502	+44 1224 295511	R.Fryer@marlab.ac.uk
Philippe Gros	IFREMER DEL/AO BP 70 29280 Plouzané France	+33 2 98 22 4337	+33 2 98 22 4555	philippe.gros @ifremer.fr
Mike Nicholson	CEFAS Lowestoft Laboratory Pakefield Road Lowestoft Suffolk, NR33 OHT United Kingdom	+44 1502 562244	+44 1502 513865	m.d.nicholson@cefas.co.uk
Steffen Uhlig (Chair)	quo data Bruhmstr. 4 D-01465 Dresden-Langebrück Germany	+49 35201 70387	+49 35201 80687	Uhlig@quodata.de
Werner Wosniok	Institute of Statistics University of Bremen P.O.Box 330440 D-28334 Bremen Germany	+49 421 218 3471	+49 421 218 4020	Werner@math.uni-bremen.de

ANNEX 3: OSPAR REQUEST FOR WORK ON TREND DETECTION METHODS FOR INPUT DATA

2.1 Fine tuning of the Trend-Y-tector for trend detection in inputs

2.1.1 Consider type and specification of the Smoother. The current LOESS Smoother seems to be quite sensitive for outliers in the first and last part of the data series. ICES is asked to consider possible alternatives, e.g., the Spline Smoother.

2.1.2 Consider the calculation of residuals. Due to the risk that over-fitting of normal residuals may lead to underestimation of the standard deviation, ICES is asked if residuals based on cross validation may be a sensible alternative.

2.1.3 Consider the calculation of standard deviation. The current guidelines use the L moments for estimating the standard deviation. ICES is asked to consider the statistical consequences of using this or other robust estimators combined with corresponding critical values calculated by simulation studies.

2.1.4 Examine alternatives of the current smoother test. This test does not seem to be accurate. ICES is asked for possible alternatives, e.g., splitting the trend into a linear and a non-linear component and then testing the linear part.

2.1.5 Establish the link between overall significance and individual significance levels.

2.1.6 Specify the procedure for power calculation, and especially the post hoc power as an integral part of the trend assessment.

2.2 Adjustment of loads

2.2.1 Consider the general procedure of adjustment and trend assessment as outlined in INPUT(2) 98/5/3 and 98/5/4 with regard to its statistical implications.

2.2.2 Consider and examine the choice of the statistical model and the underlying variables for adjustment.

2.2.3 Give advice with respect to the choice of multiplicative or additive adjustment.

2.2.4 Give advice on how to measure the gain of adjustment.

2.2.5 Examine whether and when there is a risk of ‘over-adjustment’.

2.2.6 Consider whether the use of annual adjusted loads in the trend analysis of monthly loads may become redundant.

2.3 The use of monthly data

2.3.1 Development of provisions for the use of monthly data in these trend detection methods (taking into account that any recommendations should be based on real need and best scientific judgements and should not be driven purely by statistical considerations).

ANNEX 4: TREND DECTION METHODS FOR OSPAR INPUT

Report from OSPAR INPUT

To ICES WGSAEM meeting in Nantes
From Otto Swertz
27 March 2000

At the 2000 meeting of INPUT, there was good discussion the trend detection methods. On Monday, the revised Trend-y-tector was used in an assessment. People liked the programme very well. Later, the ICES work was discussed in detail. On the next two pages you can read the text that ended in the Summary Record.

The following summarises what work is expected from WGSAEM:

1) Finalise advice on load adjustment taking into accounts remarks from INPUT.
2) Decide whether the group has any comments on the work from last year, since The Netherlands (read Otto Swertz) will use this text as the basis for draft guidelines. Forward the comments to Richard Duin.
3) Advise on whether ICES wants to incorporate guidelines for adjusting data within the guidelines or as an annex. The latter has the advantage that adjusting concentration data can be added as annexes to the guideline in an easy way in a later stage.

Any comments on the currently partly updated Trend-y-tector are very welcome. Realize that we want to promote the Trend-y-tector as an application of the future OSPAR trend assessment guidelines. This leaves other applications open.

About the Trend-y-tector. It is partly updated. The current version can be used in batch mode.

So one can assess many data records at once and one can use the programme easily from other programmes. We intend to do some work on post-power analysis and taking account of variation wthin the estimated yearly value.

Looking forward to your work already,

Regards,

Otto

ANNEX 5: L1 METHOD

Adjustment of inputs

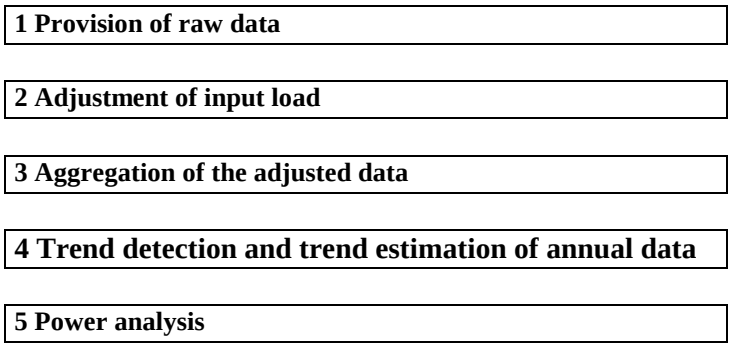
Steffen Uhlig

1 Introduction

Political targets on a substantial reduction of the quantity of nutrients and toxic substances reaching the aquatic environment of the North Sea also require checking whether reduction measures taken at the point and diffuse sources were effective in reducing the inputs into the North sea or not. This could be done on the atmospheric inputs (as deposition) and waterborne inputs (as riverine loads). However, the input data are highly dependent substance-specifically on climatic influences (precipitation, flow rate, temperature). In order to prevent that climatic influences deteriorate the trend detectability, an appropriate adjustment of annual and monthly input data prior to the analysis of temporal trends is necessary.

According to the decision of OSPAR INPUT, the steps of statistical trend analysis are as described in Figure A5.1. This concept allows use of the procedure for trend analysis based on annual data described in Section 6.4 of the WGSAEM 1999 report also for aggregated adjusted loads.

Figure A5.1. Steps of statistical trend analysis on riverine and atmospheric inputs.



The aim of adjustment is to compensate for the effect of varying runoff or precipitation. Compensation may be accomplished additively or multiplicatively, based on dynamic or non-dynamic estimation methods, but multiplicative adjustment based on a dynamic estimation method is preferred (WGSAEM 1999 Report).

The general procedure of multiplicative adjustment is as follows: Let c_{ij} denote the measured concentration of the *j*th sample in year *i*, taken at time t_{ij}. Let q_{ij} denote the actual runoff at that sampling time t_{ij} and q_{0ij} the corresponding long-term average runoff. Assume that there is a CQ function $c_{ij}(q_{ij})$ (concentration runoff function) describing the influence of current runoff and possibly further influence variables, so that the measured concentration can be well described by the statistical model

$$c_{ij} = c_{ij}(q_{ij}) + \varepsilon_{ij,}$$

where ε_{ij} denotes the respective random deviation caused by measurement error and other other natural and anthropogenic influences.

The CQ function may also be used to estimate the "measured load"

$$L_{ij} = c_{ij}\, q_{ij}$$

by the "estimated load" $L_{ij,e} = c_{ij}(q_{ij})\, q_{ij}$.

or – in case that the current runoff q_{ij} is not available – by the "mean load"

$L_{ij,m} = c_{ij}(q_{0ij})\, q_{0ij}$.

Now the "adjusted load" can be calculated

$$L_{ija} = \frac{L_{ijm}}{L_{ije}} L_{ij} = \frac{c_{ij}}{c_{ij}\left(q_{ij}\right)} L_{ijm},$$

i.e., the adjusted load can be derived from the measured load by multiplying with the correction factor $\frac{L_{ijm}}{L_{ije}}$ or from the mean load by multiplying with the correction factor $\frac{c_{ij}}{c_{ij}\left(q_{ij}\right)}$.

In the following section the concept of adjustment will be investigated in detail, focusing on its interpretation. In Section 3 the effect of lagged runoff is considered, and Section 4 describes the calculation of the power function. The statistical estimation methods are presented in Section 5 and the results are described in Section 6.

The report focuses on the adjustment of riverine loads, but the concept and the statistical methods can also be applied to atmospheric deposition.

2 Interpretation of adjusted inputs

The measured input loads are composed of many different contributions, which can be seen as caused by a set of more or less fixed sources, such as households, industry, farming, area geology etc. Adjustment of inputs according to natural variations has to take into account the different nature of these sources. If, e.g., 80 % of the load is not affected by natural variations since it is caused by point sources, only for the remaining 20 % of the load is an adjustment sensible. For special cases the adjusted input can be calculated as follows:

2.1 Adjustment of loads not affected by climatic variation

Assume that there are only point sources, i.e., assume that inputs are not affected by climatic variation. Then no adjustment is required, i.e., the adjusted load $L_{ij,a}$ should be equal to the actual load, i.e.

$L_{ij,a} = L_{ij.}$

2.2 Adjustment of loads with constant concentration

Assume that there are only diffuse sources, causing constant concentrations not depending on the runoff. Then the load L_{ij} is proportional to the runoff and it can be adjusted by multiplying with q_{0ij} / q_{ij}, i.e.

$L_{ij,a} = L_{ij}\, q_{0ij} / q_{ij}$.

2.3 Adjustment of a mixture of loads from two sources

Assume that 30 %, say, of the load is caused by point sources and 70 % by loads with constant concentration, then the aggregated adjusted load should be the sum of the single adjusted loads, i.e.

$L_{ij,a} = 0.3\, L_{ij} + 0.7\, L_{ij}\, q_{0ij} / q_{ij} = (0.3 + 0.7\, q_{0ij} / q_{ij})\, L_{ij}$

If the percentages are known, no further calculation is needed. Otherwise the question arises how to calculate the percentage of the load due to diffuse sources and that due to point sources.

If the CQ function is known, the percentages can be calculated as follows: Under the conditions prescribed, the CQ function can be written

$$c_{ij}(q) = \frac{\alpha_{ij}}{q} + \beta_{ij}$$

and therefore the LQ function (load runoff function) equals

$$c_{ij}(q)q = f_{ij}(q) = \alpha_{ij} + \beta_{ij} q$$

where α_{ij} denotes the load due to point sources and $\beta_{ij} q$ denotes the load due to diffuse sources. Hence the percentage of load due to diffuse sources can be calculated

$$\frac{\beta_{ij} q}{\alpha_{ij} + \beta_{ij} q}$$

and the percentage of load due to point sources can be calculated

$$\frac{\alpha_{ij}}{\alpha_{ij} + \beta_{ij} q}.$$

According to this approach, the adjusted load can formally be written

$$L_{ij,a} = \frac{\alpha_{ij}}{\alpha_{ij} + \beta_{ij} q_{ij}} L_{ij} + \frac{\beta_{ij} q_{ij}}{\alpha_{ij} + \beta_{ij} q_{ij}} L_{ij} \frac{q_{0j}}{q_{ij}} = \frac{\alpha_{ij} + \beta_{ij} q_{0j}}{\alpha_{ij} + \beta_{ij} q_{ij}} L_{ij} = \frac{L_{ijm}}{L_{ije}} L_{ij}.$$

This formula means that the adjusted load simply can be calculated by multiplying the actual load with the ratio of the estimated load at mean runoff and the estimated load at actual runoff, and in case of 70 % load due to diffuse sources this multiplication factor equals (0.3+ 0.7 q_{0j} / q_{ij}).

Note that this is equivalent with to the general definition of the multiplicatively adjusted load given in Section 1.

2.4 Adjustment of loads due to a non-linear load-runoff function

Assume that the load follows the LQ function

$$L_{ij} = f_{ij}(q_{ij}) = \gamma_{ij} q_{ij}^d$$

Adjustment means that the actual runoff q_{ij} will be replaced by the long-term mean runoff q_{0jj}. Since $\gamma_{ij} = \frac{L_{ij}}{q_{ij}^d}$, this leads to

$$L_{ij,a} = \gamma_{ij} q_{0j}^d = \frac{q_{0j}^d}{q_{ij}^d} L_{ij} = \frac{f(q_{0j}^d)}{f(q_{ij}^d)} L_{ij}.$$

This means again that the adjusted load simply can be calculated by multiplying the actual load with the ratio of the estimated load at mean runoff and the estimated load at actual runoff.

2.5 Adjustment of loads due to a combination of three sources

Assume that 30 %, say, of the load is caused by point sources, 50 % by sources causing constant concentration and 20 % by sources causing a non-linear load-runoff function, then the aggregated adjusted load should be the sum of the single adjusted loads, i.e.

$$L_{ij,a} = 0.3\, L_{ij} + 0.5\, L_{ij}\, q_{0j} / q_{ij} + 0.2\, L_{ij}\, (q_{0ij} / q_{ij})^d$$

If the percentages are known, no further calculation is needed. Otherwise the question arises how to calculate them. As in Section 2.3, if the CQ function is known, the percentages can be derived from the CQ or the LQ function, respectively: Under the conditions prescribed, the LQ function can be written

$$c_{ij}(q)q = \alpha_{ij} + \beta_{ij} q + \gamma_{ij} q^d$$

where α_{ij} denotes the load due to point sources, $\beta_{ij} q$ denotes the load due to diffuse sources (linear), and $\gamma_{ij} q^d$ denotes the load due to the non-linear component. Then the percentages of load due to point sources can be calculated

$$\frac{\alpha_{ij}}{\alpha_{ij} + \beta_{ij} q + \gamma_{ij} q^d},$$

the percentage of load due to diffuse sources can be calculated

$$\frac{\beta_{ij} q}{\alpha_{ij} + \beta_{ij} q + \gamma_{ij} q^d}$$

the percentage of load due to the non-linear component can be calculated

$$\frac{\gamma_{ij} q^d}{\alpha_{ij} + \beta_{ij} q + \gamma_{ij} q^d}$$

According to this approach the adjusted load can formally be written

$$L_{ij,a} = \frac{\alpha_{ij}}{\alpha_{ij} + \beta_{ij} q_{ij} + \gamma_{ij} q_{ij}^d} L_{ij} + \frac{\beta_{ij} q_{ij}}{\alpha_{ij} + \beta_{ij} q_{ij} + \gamma_{ij} q_{ij}^d} L_{ij} \frac{q_{0ij}}{q_{ij}} + \frac{\gamma_{ij} q_{ij}^d}{\alpha_{ij} + \beta_{ij} q_{ij} + \gamma_{ij} q_{ij}^d} L_{ij} \left(\frac{q_{0ij}}{q_{ij}} \right)^d$$

$$= \frac{\alpha_{ij} + \beta_{ij} q_{0ij} + \gamma_{ij} q_{0ij}^d}{\alpha_{ij} + \beta_{ij} q_{ij} + \gamma_{ij} q_{ij}^d} L_{ij} = \frac{L_{ijm}}{L_{ije}} L_{ij}.$$

Again, this is in accordance with the definition of the adjusted load in Section 1. Similar results can be obtained with more than three sources.

3 Effects of lagged runoff

Adjustment of inputs aims at reducing the interannual variability of loads. However, adjustment is performed at the level of single measurements, whereas the interannual variability is based on aggregated values. As long as the CQ function is not dependent on lagged runoff, this should not cause any problem. But what happens if there are lag effects?

In order to obtain a better understanding of possible lag effects, a simulation study applying a very simple simulation model was performed. It is assumed that daily measurements are available, and that the daily runoff is Normally distributed according to the model

$$q_{ij} = 6 + u_i + v_{ij,}$$

where u_i and v_{ij} denote random variables reflecting variations between years and within years. The measured load follows the model

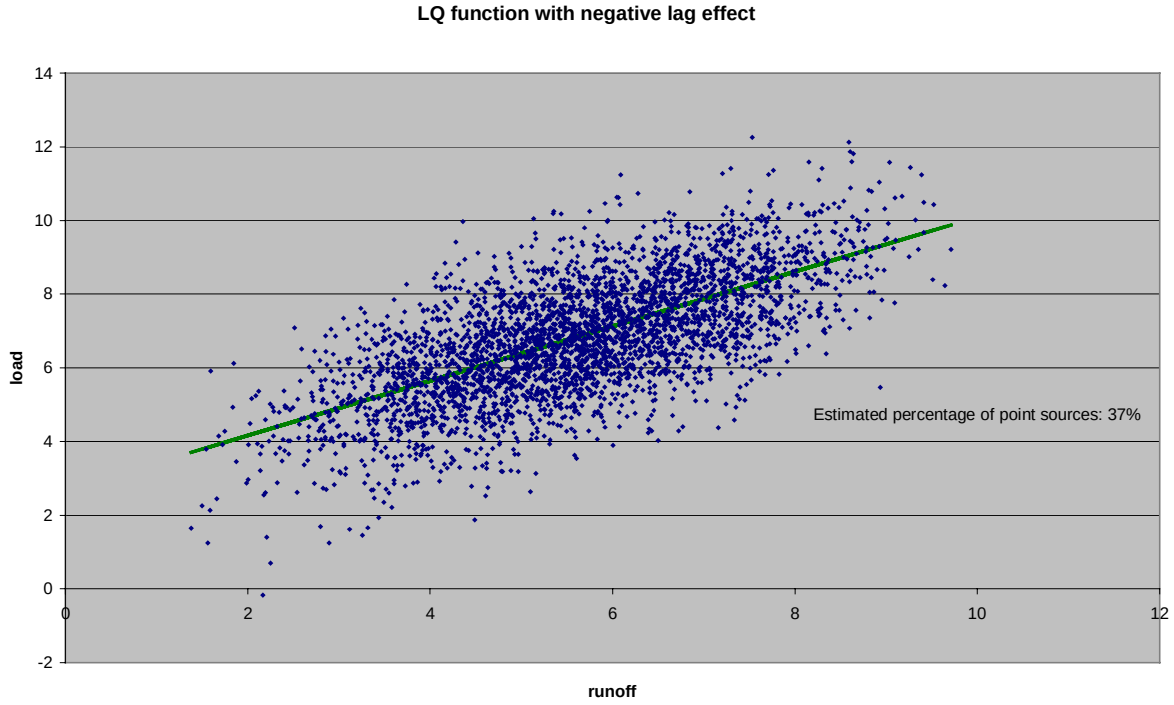
$$L_{ij} = 4 + q_{ij\,-} \; p^* q_{i,j\,-\,1} + w_{ij,}$$

where p denotes the lag effect of the runoff of the day before. All random variables u_i, v_{ij} and w_{ij} are stochastically independent and standard Normally distributed. Note that the mean load is constant over the years.

The simulation study was performed for 10 years and several p's. In each case the model parameters were calculated by using a simple regression approach, and then the adjusted loads were determined. In case that there is no lag effect (p = 0), this approach leads to substantial reduction of the interannual variability.

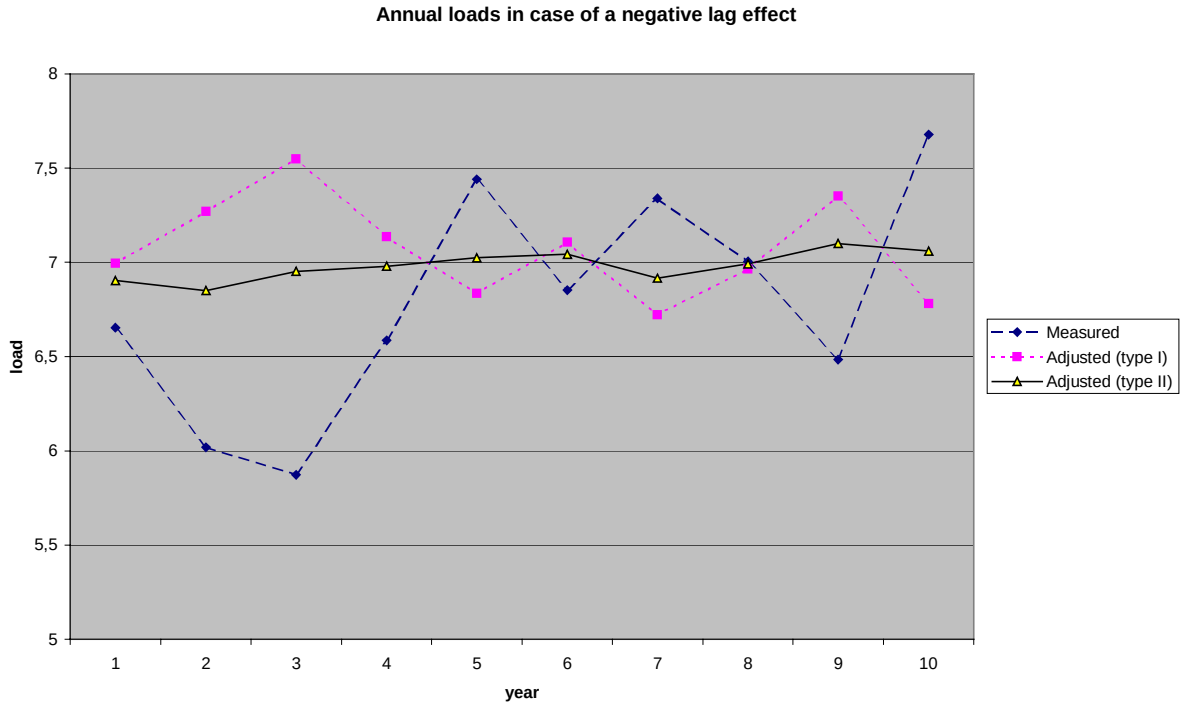
But for very large lag effects this method may fail. Figure A5.1 shows the resulting 3650 pairs of load and runoff obtained in the simulation study for p = –0.5:

Figure A5.1



The straight line represents the linear LQ function obtained from simple linear regression, not taking into account the negative lag effect. The resulting annual adjusted load is presented in Figure A5.2 (type I). It is clearly "overadjusted". This can be explained by the fact that the percentage of diffuse sources is overestimated when the negative lag effect is ignored. A better result can be explained when the lag effect will be taken into account in the regression model. The resulting adjusted load (type II) is almost constant over the years.

Figure A5.2



It can be concluded that lag effects of the runoff may lead to overadjustment (in case of negative lag effects) as well as to underadjustment (in case of positive lag effects), and in both cases the gain of adjustment may be deteriorated substantially. Therefore in case of non-satisfying reduction of the interannual variability it is strongly recommended to check possible lag effects by calculating the correlation function between the residuals c_{ij}–c_{ij} (q_{ij}) and the lagged runoff $q_{i,j-r}$ for r=1,2,3,… If there is a significant correlation, the statistical model should be extended in a suitable manner.

4 The power function

In order to assess the gain of adjustment, simply interpretable characteristics describing the variability of the series of annual adjusted loads are needed. In this section a procedure for calculating the post-hoc power based on the LOESS smoother is described. It is calculated under the assumption that the linear trend is tested on the basis of the 7-year LOESS smoother. It should be noted that this test should be regarded as an approximate test, since it is required that the annual loads have to be stochastically independent and Normally distributed.

The test statistic can be written

$$T = b / (c\, s_{RES})$$

where

b = estimated slope (linear regression)

c = const, and

$$s_{RES} = \sqrt{\frac{1}{n-df} RSS} \quad = \text{Residue standard deviation.}$$

RSS denotes the Sum of Squares of the Residuals based on the LOESS smoother, and df the trace of the smoother matrix; n denotes the number of years and c can be calculated

$$c = \frac{1}{\sum_{i=1}^{n}\left(i - \frac{n+1}{2}\right)^2},$$

It represents the variance of the estimator for the slope under the assumption that the variance of the yearly (adjusted) loads equals 1.

If these loads are stochastically independent and Normally distributed, the test statistic T is non-central t-distributed with ($n - df$) degrees of freedom and non-centrality parameter

$$\delta = \frac{\beta}{\sigma}\sqrt{\sum_{i=1}^{n}\left(i - \frac{n+1}{2}\right)^2}$$

If there is no trend, the non-central parameter vanishes and T is simply t-distributed with ($n - df$) degrees of freedom. Hence for testing the hypotheses

H_0: slope = 0

H_1: slope < 0

the null hypothesis can be rejected at significance level α if $T < -t_{n-df,1-\alpha}$. In case the slope is negative, the power function, i.e. the probability of rejection of the null hypothesis, can be calculated

$$p = F_{T_{n-df,\delta}}\left(-t_{n-df,1-\alpha}\right),$$

where $F_{T_{n-df,\delta}}$ denotes the cumulative distribution function of the non-central t-distribution with non-centrality parameter δ.

For the calculation of the power function, the following assumptions are made: the number of years is $n = 10$ and the significance level is $\alpha = 0.05$. If within this time span the reduction of inputs is 20 %, the non-centrality parameter equals

$$\delta = \frac{-0.02}{s_{Res}/\text{Level}}\sqrt{82.5} = \frac{-0.02}{s_{Res}/\text{Level}}9.083$$

and for a reduction of 50 % the non-centrality parameter equals

$$\delta = \frac{-0.05}{s_{Res}/\text{Level}}\sqrt{82.5} = \frac{-0.05}{s_{Res}/\text{Level}}9.083,$$

where *Level* denotes the mean level of the series of annual loads.

5 Estimation methods

According to the concept described in Section 1, a statistical model for load and concentration is needed in order to estimate concentration and load as functions of runoff. Figure A5.3 contains the load-runoff diagram for nitrate measured at Lobith/Rhine biweekly 1955–1995. Apparently there is an approximate linear relation between load and runoff.

Figure A5.3

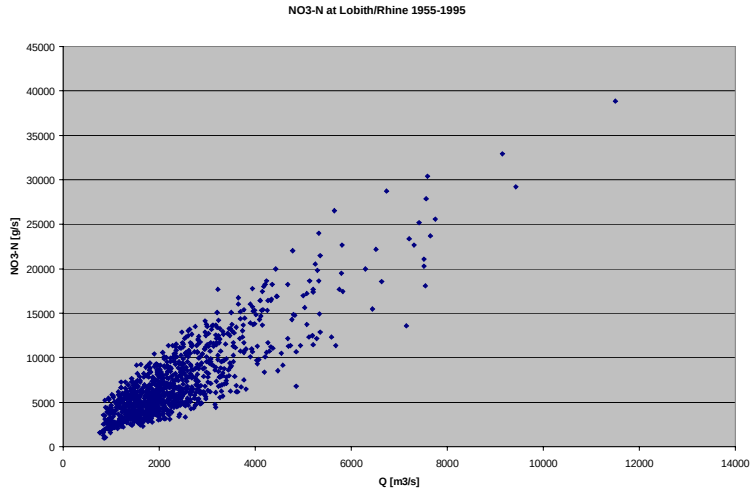
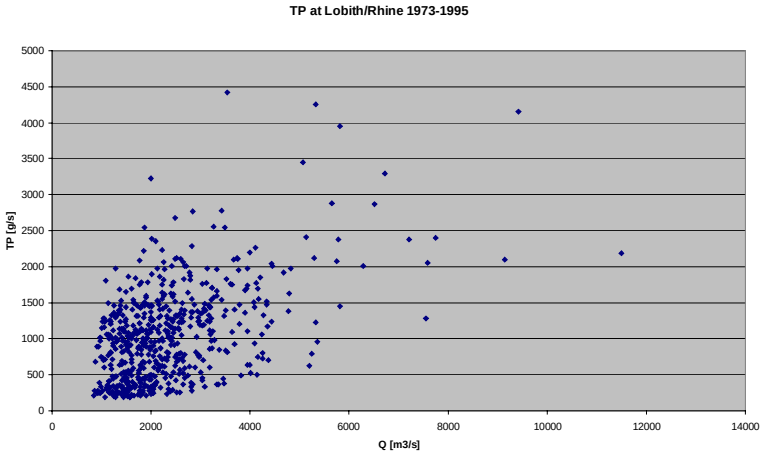


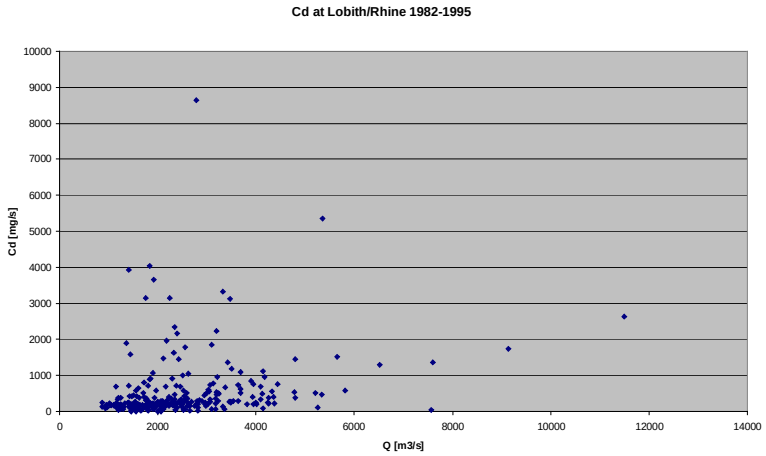
Figure A5.4 contains the results for Total P. Again the relation between runoff and load could be modelled linearly, but there is more variability and heteroscedasticity.

Figure A5.4



For other rivers and other nutrient parameters similar diagrams can be obtained. For heavy metals the relation between runoff and load is frequently less clear. Figure A5.5 shows the load-runoff diagram for cadmium.

Figure A5.5.



A linear LQ function is equivalent to the assumption that the CQ function may be modelled as a linear function of the reciprocal runoff. This is the basic assumption of the methods N, H, L1, L2, L3, and L4 which are described in the following sub-sections. A non-linear LQ function is used by the methods S1 and S2 described in Sections 5.7 and 5.8.

5.1 Method N: Smoothing with non-parametric constraints

Modifying an approach considered in the OSPAR HARP Guidelines, the following dynamic model is considered:

$$L_{ij} = \alpha_{ij} + \beta_{ij} q_{ij} + \varepsilon_{ij}$$

with L_{ij} = load in year i and season j

q_{ij} = runoff in year i and season j

$i = 1,...,T$

$j = 1,...,M$.

According to this model the multiplicatively adjusted load can be calculated

$$L_{ij,a} = \frac{\alpha_{ij} + \beta_{ij} q_{0j}}{\alpha_{ij} + \beta_{ij} q_{ij}} L_{ij} = \frac{\frac{\alpha_{ij}}{\beta_{ij}} + q_{0j}}{\frac{\alpha_{ij}}{\beta_{ij}} + q_{ij}} L_{ij},$$

i.e., the adjustment step is only dependent on the ratio α_{ij} / β_{ij}.

The model parameters α_{ij} and β_{ij} are estimated by minimising an expression of the form:

$$S(\alpha,\beta)=\sum_{i,j}(L_{ij}-\alpha_{ij}-\beta_{ij}q_{ij})^2+\lambda_1\sum_{i,j}(\alpha_{ij}-\frac{\alpha_{i+1,j}+\alpha_{i-1,j}}{2})^2+\lambda_2\sum_{i,j}(\alpha_{ij}-\frac{\alpha_{i,j+1}+\alpha_{i,j-1}}{2})^2,$$
$$+\lambda_1\overline{q^2}\sum_{i,j}(\beta_{ij}-\frac{\beta_{i+1,j}+\beta_{i-1,j}}{2})^2+\lambda_2\overline{q^2}\sum_{i,j}(\beta_{ij}-\frac{\beta_{i,j+1}+\beta_{i,j-1}}{2})^2$$

with $\overline{q^2}=\frac{1}{N}\sum_{ij}q_{ij}^2$ and penalty factors λ_1, λ_2 to define a desired compromise between overfitting and specification errors. Suitable levels of the penalty factors can be established by undertaking a cross-validation study of relationships between L_{ij} and q_{ij}. In the examples presented here further restrictions are posed: the generalised degrees of freedom of the model are constant and $\lambda_1 = \lambda_2$.

5.2 Method H: Approach of Hebbel

According to a general approach of Hebbel (1992) the following decomposition of the concentration series into trend, season and exogenous effects may be considered:

$$c(t) = u(t) + s(t) + \beta / q(t) + \varepsilon(t)$$

or equivalently

$$L(t) = q(t) \times (u(t) + s(t) + \beta / q(t) + \varepsilon(t))$$

where

$c(t)$ = concentration at time t,

$L(t)$ = load at time t,

$q(t)$ = runoff or precipitation at time t,

$u(t)$ = trend,

$s(t)$ = season,

β = effect of runoff/precipitation on the concentration.

According to this model the multiplicatively adjusted load can be written

$$L_a(t)=\frac{\beta+\big(u(t)+s(t)\big)q_0(t)}{\beta+\big(u(t)+s(t)\big)q(t)}L(t)=\frac{\frac{\beta}{u(t)+s(t)}+q_0(t)}{\frac{\beta}{u(t)+s(t)}+q(t)}L(t),$$

where $q_0(t)$ denotes the long-term mean in the month corresponding to t.

The sampling times $t_1,...,t_n$ are not necessarily equidistant. In order to describe the estimation procedure, matrix notation will be used. Let $c = (c(t_1),...,c(t_n))'$ denote the concentration vector, $u = (u(t_1),...,u(t_n))'$ and $s = (s(t_1),...,s(t_n))'$ the vectors of trend and season, and $A = (1/q(t_1),...,1/q(t_n))'$ the design matrix belonging to β. Under the assumption that β is known, trend u and season s can be estimated with an appropriate linear smoother according to the equation

$$\hat{u} + \hat{s} = S(c - A\beta), \qquad [1]$$

where *S* denotes the smoother matrix. Under the assumption that *u* + *s* is known, β can be estimated by linear regression according to the equation

$$\hat{\beta} = (A'A)^{-1} A'(c - (u + s)). \qquad [2]$$

Letting the estimates equal to the estimated parameters, the equations [1, 2] are equivalent to

$$\hat{u} + \hat{s} = S(c - A\hat{\beta})$$

with $\hat{\beta} = (A'(I - W)A))^{-1} A'(I - W)c)$.

It should be noted that in the model applied, β is a constant. However, due to reduction of inputs β may change over time, and therefore it is recommended to apply the procedure locally for a time window of a given length. In the calculations presented a seven-year window was applied.

The estimation method works with every linear smoother. In the calculations presented the method described by Hebbel (1997) is applied. According to this method, trend and seasonality are estimated by minimizing a roughness functional. This functional depends on parameters *p* for the trend, *q* for the order of seasonality, and σ^2 for the smoothness. In the calculations presented the parameters are *p* = *q*= 2 (i.e., linear trend and second order seasonality). The smoothness parameter is determined so that the generalized degrees of freedom are about 12 per seven years.

5.3 Method L1: Local regression with seasonality

If in method H the smoothness parameter σ^2 is tending to infinity, the estimation method becomes equivalent to a linear regression analysis based on the following model:

$$c(t) = \frac{\alpha}{q(t)} + \beta + \delta t + \gamma_1 \sin\frac{2\pi t}{m} + \gamma_2 \cos\frac{2\pi t}{m} + \gamma_3 \sin\frac{2\pi t}{2m} + \gamma_4 \cos\frac{2\pi t}{2m} + \varepsilon(t),$$

Here denotes *m* the length of the year. According to the notation of method H the trend is described by

trend = β + δ*t*

and the season is described by

$$\text{season} = \gamma_1 \sin\frac{2\pi t}{m} + \gamma_2 \cos\frac{2\pi t}{m} + \gamma_3 \sin\frac{2\pi t}{2m} + \gamma_4 \cos\frac{2\pi t}{2m}.$$

The estimation procedure is applied locally for a time window of seven years, i.e., for every sampling time *t* the parameters α, β, δ and γ_1, γ_2, γ_3, γ_4 are locally estimated with all data of the corresponding time window.

With this model the multiplicatively adjusted load can be written

$$L_a(t) = \frac{\alpha + (\beta + \delta t + season) q_0(t)}{\alpha + (\beta + \delta t + season) q(t)} L(t).$$

5.4 Method L2: Local regression with seasonality and lagged runoff effect

If the load is not only depending on the current runoff but also on lagged runoffs, the model used for method L1 can be extended as follows:

$$c(t) = \frac{\alpha}{q(t)} + \frac{\gamma}{q(t-1)} + \beta + \delta t + \gamma_1 \sin\frac{2\pi t}{m} + \gamma_2 \cos\frac{2\pi t}{m} + \gamma_3 \sin\frac{2\pi t}{2m} + \gamma_4 \cos\frac{2\pi t}{2m} + \varepsilon(t)$$

$q(t-1)$ denotes the runoff of the day before (which is frequently available). With this model the multiplicatively adjusted load can be written

$$L_a(t) = \frac{\alpha + \gamma + (\beta + \delta t + season) q_0(t)}{\alpha + \gamma \frac{q(t)}{q(t-1)} + (\beta + \delta t + season) q(t)} L(t)$$

where $q_0(t)$ denotes the long-term mean in the month corresponding to t. The parameters are estimated using local regression. In order to improve the stability of the estimated parameters, a time window of eight instead of seven years is used.

5.5 Method L3: Local regression with lagged runoff effect and water temperature

In order to simplify the CQ function, one idea is to replace the second-order seasonal component (four unknown parameters) by a temperature effect which requires only one unknown parameter:

$$c(t) = \frac{\alpha}{q(t)} + \frac{\gamma}{q(t-1)} + \beta + \delta t + \eta w(t) + \varepsilon(t)$$

where $w(t)$ denotes the water temperature on day t. Estimation of the unknown parameters is performed with local linear regression with an eight-year time window.

5.6 Method L4: Local regression with seasonality, lagged runoff effect and water temperature

Combining the models applied for the methods L2 and L3 results in the following CQ model:

$$c(t) = \frac{\alpha}{q(t)} + \frac{\gamma}{q(t-1)} + \beta + \delta t + \eta w(t)$$

$$+ \gamma_1 \sin\frac{2\pi t}{m} + \gamma_2 \cos\frac{2\pi t}{m} + \gamma_3 \sin\frac{2\pi t}{2m} + \gamma_4 \cos\frac{2\pi t}{2m} + \varepsilon(t)$$

Estimation of the unknown parameters is performed with local linear regression with an eight-year time window.

5.7 Method S1: CQ Spline

The methods described above assume that the LQ function is approximately linear. If this is not adequate, the CQ function may be modelled with a cubic spline. If s(.) denotes the spline function, the CQ function can be written

$$c_t = s(q_t) + \varepsilon_t$$

and the accordingly adjusted load equals

$$L_a(t) = \frac{s\big(q_0(t)\big)q_0(t)}{s\big(q(t)\big)q(t)} L(t),$$

i.e., the adjustment is determined by (1) the ratio of current runoff and long-term mean runoff, and (2) the ratio of the respective spline values.

The smoothing parameter of the spline function is determined so that the generalized degrees of freedom equals four.

According to method S1 the spline function is calculated on the basis of all available data, i.e., the adjustment step is not accommodated for temporal changes.

5.8 Method S2: CQ Spline locally calculated

In order to accommodate the CQ function for temporal changes, the spline may be calculated locally with a 3-year time window. This means that the spline s(.) is recalculated for every sampling time *t*.

6 Results and Discussion

The methods described in the preceding section were applied for seven parameters (NO_3-N, NH_4-N, total P, ortho-P, Cd, Pb and undissolved particles), measured biweekly in the river Rhine (Lobith) and monthly in the river Ems (Herbrum).

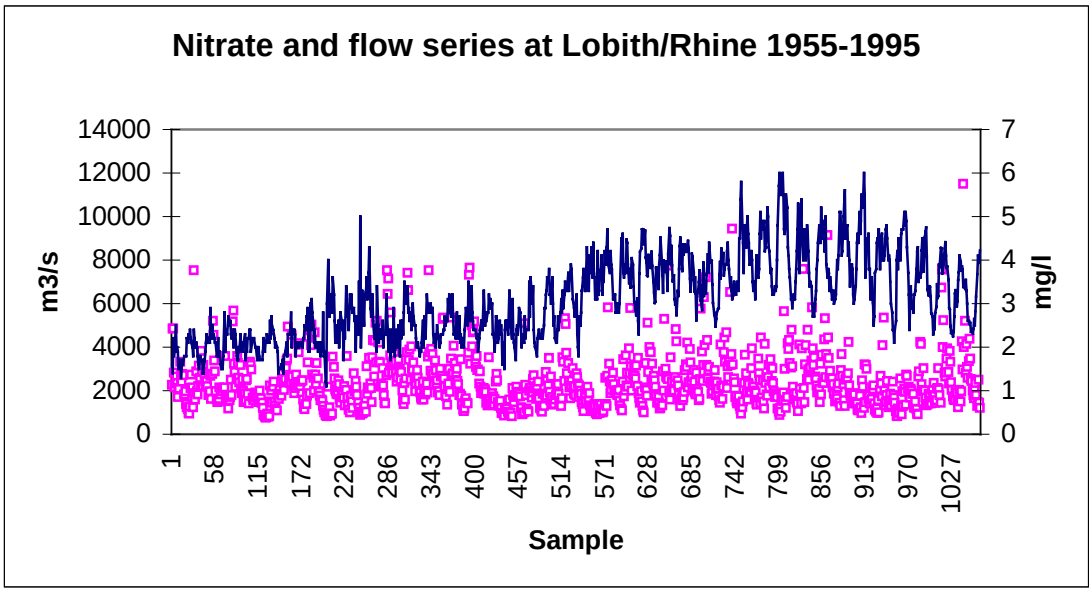
The aim is to identify a method which allows efficient adjustment, i.e., an adjustment which results in annual indices not influenced by climatic variation. Therefore the interannual variation of these indices should be smaller than that for the unadjusted load. It is quantified by the power for detecting a linear trend according to the method described in Section 4. It should be noted that this quantification is based on the assumption that there is no autocorrelation in the series of adjusted loads. Therefore it has to be guaranteed that autocorrelation induced by the estimation step can be neglected. Since the methods considered here are based on sparse statistical models, this assumption holds.

The following section contains the results of the analyses for nitrate at Lobith/Rhine. An overview of the estimated power of the methods for the different parameters is given in Section 6.2.

6.1 Nitrate at Lobith/Rhine

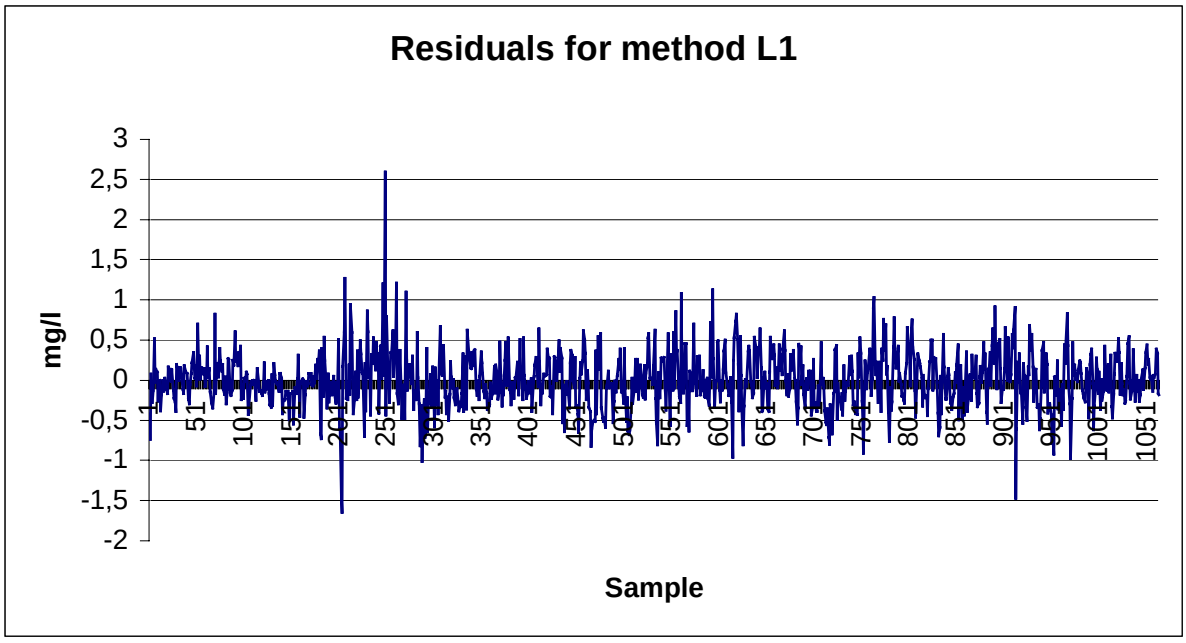
Figure A5.6 contains NO_3-N concentrations (solid line) 1955–1995 at Lobith/Rhine and the corresponding runoff series (circles).

Figure A5.6



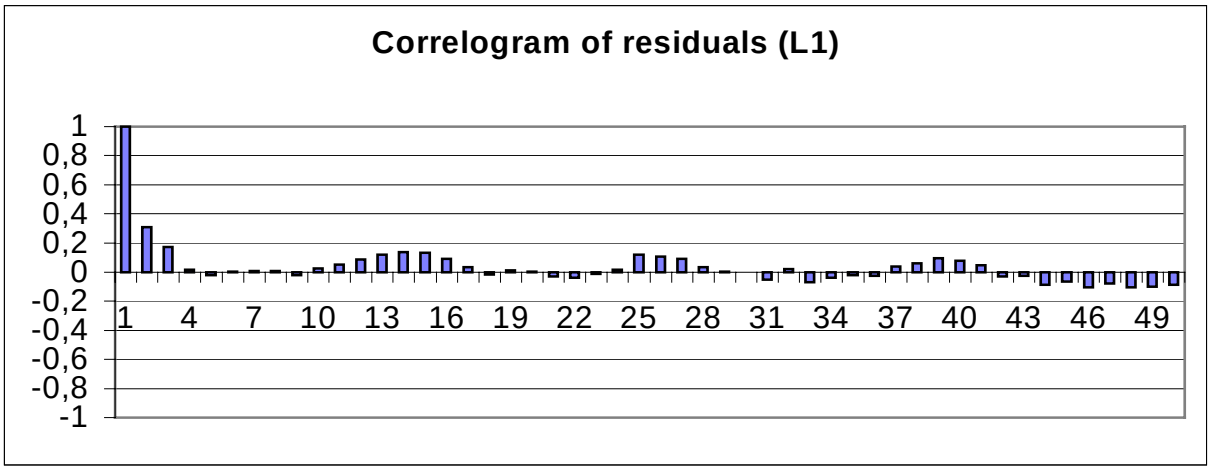
A model fit for the concentration series based on method L1 explains 86.6 % of the variance, and the residue standard deviation is 0.35 mg l^{-1}. Figure A5.7 contains a plot of the residuals.

Figure A5.7



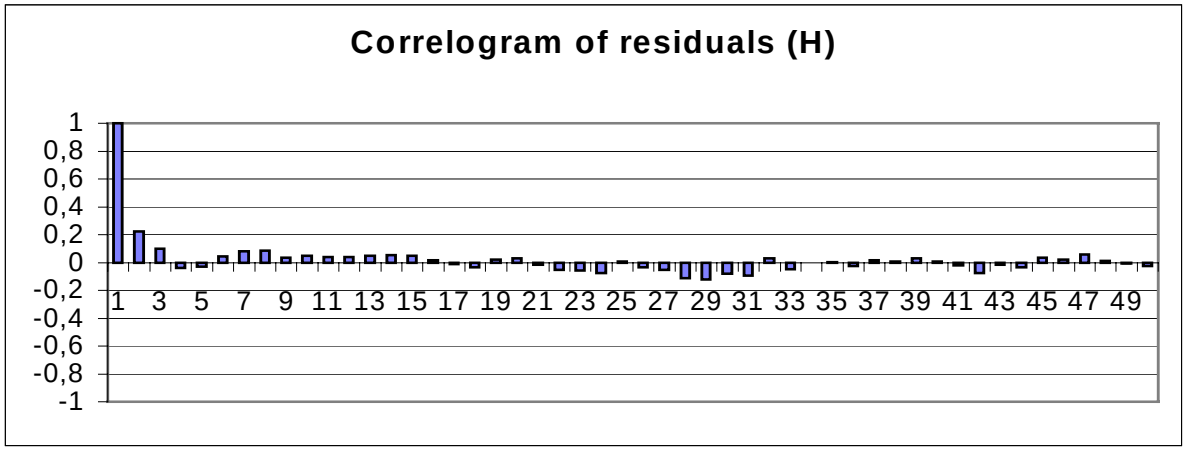
In order to assess the influence of lagged runoff, the correlation between the series of residuals and lagged runoff is calculated, but no clear dependency could be detected. There is some autocorrelation in the series of the residuals, and the Durbin-Watson statistics is 1.37. The correlogram is contained in Figure A5.8.

Figure A5.8



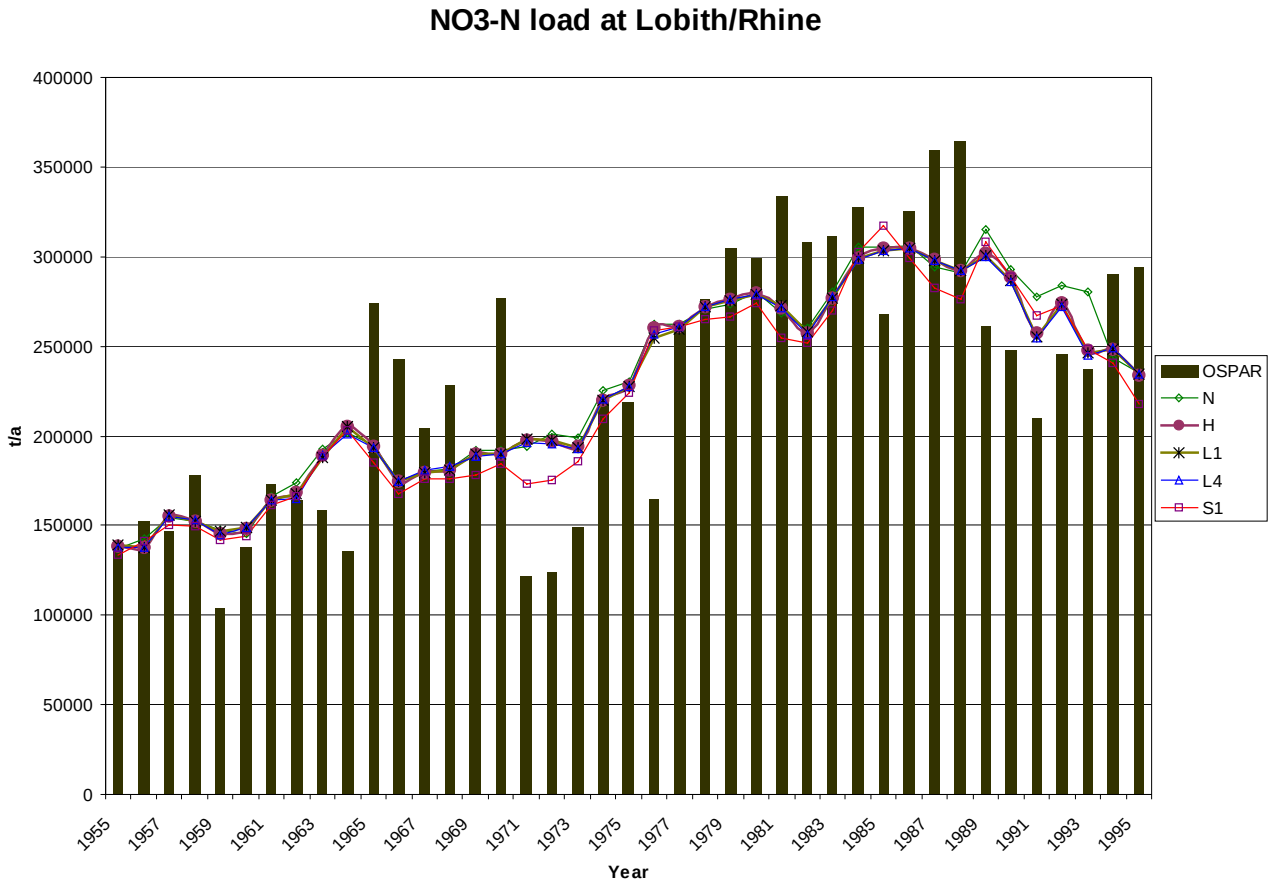
Apparently this is not a white-noise process and therefore it can be concluded that method L1 does not allow an optimal fit of the data. The fit could be improved by a time window of 5 instead of 7 years, or by use of method H. With method H 88.6 % of the variance can be explained, and the corresponding residue standard deviation is 0.32 mg l^{-1}. The correlogram is given in Figure A5.9 and the Durbin-Watson statistic is 1.55.

Figure A5.9



One can conclude that regarding the model fit, method H is clearly better than method L1. However, regarding the annual adjusted loads the differences between these methods can be neglected. The averaged relative deviation between these loads is 0.3 %, whereas the averaged deviation to the load calculated with the OSPAR formula is about 15 %. Figure A5.10 contains the annual adjusted loads according to methods N, H, L1, L4, S1 and the load according to the "OSPAR formula". Apparently the differences between the adjusted loads are small compared to the differences to the OSPAR load. Similar results were obtained for the other adjustment methods.

Figure A5.10



6.2 Results and discussion

Table A5.1 and Table A5.2 contain the estimated power values for a 20 % reduction and a 50 % reduction, respectively, of inputs for seven parameters measured in the rivers Rhine and Ems.

The power was calculated

- for all adjustment methods described before,
- for the annual mean values of the measured concentrations, i.e.,

$$\frac{1}{M}\sum_{j=1}^{M} c_{ij} \,,$$

- and for the long-term runoff-corrected load

$$\frac{q_0}{\sum_{j=1}^{M} q_{ij}} \sum_{j=1}^{M} c_{ij} q_{ij} \qquad \text{(method A0).}$$

Here, M denotes the number of samples in year i, and q_0 denotes the long-term mean runoff.

It should be noted that the length of time series is between 5 years and 41 years, and for short series with less than 12–15 years, the results presented are very crude estimates of the power.

For the river Rhine the use of adjusted loads instead of the OSPAR load increases trend detectability for nitrate, total P and undissolved particles considerably, whereas for the other substances only small differences can be observed. Using concentration mean values reduces the power substantially. Only for nitrate and undissolved particles do concentration mean values behave better than the OSPAR load (although worse than adjusted loads). For the river Ems the use of adjusted loads instead of the OSPAR load increases trend detectability for all nutrients. The same holds for the annual mean concentration values. The power for the concentration mean values is high since the runoff of the river Ems underlies large seasonal effects with a maximum in winter. Hence the variability of loads (OSPAR load and adjusted loads) is dominated by the variability of the measured concentrations in winter, whereas for the concentration mean these fluctuations are smoothed over the year. However, one should note that in this case of strong seasonality the annual concentration mean may not reflect properly temporal changes of inputs.

It turns out that there is no method which is optimum for every river and every substance. However, method L1 performs reasonably well in terms of smoothness for nutrients and partly also for heavy metals. Method L2 has slight advantages in cases where there is a significant effect of the lagged runoff, and methods S1 and S2 have advantages in cases where the LQ function is not linear. The use of the annual concentration mean cannot be recommended since there are cases where this method is very poor and where adjustment leads to much better results (e.g., for nutrients in the river Rhine).

Table A5.1. Power of the models tested at 20 % reduction within 10 years.

Station	Parameter	Time series	Concentration	OSPAR	A0	N	H12	L1	L2	L3	L4	S1	S2
Lobith	NO3-N	1955–1995	85%	25%	85%	97%	97%	98%	98%	98%	98%	86%	78%
Lobith	NH4-N	1955–1995	13%	28%	13%	25%	26%	25%	29%	29%	30%	27%	24%
Lobith	TP	1973–1995	30%	44%	29%	64%	59%	61%	61%	62%	62%	59%	59%
Lobith	OP	1973–1993	25%	38%	22%	41%	37%	39%	38%	40%	39%	36%	33%
Lobith	CD	1982–1995	10%	10%	10%	10%	12%	12%	12%	13%	12%	15%	14%
Lobith	PB	1982–1995	29%	34%	29%	30%	29%	31%	31%	31%	31%	27%	33%
Lobith	Undissolved	1991–1995	25%	13%	15%	54%	21%	24%	22%	15%	11%	21%	70%
Herbrum	NO3-N	1982–1997	66%	17%	35%	34%	58%	60%	59%	57%	57%	39%	31%
Herbrum	NH4-N	1982–1997	16%	8%	11%	18%	18%	14%	14%	14%	14%	17%	15%
Herbrum	TP	1982–1997	27%	12%	23%	25%	27%	27%	27%	27%	27%	29%	28%
Herbrum	OP	1984–1997	13%	11%	12%	12%	13%	13%	13%	14%	13%	12%	15%
Herbrum	CD	1993–1997	8%	12%	13%	9%	15%	11%	10%	10%	8%	9%	9%
Herbrum	PB	1993–1997	66%	16%	15%	7%	30%	48%	40%	66%	25%	47%	52%
Herbrum	Undissolved	1982–1997	22%	13%	14%	22%	22%	21%	22%	21%	22%	23%	19%

Table A5.2. Power of the models tested at 50 % reduction within 10 years.

Station	Parameter	Time series	Concentration	OSPAR	A0	N	H12	L1	L2	L3	L4	S1	S2
Lobith	NO3-N	1955–1995	100%	77%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Lobith	NH4-N	1955–1995	37%	84%	38%	78%	81%	77%	86%	86%	87%	82%	75%
Lobith	TP	1973–1995	87%	98%	86%	100%	100%	100%	100%	100%	100%	100%	100%
Lobith	OP	1973–1993	78%	95%	71%	97%	95%	96%	96%	97%	96%	94%	91%
Lobith	CD	1982–1995	23%	22%	24%	23%	33%	31%	32%	34%	32%	47%	39%
Lobith	PB	1982–1995	85%	92%	86%	88%	86%	89%	88%	89%	89%	83%	91%
Lobith	undissolved	1991–1995	77%	35%	46%	100%	66%	75%	69%	43%	27%	68%	100%
Herbrum	NO3-N	1982–1997	100%	54%	93%	92%	97%	100%	100%	100%	100%	96%	89%
Herbrum	NH4-N	1982–1997	48%	17%	28%	57%	57%	42%	42%	39%	40%	53%	44%
Herbrum	TP	1982–1997	82%	32%	73%	79%	82%	81%	81%	82%	82%	86%	84%
Herbrum	OP	1984–1997	36%	29%	31%	33%	35%	34%	37%	40%	38%	31%	45%
Herbrum	CD	1993–1997	17%	32%	38%	18%	45%	29%	23%	23%	17%	20%	19%
Herbrum	PB	1993–1997	100%	50%	45%	11%	87%	99%	97%	100%	77%	99%	99%
Herbrum	undissolved	1982–1997	69%	35%	42%	69%	69%	68%	70%	68%	69%	72%	60%

References

Hebbel, H. 1992. Trend und Saison im linearen Modell. *In* Data Analysis and Statistical Inference. Eul., 513-520. Ed. S. Schach and G. Trenkler.

Hebbel, H. 1997. Verallgemeinertes Berliner Verfahren VBV. *In* Analyse saisonaler Zeitreihen. Physica, 83-93. Eds. K. Edel, K.-A. Schäffer and W. Stier.

Stalnacke, P. 1996. Nutrient Loads to the Baltic Sea. Linköping University, Linköping, Sweden.

ANNEX 6: MODEL OF THE RELATIONSHIP BETWEEN CONCENTRATION, UPTAKE AND EXCRETION RATE OF AN ORGANISM AND AMBIENT CONCENTRATION

Superbeast II – in search of a clean pig

Mike Nicholson and Rob Fryer

Summary

A simple uptake and excretion model is used to generate concentration-size relationships in marine organisms, and to show how these might vary due to changes in ambient concentrations. We find some tentative support for this model from concentration-size relationships reported in the literature.

1 Introduction

The concentrations in marine organisms of unregulated heavy metals such as mercury often vary with some size covariable such as length or weight. Concentrations may also vary with age, which may simply be a surrogate for size. Alternatively, age may be a measure of exposure, an indicator of different levels of exposure encountered at different life stages, or age-related changes in biological processes. Resolving this murky mess from data collected in monitoring programmes is difficult. The associations between length, weight, and age, and the uncertain life histories of many marine organisms make interpretation difficult and ambiguous. Despite these problems, any clarification would help in the selection of appropriate monitoring organisms, and in the interpretation of monitoring data.

Given such a potentially complex system, it is perhaps not surprising that the relationships between concentration and size often vary over time. For example, Nicholson and Wilson (1987) found an overall significant positive increase of log mercury concentrations in muscle with fish length for several species, but both significantly higher and lower (including negative) regression coefficients in particular years. Warren (1993) suggested that varying regression coefficients may reflect changing availability of the contaminant in the environment and, using a state-space model, found evidence that these regression coefficients tended to evolve smoothly from year to year. However, there are few data that could allow us to piece together all the bits and pieces of this puzzle in practice. Parallel time series of concentrations in the environment and in biota are as rare as a clean pig.

Here we suggest a simple mathematical model that provides insight into how processes within an organism and temporal trends in ambient concentration might interact to affect the

- temporal trends induced in the organism,
- perceived concentration-size relationships in the organism.

The model is derived in Section 2, and extends a simple first-order uptake and excretion model described by Nicholson and Fryer (1993). Section 3 explores the theoretical implications of the model, and Section 4 illustrates the theory graphically. Section 5 then relates these results to observed patterns reported in the literature.

2 A simple first-order model of contaminant uptake and excretion

Let

- $c(a,t)$ be the contaminant concentration in an organism aged a at time t,
- $A(t)$ be the ambient concentration at time t,
- λ_u be the rate of uptake, assumed constant over time,
- λ_e be the rate of excretion, assumed constant over time.

Then a simple model for the rate of change in concentration in the organism is

$$\frac{d}{dt}c(a,t) = \lambda_u A(t) - \lambda_e c(a,t),$$

which we can rewrite as

$$\frac{d}{dt}c(a,t)e^{\lambda_e t} = \lambda_u A(t)e^{\lambda_e t}.$$

Assuming the organism has zero concentration when it first emerges into the big wide world (at time $t–a$), we have

$$c(a,t)e^{\lambda_e t} = \int_{t-a}^{t} \lambda_u A(s)e^{\lambda_e s} ds$$

and hence

$$c(a,t) = \lambda_u e^{-\lambda_e t} \int_{t-a}^{t} A(s)e^{\lambda_e s} ds .$$

3 The effect of different uptake and excretion rates and trends in ambient concentration

We model trends in ambient concentration as

$$A(t) = e^{\mu(t)}$$

so that log concentration varies according to some simple trend $\mu(t)$. We then obtain

$$c(a,t) = \lambda_u e^{-\lambda_e t} \int_{t-a}^{t} e^{\mu(s)+\lambda_e s} ds .$$

We can now see how concentrations in the organism will vary with age and time for specific trends in ambient concentration.

Pattern 1 $\mu(t) = \mu = constant.$

This gives

$$c(a,t) = \frac{\lambda_u}{\lambda_e} e^{\mu}(1 - e^{-\lambda_e a})$$

and hence

$$\log c(a,t) = \log \frac{\lambda_u}{\lambda_e} + \mu + \log(1 - e^{-\lambda_e a})$$

As would be expected, for any given age, the concentration in the organism is constant over time. Further, the relationship between log concentration and age is

- constant over time,
- non-linear,
- flatter when the excretion rate λ_e is large.

Pattern 2 $\mu(t) = \alpha + \beta t$.

This gives

$$c(a,t) = \frac{\lambda_u}{\beta + \lambda_e} e^{\alpha+\beta t}(1 - e^{-(\beta+\lambda_e)a})$$

and hence

$$\log c(a,t) = \log \frac{\lambda_u}{\beta + \lambda_e} + \alpha + \beta t + \log(1 - e^{-(\beta+\lambda_e)a}).$$

Thus, for any given age, the temporal trend in the organism is the same as the temporal trend in ambient concentration. Further, the relationship between log concentration and age is

- constant over time,
- non-linear,
- flatter when the excretion rate λ_e is large,
- flatter when ambient concentration is increasing (since all ages are dominated by the current ambient concentration),
- steeper when ambient concentration is decreasing (since older ages have a high residual concentration from earlier times, whereas younger ages are relatively uncontaminated).

Pattern 3 $\mu(t)$ = any pattern of change.

To make progress here, we assume we can linearise $\mu(t)$ in the region $t-a \le s \le t$, by writing $\mu(s) = \mu(t) + (s-t)\mu'(t) \quad t - a \le s \le t$. This gives

$$c(a,t) = \frac{\lambda_u}{\mu'(t) + \lambda_e} e^{\mu(t)}(1 - e^{-(\mu'(t)+\lambda_e)a})$$

and hence

$$\log c(a,t) = \log \lambda_u + \mu(t) - \log(\mu'(t) + \lambda_e) + \log(1 - e^{-(\mu'(t)+\lambda_e)a}).$$

For any given age, the temporal trend in the organism is related to, but not the same as, the temporal trend in ambient concentration. Further, the relationship between log concentration and age is no longer stable over time.

4 Examples

We now illustrate the theory by using specific values of the uptake and excretion rates, and trends in ambient concentration. Assume that we monitor an organism that lives to age 5, and that we monitor for a twenty-year period, in years 1…20 say. Figures A6.1, A6.2, A6.3, and A6.4 correspond to the following combinations of λ_e and $\mu(t)$, with λ_u equal to 1 throughout.

1. $\lambda_e = 10$ $\mu(t) = 0.1t$ high excretion rate, linear trend (on log scale)
2. $\lambda_e = 0.1$ $\mu(t) = 0.1t$ low excretion rate, linear trend
3. $\lambda_e = 0.1$ $\mu(t) = 1 - e^{-0.2t}$ low excretion rate, non-linear monotonic trend
4. $\lambda_e = 0.1$ $\mu(t) = \sin(t/2)$ low excretion rate, oscillating trend.

In each figure,

- the top picture shows the ambient concentration over the 20-year period (and in the five years before that, to show the exposure history of all organisms sampled in the monitoring period),
- the middle picture shows log concentration plotted against log age (a sort of surrogate for length) each year,
- the bottom picture shows the induced temporal trend in the organism for ages 1…5.

Figure A6.1 (high excretion rate, linear trend) shows

- no evidence of an age effect; this is because the excretion rate is large, so all ages reflect the current ambient concentration,
- an identical trend in the organism and in the environment.

Figure A6.2 (low excretion rate, linear trend) shows

- an age effect with the same relationship between log concentration and log age in all years,
- identical trends in the organism and the environment, although the level increases with age.

Figure A6.3 (low excretion rate, non-linear monotonic trend) shows

- an age effect that evolves over time,
- similar trends in the organism and the environment, but with a small suggestion of age-dependent trends.

Figure A6.4 (low excretion rate, sinusoidal trend) shows

- an age effect that evolves quite markedly over time: the age effect is strongest when ambient concentrations are falling, since the older organisms are still affected by the relatively high concentrations prevalent in their youth,
- similar trends in the organism and the environment, but with strong evidence of age-dependent trends,
- peaks and troughs in the concentration in the organism lag behind peaks and troughs in ambient concentration; the time lag increases with age.

5 Observed concentration-size relationships

To seek support for the uptake and excretion model, we looked at the concentration-length relationships from the four time series of mercury concentrations measured in fish muscle described in Fryer and Nicholson (2000). For each time series, we constructed annual indices of log-concentration in "small" and "large" fish, and summarised

- the average *level* of mercury each year by the mean of the two indices,
- the *slope* of the concentration-length relationship each year by the difference between the large and small indices.

Figure A6.5 shows the values of *level* and *slope*, with a smoother fitted through them, for all four time series.

Average levels have, if anything, been declining over time in all four time series. For North Sea cod and Irish Sea whiting, there is no evidence of any systematic changes in slope. However, for North Sea plaice and Irish Sea plaice, the slope of the concentration-length relationship appears to be increasing over time. (These graphical impressions are supported by more formal analyses of variance which, in particular, show a time-length interaction for the plaice time series, but not for the cod or whiting time series (Fryer and Nicholson, 2000).) These results are consistent with the uptake and excretion model, which suggests that declining levels should be accompanied by constant or increasing slopes (as in Figure A6.4).

A potentially more incisive display is a plot of *slope* against *level* (van der Meer, 1994). The top panel of Figure A6.6 shows such a display for the stylised scenario of Figure A6.4. We have taken the average *level* to be the log concentration of fish of age 3, and the *slope* to be the difference in log concentration between fish of age 5 and of age 1. The plotting symbols denote the year. As the ambient concentration follows its sinusoidal trend, the points in the *slope-level* plot move in an anti-clockwise direction. The other two panels in Figure A6.6 correspond to Irish Sea plaice and North Sea plaice. To emphasise any pattern in the data, we have also plotted the smoothed values of *slope* and *level* shown in Figure A6.5, and joined these up as they move forward through time. The plot for Irish Sea plaice is consistent with the top right hand corner of the stylised plot, corresponding to declining levels and increasing slopes. The plot for North Sea plaice is less convincing, but a couple of beers help the suggestive process. Of course, any simple pattern would be consistent with some part of the stylised plot, and longer time series with more contrast would be necessary to demonstrate the utility of these displays.

6 Conclusions

We have suggested a simple model that can induce variation in concentration-size relationships, and that allows us to predict how these relationships will evolve in response to changes in ambient concentrations.

It has been difficult to collate sufficient data to test this model. The OSPAR monitoring programme does not currently generate parallel time series of concentrations in organisms, sediments and sea water. Although these are now collected within the UK national programme, the time series are currently very short. Also, there is little information about the uptake and excretion rates of contaminants in the field. Nicholson and Fryer (1993) exploited estimates of uptake and excretion rates for several metal isotopes in fish and shellfish tissue, but the behaviour of non-radioactive contaminants may be very different. We have therefore taken the simple approach and looked for indirect support by observing trends in concentration-size relationships in time series that we could easily lay our hands on. We see smooth patterns reminiscent of those suggested by our model, which, at this stage, is as good as it is likely to get. For more progress towards selection of appropriate monitoring organisms, more information is required from the laboratory about uptake and excretion rates, and from the field about exchanges between media and fish behaviour.

References

Fryer, R.J., and Nicholson, M.D. 2000. Using smoothers for assessments of covariate-dependent contaminant times series in the marine environment. Just about submitted to the ICES Journal of Marine Science.

Nicholson, M.D., and Fryer, R.J. 1993. Selecting and organism for contaminant trend monitoring. In search of Superbeast. The Working Group on Statistical Aspects of Environmental Monitoring. ICES CM 1993/Env:6.

Nicholson, M.D., and Wilson, S. 1987. Comments on the statistical analysis of CMP data on trends in fish muscle tissue. The Working Group on Statistical Aspects of Trend Monitoring. ICES CM 1987/E:24.

van der Meer, J. 1994. An exploratory graphical display for the analysis of covariance. The Working Group on Statistical Aspects of Environmental Monitoring. ICES CM 1994/Env:6.

Warren ,W.G. 1993. Modelling temporal trends in marine populations: Pollution concentrations as a function of fish length. The Working Group on Statistical Aspects of Environmental Monitoring. ICES CM 1993/Env:6.

Figure A6.1. Example with high excretion rate and linear trend.

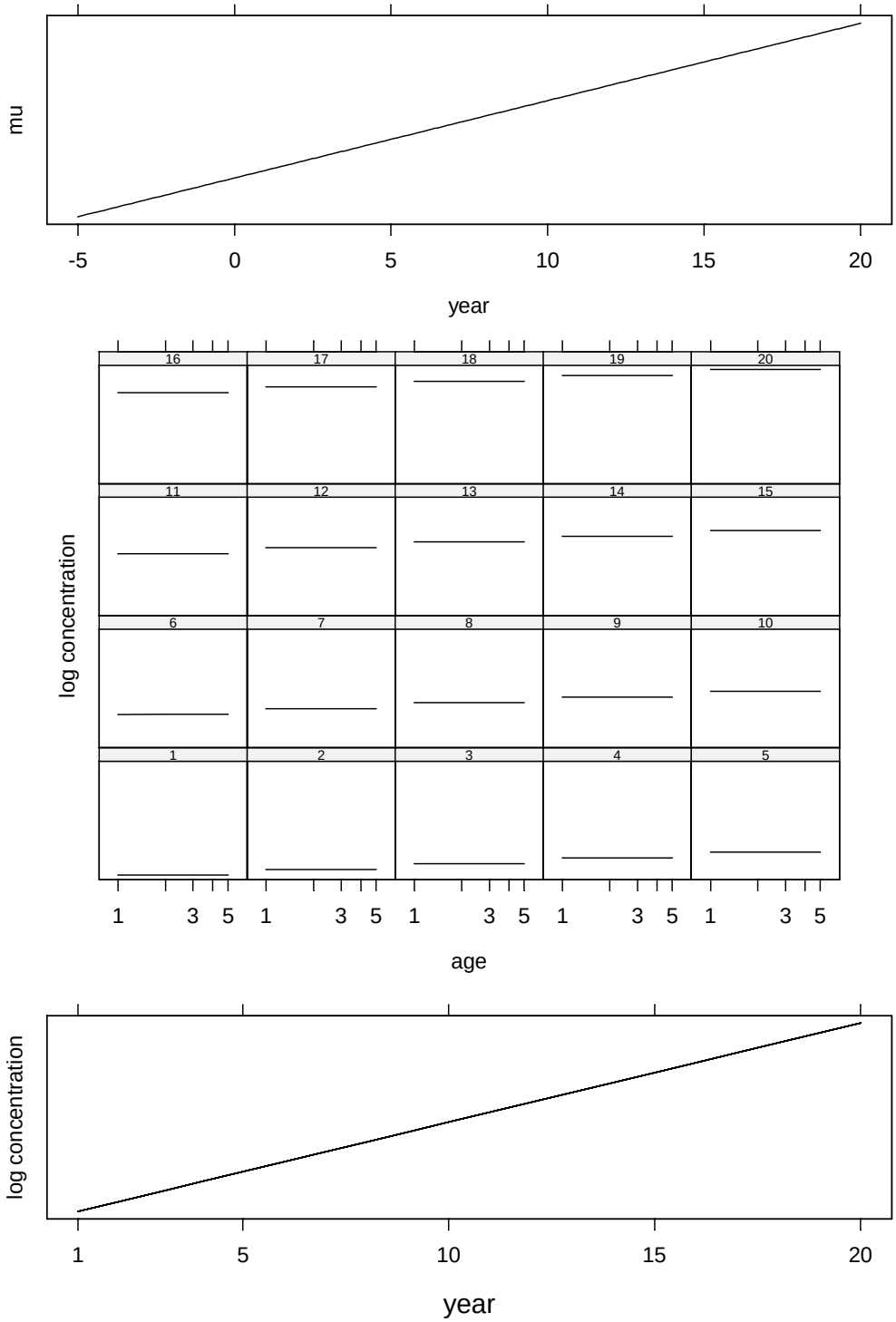


Figure A6.2. Example with low excretion rate and linear trend.

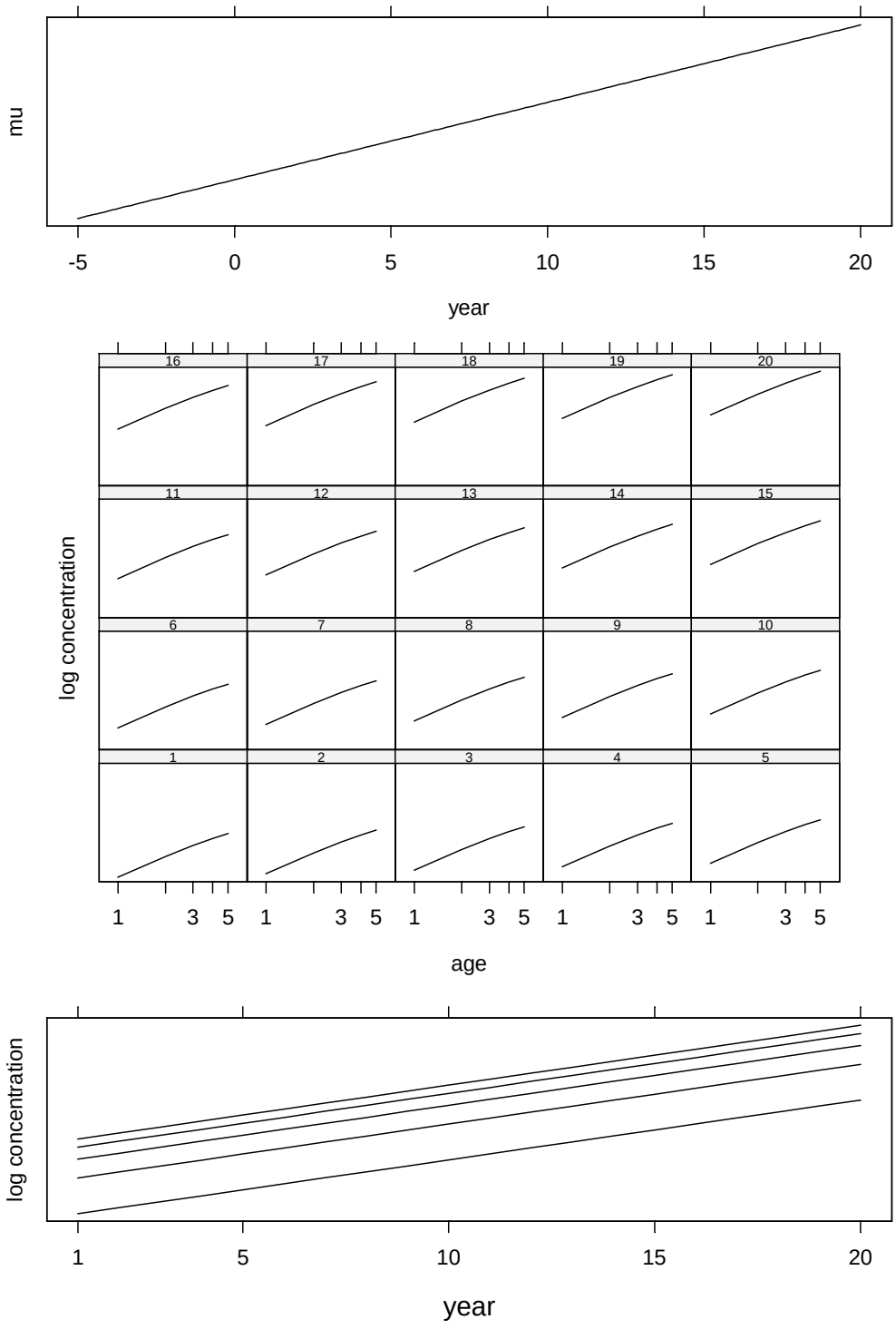


Figure A6.3. Example with low excretion rate and non-linear monotonic trend.

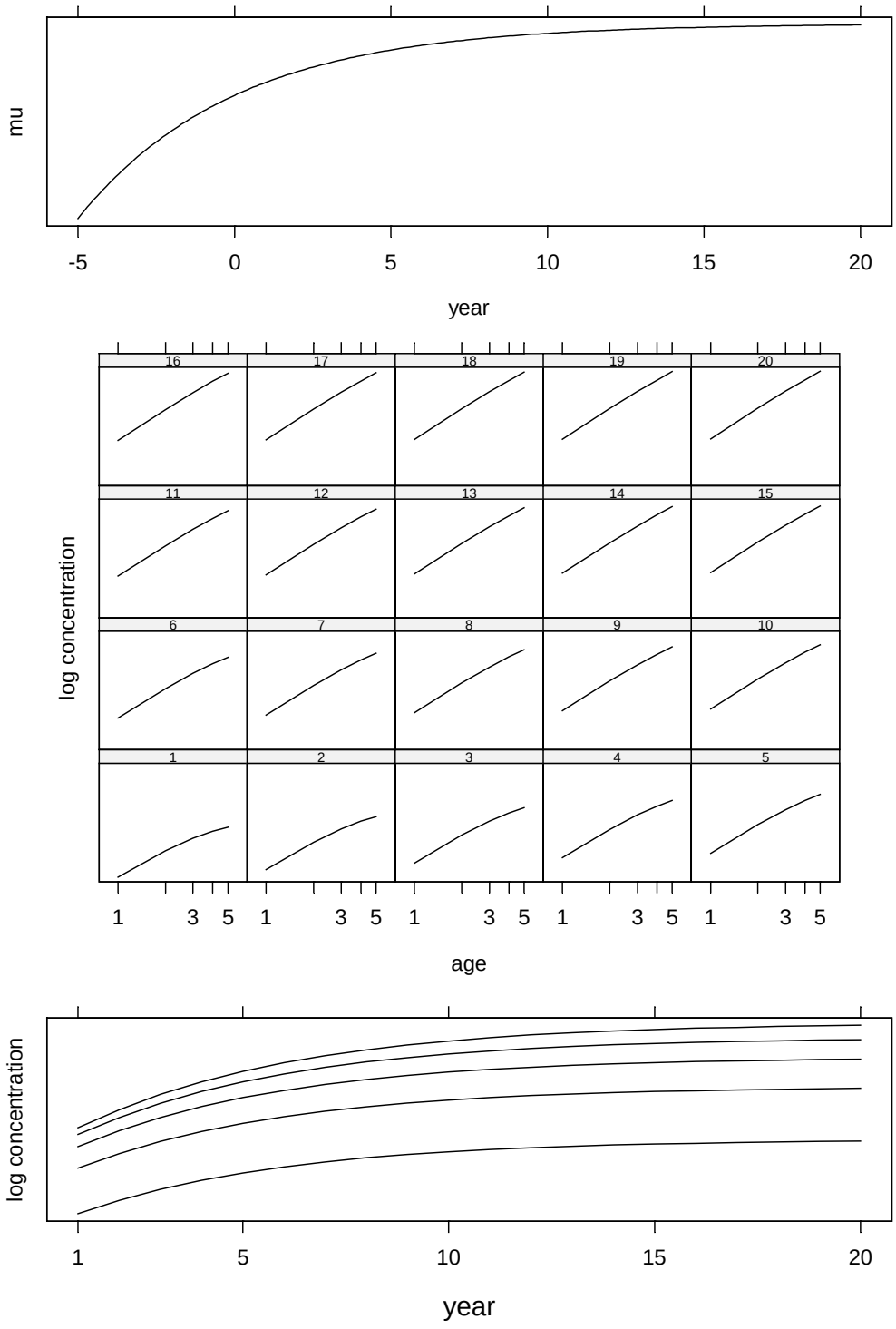


Figure A6.4. Example with low excretion rate and sinusoidal trend.

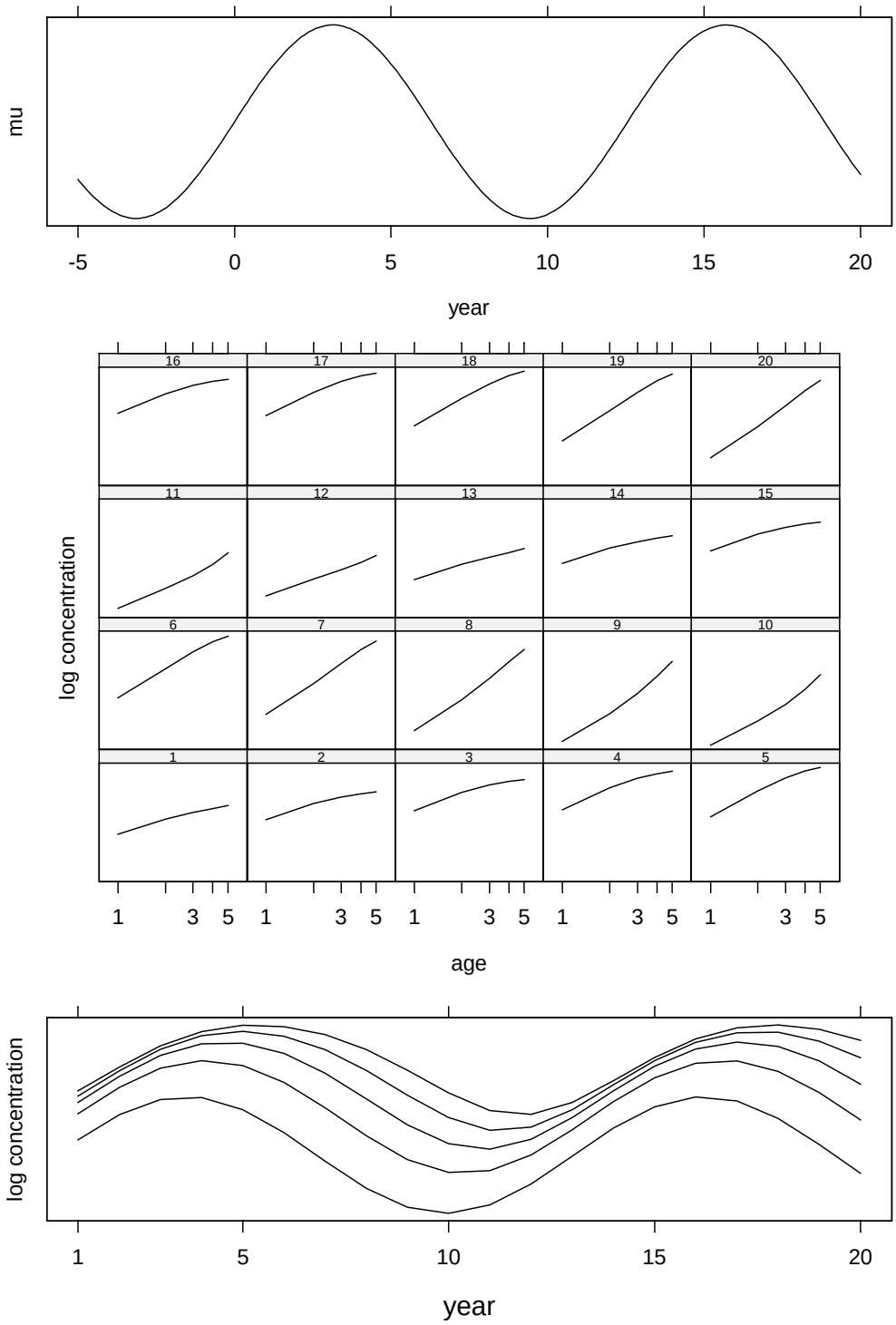


Figure A6.5. Results of the analysis of four time series of mercury concentrations in fish. See text for explanation.

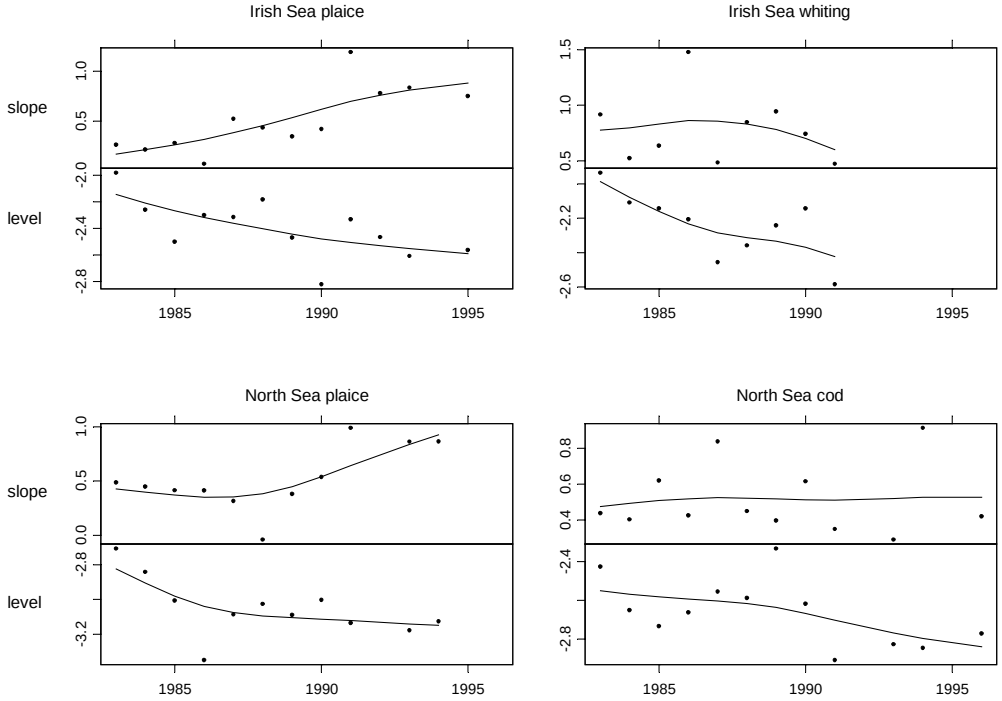
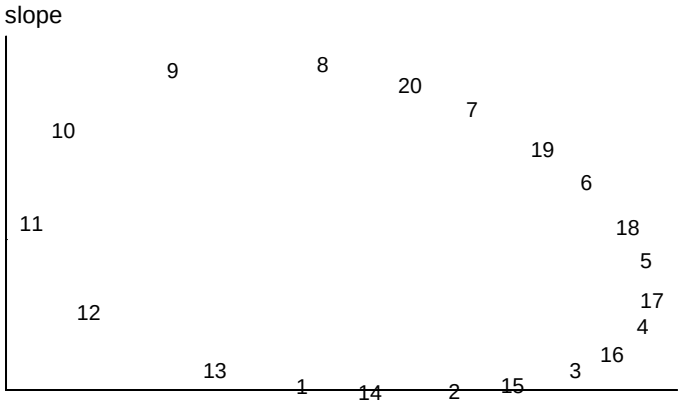
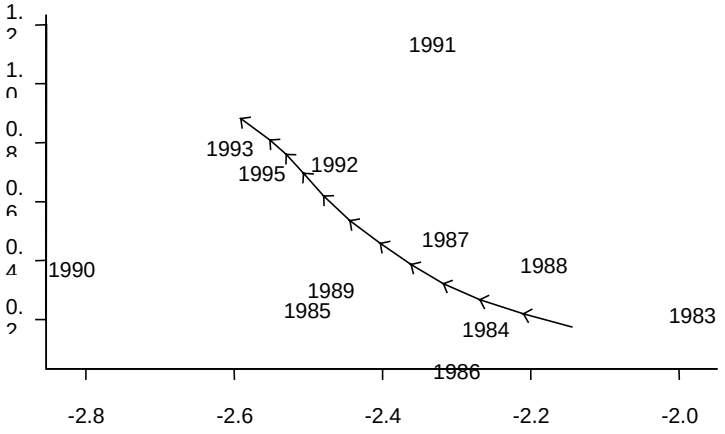


Figure A6.6

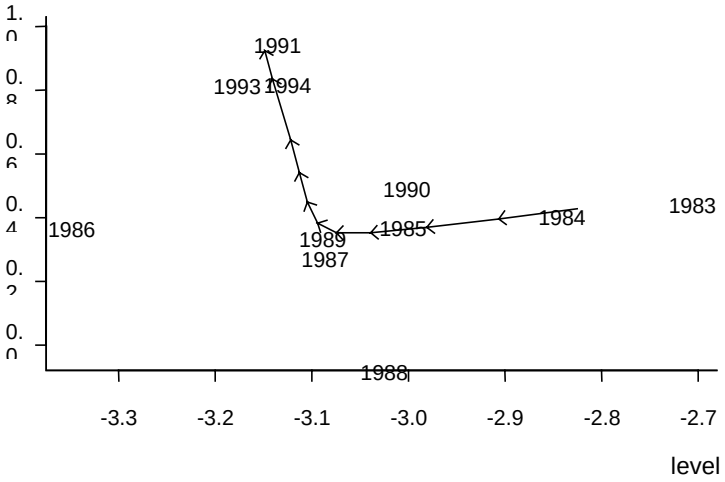
stylised plot based on Figure A6.4



Irish Sea plaice



North Sea plaice



ANNEX 7: VIC DATA ON CONTAMINANTS IN BIOTA

Annex 4

SIME 00/4/11-E (L)
Original: English
English only

OSPAR CONVENTION FOR THE PROTECTION OF THE MARINE ENVIRONMENT OF THE NORTH-EAST ATLANTIC

WORKING GROUP ON CONCENTRATIONS, TRENDS AND EFFECTS OF SUBSTANCES IN THE MARINE ENVIRONMENT (SIME)

STOCKHOLM: 21-25 FEBRUARY 2000

The Voluntary international contaminant-monitoring (VIC) for temporal trends with the aim to test sampling strategies for a co-operative revision of guidelines by 1999

Draft 22.2.2000

Presented by Norway

Norwegian contact point and author:
Birger Bjerkeng
Norwegian Institute for Water Research (NIVA)
PB 173 Kjelsås
N-0411 Oslo
Tel: 0047 22 185 100/151
e-mail: birger.bjerkeng@niva.no

Summary

Background

The background for the VIC program was a recognised need for more cost-effective sampling and analysis strategies for monitoring temporal trends of contaminants in biota and sediments (references). The purpose was to get a basis for assessing the various variance components that contribute to the noise in the series of yearly means, and to compare different sampling strategies regarding pooling and multiple sampling within year and area. The goal is to reduce the total noise component in the data and thereby increase the power to detect time trends. The VIC sampling program should be carried out within the constraints of existing guidelines, by taking supplementary samples for selected areas over a few years.

Preliminary conclusions

Netherlands, Norway and Sweden have submitted data concerning metals and organic contaminants in fish for statistical analysis.

Netherlands has submitted data for flounder, collected from one area over the years 1996-1998. Each year two length-stratified samples of 35 fish were taken at the same time and place and analysed individually or in pairs on different sets of contaminants. In addition 4 pooled samples of 10 fish one size group were taken, one as a parallel sample (same time and place), two at about the same time from other locations within the area, and one from the same location a month later (1997-98 only). Organochlorine contaminants in liver are clearly related to fat fraction. The data indicate with 90 % confidence that for HCB the ratio of variance components due to small-scale temporal/spatial variation to the residual variance for the pooled sample was 1.2 for pooled samples of 10 fish. The same may be true for CBΣ7, but with lower confidence. Analysis of data on metals are yet to be completed.

Sweden has collected herring for individual analysis from 4 sites in the south-western part of the Baltic (only 1997), 2 sites in Kattegat (1997 + additional sample in 1998 at one site) and 3 sites in Skagerrak (only 1998), most samples consisting of 12 specimens. The most successful covariate for reducing residual variance turned out to be the difference between actual fish weight and expected weight given length, but this varied between contaminants. For organic contaminants adjusted for physiological differences, between-sample to residual variance component ratios estimates range from 0.4 to 0.8, with 90 % confidence lower limits from 0.15 to 0.4. For metals the variance component ratios are higher, with mean estimates ranging from 0.65 to 2.5, with 90 % confidence lower limits from 0.3 to 1.35. Even with a low variance ratio of 0.15, taking 3 samples of 2 fish each would give somewhat better precision than sampling 12 fish at one occasion. If the 12 fish were sampled in 3 catches of 4 fish each instead of in one sample, the sampling-dependent variance would be reduced by 40 %.

Norway has taken multiple samples of Cod and flounder each year for one station in the Oslofjord and two stations in Sørfjord and Hardangerfjord. Results for organic contaminants in cod indicate with 90 % confidence that the between-site and time variance component is at least 20 % of the between-specimen variance for samples taken at the same site and time. This could mean that instead of sampling 25 fish at one occasion, it might be somewhat better to take 2 samples of 5 fish each at different sites or with a time difference, or 3 samples of 3 fish each. By distributing the 25 fish on 3 samples, the variance of the yearly mean estimate could be reduced by 65 %. Results for metals in Norwegian Cod do not indicate any significant small-scale variation, in this case the total number of fish is most important.

Way forward

- Complete draft and circulate to contact points by mid March
- Contracting Parties to designate contact points to Birger Bjerkeng by March 1
- Circulate current document to ICES Working Group on the Statistical Aspects of Environmental Monitoring (WGSAEM)
- Ask WGSAEM to consider results at their meeting in 27-31 March 2000.

Action requested to SIME 99

SIME is invited to note the Norwegian assessment of VIC data and to comment as appropriate

SIME is also invited to support suggested way forward.

Background

Tissue levels of metals and organochlorine compounds in various fish species have been used in spatial and temporal monitoring for over 30 years. According to ICES (1996) and OSPAR (1990, 1997) guidelines, 25 individual fish within a predetermined size range should be sampled at any given site.

An OSPAR Commission (OSPAR) Ad Hoc Group on Monitoring (MON) sub-group proposed a simple international programme called Voluntary International Contaminant-monitoring for temporal trends with the aim to test sampling strategies for a co-operative revision of guidelines by 1999 (VIC. cf. SIME 1996, 1997a). The background for the VIC program was a recognised need for more cost-effective sampling and analysis strategies for the monitoring of temporal trends of contaminants in biota and sediments (references).

A simple 3-year sampling program was set up as an extension to the normal monitoring program, with countries participating on a voluntary basis. The countries that expressed interest in contributing to the program were Germany, Netherlands, Norway and Sweden. The purpose was to get a basis for assessing variance components within year and station compared to residual within-sample variance, and to compare different sampling strategies regarding pooling for analysis and multiple sampling within year and area. The basic structure of testing is the same though the details may vary from country to country depending on practicalities. So far, data from Netherlands, Norway and Sweden have been submitted for statistical analysis. This paper presents the results of this analysis.

Problem definition

The goal in monitoring for temporal trends is typically to detect permanent changes or long-term variations in levels by analysing a series of yearly means. Estimated yearly means will vary both as function of long-term changes, which we want to detect, and due to short-term or small-scale irregular variation, which will act as noise in the data. To the degree

that small-scale variations cannot be controlled by choosing sampling times or specific sites, there will be a statistical error on the sample level in addition to error on the specimen level.

In order to achieve statistical power for detecting trends in contaminant levels in fish in as cost-effective way as possible, it is important to identify and estimate the different variance components involved. If the estimated yearly means are based on catching a number of fish at one occasion each year, the within-year variation between samples (i.e. between subpopulations of fish) will be included in the between year residual variance in the data series. Even if the design is to collect one sufficiently large sample each year at a specific location, this may not always be possible. In order to get a sufficient number of fish it may sometimes be necessary to move to a nearby location, do more than one catch or use a sample of fish deviating from the specified size distribution. This may affect the residual variance in the estimated yearly means that are used to detect trends.

Sampling from a single location at one time each year will often mean sampling from a sub-population that does not represent the true yearly mean. With such sampling, where a sub-population is selected randomly by choosing time and place, and then sampled by catching and selecting specimens for analysis, the total within-year residual variance will be composed of two components, first the variance between sub-populations and second the variance between individuals in each sub-population. The effect of the last component can be reduced by sampling more individuals at an occasion, but the first component can only be reduced by sampling more than once, or at more than one place.

If the between-site or seasonal variation is systematic, the best sampling scheme will be to keep to a specific location and a specific season. If, on the other hand, there is a random variation between locations and sampling times due to fish migrating within the area, this variation is impossible to control. In that case, changing the sampling program into collecting a smaller number of fish from each of a number of locations, and/or at different times, may increase the precision of the yearly mean estimate.

In this context, it is also important to consider how physiological measurements on the fish can be used as covariates to reduce residual variance. If variations in contaminant levels are due to physiological variations over time or between fish sub-populations, the adjustment for such factors may reduce residual variance and increase the ability to detect trends. Such normalisation could influence the variance component ratios, and affect conclusions regarding repeated sampling. It could also change the cost/benefit balance between individual and pooled samples.

Statistical theory

Basic model of variance structure of yearly means

Assume that $S \cdot I$ fish are collected each year from a station or area, distributed equally among S sub-samples, and analysed individually. The samples could be from different locations in the station area and/or at different times within the designated sampling period (i.e. outside the normal spawning season).

The linear statistical model for some suitable transformation of the contaminant level in fish no. $i=1,\ldots,I$ of sub-sample s in year y from an area is

$$x_{ysi} = f(t) + \alpha_y + \beta_{s(y)} + \varepsilon_{(sy)i} \qquad (1)$$

where $f(t)$ is the trend function we want to detect common to a station or geographical area, α_y is the irregular variation between years and $\beta_{s(y)}$ is the deviation of the mean value of sample s from the yearly mean, while $\varepsilon_{(sy)i}$ is the residual variation of specimen i from the sub-sample mean. The parentheses in the indices show that sample index is nested within years (y) and specimen index is nested within sample (sy). Variation between samples can be both related to differences in space or in time within sampling season, and may have a more or less "random" (=irregular) character

The problem addressed here is sub-sampling from a large and heterogeneous population, where the within-year variation cannot be controlled by stratified sampling. The sub-sample variation must accordingly be treated as a random component. The variance components involved are:

σ_Y^2 = irregular variance between years, unrelated to the trend function

σ_S^2 = variation between samples (a function of small-scale variance in time and/or space)

σ_I^2 = variation between individual fish within samples, including analysis error.

The variance of the level in a randomly selected specimen around the trend function $f(t)$

$$V\left(x_{ysi}\right)=\sigma_Y^2+\sigma_S^2+\sigma_I^2 \tag{2}$$

The yearly sample mean based on these values has variance around the true expectation

$$V\left(\bar{x}_{y\bullet\bullet}\right)=\sigma_Y^2+\frac{\sigma_S^2+\frac{\sigma_I^2}{I}}{S} \tag{3}$$

where • indicates averaging over the corresponding index. The sampling design should aim at reducing this variance as much as possible.[1] If σ_S^2 is not negligible compared to σ_I^2, one may achieve a better precision by reducing $S{\cdot}I$, and analyse fewer fish totally, but increase the number of sub-samples. The effect of the yearly variance component σ_Y^2 cannot be reduced, and the size of that component will limit what can be achieved by improving within-year sampling.

The equation can be written:

$$V\left(\bar{x}_{y\bullet\bullet}\right)=\sigma_Y^2+\sigma_I^2\frac{1+I\cdot\sigma_S^2/\sigma_I^2}{S\cdot I} \tag{4}$$

How much is achieved by increasing number of samples (S) and decreasing number of individuals within sample (I) with fixed total number of fish $S{\cdot}I$ depends only on the ratio σ_S^2/σ_I^2. Consequently, the interest is primarily on estimating this variance ratio. This can be done through analysis of variance as described below.

Estimation of confidence limits for variance component ratios

The available VIC data sets on individual data can be analysed according to the following linear model:

$$x_{ysi}=\left(AREA*YEAR\right)_y+SAMPLE\left(AREA*YEAR\right)_{s(y)}+\left(\text{Covariate regression}\right)+\varepsilon_{(sy)i} \tag{5}$$

where indices are as in equation (1) in chapter 0. The between-year component has been generalised into also covering variation between areas, so that data sets covering more than one area can be analysed as a whole to estimate common variance components on the sample and specimen level. The within-area_and-year variance across samples is captured by a random SAMPLE effect, nested within the AREA*YEAR factor. The covariate regression models may be different for different contaminants, based on a closer investigation of the different data sets.

In order to fulfil the assumption of linear effects and homogeneous variances one may have to transform the data. For the VIC data a log-transformation seems in most cases to be required, so x=ln(concentration).

The common feature of ANCOVA analyses based on variants of this model is that they extract largely independent estimates of between-sample and within-sample (residual or error) mean squares. The expected mean squares are:

$$EMS_{sample}=n\sigma_S^2+\sigma_I^2 \tag{6}$$

$$EMS_{error}=\sigma_I^2 \tag{7}$$

where σ^2_S and σ^2_I are the variance components due to variation between samples and individual within sample, respectively.

The coefficient n is equal to the number of fish in each sample for a balanced data set with equal number in all samples, and in that case the variance of sample geometric means is $\sigma_s^2+\sigma_\varepsilon^2/n$. Otherwise n is a weighted mean of these

[1] If the analysis of time trends is done on yearly means, the residual degrees of freedom will be independent of within-year sampling.

numbers. The numbers used here are calculated by the VGLM[2] module in Statistica v. 5.5 according to Satterthwaite, refer to Milliken and Johnson 1992).

In general, if we have independent estimates s^2_A and s^2_B with ν_A and ν_B degrees of freedom of two variances σ^2_A and σ^2_B, the ratio of the estimates follow the $F(\nu_A, \nu_B)$ distribution under the null hypothesis of equal variances, and this is used to test for significant difference between the variances in the ordinary F test.

The normalised ratio $(s^2_A/ s^2_B)\cdot(\sigma^2_B/\sigma^2_A)$ is F distributed by definition, and this can be used to determine "exact" confidence limits for the ratio between the two variances. For a chosen two-sided significance level α, the two-sided confidence interval of the F-distributed adjusted ratio is defined by the upper- and lower $\alpha/2$-percentage points of the F distribution: $F_{\alpha/2}(\nu_A, \nu_B)>1$ and $F_{1-\alpha/2}(\nu_A, \nu_B)<1$.

The estimated ratio (s^2_A/ s^2_B) will then with confidence $1-\alpha$ be found between the numerically unknown limits:

$$\frac{\sigma_A^2}{\sigma_B^2} F_{(1-\alpha/2),\nu_A,\nu_B} \le \frac{s_A^2}{s_B^2} \le \frac{\sigma_A^2}{\sigma_B^2} F_{(\alpha/2),\nu_A,\nu_B} \tag{8}$$

These relations can be inverted into a confidence interval for the unknown ratio between variances:

$$\frac{s_A^2/s_B^2}{F_{(\alpha/2),\nu_A,\nu_B}} \le \frac{\sigma_A^2}{\sigma_B^2} \le \frac{s_A^2/s_B^2}{F_{(1-\alpha/2),\nu_A,\nu_B}} \tag{9}$$

If we assume that $\sigma_A^2 = k\sigma_C^2 + \sigma_B^2$, as in test of random effects in ANOVA models, this becomes:

$$\left(\frac{s_A^2/s_B^2}{F_{(\alpha/2),\nu_A,\nu_B}} - 1 \right) \frac{1}{k} \le \frac{\sigma_C^2}{\sigma_B^2} \le \left(\frac{s_A^2/s_B^2}{F_{(1-\alpha/2),\nu_A,\nu_B}} - 1 \right) \frac{1}{k} \tag{10}$$

If an effect is found significant, both limits of this last interval are positive, and the interval can then be used to indicate the precision of the variance ratio. Note that the actual confidence level of this interval is strongly dependent on residuals being normally distributed, or on having a large data set.

Adapted to the sample and residual mean square estimates from ANCOVA analyses on the VIC data, this leads to the following confidence limits for the ratio of variance components can be estimated as:

$$\left(\frac{EMS_{sample}/EMS_{error}}{F_{(\alpha/2),\nu_{sample},\nu_{error}}} - 1 \right) \frac{1}{n} \le \frac{\sigma_S^2}{\sigma_I^2} \le \left(\frac{EMS_{sample}/EMS_{error}}{F_{(1-\alpha/2),\nu_{sample},\nu_{error}}} - 1 \right) \frac{1}{n} \tag{11}$$

A conservative estimate of what may be achieved by multiple sampling, is found be using the lower confidence limit. A 90 % one-way confidence level may than be reasonable, in that case one should use $\alpha=0.2$ in the formulas above.

Pooled sample variance compared to variance of mean values based on data for individual specimens

The problem is in principle the same whether the fish are analysed individually and 'pooled' by calculation into a mean value or pooled before analysis. The variance components involved may be different because analysis of individual specimens give more freedom to choose how to combine specimen values into means (arithmetic or geometric means, weighting according to variance structure etc.).

To be completed

log-normal distribution

---

2 Module for visual linear general models

Contaminant levels in biota often show quite large variation between specimens, and typically have skewed distributions with an absolute lower boundary of 0 and a long tail towards high values. A log-transformation often gives more symmetric distributions and more homogeneous variance between samples and across covariates. Thus, the analyses here are done with log-transformed contaminant levels. All continuously measured physiological covariates, like length, weight etc. are also log-transformed.

To be completed

Dutch data

Sampling and analysis program

Sampling for the VIC program was done through 1996, 1997 and 1998 on flounder. Each year, 4 samples were taken from the same area for comparison. The samples represent different sampling strategies.

Strategy 1: Under this strategy 2*35 individual fish are collected, distributed over 5 size classes, and analysed as shown in the table. Physiological data (length, weight, liver weight etc.) are available for individual fish under strategy 1, and water content and lipid content for part of the analysed samples.

<table>
<tr><th>Size class</th><th>Actual size range</th><th colspan="2">Analysis program</th></tr>
<tr><td>1</td><td>19.8–23.2 cm</td><td rowspan="2">20 fish of each size class, pooled pairwise into 10 analysis samples</td><td rowspan="5">5 samples from each size class are analysed for mercury in muscle and cadmium in liver.
The other 5 samples are analysed for organic contaminants in liver</td></tr>
<tr><td>2</td><td>22.5–25 cm</td></tr>
<tr><td>3</td><td>25–27.9 cm</td><td rowspan="3">10 fish of each size class, analysed individually</td></tr>
<tr><td>4</td><td>28–31.5 cm</td></tr>
<tr><td>5</td><td>31.2–34.5 cm</td></tr>
</table>

Strategy 2: One pooled sample of 10 fish, collected at the same time and part of area as the sample for strategy 1. All fish are from size class 3.

Strategy 3: Two pooled samples of 10 fish (Strategies 3-1 and 3-2), collected at the same time as sample 1 and 2, but from other parts of the same area. All fish are from size class 3

Strategy 4: One pooled sample of 10 fish, collected from the same area a month later than the other samples. Samples are collected by a local fisherman, and are also supposed to be from the size class 3, although possibly not so stringent. This strategy is missing for 1996.

Only water and fat fractions are given as physiological information for the pooled samples.

Scheme of statistical analysis

The goal in this context is to assess if there is significant variation between samples within year and area, and to compare different sampling strategies regarding pooling for analysis and multiple sampling within year and area.

Since there is only one statistical sample each year of individually analysed fish, the within-year*area variation must be assessed based on pooled samples. The samples under strategy 1 can for this purpose be used to construct a hypothetical pooled sample of the fish from size class 3, assuming equal amounts from each fish, and neglecting differences in how analytical errors contribute.

The individual specimens analysed under strategy 1 can be used to estimate variance components for variation with physiological variables as a background for the study of the pooled samples. They can also be used to assess how the adjustment for physiological covariates can make individual specimen analysis more cost-effective for detecting trends.

Comments on the data

Contaminants analysed in flounder are mercury in muscle, cadmium in liver, HCB and a series of PCBs.

HCB has three cases of below detection limit values in the strategy 1 sample from 1998. The detection limit (<1.0) for two of these cases, both in size group 5, are higher than about 20 % of the other HCB values, probably due to normalisation against fat weight. The detection limits are used as values in the analysis - this should tend to inflate the mean estimate for this sample, and decrease variance estimates slightly.

The 7 components included in CBΣ7 (CB28, CB52, CB101, CB118, CB138[3], CB153, CB180) are present for all samples and only in one sample given as below detection limit for the quite minor component CB28. Thus CBΣ7 can be analysed statistically without problems, and the PCB analysis is restricted to CBΣ7.

Physiological data - covariates

Fat in liver are given as Fat-soxhlet and Fat-B&D. For the 69 samples where both are measured, they are strongly related by a function

$$Fat_{soxhlet} = \frac{1.23 Fat_{BD} - 25.4}{1 + 0.000564 Fat_{BD}} \qquad (12)$$

In cases where both are measured, the geometric mean of $Fat_{soxhlet}$ as measured and as calculated from $Fat_{B\&D}$ is used. In the cases where only $Fat_{B\&D}$ are measured, the formula is used to calculate $Fat_{soxhlet}$, and where only $Fat_{soxhlet}$ is measured, that value is used.

The fat content of the liver is related to the water content and size of the liver.

The relation between different physiological measures have been analysed on samples in strategy 1. For the small size classes, the analysis uses sample values, that is mean values for pairs of fish. Since the various parameters are strongly correlated, the relations are analysed by deriving a sequence of linearly independent measures from the variables, each one found by regression one the derived measures found so far. This avoids the colinearity problem, and may give more meaningful measures than for instance using PCA. The relations are expressed by measures relative to sample means, which gives both better numerical precision and also takes care of the unit issue.

A log-log linear regression between weight (W) and length (L) yields an expression for the geometric mean estimate:

$$\hat{W} = W_0 \cdot \left(L / L_0 \right)^{2.88} \quad \text{with geometric means } L_0 = 36.32 \text{ cm}, W_0 = 222.6 \text{ g}. \qquad (13)$$

This model captures around 97 % of the variance in log(W). The deviation of the weight from the length-based normal value can be used as a secondary size measure, and is expressed as length-normalised dimensionless weight ratio

$$\omega = \frac{W}{\hat{W}} \qquad (14)$$

Next, the liver weight (W_{liver}) can be related to fish length and weight deviation, again by log-log-linear regression. The resulting geometric mean model estimate is

$$\hat{W}_{liver} = W_{liver,0} \left(\frac{L}{L_0} \right)^{3.93} \omega^{1.70} \quad \text{with sample geometric mean } W_{liver,0} = 3.4 \text{ g} \qquad (15)$$

This relation captures about 91 % of the total variance in log(W_{liver}). The size-normalised liver weight can be used as a third physiological measure, statistically independent of length and weight:

[3] The CB138+163 component listed in strategy 1 data from 1998 is assumed to be the same as the CB138 component in the rest of the data.

$$\omega_{liver} = \frac{W_{liver}}{\hat{W}_{liver}} \quad (16)$$

For the liver there is a strong correlation between variations in water content and lipid content. The pattern of variations is shown in the ternary plot in Figure **A7.1**A7.1. The ternary plot shows a trend of a constant ratio between water and non-fat dry matter around 4:1, with fat content varying from 2.5 to 37 %.

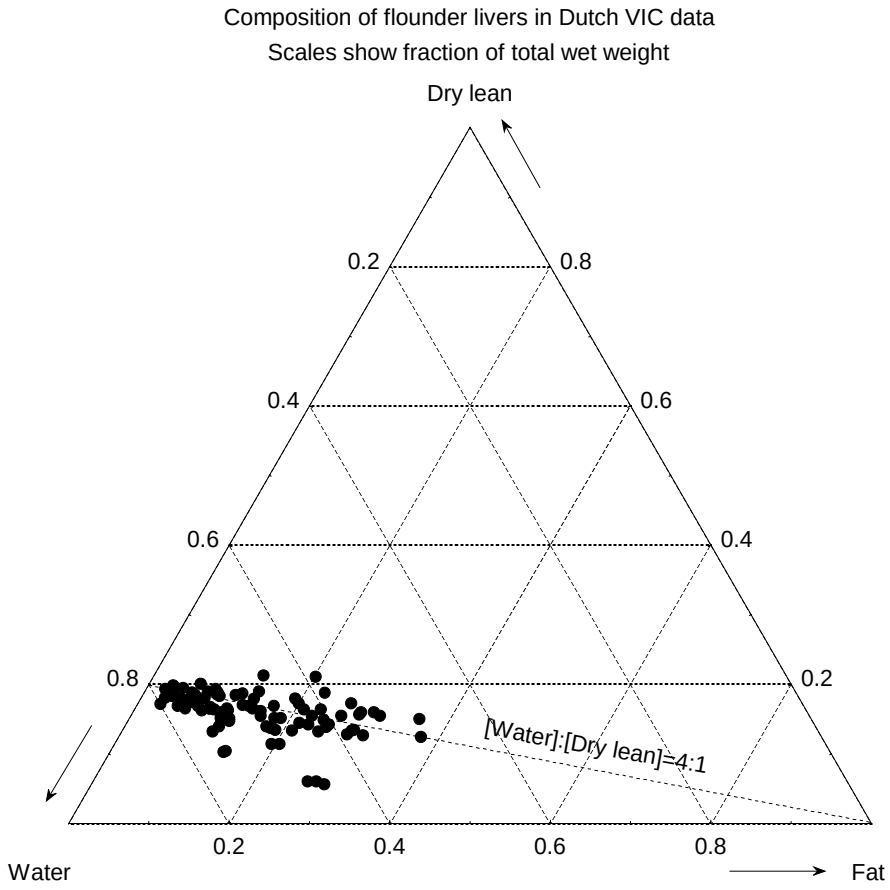


Figure A7.1. Composition of livers in flounder, Dutch VIC data.

It seems reasonable to use the fat-% as the main measure of liver condition, and try to relate that to the three size measures above. A log-log-linear regression of fat fraction against the three size measures (L, ω, ω_{liver}) captures 69.5 % of the fat-% variation, with the geometric mean estimate of fat fraction (F_{liver}) given by

$$\hat{F}_{liver} = F_{liver,0}\left(L/L_0\right)^{3.06} \omega^{2.38}\omega_{liver}^{1.19} \text{ with sample geometric mean } F_{liver,0} = 11.9\ \% \quad (17)$$

The deviation from normal is then

$$\phi_{liver} = \frac{F_{liver}}{\hat{F}_{liver}} \quad (18)$$

Alternatively, fat fraction (or fat as fraction of dry matter) could be used as the primary description of the liver, and regressed on length and size to derive a deviation from normal fat fraction given size. Liver weight could then be seen as a function of fish size and liver fat %. That might make more sense biologically (?), but the main purpose here is to derive log-linearly independent physiological measures that allow for contaminant-physiology relations to be studied.

The dry-lean:water ratio can be used as a secondary measure of liver composition.

Water content in muscle varies between 76.4 and 84.5 % except for an outlier where it is given as 62.3 %. This value might be due to an error.

Variance analysis on the different physiological measures with year and direction (Left/Right) shows that there is a significant difference in fat-% in livers between years, with generally higher levels in the 1997 sample, and also differences in the dry-lean:water ratio. This is mainly due to lack of low-fat livers in 1997, and the occurrence of the apparent outliers with low dry:water ratio.

Organic contaminants in liver

Individual samples - size and physiological as covariates

Log-transformed CBSUM7 and HCB have been analysed in step-wise ANCOVA, using year and size groups as factors and log-transformed physiological covariates. Finally only logtransformed fat-% in liver has been included as covariate. The covariate reduces residual variances within groups (*year*sizegroup*) by about 65 %, to 0.075 and 0.090, respectively, for the two contaminants.

For fat-adjusted CBΣ7 the *year*sizegroup* interaction term is clearly significant for CBΣ7 (p=0.009), and there are no indications of significant main effects between years or between size groups compared to the interaction.

For fat-adjusted HCB the *year*sizegroup* interaction term is not quite significant (p=0.093). The variation between years as a mean over size groups is significant both compared to the residual ($p=1.5 \times 10^{-6}$) and the interaction term (p= 0.0046). There is no indication of significant differences between size groups.

The regression coefficients on log(fat-%) are close to 1 for both contaminants (0.86±0.08 and 0.95±0.09 with 95 % confidence), which means that concentrations are proportional to fat %. Presumably the analyses have been done on extracted fat, and the concentrations on wet weight basis calculated by multiplication with fat fraction. The proportionality reflects this, indicating that concentrations in the extracted fat vary independently of the amount of fat extracted from the liver.

Residual variance in pooled samples estimated from data on individuals

In order to estimate the within-year spatial or temporal variance components, we need to compare mean squares for variance between pooled samples with estimates of residual effect of the variation between individuals. The residual variance can only be estimated by looking at the data on individuals from strategy 1. For this purpose we look at data from size group 3 only, since the pooled samples are (mainly) from this size group, and estimate residual effect in pooled samples by numerical simulation of pooling. The only available covariate for the pooled samples is pooled fat % in liver.

Whole livers are pooled, and liver weights vary by a factor of at least 2 within sample, additionally both fat fraction and contaminant levels vary significantly with liver weight. For individual concentrations C_i on wet weight basis and liver wet weights w_i, the concentration in the pooled sample will be:

$$C_p = \frac{\sum_{i=1}^{n} w_i C_i}{\sum_{i=1}^{n} w_i} \qquad (19)$$

The same equation holds for fat and water fractions of the pooled sample. The between-sample variance of C_p and its relation to fat fraction in a pooled sample of 10 specimens will be given by the covariance structure of concentration, weight and fat fraction.

Both contaminants are log-linearly related to log(fat-%) with regression coefficients around 1.0 ($p \leq 0.0008$) for individuals in size group 3. The residual variance is around 0.12 for both contaminants. There is no significant between-year effect, therefore the covariance structure is analysed for the data from all three years as a whole.

The effect of between-specimen variation on variation between pooled samples has been estimated by numerical simulation of 1000 pooled samples, based on the covariance structure between specimen found by Factor Analysis on data for size group 3 from strategy 1. Table A7.1 shows the parameters for distributions of variables and the variance structure in terms of correlation with common factors. The two rightmost columns show variance of the simulated pool sample values, before and after adjustment for the pooled values of the simulated fat fractions. These variance estimates ($\sigma_{I,P}^2$) are the expected residual effect of the random variation between individuals on the variance between pooled samples. By comparing columns σ^2 and $\sigma_{I,P}^2$ (unadjusted) one will see that the pooling reduction factors on the variances are 6 to 7, instead of 10 as would be obtained for equal amounts of tissue samples and symmetrically and identically distributed concentrations. The same holds for fat-adjusted levels of both contaminants, where the pooled samples have a residual variance of 0.018 versus the variance 0.12 for individual specimen in size group 3.

Table A7.1. Distribution parameters for log-normal distributions and covariance structure of wet weight, fat fraction, and organochlorine contaminants in livers from flounder of size group 3 in Dutch VIC data where organochlorines were measured.

The two rightmost columns show variance of simulated pool values, before and after adjustment for pooled values of the simulated fat fractions.

Variable	Log-normal distribution ln(Variable)~ $N(\mu,\sigma^2)$ size group 3		Covariance structure of ln(Variable). regression coefficients on 3 common factors distributed ~N(0,1):			residual variance	Variance of ln(variable) for simulated pooled sample of 10 fish $\sigma^2_{I,P}$	
	μ	σ^2	factor 1	factor 2	factor 3		un-adjusted	adjusted for fat fraction
liver weight *w*(g)	1.24	0.087	0.247	-0.156	-0.038	0.00068	0.0097	--
Fat fraction	-2.32	0.34	0.568	-0.048	0.082	0.016	0.54	--
ln(HCB) (ppb w.w.)	0.58	0.57	0.706	0.161	0.193	0.013	0.094	0.017
ln(CBΣ7) (ppb w.w.)	5.94	0.31	0.492	0.204	-0.169	2.3E-05	0.046	0.019

Pooled samples - between sample variance

The actual data from pooled samples from strategies 2, 3.1, 3.2 and 4 constitute a two-way ANOVA design with Year and Strategy (spatial or time difference) as crossed factors.

Table A7.2 summarises the results of variance analysis on these data for HCB and CBΣ7, both on wet-weight levels (left side) and on levels adjusted for fat fraction regression (right side, ANCOVA analysis).

The interaction between Year*Strategy is the irregular between-sample component. The F tests on year and strategy effects show comparisons of effects with expected error due to the irregular variation, and it is clear that there are no indications of any over-all variation between years or persistent differences between locations or times within year.

The bottom part of the table for each contaminant compares the interaction term with the residual variance between pooled samples due to variation between specimens that were estimated by simulation in 0. These residuals are listed with degrees of freedom in the source data (strategy 1, size group 3) that were used for the simulation. The statistics for variance component ratios have been estimated by the formula given in Chapter 0, using a value k=1, since both the interaction terms and the residual terms are estimated for variance between single pooled values.

Table A7.2 indicates that for pooled samples, the dominant variance component is probably the small-scale variation between sites or between months within year (mean ratios 1-5). HCB concentrations adjusted for pooled fat fraction appear to almost certainly have a considerable between-site/time variance, since the 90 % lower confidence limit for the variance component ratio is >1. The practical conclusion is that precision of yearly mean estimates can be improved considerably by taking more than one pooled sample, or by pooling specimens from different catches of fish. The same may be true for CBΣ7 (mean>1), but a between-site/time effect can only be stated with 85 % confidence in this case.

Table A7.2. Analysis of variance for pooled samples, strategy 2 - 4, for HCB and CBΣ7. One-way confidence limits for variance component ratio (Strategy*Year)/Residual are based on the numerical simulation described in Table A7.1 of the residual error of pooled samples.

ln(HCB):

		unadjusted				adjusted for fat fraction			
Effect	Error term	df	MS	F	p	df	MS	F	p
Strategy	Strategy*Year	3	0.025	0.118	0.946	3	0.023	0.24	0.865
Year	Strategy*Year	2	0.229	1.075	0.409	2	0.084	0.87	0.486
Strategy*Year ($\sigma_S^2+\sigma_{I,P}^2$)	Residual	5	0.213	2.269	0.12	4	0.096	5.662	0.012
Residual of pooled samples ($\sigma_{I,P}^2$)		11	0.094			10	0.017		
Variance component ratio $\left(\frac{\text{Strategy}*\text{Year}}{\text{Residual}}\right)$	90 % lower			-0.07				1.2	
	mean			1.3				4.7	
	90 % upper			7.5				22	

ln(CBΣ7):

		unadjusted				adjusted for fat fraction			
Effect	Error term	df	MS	F	p	df	MS	F	p
Strategy	Strategy*Year	3	0.009	0.054	0.981	3	0.017	0.409	0.756
Year	Strategy*Year	2	0.156	0.941	0.45	2	0.179	4.336	0.1
Strategy*Year ($\sigma_S^2+\sigma_{I,P}^2$)	Residual	5	0.166	3.604	0.036	4	0.041	2.178	0.145
Residual of pooled samples ($\sigma_{I,P}^2$)		11	0.046			10	0.019		
Variance component ratio $\left(\frac{\text{Strategy}*\text{Year}}{\text{Residual}}\right)$	90 % lower			0.5				-0.16	
	mean			2.6				1.2	
	90 % upper			12				9	

Metals

Individual samples - size and physiological as covariates

For the metal analysis on individual fish in strategy, analysed on a different set of fish than the organic contaminants, the water contents of liver and muscle are available as covariates for part of the samples in addition to the size measures given for all samples. Neither appears to be very useful as covariates for the metals.

Mercury in muscle is clearly related to length and weight as covariates; there is no significant variation between size groups or interaction year*sizegroup when the covariates are adjusted for. The residual variance of the size-adjusted contaminant levels on natural log-scale is 0.10. The covariate relation is an overall increase with length, and a secondary, but clear decrease with increasing weight for given length.

To be completed

Norwegian data

Sampling program

The Norwegian VIC includes monitoring of cod and flounder from Sørfjord and Hardangerfjord since 1996 as well as from the Oslofjord since 1997. It involves supplemental analyses to OSPAR's Joint Assessment and Monitoring Programme (JAMP) to obtain better quantitative information of the variability in time and space within the guidelines of the sampling strategy. In this paper, only data on cod are analysed.

The VIC program focuses on station 30B in the inner Oslofjord and stations 53B and 67B in the Sørfjord and Hardangerfjord system (Table A7.3 for cod). The contaminants analysed so far are CB153, pp'DDE, Cd and Zn in liver, and Hg in muscle. Both unadjusted and adjusted concentrations have been analysed. Bjerkeng and Green (1999) give a more detailed presentation of the results.

Table A7.3. Summary of sampling program for Atlantic cod carried out under the Norwegian VIC programme for monitoring years 1996, 1997 and 1998. For each sample, the table lists sampling date as month.day (mm.dd) followed by the number of specimens analysed within parentheses (n). Footnotes are found at the bottom of the page.

JAMP YEAR Calendar year		**JAMP Stations with Area names and Locations**					
		30B Inner Oslofjord			**53B Sørfjorden**		**67B Hardangerfjord**
		(Måsene)	**(Håøya)**	**(Svestad)**	**(Tyssedal)**	**(Edna)**	**(Strandebarm)**
		mm.dd (n)	*mm.dd* (n)	*mm.dd* (n)	*mm.dd* (n)	*mm.dd* (n)	*mm.dd* (n)
1996	1996				*08.07* (15)[4]	*08.14* (15)[5]	
							08.17 (25)[6]
							10.31 (10)
					12.01 (10)	*12.02* (10)[7]	
	1997	*01.15* (10)	*01.16* (10)	*01.18* (10)			
		01.22 (10)					
		02.03 (10)					
1997	1997				*09.30* (15)[8]	*10.04* (15)	*09.30* (25)[9]
	1998	*01.15* (10)	*01.16* (10)	*01.17* (10)			
		01.21 (10)					
		02.02 (10)					
1998	1998				*10.24* (15)	*10.24* (15)[10]	*10.28-11.05*(25)
	1999	*01.14* (10)					
		01.21 (10)	*01.18* (10)	*01.21* (10)			
		01.28 (10)					

* should read "JAMP stations with area names and locations"

Relations between contaminant levels and physiological parameters.

The relations between contaminant levels and physiological parameters have been analysed to see how adjusting concentrations for physiological variation might influence the ability to detect trends. The physiological covariates

4 Lipid % and organic contaminants missing for 5 specimens.

5 Lipid % and organic contaminants missing for one specimen.

6 Lipid % only in 7 specimens, organic contaminants in 5 specimens

7 Lipid % in liver was not measured, so adjusted organic contaminants can not be analysed.

8 Lipid % and organic contaminants measured for only 7 of 15 specimens, dry wt % missing for one of the others.

9 Lipid % in liver and organic contaminants was measured for only 10 of the 25 specimens

10 Data on liver composition and organic contaminants for 13 of the 15 specimens

investigated are sex, age, length, fish wet weight and liver wet weight, as well as dry-wt % and extracted lipid % in liver.[11] All covariates except age were analysed on log scale, as were the contaminant concentrations.

The physiological measures are correlated to each other to various degrees, and using them as covariates directly will not give well-defined relations. A model with more precise regression coefficients was achieved by using instead a derived set of more linearly independent parameters. The derived parameters represent different combinations of the original parameters (e.g. ratio between actual weight and mean weight for given length).

The measurements of liver composition before 1990 seem to be very uncertain, varying much more around the ordinary trend than in later data. Thus only data from 1990 or later have been used in the analysis. A few specimens from after 1989 with strongly deviating liver composition data, possibly showing analytical errors, were not used in the covariate analysis. Further, some cases with levels below the detection limit for Cd in liver or for Hg in muscle were ignored for these contaminants. Obviously, cases with missing values for some of the covariates used in the final ANCOVA model are also excluded from the covariate analysis.

The main features of the results for cod are presented below. Bjerkeng and Green (1999) give a more detailed presentation of the results.

CB153 in liver on a wet weight basis increases significantly with age, and with length for given age, but decreases with length-adjusted fish weight. For fish of given age, length and weight concentrations vary with liver composition and size. The variation is too complex to be reduced to normalising the concentration on a single parameter such as lipid content. The covariate adjustment reduces residual within-sample variance by about 30%, while the differences between samples stays about the same. The covariate adjustment thus increases the signal/noise variance ratio by about 30%, so the power for detecting trends should increase.

pp'DDE in liver shows much of the same picture as CB153, with similar sums of squares and regression coefficients for the different covariates. Correcting concentrations for covariate variation reduces residual variance within samples by one third, while the between-sample variance increases by about 6%, increasing the signal/noise variance ratio by about 40%.

Cadmium in liver shows a tendency to increase with age but decrease both with age-adjusted length and length-adjusted weight. The dominant physiological factor is a decrease with increasing dry-weight fraction in the range above 30 %, i.e., with reduced lean fraction. Both within- and between-sample variances are reduced considerably by the covariate adjustment (by 46 % and 25 % respectively). This can be seen as a strong confirmation that the variation between samples is really in part caused by physiological variations and not by differences in exposure. In particular there is a tendency that covariate adjustment reduces the between-sample (between-year) variation within each station. The signal/noise ratio increases by about 40 %, again leading to improved trend detection ability.

Mercury in muscle is mainly related to age. The concentrations tend to be higher in older fish, and in particular if the fish is long for its age. Weight given length and age is relatively unimportant. The concentrations in muscle are also related to liver conditions, presumably only as indicators of fish condition. Fish with high dry weight fraction in livers, i.e. with fat livers, tend to contain less mercury in the muscle. The covariate adjustment reduces the within-sample variance by 28 %, but the between-sample variance simultaneously increases by about 30 %. A further inspection shows that this is because the covariate adjustment tends to reduce further levels at stations with low concentrations (st. 10B, 43B, 46B, 92B and 98B) and increase levels at stations with high unadjusted levels (station 53B, 67B, 77B). The between-sample variation at each station is, however, reduced by about 20 %, so the power of detecting time trends at one station should increase. It remains an open question whether the adjusted concentrations give differences between stations that better represent exposure differences, or whether the adjustment introduces artefacts into the mercury data for comparison between stations. The artefacts could be due to incidental geographical correlations between exposure levels and physiological conditions following a different pattern than the within-sample variation.

Analysis of variance components in Norwegian VIC data.

A preliminary statistical analysis has been made based on most of the data up to and including the monitoring year 1998, mainly for individually analysed Atlantic cod (*Gadus morhua*).

---

[11] Dry-weight % and lipid % in tissue have also been analysed, but varies little, and has not been included in this analysis.

Both the raw and the covariate-adjusted contaminant data from the Norwegian VIC program for cod have been analysed by ANOVA models to estimate the variance between samples within station and year. The covariate adjustments are based on the models derived from all data since 1990, as described above.

For station 30B, location Måsene was sampled three times in January/February each of the three years. Thus, data from this location can be used to estimate between-time variance for a given location. The other two locations were sampled once each year, at approximately the same time as one of the other samples. This allows for testing for systematic variation between locations.

The analysis indicates that organic contaminants (CB153 and pp'DDE) have a strongly significant variation between samples from the same location repeated with weekly intervals in winter ($p<10^{-4}$). The variance component due to times (weekly intervals) within-year is estimated to be about 50 % of the within-sample residual variance for unadjusted values, and around 100 % for covariate adjusted values[12]. The covariate adjustment based on all data after 1989 increases the total variation between samples in the VIC data from station 30B, Måsene, without affecting residual within-sample variance. This may indicate that the covariate model is not generally applicable, but because the samples are so few it could also be just incidental. The ratio of between-sample to within-sample variance component can with 90 % confidence be said to be at least 0.2. This could mean that instead of sampling 25 fish at one occasion, it might be somewhat better to take 2 samples of 5 fish each at different sites or with a time difference, or 3 samples of 3 fish each. By distributing the 25 fish on 3 samples, the variance of the yearly mean estimate could be reduced by 65 %.

Cadmium in liver does not appear to vary significantly between times within year (p=0.12 unadjusted, p=0.3 adjusted). The variance component between times may be of some importance, with a 90 % confidence upper limit of 20–40 % of the within-sample variance for unadjusted Cd, but the upper limit is only 5% for covariate-adjusted values.

Mercury in muscle has no significant variation between times within a year for the mean of the unadjusted values (p=0.87), and only a weak indication of significant variation if we rely on the adjusted values (p=0.075). The covariate adjustment for Hg reduces the difference between years, but increases the differences between times within years. The samples are so few that this change in variance decomposition between and within years may well be just incidental. For adjusted mercury the estimated variance component between repeated samples within year may be around 15 % of the residual within-sample variance (90 % confidence upper limit), but is probably much smaller.

The other two locations at station 30B have only been sampled once per year, so for these data only the between-location effect can be tested. The data from VIC sampling times 6131, 7131 and 8132 constitute a balanced data set across the locations of stations 30B.

For metals in liver the design is complete, but for organic contaminants and mercury in muscle the data from 1998 are not available from locations Håøya and Måsene, so for these contaminants only 1996 and 1997 can be used. For cadmium in liver there is a significant mean difference between locations compared to the within-sample variance, indicating a between-location variance component that is about 30 % and 40–50 % of the residual variance for unadjusted and adjusted values, and p-values of 0.0008 and 0.0006, respectively. However, the difference between locations is not significant when compared with the interaction between years and locations, so it might be just a residual effect of irregular variations between samples. The interaction term, although not significant (p=0.1 to 0.2), is of about the same size as the variance between times within year for the analysis of location Måsene. The data have too few degrees of freedom to establish the variance components with any certainty. However, the indication of systematic location differences could be taken as a warning against using samples from different locations, even within an area of a few kilometres. For zinc there is no indication of significant differences between locations at station 30B compared to the residual within-sample variance or the interaction. For covariate-adjusted pp'DDE there is a weak indication of significant variation between locations, both compared to the residual within-catch variance and the interaction between year and location (p=0.06). The estimated location variance component is about 15 % of the residual variance. For mercury in muscle the two years of data indicate that there may be a significant difference between locations in the adjusted values. It is not present in the unadjusted values, so it may possibly just reflect an artefact introduced by the adjustment for physiological conditions, due to incidental between-sample correlation between conditions and exposure.

For station 53B, with two locations, repeated samples during the sampling year were only made in 1996. The sampling was done in different months each year, and the difference in time between the two sampling locations in 1996 is larger than the difference in season between years. It is therefore not possible to use these data by themselves to test for

---

[12] The results presented by Bjerkeng and Green (1999) differ in details from the ones given here, because full data for monitoring year 1998 were then not available.

within-year vs. between-year variance. The data can, however, be used to check for systematic differences between the two locations. Due to imbalance in lipid and organic contaminant data and missing data at the time of the analysis (see notes to Table A7.3) only metals in liver have been analysed. The results show significant variation between sampling times (year and/or season) compared to residual variance, but no significant variation between locations. Station 67B has little data, and they have not been analysed in this paper.

The main conclusion from these analyses of the Norwegian data appears to be that for the organic contaminants there appears to be a considerable short-term variation between times within year, maybe of the same size as the within-sample variation. For metals there are no clear indications, but the between-sample variance might still be as high as 20 % of the within-sample variance. These results indicate that it may be of advantage to base yearly means on a number of repeated samples from each station, rather than using only one sampling time.

The data from the Oslofjord give weak indications of possible systematic differences between locations, particularly for Cd, and quite uncertain for organic contaminants and Hg in muscle. Although the effect is uncertain, it might be wise to avoid taking subsamples from different locations if possible, and rather do repeated samplings in time.

These conclusions, based on data from enclosed fjords, may of course be quite area specific.

Swedish data

Sampling scheme

The available Swedish data, on contaminants in herring (*Clupea harengus*), are from 1997 and 1998. There is one subset of data from two rather close sites in the Kattegat, and one subset from four sites in the southern Baltic Proper. Table A7.4 shows the spatial and geographical distribution of sampling in the Swedish VIC program, and locations are shown on the map in Figure A7.2. Each sample consists of 10–20 specimens, analysed individually. There are no analyses of pooled samples in these data.

From the Baltic there are only data from 1997. Samples from three different sites (POL1, POL2, POL3) at approximately the same time (Week 35, August/September) could be used to analyse specific between-site variance. The samples from locations HAV4 and POL3, which are relatively close, is an example of data that could be used to get an indication of between-month variance (Week 35 vs. Week 46). However, one pair of samples is not sufficient for achieving a useful estimate.

The data from the Kattegat consist of two samples at the same time from two close sites (HAV6, OSP1), and also a repeated sampling, but from another year at one site (HAV6).

From the Skagerrak, samples were collected in 1998 almost simultaneously from two sites that are fairly close together (OSP3, OSP4), and only 5 days later from a somewhat more distant site (HAV7). These data could be used to test for variance between sites.

The data mainly consist of samples from different sites at approximately the same time. The only two cases of repeated sampling at the same spot in these data are from different years and cannot be used to test within-year temporal variation. The data are analysed for between-sample variance in general within region and year, not trying to distinguish between variation due to site and variation due to time. This may be reasonable if the samples are seen as random cluster sampling among schools of fish moving about irregularly.

Table A7.4. Sampling scheme in Swedish VIC program. Sampling occasions are indicated by numbers (8, 10, 12 or 20), giving the number of individuals in each sample.

		Position (deg., min., dir)		1997					1998				Year
				35		37	38	46	36		37		Week
Area	Station:	Latitude	Longitude	1	7	4	4	5	3	4	1	2	Day
Baltic:	POL1	55 22 N	14 44 E	**12**									
	POL2	55 45 N	15 03 E		**12**								
	POL3	56 00 N	15 82 E		**12**								
	HAV4	56 05 N	15 55 E					12					
Kattegat:	HAV6	57 10 N	11 50 E			**12**					12		
	OSP1	57 01 N	12 12 E			**10**							
Skagerrak	HAV7	58 33 N	10 58 E				20					**20**	
	OSP3	57 52 N	9 56 E							**8**			
	OSP4	58 00 N	9 56 E						**8**				

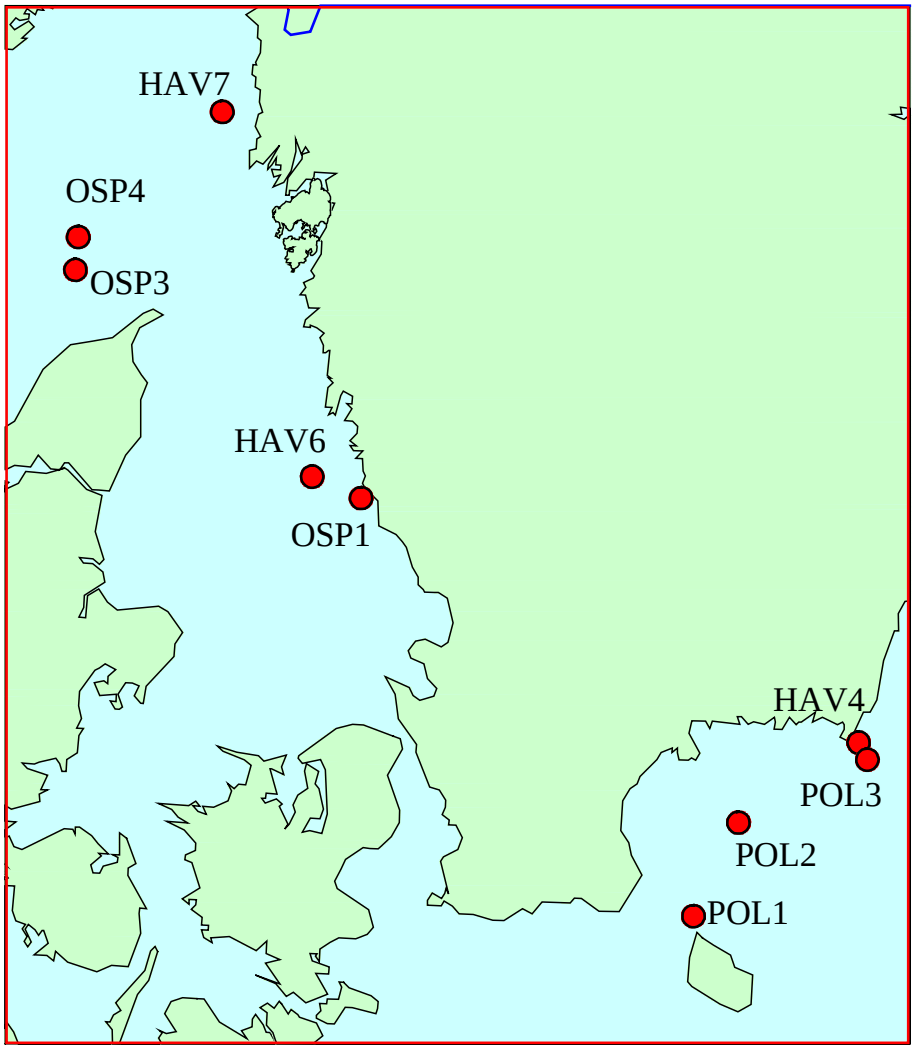


Figure A7.2. Map showing location of the Swedish sampling locations.

Physiological data

Physiological covariates are length, weight, age, and lipids in muscle for all specimens analysed except two of the specimens from OSP1. Dry weight % in liver and muscle and liver weight are not included in the analysed data set for the samples from OSP1, OSP3 and OSP4.

Age is not useful as a covariate in these data, since most samples are from one age group only - and the variation with age is confounded with variation between areas and sampling years. Fish in the Baltic, mostly caught in 1998, are both older and smaller in size than fish from the Kattegat and Skagerrak.

The relation between physiological measures has been analysed by log-linear regression.

There is a clear log-linear regression between length and weight:

$$\hat{W} = W_0 \cdot \left(L / L_0 \right)^{\alpha} \quad \alpha = 2.89_{\pm 0.2} \tag{20}$$

Within-sample regression gives significant differences in W_0 between samples. This is mainly a difference between areas; samples from the Kattegat and Skagerrak show a better log-linear fit than the Baltic samples, and generally higher weight/length ratio. The differences are expressed by an adjusted weight, or weight deviation:

$$\omega = \frac{W}{\hat{W}} \tag{21}$$

Liver weight is clearly related to size. It turns out to be sufficient to normalise it on total fish weight, the expected geometric mean liver weight given fish weight is

$$W_{liver} = W_{liver,0} \left(\frac{W}{W_0} \right)^{\beta} \quad \text{where } \beta = 0.87_{\pm 0.16} \tag{22}$$

Muscle conditions are described by total dry-weight % and lipid %. The lipid fraction of muscle is clearly positively correlated to dry weight fraction, and both are positively related to weight. Muscle with very low fat levels have about 1.7 % dry-weight, with increasing fat % up to about 10, the dry weight % increases to about 35. The lipid fraction (lf_m) in muscle is related to the weight by a relation

$$\hat{lf}_m = lf_{m,0} \cdot \left(W / W_0 \right)^{1.6} \tag{23}$$

and the deviation $\lambda = lf_m / \hat{lf}_m$ is used to express variation in lipid content largely independent of length or weight. The dry-weight fraction can be normalised on length, weight and lipid weight, the result is

$$\hat{D}_m = D_{m,0} \cdot \left(\frac{L}{L_0} \right)^{0.86} \omega^{0.45} \lambda^{0.13} \tag{24}$$

and the deviation $\delta_m = D_m / \hat{D}_m$ can be used to express deviations in dry weight % relative to expected value given length, weight and lipid content.

Preliminary covariate analysis

First an ANCOVA analysis was performed on the selected contaminants (all the metals, pp'DDE, sDDT, sPCB, γ-HCH, HCB and CBΣ4 with sample (cells in Table A7.4 with data)) as random factor, and all continuous physiological measurements as covariates. Both contaminants and covariates are analysed on a log scale.

By comparing residual variance of the full covariate models with simple one-way ANOVA models that use sample as factor and without any covariates, it was found that all the contaminants except lead have strongly significant covariate within-sample regression.

For all but one contaminant, the between-sample variance[13] was reduced when individual levels were corrected for this within-sample covariate regression, the reduction varied from 20 % to 90 %. When the between-sample variance is strongly significant compared to the within-sample variance, this may indicate that covariate relations are of a general nature, describing both the variations between individuals in a sub-population, and the variations between sub-populations with varying distributions of physiological adjustments. The exception was zinc in liver, where differences between samples increased with covariate correction, but in this case the covariate regression has overwhelming significance ($p<10^{-18}$), and inspection revealed a clear within-sample log-linear negative regression on dry weight fraction. The increased between-sample variance is due to area differences, with herring from the Baltic having both a lower dry-weight fraction and a lower zinc content.

Statistical analyses for each contaminant

For each selected contaminant the variations are analysed by a log-linear GLM model as described in Chapter 0, equation (5). The variation between areas or years is captured by the AREA*YEAR factor, with levels *Baltic*1997, Kattegat*1997, Kattegat*1998, Skagerrak*1997* and *Skagerrak*1998*. The within-area_and-year variance across samples is captured by a random SAMPLE effect, nested within the Area*Year factor. Through this design, the isolated samples from Kattegat in 1998 and Skagerrak in 1997 are included in the residual variance estimate, but ignored for the estimation of the other variance components. The covariate regression models are different for different contaminants, as described below, and results are presented both for adjusted and unadjusted levels of each contaminant where covariate regression is significant. Confidence limits on the ratio of sample variance over specimen variance is determined as described in Chapter 0. Results are presented below for each contaminant, and discussion focuses on the 90 % lower confidence limit, so that conclusions are conservative with respect to what is achieved by multiple sampling.

PCBs in muscle

The PCBs for the Swedish data are reported in ppm on a lipid basis.

Of the PCBs, only the components used in calculation of CBΣ7 are included in the data. The components CB28, CB52 and CB180, are given as below detection limit for a considerable number of records, and CB28 is missing from around 30 records. Because of this, the sum of only 4 main components is used in the statistical analysis instead of CBΣ7. This variable, denoted here CBΣ4 and calculated as CB101 + CB118 + CB138 + CB153, makes up from 85 % to 89 % of CBΣ7 in almost all cases where all components are given as full analysis values.

A preliminary ANCOVA analysis on only the data that have complete physiological covariates, and including sample number (time and location) as random factor, indicate that total length and weight are the most important covariates. There is no correlation with lipid content of muscle, and this confirms that the normalisation to lipid basis serves to minimise variability for these data.

A new analysis with only length and weight as covariates on all samples shows that they have coefficients corresponding closely to the length/weight trend, which means that the effect can be expressed simply be using the length-adjusted weight measure ω as the only covariate. The residual variance is brought down from 0.133 to 0.082 by the covariate regression. The regression coefficient is $b = -2.75 \pm 0.3$, and none of the samples show significantly different covariate regression. The within-sample regression also reduces both the between-sample and the YEAR*AREA mean square by about 70 %. Both effects are still significant, the variations between year*area is also significant compared to the variation between samples.

For covariate-adjusted contaminants, the sample variance component within year*area is estimated at 0.03. The ratio between sample and specimen variance components is estimated at about 0.35, with a 90 % confidence lower limit of 0.15. The estimated ratio indicates that for samples of more than 5 fish, the variation between geometric means will mainly be a sample effect and not due to variation between specimens within each sample. If this is right, very little is achieved by increasing the number of fish per sample beyond 10. Sampling 5 fish 3 times from an area within a year would reduce the sampling-dependent between-year variance by about 65 % compared to one sample of 15 fish, and even sampling only 2 fish 3 times would reduce variance by 20-30 %.

[13] Including mean effect of specimen variance, cf. page 74.

Even if the variance component ratio is as low as the 90 % confidence lower limit, it turns out that 4 fish sampled 3 times still reduce between-year variance by 45 % compared to 12 fish sampled once, and 3 fish sampled 3 times will reduce between-year variance by 15 %. For unadjusted contaminants the advantage of repeated sampling is even larger.

The estimates are rather imprecise due to the few degrees of freedom, and the results should only be taken as a crude indication. Individual comparisons (LSD post-hoc test) on adjusted means show that the sample variance within year*area is only due to the 1998 sample from HAV7 having significantly a lower geometric mean than the data from OSP3, and from HAV7 in 1997. Except for that, there are no significant differences between samples within an area. Thus, results rest upon a single deviating sample, and it could be that with more data, the result would be different. The deviating sample, may, however, indicate that multiple sampling coupled to a robust estimation could improve precision.

pp'DDE in muscle, expressed on a lipid basis

Covariate analysis on data with the physiological covariates that are present for all samples (length, weight and lipids in muscle) shows that log(pp'DDE) is significantly related to all of them ($p<0.005$). The covariate within-sample regression is captured well by using only length-adjusted weight ω and weight-adjusted lipid content in muscle. When concentrations are corrected for the covariates, the within-sample variance is reduced from 0.135 (df=125) to 0.086 (df=122), and the between-sample mean square is reduced from 1.05 to 0.65 (df=6).

The ratio between sample and specimen variance components for adjusted contaminants is estimated to be about 0.8, with a 90 % lower confidence limit of 0.4. This is a high ratio, with the lower limit indicating that 2 samples of 2 fish each might be about as good as one sample of 15 fish. However, further inspection shows that the between-sample variance is due to only one sample from the Baltic, POL3, with much smaller mean than the three other samples from the Baltic. If this sample is excluded, there is no significant variation between samples beyond the expected residual effect from within-sample variance. The conclusion is that for pp'DDE in herring, there is no strong indication that repeated sampling is generally important. However, a sampling design with 3 sub-samples per area and year coupled to a robust-statistics mean estimation would protect against deviations such as that displayed by the POL3 sample.

Lead in liver

This contaminant shows no significant variation with the physiological covariates. The residual variance within sample is 0.13 (df=104), with no signs of heterogeneous variance neither across samples in general nor across areas. There is a significant difference in mean values between samples within area ($p=10^{-7}$), and a significant difference between areas, when tested against the variation between samples within area. The ratio of between-sample to within-sample variance components is estimated at 0.65, with a 90 % confidence lower limit of 0.3. Multiple comparisons by Scheffes method show that 2 out of 6 pairwise comparisons between samples within area and year (between the Baltic samples) are significantly different with $p<0.02$. This may indicate a generally significant variation between samples. Based on the lower limit ratio, one may state with 90 % confidence that 3 samples of 2 fish each is as good as one sample of 15 fish.

Nickel in liver

The nickel levels are related to a set of covariates that explain about 30 % of within-sample variance. The levels tend to increase with age, and decrease with increasing dry weight % (which largely expresses variations in fat %) and with the deviations of liver weight and fish weight from expected values given length. None of the covariates can be selected as the dominant one. Adjusting for the within-sample covariate regression reduces the between-sample variance within year and area by 25 %. The variance ratio σ_S^2/σ_I^2 is estimated at about 1.3, with a 90 % confidence lower limit of 0.55, about the same for unadjusted and adjusted levels.

Cadmium in liver

Cadmium increases somewhat with age for fish of normal weight/length ratio, but decreases with increased weight deviation ω. The covariate regression reduces within-sample variance and between sample variance about equally, with 20–30 %. For the adjusted levels 6 out of 10 pairwise comparisons between samples within year and area are significant with $p<0.02$ by the conservative Scheffes test. This indicates clearly a significant sample variation as a general feature of the data. The sample/specimen variance component ratio of adjusted levels is estimated at 2.5, with a 90 % confidence lower limit of 1.35. For the unadjusted values, the ratios are about 10 % higher.

Mercury in muscle

Mercury in muscle decreases significantly with deviation from normal of fish weight and lipid content in muscle. The within-sample covariate adjustment reduces within-sample variance by 15 %, and the between-sample variance by merely 5 %. The between-sample variation is clearly most important for sample sizes around 10 specimens. Sample/specimen variance component ratio of adjusted levels is estimated at 1.75, with a 90 % lower confidence limit of 0.75. For the unadjusted values, the ratios are about 10 % lower.

Preliminary assessment of possible effects of changing within year*station sampling design

Assume a sampling scheme for one station where a total of *K* fish are collected each year, distributed as equally as possible among *C* samples, and analysed individually. The samples could be distributed between different locations in the station area and/or over different points in time within the designated sampling period (i.e. outside the normal spawning season when physiological changes are stable). If there are systematic differences between locations, as indicated by the analysis above, the samples should represent different sampling times for a single selected location.

The variance components involved are:

σ_c^2 = variation between samples (a function of small-scale variance in time and/or space)

σ_k^2 = variation between fish within samples, including analysis error.

The linear statistical model for the contaminant level in the individual fish (*k*) for a certain year (*y*) is

$$x_{yck} = \mu_y + \varepsilon_{c(y)} + \varepsilon_{(cy)k} \tag{25}$$

The variance of each value around the "true" expectation for the year

$$V\left(x_{yck}\right) = \sigma_c^2 + \sigma_k^2 \tag{26}$$

The yearly mean estimate based on these values has variance around the true expectation

$$V\left(x_{y\bullet\bullet}\right) = \frac{\sigma_c^2}{C} + \frac{\sigma_k^2}{K} \tag{27}$$

If *C* is increased by δC, the variance of the yearly mean will be roughly the same if *K* is simultaneously reduced by δK equal approximately to

$$\delta K = \frac{R \cdot K^2 \cdot \delta C}{R \cdot K \cdot \delta C + C\left(C + \delta C\right)} \text{ where } R = \frac{\sigma_c^2}{\sigma_k^2} \tag{28}$$

assuming that the reduced number of fish are also distributed as equally as possible on the *C* samples.

Table A7.5 summarises the results for a few sets of alternative scenarios, with the ratio (*R*) between variance components defined in eq. 28 ranging from 0.05 to 0.5 (column I). This range is somewhat on the conservative side of what the results of analyses in the previous sections indicate.

Table A7.5 shows for each *R* a series of values of *C* from 1 to 3 or 4 subsamples (column II). Column III gives for each *C* the *K* value that gives approximately the same variance as *C*=1, *K*=25. Column IV shows the corresponding variances, scaled relative to within-sample variance. Finally, column V shows the variances that would be achieved if *C* were increased as indicated by column II, but the total number of fish *K* was kept at 25, i.e., with the same analysis effort, but a different sampling procedure.

Table A7.5. Comparison between alternative sampling scenarios giving approximately unchanged variance of yearly means..

(I)	(II)	(III)	(IV)	(V)
R = assumed ratio of 'Between sample' to 'Within sample' variance	**C = number of samples per year**	**K = total number of fish**	**Variance of yearly means relative to variance between fish within sample** $V(x_{y\bullet\bullet})/\sigma_k^2$ **For K = value in column III**	**At fixed K = 25**
0.05	1	25	0.090	0.090
	2	16	0.088	0.065
	3	14	0.088	0.057
	4	13	0.089	0.053
0.1	1	25	0.140	0.140
	2	12	0.133	0.090
	3	10	0.133	0.073
	4	9	0.136	0.065
0.2	1	25	0.240	0.240
	2	8	0.225	0.140
	3	6	0.233	0.107
0.5	1	25	0.540	0.540
	2	4	0.500	0.290
	3	3	0.500	0.207

The table shows that even for a very low "between-sample"/"within-sample" variance ratio (R = 0.05), unchanged precision may be obtained with the total number of fish analysed individually cut by 50 % if they are distributed equally between 3 or 4 different samples instead of coming from a single sample. Unchanged precision means an unchanged sampling-dependent variance component in the yearly means. If the variance between samples is 50 % of the within-sample variance, the same precision would be achieved by analysing three fish altogether, each caught at a different time, as by sampling 25 fish at a single time. This variance component ratio is still lower than many of the estimates in the preceding chapters,

However, the estimates of between-sample variances are still quite uncertain, and one should not rely too much on them. Reducing the total number of fish that are analysed and increasing number of sub-samples with the aim of keeping variance unchanged must be based on reasonably certain lower limits for the ratio of "between-sample" to "within-sample" variance. Further analysis is required in order to make recommendations.

A more cautious approach would be to keep unchanged the total number of fish analysed, but draw them from an increased number of sampling times, and thus achieve an improved precision, rather than aiming at cutting the costs to keep a certain precision. This is illustrated in the rightmost column of Table A7.5. For R = 0.1 one can achieve a doubling of the precision, with a corresponding increase in trend detection power, by sampling 25 fish altogether, separated into 8 or 9 fish at each of three different occasions, rather than catching all 25 fish at once.

Such a change would represent a smaller or larger improvement in trend-detection power depending on the ratio of between-sample over within-sample variance, but should never lead to increased yearly mean variance, as long there are no systematic seasonal variation across years during the sampling period. And also provided that the sampling period is outside the spawning time and when the fish are in a stable physiological state OSPAR (1997).

The results differ for different contaminants, species and areas. In some cases, the between-sample variance appears to be negligible and the trend-detection power then depends mainly on total number of fish, independently of how the catches are distributed in time. A way to preserve trend detection power and still achieve a reduction of total cost might

be to catch the same number of fish as today, but distributed over a number of times, and then adjust the analysis effort for different contaminants according to their variance patterns. For contaminants with a high between-sample/within-sample variance ratio, one could analyse only smaller number of fish from the total sample, drawn randomly or by stratified sampling.