3 Statistical Considerations

The aim of this chapter is to review the statistical requisites that must be satisfied if a monitoring programme is to meet the objectives set out in Chapter 1. However, objectives at that level will not help to answer questions such as: *How many samples do we need to take? For how long a period do we need to continue monitoring? How frequent should we sample?* Furthermore, we need to specify the magnitude of the changes or differences we have to detect. The risks of reaching the wrong conclusions (e.g. to conclude that there is a trend when there is not or to miss a true trend) have also to be considered.

3.1 Quantitative objectives

Describing and carefully defining the objectives are the most crucial step in planning and organizing monitoring activities. It includes the choice of sampling matrices and strict definitions of sampling units and a description of what they represent in time and space. This description is a prerequisite for an appropriate interpretation of the results. However, in order to properly estimate, for example, the number of samples per sampling occasion, length of the time-series, sampling frequency etc, required for the investigation, quantitative objectives have to be defined. Quantitative objectives imply that the required sensitivity of the programme is stated, i.e. that the smallest change for temporal studies or smallest difference between areas for geographical studies is specified together with the required statistical power to detect such a difference at a specified significance level.

A quantified objective for temporal studies could thus, for example, be stated as follows:

To detect a 50 % decrease within a time period of 10 years with a statistical power of 80 % at a significance level of 5 %. (A 50 % decrease within a time period of 10 years corresponds to an annual decrease of about 7 %).

And for spatial studies, for example as follows:

To detect differences of a factor 2 between sites with a power of 80 % at a significance level of 5 %.

A significance level of 5% means that we are prepared to accept a risk of 5% to conclude from our data that there is a trend or difference when there actually is not. Similarly, a power of 80% means that we accept a risk of 20% to conclude that there is no trend or difference when it really is one. Statistical power and methods to estimate power are discussed in detail in Cohen (1988).

It had to be stressed, however, that statistically significant trends do not guarantee that detected temporal trends are a result of a causal relation between concentration and time. If the samples are biased, not comparable over time or if relevant confounding co-variants are not accounted for, “false-trends” may well occur.

Furthermore, in order to calculate, for example, the number of samples and the sampling frequency required to fulfil those objectives, an estimate of the sample variance is needed. Expected variance estimates could, perhaps, be extracted from similar ongoing monitoring programmes or, what is more reliable, be assessed from a pilot project using the same sampling strategy, sampling matrices etc as the currently planned monitoring programme. In order to optimise the programme from a cost-benefit point of view, all costs, for example, for sampling, sample preparation and chemical analysis must be specified.

3.2 Representatives

It is essential that the suggested matrices are thoroughly described concerning what they represent in relation to contaminant load or exposure. Apart from factors like availability, sampling costs etc information on, for example, concentration factors, bioaccumulation rates, metabolic capacity, and excretion rates would be useful. Various tissues within the same species vary considerably with respect to the above-mentioned factors i.e. they may represent totally different ranges of time and they may react to changes in the environment very differently.

Even though these questions are not purely interesting from a statistical point of view they will constitute invaluable pieces in the building of a modelling framework to enable an integrated assessment of contaminant load and exposure from various matrices.

Using mammals or species with a more or less developed capacity to degrade POPs may lead to spurious results. Elevated levels of one POP may trigger and enhance the metabolic capacity to degrade other POPs. This may cause a problem, for example, to evaluate spatial differences in POP exposure from human milk (Weiss *et al*., 2003).

Monitoring contaminants on the global scale will inevitable raise question such as: *How many sampling sites do we need to appropriately represent a region?* Any firm advice from a statistical point of view needs estimates on spatial heterogeneity. For spatial studies the objectives have to be clearly specified (e.g. spatial trends, differences between regions etc) and made quantitative. A variogram (Fig 3.1) may be used to describe the spatial correlation structure (Cressie, 1993; Davis, 1986). A sampling site does not represent a point outside the radius where the correlation with other stations ceases to exist better than any other sampling site outside this radius and thus hardly represent an area larger than an area confined by the perimeter of this radius.

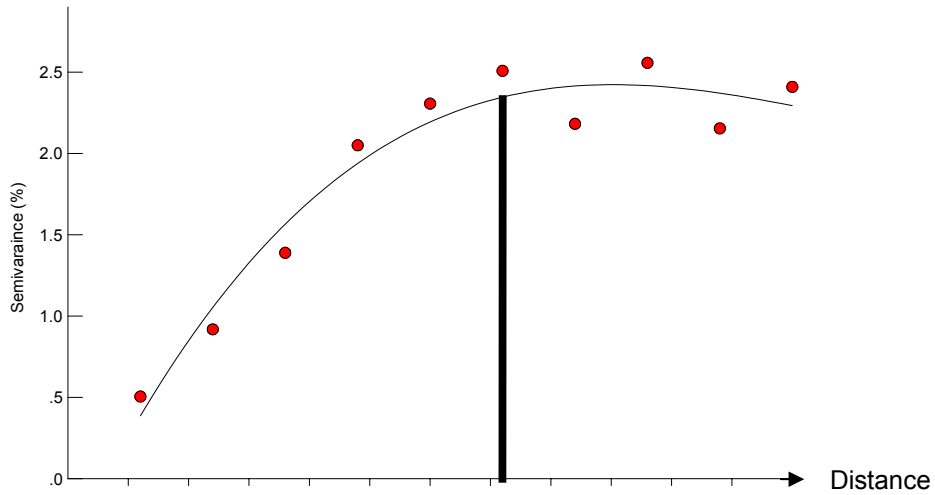

Figure 3.1. Showing an example of a variogram where the differences of concentrations between neighboring sites increases with distance up to a certain distance.

From a temporal trend perspective, a focus on well-defined strata of the monitored population/region will decrease the variance and improve the likelihood to detect changes over time.

When time-series are available from several sites within a region, statements about the presence or absence of trends in the same direction within the region are interesting. The homogeneity of trends can be

checked using methods described in most standard text books in statistics (e.g. Dixon & Massey, 1969; Snedecor & Cochran, 1968). Van Belle and Hughes (1984) propose a method for testing homogeneity among trends derived from the non-parametric Mann-Kendall trend test. Also methods from the fast growing field of meta-analysis can be of value when interpreting trends from several sites within and among regions (for example Hunter & Smith, 1990).

3.3 Sources of variation

There are numerous factors that affect measured concentration in environmental samples other than those of anthropogenic origin. For monitoring programmes that are designed to assess the effects of measures taken to reduce discharges of contaminants from industrial activities or control by means of pesticides, these factors can be considered as confounding factors. Avoiding or adjusting for confounders can improve statistical power in monitoring programmes considerably (Grimås *et al*., 1985; Nicholson *et al*., 1991b; Bignert, 2002).

Seasonal variation for several POPs (e.g. PCB, PCDD/PCDF, DDTs and HCB) has been demonstrated. The reasons could be both a seasonal variation in the discharge pattern from the sources and be due to, for example, physiological factors. If the main objective is to monitor the mean change in pollution load rather than to investigate the seasonal pattern in the discharges, sampling should be restricted to one season (the most favourable season from a minimum random variation point of view) in order to gain statistical power. The same arguments could be used if a diurnal pattern is discernible for fast changing matrices such as air.

Fat content and composition in human milk changes dramatically during the first weeks after birth, which leads to variation also in analysed POPs (e.g. Weiss *et al*., 2003). In order to reduce random variation, sampling should preferably be carried out during a well defined period three weeks after birth (Also the fat content varies considerably depending on whether sampling is carried out in the beginning or at the end of the feeding session).

Other known or suspected confounding factors for which control is possible at sampling should be specified in the monitoring guidelines.

The use of narrow sampling unit definition implies that a smaller part of the studied population is represented. Often, this leads to unfounded assumptions of similar trends, for example, for both sexes or for various age classes. To improve representativity, if economy permits, stratified sampling should be applied rather than allowing for a wider definition of the sampling unit. General aspects of sampling design, applicable also for monitoring, are discussed, for example, by Underwood (1993, 1994, 1996).

The precision of chemical analysis is generally believed to constitute only a minor part of the total variance in monitoring time-series of environmental data where sample variation is expected to be large, much larger compared to laboratory precision. That is true if the same accredited laboratory is used through the whole series. However, if, from year to year, different laboratories carry out the analysis, it could seriously decrease or disable the possibility to evaluate time-series of, for example, POPs. The same is true if the same laboratory changes its methodology and, for example, co-elutions are resolved leading to a decrease in estimated concentrations unless measures are taken to compensate for them. If detection limits are improved, i.e. analytes are now found where they were not detected before, that may lead to similar problems depending on how results below the detection limit (‘less-than-concentrations’) are treated.

Provided that individual samples are taken and that appropriate confounding variables are registered or measured at the chemical analysis, the concentrations may be adjusted for varying covariates by means of,

for example, ANCOVA (Analysis of Covariance). This may improve the power to detect changes over time or differences among sites considerably (Bignert, 2002). Furthermore, the detection and possible elimination of erroneous extreme values would also noticeably improve the power (Barnett and Lewis, 1994; Nicholson *et al*., 1998; Bignert, 2002).

For temporal trends, the between-year variation may be expressed as the standard deviation of the residuals from a regression line on a log-scale or as a Coefficient of Variation (CV, %). The Coefficient of Variation found in time-series of contaminants in biological samples, including human milk, will most probably be over 35%, even if the between-year variation can be considered extremely low.

3.4 Length of time-series

It can be shown that several well-established monitoring programmes have surprisingly low power to detect temporal changes of significant importance (Nicholson and Fryer, 1991; Bignert *et al*., 2004). It is naïve to expect monitoring time-series of POPs to reveal changes with any confidence within a sampling period of five years unless the changes are very large. More likely, we would expect a period of at least 10-15 years to detect changes of moderate size (5 % /year).

A study would need at least 4-5 years of monitoring to give reliable estimates of random within- and between-years variation and other components of variance. This information would be invaluable for the improvement and fine-tuning of the on-going monitoring activity. It should be stressed that even for spatial studies a few years of sampling is not enough but can lead to spurious results (Bignert *et al*., 1994).

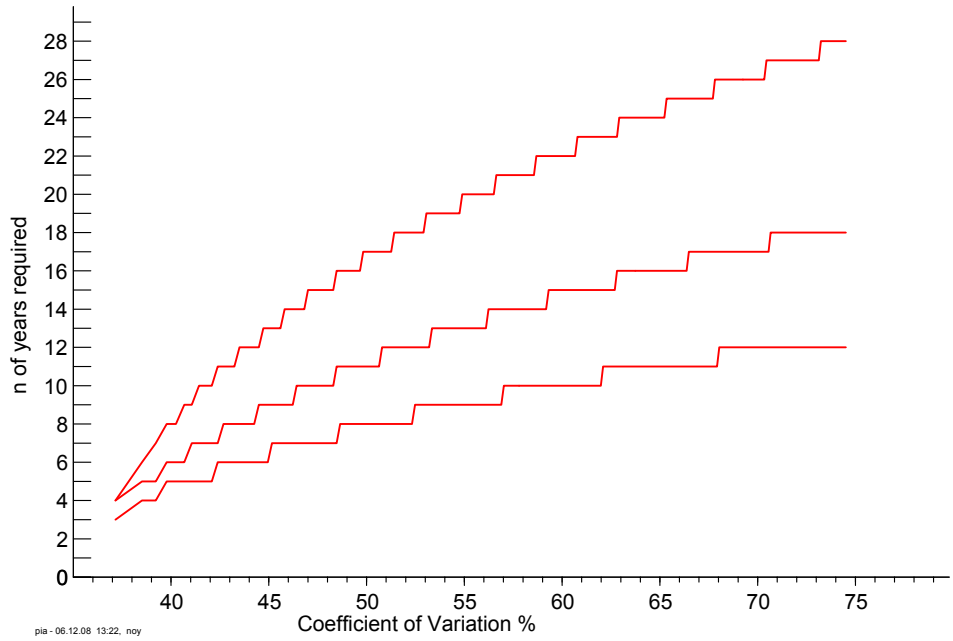

Figure 3.2. Number of years required to detect a 5 (left/above), 10 and 20% change per year respectively, at a power of 80% at a significant level of 5% applying a simple two-sided regression analysis for various magnitudes of between-years variation expressed as Coefficient of Variance (%) assuming single annual mean concentrations (or one pooled sample per year).

3.5 Number of samples needed

Larger samples provide more precise and reliable estimates of mean concentrations and variance. However, the contributions from additional samples depend to a very high degree on the sampling strategy.

To estimate the number of samples needed in an appropriate way for a certain situation, quantitative objectives must be defined and information on expected variance must be available (see above). The standard formulae for calculating the number of samples needed assume independent observations. In many typical monitoring situations this assumption is not altogether true. On a large scale, the weather situation one particular year at a sampling station may affect all the individual samples in the same direction.

Small-scale variation in time and space may not be covered by the sampling scheme which leads to an underestimated variance and increased between-year variation, for example, Bjerkeng (2000) showed that by sampling at three occasions during the sampling period instead of one and using the same number of samples or less, the yearly mean variance estimate could be reduced by up to 65%. Furthermore, stratified sampling and the choice between individual and pooled samples will affect the estimates of the required number of samples. Without the information mentioned above, no optimal figures on the required number of samples can be calculated.

Using pooled samples of several specimens will decrease the number of chemical analyses required to estimate a reliable mean concentrations compared to one or a few individual samples, since a larger proportion of the total population is represented. Disadvantages with pooled samples are that extreme values from single specimens may influence the concentration of the pool without being revealed, and that the possibility to adjust for confounding variables or correlate with biological effects disappears. Information on individual variance within a year has also a value in itself. An increased variance is often the first sign of elevated concentrations. In particular in the first stage of establishing a new sampling site, individual samples could help to reveal possible sources of variation. A more detailed discussion of advantages and disadvantages with individual versus pooled samples is given by Bignert *et al.* (1993).

3.6 Expected trends

Concentrations of pesticides can be expected to decrease relatively fast in environmental samples directly after a ban or other measures taken to reduce discharges, often with a magnitude of about 10 – 20 % per year. Similar trends have been measured in biota from terrestrial, freshwater and marine environments (Bignert *et al.,* 1998 a, b, c). That is, if a source disappears, the bio-available amount of hazardous persistent substances decreases much faster than that which may be expected from their estimated half-times. From a statistical point of view, this will enhance the possibilities to detect changes due to measures taken to reduce discharges, at least for persistent pesticides. For POPs such as PCB or others that are found in many different products in the techno-sphere the decrease would probably be lower, about 5-10 % per year. This means that the minimum trend possible to detect with a reasonable power (80%) should be smaller than 20% and preferably smaller than 10%. Assuming an appropriate sampling design, annual sampling for a period of ten years would probably be enough to detect trends in human milk/blood of 10% per year at a statistical power of 80% for pesticides and other POP's. Temporal trend analyses for air samples will preferably be treated with other methods (Chapter 4.1) that will affect the power calculation.

3.7 Expected sensitivity to detect trends

For a proper estimate of sensitivity, a pilot study should be carried out. It depends very much on the sampling strategy, choice of matrix, how well sampling follows the guidelines, whether the same laboratory is undertaking the analyses from year to year or not etc. The sensitivity will also differ between various POPs. For human milk the sensitivity could be expected to be, around 5% per year, assuming relatively large pooled (consisting of 25 individual samples) or individual samples of the same number following the guidelines in Section 4.4.

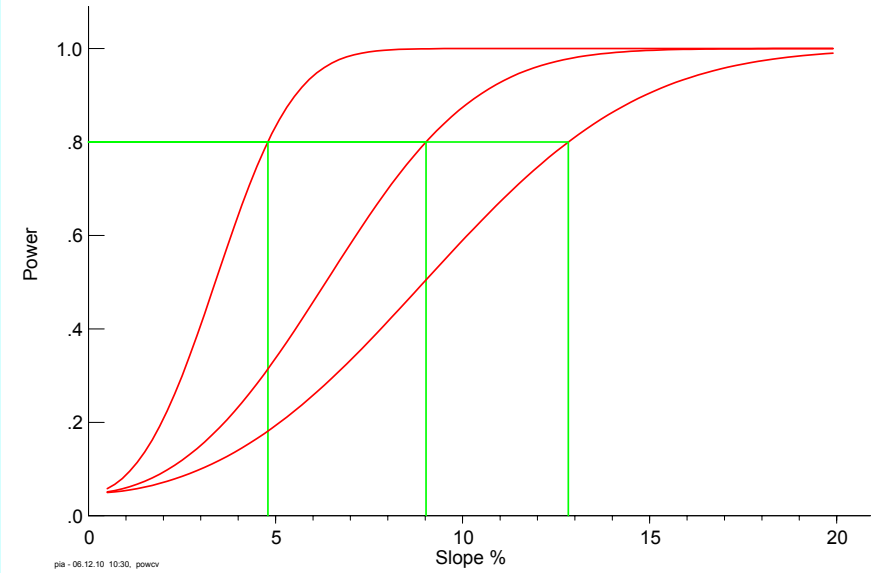

Figure 3.3. Power as a function of the minimum annual change possible to detect, after a sampling period of 12 years at a significant level of 5% applying a simple two-sided regression analysis for various magnitudes of between-years variation, expressed as Coefficient of Variance from left: 20, 40 and 60% respectively, assuming single annual mean concentrations (or one pooled sample per year).

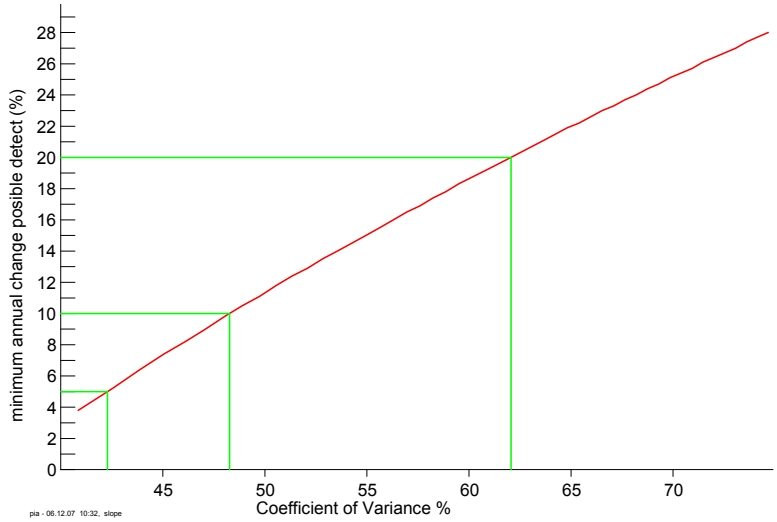

Figure 3.4. Minimum annual change possible to detect at a power of 80%, a significant level of 5% applying a simple two-sided regression analysis for various magnitudes of between-years variation expressed as Coefficient of Variance (%)assuming single annual mean concentrations (or one pooled sample per year).

3.8 Sampling frequency for temporal trend studies

To determine an appropriate sampling frequency, the required temporal resolution has to be specified. To monitor certain events or incidents with a short time lapse, sampling may have to be carried out very often during certain periods. Considering, for example the half time for POPs in biological tissues, analytical cost etc., sampling once or, at most, twice per year is generally appropriate for monitoring of contaminants in biological samples. Sampling on several occasions during the sampling period to cover small scale temporal variation will, however, improve the mean estimate, as has been pointed out above). The examples above refer to sampling once a year. Obviously the statistical power of a trend-test is seriously reduced when sampling with a lower frequency.

If the length of a time-series is fixed, the power for various slopes at a certain between-year variation can be estimated. Figure 3.3 shows the relation between power and slope (e.g. the change in time-series of POPs measured in biota samples), estimated at sampling every, every-second, third and fourth year, respectively, at a standard deviation (between-year variation) along a regression line of 0.20 on a log-scale, corresponding to a Coefficient of Variation of 20-25%. If the desired sensitivity of the monitoring programme is to be able to detect an annual change of at least 5% per year within a time period of 12 years, the power is almost 80% for sampling each year at this standard deviation (Figure 3.3). For sampling every second, third or fourth year the corresponding power is only approximately 35, 17 and 10%, respectively.

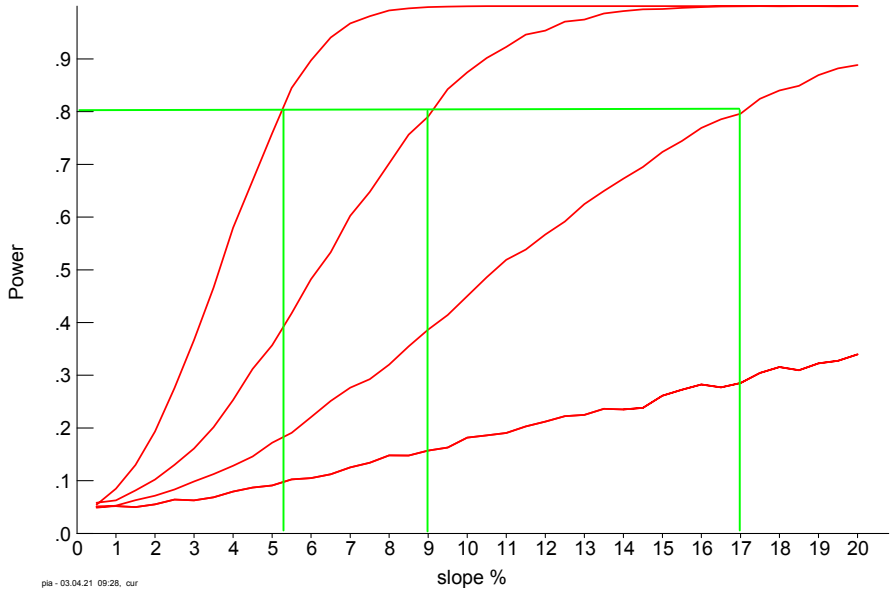

Figure 3.5 Power as a function of slope (annual change in %) at log-linear regression analysis (two-sided, α=0.05) for a sampling period of 12 years at a residual standard deviation on a log-scale of 0.20, assuming normally distributed residuals. The graphs, from left to right, represent sampling every, every-second, third and fourth year, respectively and is based on Monte Carlo simulations at 10,000 runs.

3.9 Evaluation of results

Geographic information system (GIS) and modelling will inevitably play a great role in the interpretation and evaluation of the results for spatial distribution and exposure etc. It has to be stressed, however, that the reliability of such an evaluation will depend on the validation with real data from the environment and will become poor if the number of samples is too low. For time-series analyses a robust method proposed by Nicholson *et al.* (1995) has been used during recent years for several assessments of monitoring data within OSPAR, HELCOM and AMAP. This method supplemented with a non-parametric trend test and an efficient outlier test could form a basic package to evaluate temporal trends. Parametric tests are more powerful compared to non-parametric ones if the assumptions behind the tests are fulfilled (e.g. the residuals around the regression line is normally distributed). If however, this is not the case (e.g. if the presence of outlier violates the assumption of normally distributed residuals) the non-parametric tests become more powerful compared to the parametric ones.

3.10 Examples of statistical treatment and graphical presentation

One of the main purposes of the monitoring programme is to detect trends. Examples of methods to detect trends could be simple log-linear regression analyses. The slope of the line describes the yearly change in percent. A slope of 5 % implies that the concentration is halved in 14 years whereas 10 % corresponds to a similar reduction in 7 years and 2 % in 35 years.

The regression analysis presupposes, among other things, that the regression line provides a good description of the trend. The leverage effect of points in the end of the line is also a well-known fact. An exaggerated slope, caused 'by chance' by a single or a few points in the end of the line, increases the risk of a false significant result when no real trend exists. A non-parametric alternative to the regression analysis is the Mann-Kendall trend test (Gilbert, 1987, Helsel and Hirsch, 1995, Swertz, 1995). This test has generally lower power than the regression analysis and does not take differences in magnitude of the concentrations into account, it only counts the number of consecutive years where the concentration

increases or decreases compared with the year before. If the regression analysis yields a significant result but not the Mann-Kendall test, the explanation could be either that the latter test has lower power or that the influence of endpoints in the time-series has become unwarrantably great on the slope. Hence, the eights line reports Kendall's 'tau' (see Table 3.2), and the corresponding p-value. The Kendall's 'tau' ranges from 0 to 1 like the traditional correlation coefficient 'r' but will generally be lower. 'Strong' linear correlations of 0.9 or above correspond to tau-values of about 0.7 or above (Helsel and Hirsch, 1995). This test has been recommended for use in water quality monitoring programmes with annual samples in an evaluation comparing several other trend tests (Loftis *et al.*,1989).

In order to describe non-linear trend components in the development over time some kind of smoothed line could be applied. The smoother used in the example (Fig 3.3) is a simple 3-point running mean smoother fitted to the annual geometric mean values. In cases where the regression line is badly fitted the smoothed line may offer a more appropriate description. The significance of this line is tested by means of an ANOVA (Analysis of Variance) where the variance explained by the smoother and by the regression line is compared with the total variance. This procedure is used at assessments at ICES and is described by Nicholson *et al.*, 1995, see the smoothed line in the HCB-plot in the example (Fig 3.3).

Observations too far from the regression line considering what could be expected from the residual variance around the line is subjected to special concern. These deviations may be caused by an atypical occurrence of something in the physical environment, a changed pollution load or errors in the sampling or analytical procedure. The procedure to detect suspected outliers in this example is described by Hoaglin and Welsch (1978). It makes use of the *leverage coefficients* and the *standardised residuals*. The standardised residuals are tested against a $t_{.05}$ distribution with $n-2$ degrees of freedom. When calculating the ith standardised residual the current observation is left out implying that the ith observation does not influence the slope nor the variance around the regression line.

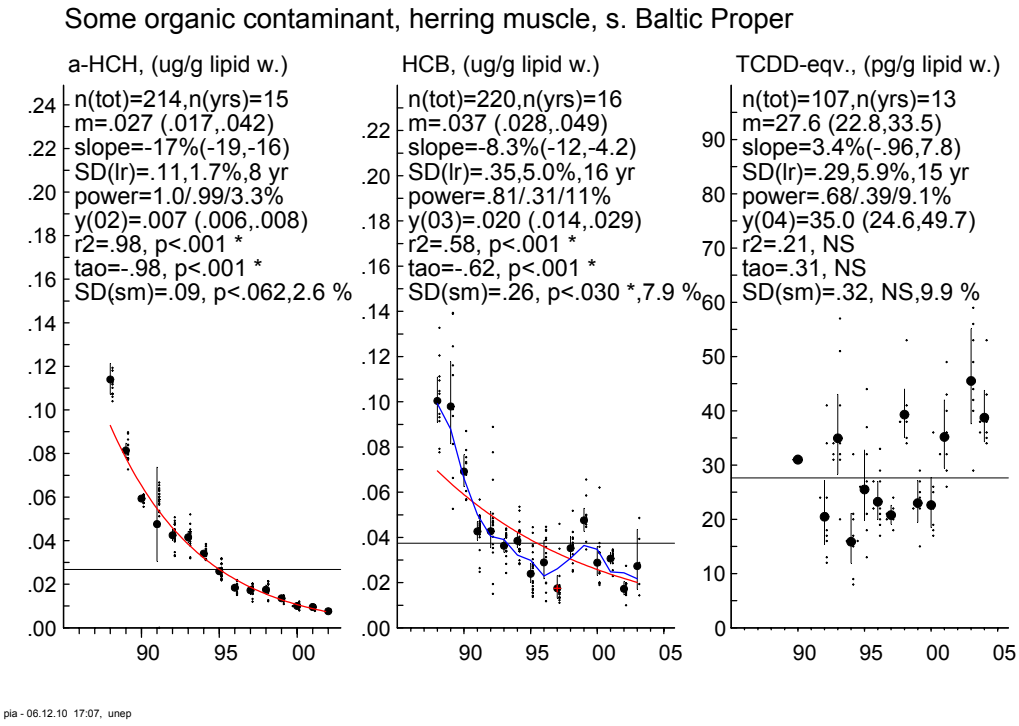

Figure 3.6 Examples of time-series; alpha-HCH, HCB and TCDD-equivalents (µg/g lipid weight) in herring muscle from the southern Baltic Proper. The legend to the figure is found in Table 3.2.

Table 3.2. Legend to Figure 3.3

The plots display the geometric mean concentration of each year (circles) together with the individual analyses (small dots) and the 95% confidence intervals of the geometric means. The overall geometric mean value for the time-series is depicted as a horizontal, thin line. The trend is presented by a regression line (plotted if $p < 0.05$, two-sided regression analysis). The log-linear regression lines fitted through the geometric mean concentrations follow smooth exponential functions. A smoother is applied to test for non-linear trend components. The smoothed line is plotted if $p < 0.05$. Below the header of each plot the results from several statistical calculations are reported:

n(tot)= Total number of analyses included together with the number of years (**n(yrs)=**).

m= The overall geometric mean value together with its 95% confidence interval (*N.B.* the number of degrees of freedom = n of years - 1).

slope= The slope, expressed as the yearly change in percent together with its 95% confidence interval.

sd(lr)= The square root of the residual variance around the regression line, as a measure of between-year variation, together with the *lowest detectable change* in the current time-series with a power of 80%, one-sided test, *alpha* =0.05. The last figure is the estimated *number of years* required to detect an annual change of 5% with a power of 80%, one-sided test, *alpha*=0.05.

power= The power to detect a log-linear trend in the time-series (Nicholson and Fryer, 1991). The first figure represents the power to detect an annual change of 5% with the number of years in the current time-series. The second figure is the power estimated as if the slope were 5% a year and the number of years were ten. The third figure is the *lowest detectable change* (given in percent per year) for a ten year period with the current between year variation at a power of 80%.

r^2= The coefficient of determination (r2) together with a p-value for a two-sided test (H_0: slope = 0), i.e. a significant value is interpreted as a true change, provided that the assumptions of the regression analysis is fulfilled.

y(02)= The concentration estimated from the regression line for the last year together with a 95% confidence interval, e.g. y(02)=0.007 (0.006, 0.008) is the estimated concentration of year 2002 where the residual variance around the regression line is used to calculate the confidence interval. Provided that the regression line is relevant to describe the trend, the residual variance might be more appropriate than the within-year variance in this respect.

tau= The Kendall's ' **tau '** as a result from the non-parametric Mann-Kendal trend test, and the corresponding p-value.

sd(sm)= The square root of the residual variance around the smoothed line. The significance of this line could be tested by means of an Analysis of Variance. The p-value is reported for this test. A significant result will indicate a non-linear trend component.

3.11 References

Barnett V., Lewis T., 1994. Outliers in Statistical Data. Third ed. Wiley and Sons Ltd.

Bignert A., Göthberg A., Jensen S., Litzén K., Odsjö T., Olsson M., Reutergårdh L., 1993. The need for adequate biological sampling in ecotoxicological investigations: a retrospective study of twenty years pollution monitoring. *The Science of the Total Environment*, 128:121-139.

Bignert A., Olsson M., de Wit C., Litzen K., Rappe Ch., Reutergårdh L., 1994. Biological variation – an important factor to consider in ecotoxicological studies based on environmental samples. *Fresenius Journal of Analytical Chemistry*, 348:76-85.

Bignert, A., Greyerz, E., Olsson, M., Roos, A., Asplund, L., Kärsrud, A.-S., 1998a. Similar Decreasing Rate of OCs in Both Eutrophic and Oligotrophic Environments – A Result of Atmospheric Degradation? Part II. *Organohalogen Compounds,* 36:459-462.

Bignert, A., Olsson, M., Asplund, L., Häggberg, L., 1998b. Fast Initial Decrease in Environmental Concentrations of OCs – A Result of Atmospheric Degradation? Part I. *Organohalogen Compounds*, 36:373-376.

Bignert, A., Olsson, M., Persson, W., Jensen, S., Zakrisson, S., Litzén, K., Eriksson, U.,Häggberg, L., Alsberg, T., 1998c. Temporal trends of organochlorines in Northern Europe, 1967-1995. Relation to global fractionation, leakage from sediments and international measures. *Environmental Pollution*, 99:177-198.

Bignert, A., 2002. The power of ICES contaminant trend monitoring. ICES Marine Science Symposia, 215: 195-201.

Bignert A., Riget F, Braune B., Outridge P., Wilson S., 2004. Recent temporal trend monitoring of mercury in Arctic biota – how powerful are the existing datasets? *J. Environ. Monit*, 6:351 - 355.

Bjerkeng, B., 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. SIME 00/4/11-E (L). Gilbert R.O., 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York.

Cohen, J. 1988. Statistical Power Analysis for the Behavioural Sciences. Academic Press, New York.

Cressie N.A.C. (1993). Statistics for Spatial Data. Wiley & Sons. 900 p.

Davis J.C. 1986. Statistics and Data Analysis in Geology. Wiley & Sons, New York, ISBN 0-471-08079-9

Dixon, J.W.,F.J. Massey. 1969. Introduction to Statistical Analysis. 3rd ed. McGraw-Hill, New York. 638p.

Gilbert O.R. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold. New York.

Grimås, U., Göthberg, A., Notter, M., Olsson, M., Reutergårdh, L., 1985. Fat Amount - A Factor to Consider in Monitoring Studies of Heavy Metals in Cod Liver. *Ambio*, 14:175 – 178.

HELCOM, 1988. Guidelines for the Baltic Monitoring Programme for the Third Stage; Part C. Harmful Substances in Biota and Sediments. HELCOM, BSEP 27C.

Helsel, D.R., Hirsch., R.M., 1995. Statistical Methods in Water Resources, Studies in Environmental Sciences 49. Elsevier, Amsterdam.

Hoaglin, D.C., and Welsch., R.E., 1978. The hat matrix in regression and ANOVA. *Amer. Stat*. 32:17-22.
Hunter, J.E. and Schmidt, F. L. 1990. Methods of meta-analysis. Correcting error and bias in research findings. Newbury Park, CA: Sage.

Loftis, J.C., Ward, R.C., Phillips, R.D., 1989. An Evaluation of Trend Detection Techniques for Use in Water Quality Monitoring Programs. EPA/600/3-89/037, p. 139.

Nicholson, M.D., Fryer., R., 1991. The Power of the ICES Cooperative Monitoring Programme to Detect Linear Trends and Incidents. In: Anon. Report of the Working Group on Statistical Aspects of Trend Monitoring. ICES Doc CM 1991.

Nicholson, M.D., Green N., Wilson S., 1991. Regression Models for Assessing Trends of Cadmium and PCB in Cod Livers from the Oslofjord. *Marine Pollution Bulletin*, 22:77-81.

Nicholson, M.D., Fryer, R., Larsen, J.R. 1995. A Robust Method for Analysing Contaminant Trend Monitoring Data. *Techniques in Marine Environmental Sciences*. ICES.

Nicholson, M. D., Fryer, R., Maxwell, D., 1998b. The influence of individual outlying observations on four methods of detecting trends. ICES CM 1998/E:8. Annex 8, pp.62-67.

Snedecor, G.W.,W.G. Cochran. 1968. Statistical Methods. Iowa 1969.

Swertz, O., 1995. Trend assessment using the Mann-Kendall test. Report of the Working Group on Statistical Aspects of Trend Monitoring. ICES CM 1995/D:2.

Underwood, A.J., 1993. The mechanics of spatially replicated sampling programmes to detect environmental impacts in a variable world. *Austr. J. Ecol.*, 18:99-116.

Underwood, A.J., 1994. Beyond BACI: sampling designs that might reliably detect environmental disturbances. *Ecol. Applic.,* 4:3-15.

Underwood, A.J., 1996. Environmental Design and Analysis in Marine Environmental Sampling. Intergovernmental Oceanographic Commission Manuals and Guides No 34, UNESCO.

van Belle, G. And Hughes J.P. 1984. Nonparametric tests for trend in water quality. *Water Resource Research 20:127-136.*

Weiss, J., Päpke, O., Bignert, A., Greyerz, E., Agostoni, C., Riva, E., Giovannini, M., Zetterström, R., 2003. Concentrations of dioxins and other organochlorines (PCB, DDTs, HCHs) in human milk from Seveso, Milan and a Lombardian rural area in Italy: a study performed 25 years after the heavy dioxin exposure in Seveso. *Acta Pediatrica*, 92: 467-472.