Not to be cited without prior reference to the authors

International Council for the Exploration of the Sea

CM 1993/E:18
Marine Environmental Quality Committee

PREDICTING FUTURE CONTAMINANT LEVELS IN RELATION TO EQOs USING TEMPORAL TREND MONITORING DATA

by

R J Fryer and M D Nicholson*

SOAFD Marine Laboratory
PO Box 101, Victoria Road
Aberdeen, AB9 8DB
Scotland, UK

*MAFF Fisheries Laboratory
Pakefield Road, Lowestoft
Suffolk, NR33 0HT
England, UK

SUMMARY

Contaminant data collected for temporal trend monitoring purposes are generally used to assess whether contaminant levels are increasing, decreasing or staying much about the same. Here, we show that such data can also be used as a management tool to predict future contaminant levels and to quantify the probability that environmental quality objectives will be met throughout the time period before the next trend assessment.

INTRODUCTION

Temporal contaminant monitoring programmes are typically used to investigate whether contaminant levels are changing with time; eg going up or down, or varying in some other systematic way. Various ways of assessing trends have been used in the past, including regression of contaminant level on time (Anon., 1989; Fryer and Nicholson, 1990), locally-weighted smoothers (Fryer and Nicholson, 1991; Nicholson and Fryer, 1993) and rank correlation methods (El-Shaarawi and Niculescu, 1992; McLeod *et al.*, 1991).

However, information about trends is only part of the picture and must be combined with information about the actual contaminant level to give effective management advice. Testing the current estimated mean contaminant concentration against some reference level is relatively straightforward (cf Rogers, 1992), but a more informative approach is to integrate information about both the current level and any trend that can be predicted from previous years. Then, for example, a contaminant showing a slow rise in concentration but which is way below levels of concern might prompt less immediate action than one whose concentration is stable but close to or above an environmental quality objective (EQO).

This approach enables the management advice to move beyond an assessment of the past and current situation to being able to quantify the risk that future contaminant levels will come close to, or exceed, EQOs.

This paper shows how temporal contaminant data can be used for management purposes by:

- identifying trends in contaminant levels,
- predicting future contaminant levels,
- quantifying the uncertainty in these predictions,
- relating these predictions to environmental quality objectives.

The paper is in two parts. Part 1 gives an informal account of the use of temporal data for predictive purposes, using two contaminant time series (mercury and copper in flounder in ICES area 31F2) as motivating examples. Part 2 provides the rigorous treatment.

PART 1: AN INFORMAL DISCOURSE

Trend Assessment

Figure 1a shows a linear trend fitted to log mercury concentrations in flounder in 31F2 between years 1-8. (The log-scale is used to satisfy various statistical assumptions.) Also shown are pointwise 95% confidence intervals for the fitted line. Clearly, not much is going on. Figure 1b shows a more general smooth (not necessarily linear) function fitted to the same data, with much the same results.

Figure 2 shows a linear trend and a more general smoother fitted to log copper concentrations. The linear fit again suggests that there is no trend in contaminant levels. However, the smoother follows the data points better and suggests that copper levels might have decreased and are now beginning to rise again.

Predicting Future Levels and Quantifying the Precision of These Predictions

The lines marked .50 in Figures 1 and 2 show predicted log mercury and log copper concentration in years 9-11, assuming that the linear trend or the smooth trend continue to apply in the future. The predictions go three years ahead, because that is the time interval between temporal trend assessments.

Also shown are pointwise upper 5, 10, 25, 50, 75, 90, 95% confidence limits for the predicted log-concentration, quantifying the uncertainty in the predictions. Loosely, there is, for example, a 95% probability that the mean log concentration in year 9 will be below the level indicated by the line marked .95 in year 9. Note that these lines fan out as we go into the future, reflecting the increased uncertainty in future predictions.

Relating Predictions to EQOs

Figures 3 and 4 convert Figures 1 and 2 to the original concentration scale. We have also added human health standards of 0.3 and 20.0 mg kg^{-1} for mercury and copper respectively (Franklin, 1987; Nauen, 1983). These figures combine both trend and prediction information to give an overall picture of what is going on. Thus, for mercury, there is no evidence of an underlying trend in concentration; however, there is an approximately 5% chance that mercury concentrations will exceed the EQO in any of the next three years. For copper, the smoother suggests that copper levels might be rising, but copper levels are way below the EQO.

Discussion

The results of large monitoring programmes of large numbers of contaminants observed in a large number of areas can be overwhelming. For example, Anon. (1990) reported analysis of 270 data sets. Information-fatigue can reduce the effectiveness with which data sets which should give cause for concern are identified and acted on. In the past (Fryer and Nicholson, 1992), we have recognised the need for simple statistical/graphical tools for presenting the results of monitoring programmes which facilitate the process of identifying problems and setting priorities.

The technique presented here offers several benefits. Providing an EQO is available, the technique provides an index that can be constructed and meaningfully compared for all contaminants. For example, from a large number of results, those contaminant/area combinations for which eg there is a 50% chance or more of exceeding the appropriate EQO within the next three years can easily be identified. Alternatively, the data sets could be ranked by the probability that the EQO will be exceeded in the next three years. An important statistical benefit is that this ranking is equally valid for short time series or for time series for which the precision of the estimated contaminant levels is poor. Uncertainty is penalised by increasing the probability that an EQO **could** be exceeded in the future. This recognises that in these cases, an important management decision is to keep on sampling.

PART 2: FUN STATISTICAL STUFF

A Model of Contaminant Levels in Biota

Consider an annual monitoring programme in which R samples have been collected at the same time each year, in years t_i, i = 1...T. Let y_i be the sample mean log-concentration in year t_i. Investigation of the ICES CMP (Fryer and Nicholson, 1990) suggests that contaminant levels in biota can often be modelled as

$$y_i = f(t_i) + \omega_i + \varepsilon_i$$

where

- f(t) is a smooth function describing the underlying trend in log-concentration over time,

- the ω_i represent random between-year variation in log-concentration, and are assumed to be independent and normally distributed with zero mean and variance τ^2,

- the ε_i represent random within-year variation in log-concentration, and are assumed to be independent and normally distributed with zero mean and variance σ^2/R; further, the ε_i are assumed to be independent of the ω_i.

Let

$$\psi^2 = \mathrm{Var}[y_i] = \tau^2 + \sigma^2/R$$

and let $\hat{\sigma}^2$ be an estimate of σ^2 obtained from the replicate within-year samples.

The random between-year variation ω_i is likely to be a mixture of genuine short term fluctuations in the mean log-concentration of the population (for example, due to varying environmental conditions or short term fluctuations in inputs) and short term bias in sampling (eg "cluster" sampling) and analytical performance. It is not possible to separate these components of variation without considerably more data. Since we wish to construct confidence intervals for the mean log-concentration of the population in the future, and since we would rather err on the side of caution by having confidence intervals that were too large rather than too small, we assume that the short term bias in sampling and analytical performance are negligible. Thus, we can write the mean log-concentration in year t_i as:

$$\mu_i = f(t_i) + \omega_i$$

Linear Trend

Suppose that

$$f(t) = \beta_0 + \beta_1 t$$

Let $\hat{\beta}_0$, $\hat{\beta}_1$ be the least squares estimates of β_0, β_1. Further, let $\hat{\psi}^2$ be an estimate of ψ^2 on T-2 degrees of freedom, obtained from the residual sum of squares, and let

$$\hat{\tau}^2 = \hat{\psi}^2 - \hat{\sigma}^2/R$$

An estimate of $f(t_i)$ is given by:

$$\hat{f}(t_i) = \hat{\beta}_0 + \hat{\beta}_1 t_i$$

and a $(1 - \alpha)\%$ confidence interval for $f(t_i)$ is given by:

$$\hat{f}(t_i) \pm t(T-2\ ;\ 1-\alpha/2) \left(\frac{1}{T} + \frac{(t_i - \bar{t})^2}{\Sigma(t_j - \bar{t})^2} \right)^{1/2} \hat{\psi}$$

(eg, p 28-29 of Draper and Smith, 1981), where

- $\bar{t} = \Sigma t_j / T$
- $t(T\text{-}2\ ;\ 1\text{-}\alpha/2)$ is the $1\text{-}\alpha/2$ percentile of Student's t-distribution on T-2 degrees of freedom.

Assuming that contaminant levels continue to follow the linear trend, the population mean log-concentration μ_0 in some future year t_0 is predicted to be

$$\hat{\mu}_0 = \hat{\beta}_0 + \hat{\beta}_1 t_0$$

and an upper $(1\text{-}\alpha)\%$ confidence limit for μ_0 is approximately

$$\hat{\mu}_0 + t(T-2\ ;\ 1-\alpha) \left\{ \left(\frac{1}{T} + \frac{(t_0 - \bar{t})^2}{\Sigma(t_j - \bar{t})^2} \right) \hat{\psi}^2 + \hat{\tau}^2 \right\}^{1/2}$$

A More General Trend

Now relax the assumption that $f(t)$ is linear and estimate $f(t)$ by a lowess smooth (Cleveland, 1979). Here we do not use the robust fitting algorithm, but extensions to this case are straightforward.

Let y be the T-vector of log-concentrations $(y_1,...,y_T)'$. The fitted (smoothed) values at the sample times can be written

$$\hat{f} = Sy$$

where S is a TxT matrix which only depends on the sampling times.

The variance ψ^2 is estimated to be

$$\hat{\psi}^2 = \frac{y' (I - S)' (I - S) y}{tr(I - S)' (I - S)}$$

on

$$\nu = \frac{\{tr(I - S)' (I - S)\}^2}{tr\{(I - S)' (I - S)\}^2}$$

degrees of freedom (Cleveland, 1979). Let

$$\hat{\tau}^2 = \hat{\psi}^2 - \hat{\sigma}^2/R$$

as before.

The covariance of the fitted values is given by:

$$Cov(\hat{f}) = \psi^2 SS'$$

Thus, assuming negligible bias in $\hat{f}$, a $(1 - \alpha)\%$ confidence interval for $\hat{f}(t_i)$ is given by:

$$\hat{f}(t_i) \pm t(\nu\ ;\ 1-\alpha/2)\ \{(SS')_{ii}\}^{1/2}\ \hat{\psi}$$

where $(SS')_{ii}$ is the ith diagonal element of SS'.

Predicting future values of mean log-concentration is difficult because no parametric curve is ever used to describe the underlying trend. One intuitively appealing method is to use the locally weighted straight line fitted at the most recent sampling time T. This provides a compromise between assuming no trend at all and assuming a more complicated polynomial trend with poor extrapolation properties. There are no guarantees that this straight line will adequately describe future contaminant levels, but provided that predictions are not too far ahead, it should provide some reasonable quantitative assessment of future log-concentrations.

Thus, let $\underset{\sim}{W}$ be the diagonal matrix whose elements are the weights used in the locally weighted straight line at t_T and let:

$$X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ t_1 & t_2 & \dots & t_T \end{bmatrix}'$$

$$Z = (X'WX)^{-1}X'W$$

Then, assuming that contaminant levels continue to follow the straight line fitted at t_T, the population mean log-concentration μ_0 in some future year t_0 is predicted to be:

$$\hat{\mu}_0 = (1, t_0)Zy$$

and an upper $(1 - \alpha)$% confidence limit for μ_0 is approximately:

$$\hat{\mu}_0 + t(\nu\ ;\ 1-\alpha)\left\{(1, t_0)ZZ'\,(1, t_0)'\hat{\psi}^2 + \hat{\tau}^2\right\}^{1/2}$$

REFERENCES

Anon. 1989. Statistical analysis of the ICES cooperative monitoring programme data on contaminants in fish muscle tissue (1975-1985) for determination of temporal trends. *ICES Cooperative Research Report 162.*

Anon. 1990. Statistical analysis of the ICES cooperative monitoring programme data on contaminants in fish liver tissue and *Mytilus edulis* (1978-1988) for determination of temporal trends. *ICES Cooperative Research Report 176.*

Cleveland, W.S. 1979. Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74, 829-836.

Draper, N.R. and Smith, H. 1981. Applied Regression Analysis, Second Edition. John Wiley and Sons, New York.

El-Shaarawi, A.M. and Niculescu, S.P. 1992. On Kendall's tau as a test of trend in time series data. *Environmetrics*, 3, 385-411.

Franklin, A. 1987. The concentration of metals, organochlorine pesticide and PCB residues in marine fish and shellfish: results from MAFF fish and shellfish monitoring programmes, 1977-1984. MAFF Aquatic Environment Monitoring Report No. 16.

Fryer, R.J. and Nicholson, M.D. 1990. The ICES Cooperative Monitoring Programme. Part 2. Testing for trends in annual mean contaminant levels. In: *Report of the Working Group on Statistical Aspects of Trend Monitoring.* ICES CM 1990/Poll: 6.

Fryer, R.J. and Nicholson, M.D. 1991. Summarising trends with locally-weighted running-line smoothers. In: *Report of the Working Group on Statistical Aspects of Trend Monitoring.* ICES CM 1991/Poll:2.

Fryer, R.J. and Nicholson, M.D. 1992. Statistical aids for interpreting trend data. In: *Report of the Working Group on Statistical Aspects of Trend Monitoring.* ICES CM 1992/Poll:1.

McLeod, A.I., Hipel, K.W. and Bodo, B.A. 1991. Trend analysis methodology for water quality time series. *Environmetrics*, 2, 169-200.

Nauen, C.E. 1983. Compilation of legal limits for hazardous substances in fish and fishery products. *FAO Fisheries Circular No. 764.*

Nicholson, M.D. and Fryer, R.J. 1993. Generalized additive models for analysing contaminant trend data. In: *Report of the Working Group on the Statistical Aspects of Environmental Monitoring.* ICES CM 1993.

Rogers, J. 1992. Assessing attainments of ground water cleanup standards using modified sequential t-tests. *Environmetrics*, 3, 335-359.

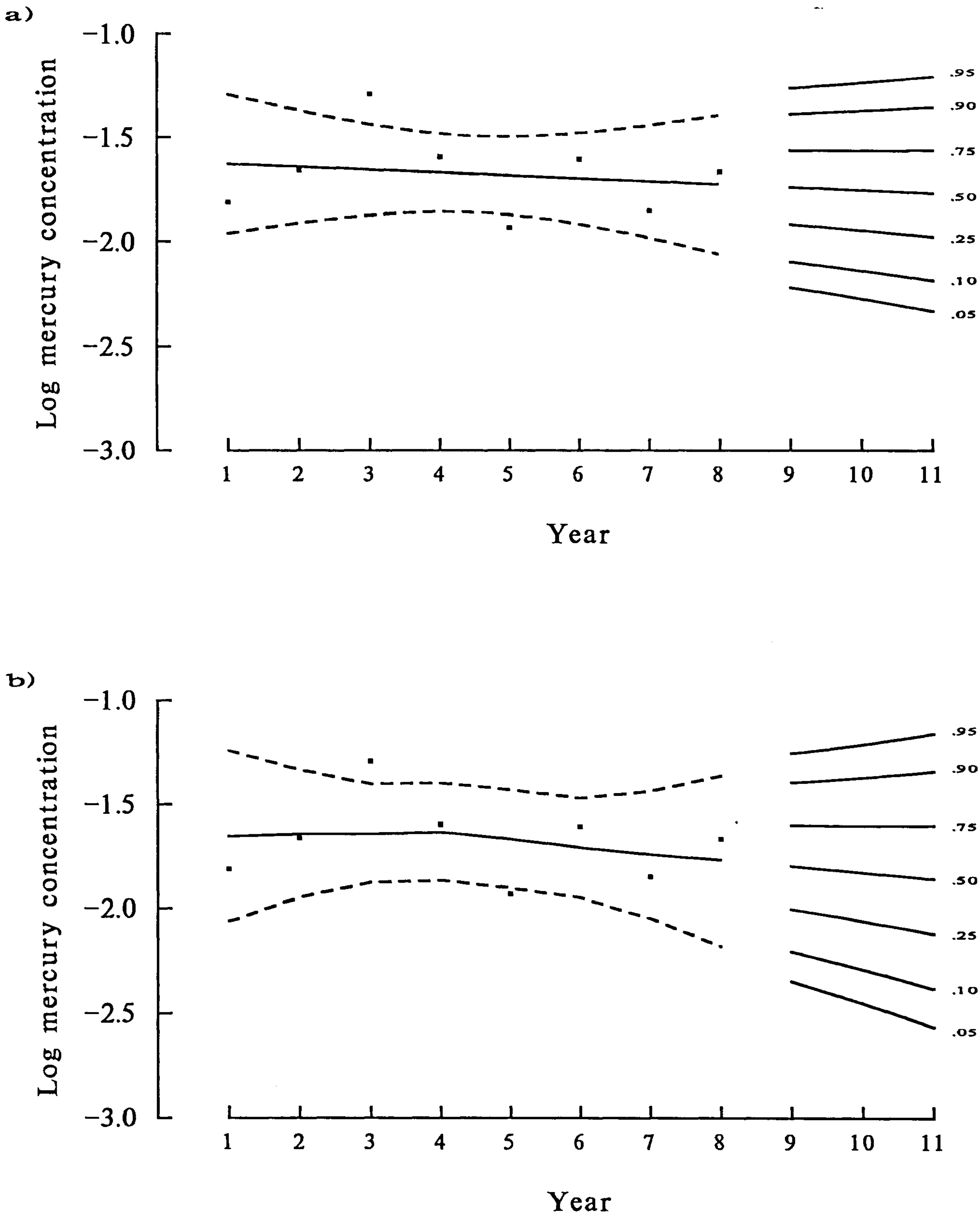

a)
Log mercury concentration
−1.0
−1.5
−2.0
−2.5
−3.0
1 2 3 4 5 6 7 8 9 10 11
Year
.95
.90
.75
.50
.25
.10
.05
b)
Log mercury concentration
−1.0
−1.5
−2.0
−2.5
−3.0
1 2 3 4 5 6 7 8 9 10 11
Year
.95
.90
.75
.50
.25
.10
.05

Figure 1

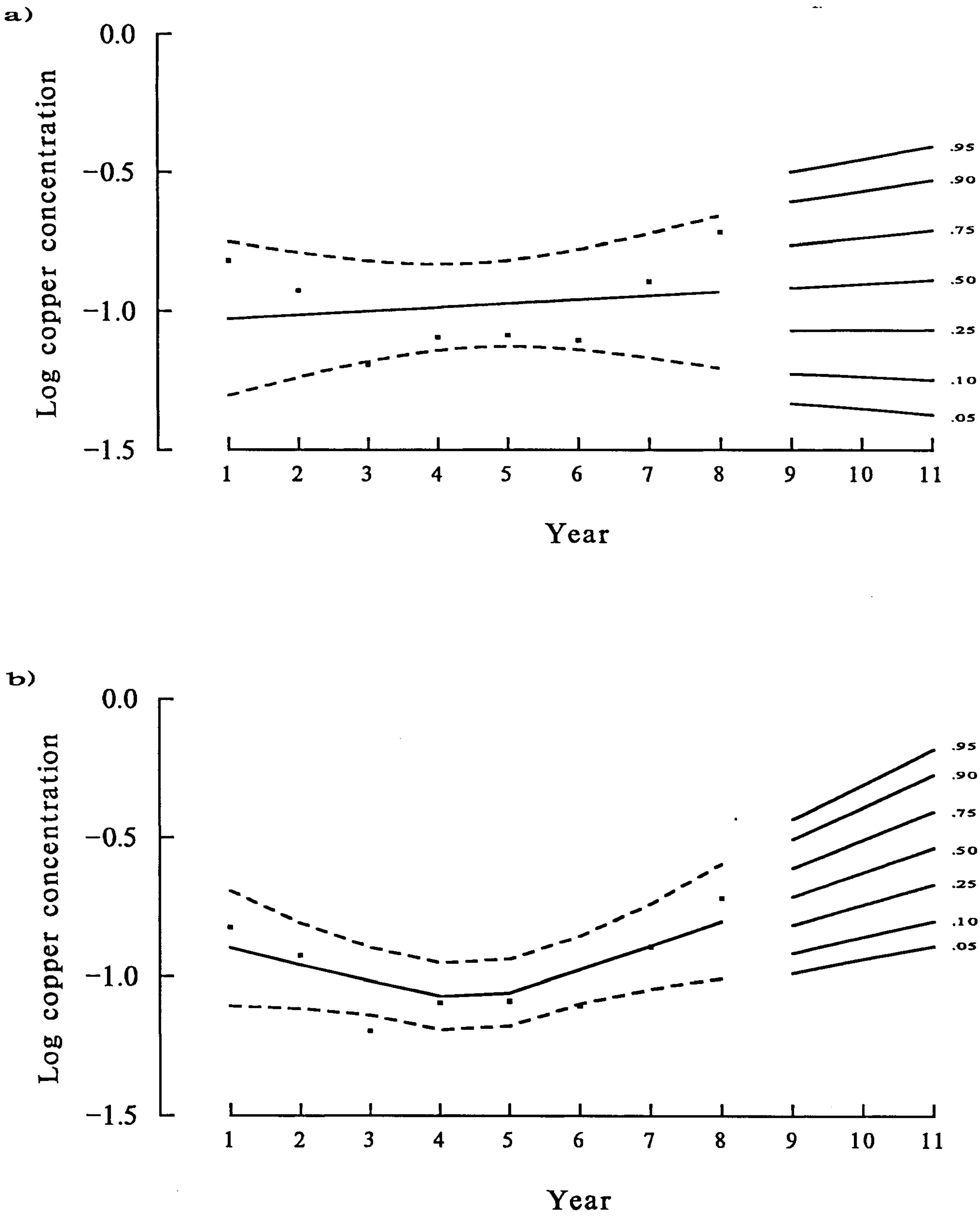

a)
0.0
−0.5
−1.0
−1.5
Log copper concentration
.95
.90
.75
.50
.25
.10
.05
1
2
3
4
5
6
7
8
9
10
11
Year
b)
0.0
−0.5
−1.0
−1.5
Log copper concentration
.95
.90
.75
.50
.25
.10
.05
1
2
3
4
5
6
7
8
9
10
11
Year

Figure 2

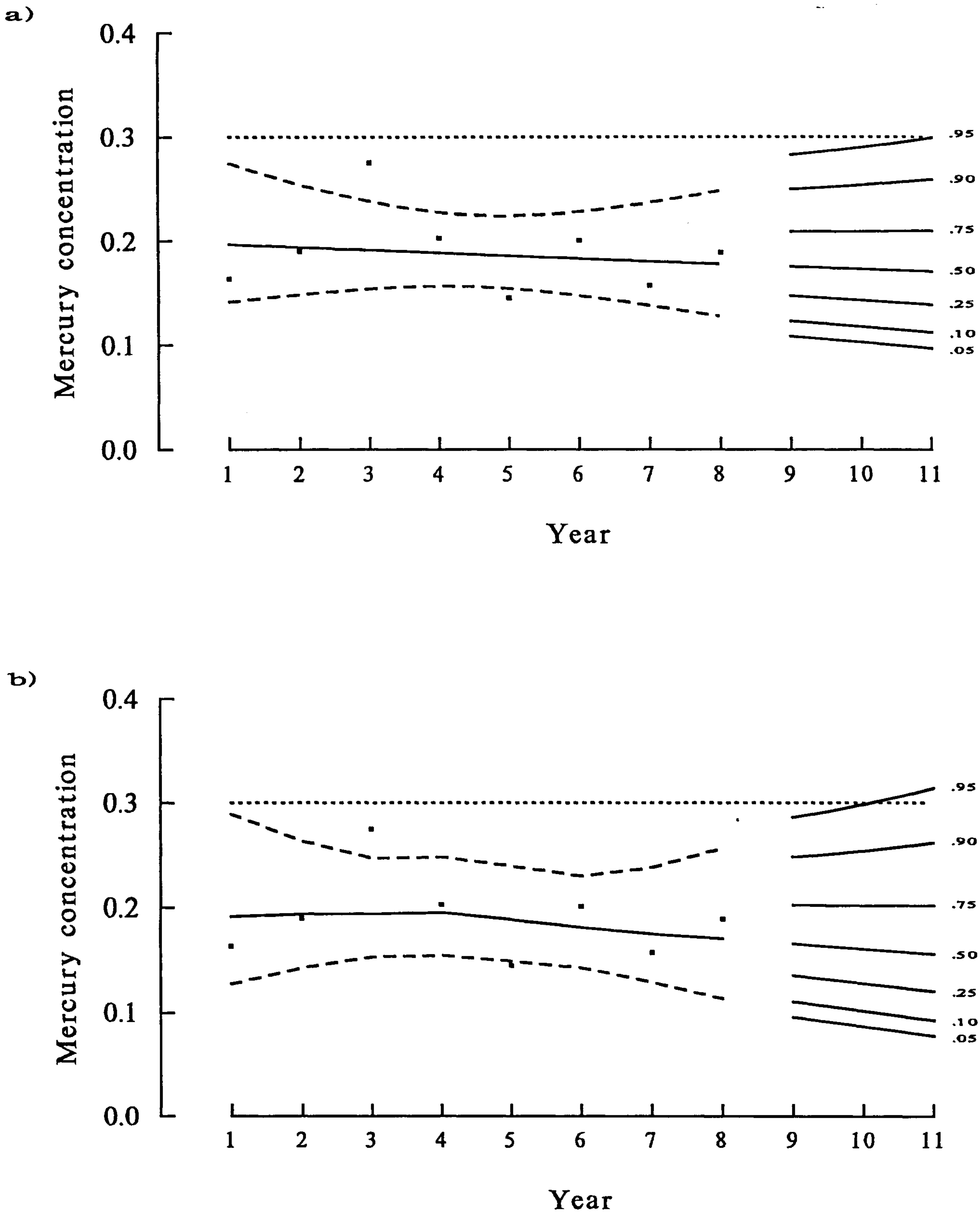
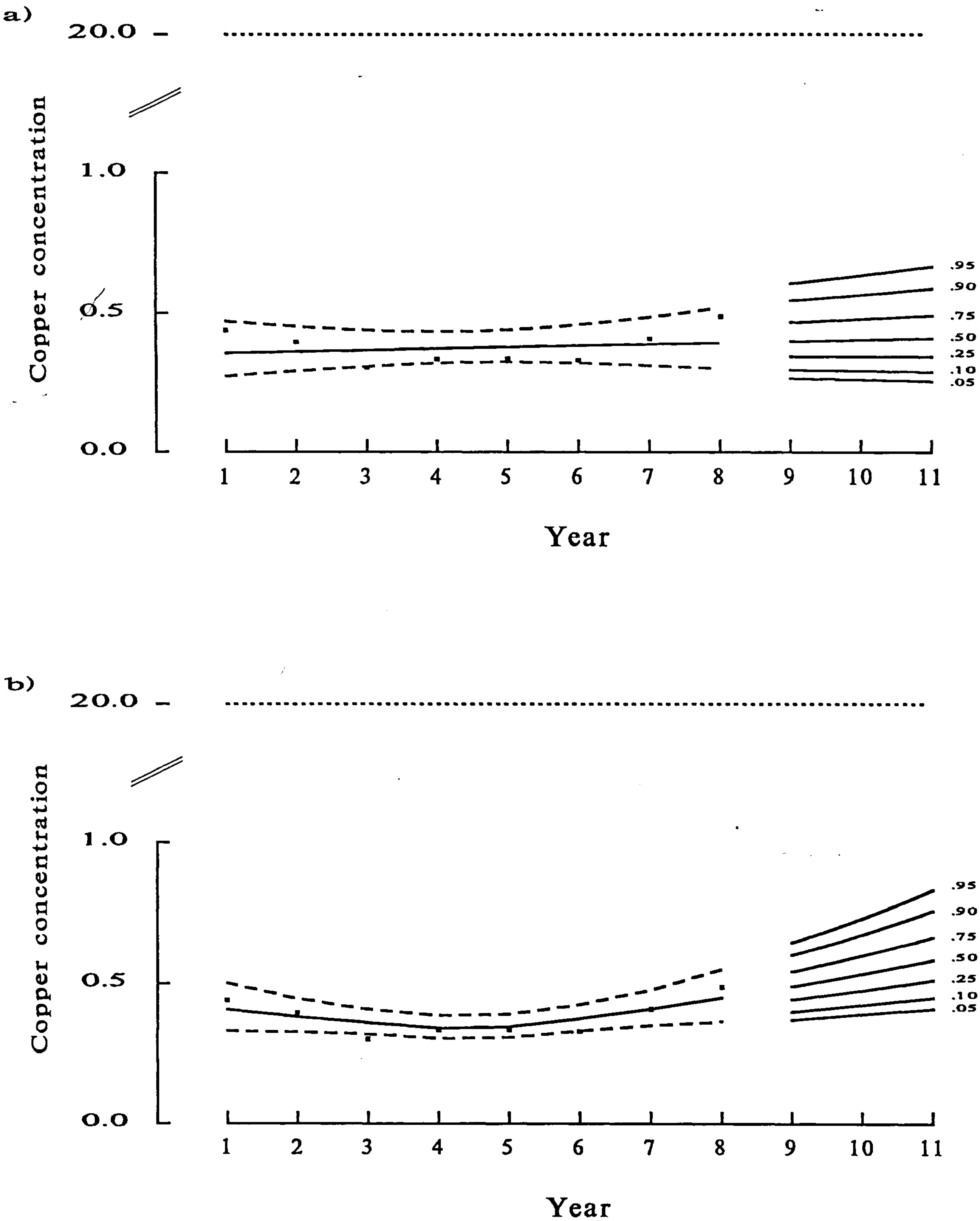

a)
0.4
0.3
0.2
0.1
0.0
Mercury concentration
1
2
3
4
5
6
7
8
9
10
11
Year
.95
.90
.75
.50
.25
.10
.05
b)
0.4
0.3
0.2
0.1
0.0
Mercury concentration
1
2
3
4
5
6
7
8
9
10
11
Year
.95
.90
.75
.50
.25
.10
.05

Figure 3



a)
20.0
1.0
0.5
0.0
Copper concentration
.95
.90
.75
.50
.25
.10
.05
1
2
3
4
5
6
7
8
9
10
11
Year
b)
20.0
1.0
0.5
0.0
Copper concentration
.95
.90
.75
.50
.25
.10
.05
1
2
3
4
5
6
7
8
9
10
11
Year

Figure 4