Reprinted from

ECOLOGICAL
MODELLING

Ecological Modelling 74 (1994) 125–135

# Probabilistic uncertainty assessment of phosphorus balance calculations in a watershed

A. Taskinen [a,*], O. Varis [a], H. Sirviö [a], J. Mutanen [b], P. Vakkilainen [a]

[a] *Helsinki University of Technology, Laboratory of Hydrology and Water Resources Management, FIN-02150 Espoo, Finland*

[b] *Mikkeli Water and Environment District Office, Mikkeli, Finland*

(Accepted 2 November 1993)

ELSEVIER

ELSEVIER

Ecological Modelling 74 (1994) 125–135

ECOLOGICAL MODELLING

# Probabilistic uncertainty assessment of phosphorus balance calculations in a watershed

A. Taskinen [a,*], O. Varis [a], H. Sirviö [a], J. Mutanen [b], P. Vakkilainen [a]

[a] *Helsinki University of Technology, Laboratory of Hydrology and Water Resources Management, FIN-02150 Espoo, Finland*

[b] *Mikkeli Water and Environment District Office, Mikkeli, Finland*

(Accepted 2 November 1993)

**Abstract**

Water quality studies in reservoirs, rivers, lakes, and entire river basins include increasingly often mass balance calculations. In comparison with hydrological studies where water balance models are traditionally frequently used, water quality investigations are typically faced with essentially higher uncertainty. This study presents an approach to the assessment and propagation of this uncertainty, with a case study on a river reach with three point source polluters, substantial throughflow, and non-point components. A computational implementation using spreadsheets was used, which allows the use of probabilistic models in close contact with environmental data bases.

*Key words:* Phosphorus; Water quality

## 1. Introduction

Mass balance calculations have become a commonplace in surface water quality studies in reservoirs, rivers, lakes, and river basins. They allow means for merging hydrologic, hydrodynamic, and water quality information. Moreover, mass balances are the core of most process models. With the growing number of applications of the mass balance approach, the problems due to the uncertainties involved deserve more attention. When comparing water quality studies to, e.g., problems encountered in the field of hydrology where the water balances are traditionally widely used, the data and knowledge on water quality components are typically essentially

* Corresponding author.


*SSDI* 0304-3800(93)E0100-H

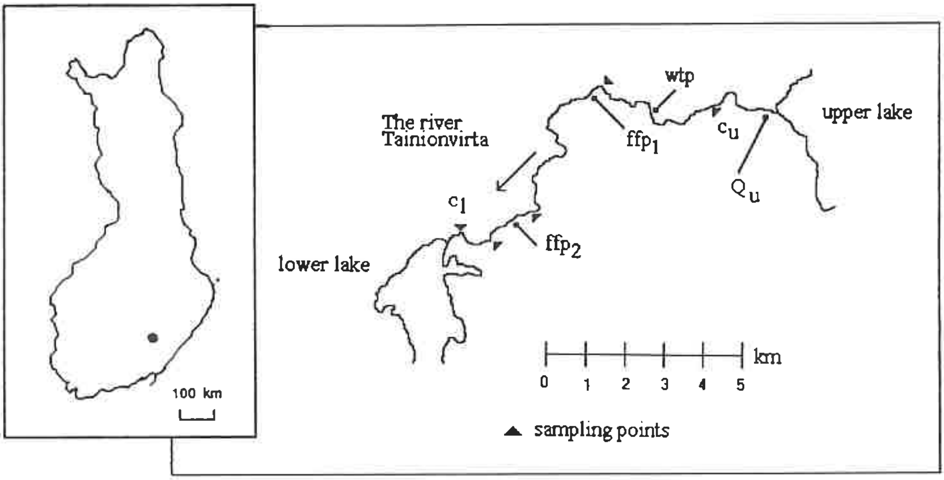


Fig. 1. The river Tainionvirta with sampling points and point dischargers.

poorer than those in hydrology. Additionally, the system complexity is usually much greater.

The evaluation and propagation of uncertainty is thus crucial in observational design, as well as in planning and management of water quality. In this study the phosphorus balances of a eutrophicated river stretch subject to loading from two fish farming plants and one town were analyzed. The probabilistic approach presented was used to assess the contribution of each mass balance component to the total uncertainty in calculations.

## 2. The site and the problem

The river Tainionvirta (Fig. 1), south eastern Finland (26°01′E, 61°34′N) connects two lakes, Jääsjärvi and Joutsjärvi, here called the Upper and the Lower Lake, respectively. The river reach has a length of 13.5 km and a mean flow of 12.5 $m^3/s$. It receives loading from two fish farming plants (denoted here as $ffp_1$ and $ffp_2$), and from a municipal wastewater treatment plant (*wtp*) which treats the sewage of 1700 inhabitants. The watershed area above the Lower Lake is 1515 $km^2$, 90 $km^2$ of which is below the Upper Lake.

During the last two decades, the river Tainionvirta as well as the whole watershed has been subjected to substantial eutrophication. Expectedly, one major reason to this unfavorable development is the nutrient load from fish farming The authorities have strained the discharge licences of the plants since the mid-1980s, and this has yielded a reduction in loading (cf. Fig. 3). Nevertheless, more information is needed on the contribution of all the dischargers – including upstream sources and non-point pollution – to eutrophication. Also the monitoring policy apparently needs revision. An essential motivation of this study rose

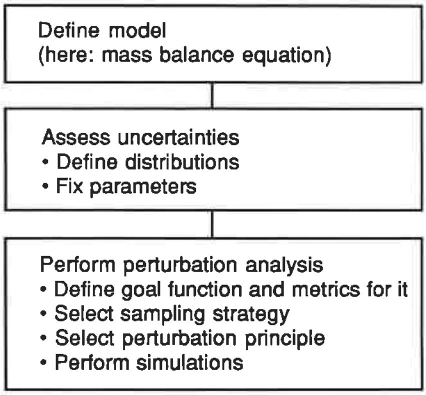


Fig. 2. The probabilistic approach to study and propagate model uncertainty.

from the need to assess and propagate the uncertainties involved in such calculations.

The river flow has been monitored daily below the Upper Lake, and four water quality sampling points are located along the river. Water quality has been observed four times annually, in March, May, August, and October. The discharges of the waste water treatment plant and the fish farming plants are monitored monthly with an aggregated sample over one day. Additionally, the fish farming plants report the monthly feed consumption and the growth in fish biomass, which are rather contradictory with the discharge data. The assessment of nutrient loading from fish farming was mainly based on feed consumption and fish growth data since they were assumed to be more reliable.

## 3. Probabilistic approach to mass balance analysis

In this case, each mass balance component was subject to considerable uncertainty. A monthly sample did not provide comprehensive and accurate information on concentration, flow, or loading in a river stretch. Additionally, the discharge data may be biased due to subjective motivations. The following probabilistic approach consisting of three steps (Fig. 2) was developed for the assessment and propagation of those uncertainties. It is based on the risk analysis philosophy (e.g., McCormick, 1981; Hertz and Thomas, 1983; Megill, 1985).

*Model definition*

The core of this study were monthly mass balance calculations for the river stretch. Originally, both total phosphorus (P) and total nitrogen (N) were included in the study, but N balances appeared excessively inadequate, and they were thus omitted. Eq. 1 was used.

$$c_u Q_u + P_{wtp} + P_{ffp_1} + P_{ffp_2} + Y = c_l Q_l, \tag{1}$$

where $c$ denotes concentration, $Q$ stands for river flow, and $P$ for P load. The sub-$u$ and sub-$l$ denote Upper and Lower Lake, respectively. $Y$ is a residual term, consisting of non-point load, sedimentation, resuspension, and errors. The unit used was kg P month$^{-1}$. The P balances are shown in Fig. 3. Notable is the big proportion of the highly uncertain residual term $Y$.

*Uncertainty assessment*

A parametric distribution, either truncated normal ($TN$) or truncated lognormal ($TL$) distribution, was chosen for the components. Empirical fitting of a distribution was impossible due to the character of the data. Instead, subjective assessment of the distribution and its parameters was performed, yielding to those shown in Table 1. The uncertainty measure used was the coefficient of variation, $\mathrm{cv} = \sigma/\mu$, where $\sigma$ is standard deviation and $\mu$ is mean. The coefficient of variation allows explicit comparability of the uncertainties of different components. Truncated distributions were used because no negative values were allowed for loads, concentrations or flows. Upper tails were cut from the value $\mu + 5\sigma$ due to properties of the software.

*Perturbation analysis: Principle*

The overall uncertainty was here assumed to be cumulated in the residual term $Y$. For the residuals of the P balance for each month, the sample mean $\mu_m$ over the years $m$ considered, weighted by the inverse of the variance $1/\sigma_m^2$

$$\mu_Y = \sum_m \frac{\mu_m}{\sigma_m^2} \Big/ \sum_m \frac{1}{\sigma_m^2}, \tag{2}$$

and the variance of the sample mean

$$\mathrm{Var}(Y) = 1 \Big/ \sum_m \frac{1}{\sigma_m^2} \tag{3}$$

were calculated. These were obtained by defining first the expected value $\beta$ for the residual term

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim N(0, \boldsymbol{V}), \tag{4}$$

where $\boldsymbol{Y}$ is a $1 \times k$ vector for monthly sample means for each year $(1, \ldots, k)$ for the residual term, $\boldsymbol{X}$ is a $1 \times k$ unit vector, and $\epsilon$ is the $1 \times k$ vector for deviations of the sample means from the expected value. This is thus equivalent to

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_k \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix} \beta + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdots \\ \epsilon_k \end{bmatrix}, \quad \boldsymbol{V} = \begin{bmatrix} \sigma_1^1 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & 0 \\ \cdots & & & 0 \\ 0 & 0 & \cdots & \sigma_k^2 \end{bmatrix}. \tag{5}$$

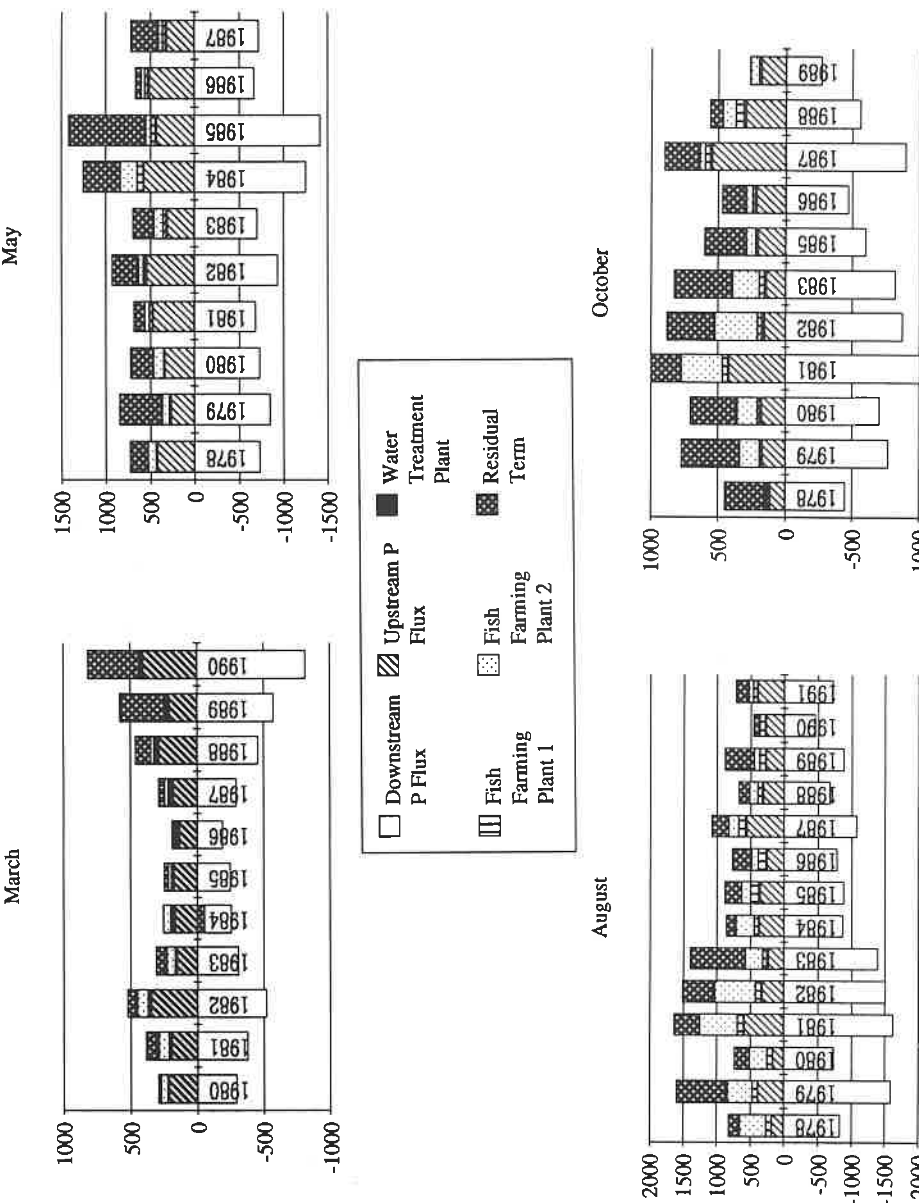


Fig. 3. Phosphorus balances at the river Tainionvirta.

Table 1
The distributions and their parameters for each component of the phosphorus balance. Additionally, for $Q_l$, $cv_j$ was assessed to be 0.2, but its impact on overall uncertainty was not studied because it was assessed through water balance, not through observations

| Component | Distribution | $\mu$ | cv |
|---|---|---|---|
| $c_u$ | *TL* | mean of observation within the month | 0.3 |
| $Q_u$ | *TN* | sum of daily observations | 0.1 |
| $c_l$ | *TL* | mean of observations within the month | 0.3 |
| $P_{wtp}$ | *TN* | data reported monthly | 0.2 |
| $P_{ffp_1}$ | *TN* | assessed on the basis of monthly feed consumption and fish biomass growth | 0.25 |
| $P_{ffp_2}$ | *TN* | assessed on the basis of monthly feed consumption and fish biomass growth | 0.3 |

Using weighted least squares (e.g., Seber, 1977) the estimate of the expected value was obtained by

$$(X^TV^{-1}X)^{-1}X^TV^{-1}Y=\sum_m \frac{\mu_m}{\sigma_m^2} / \sum_m \frac{1}{\sigma_m^2}. \tag{6}$$

This estimate is actually the mean over years weighted with the inverse of the variance. The means with low variance get the greatest weight. This estimate allows the comparability of uncertainties of different P balance components. With Gaussian data, this estimate is the minimum variance unbiased estimator (MVUE). With non-Gaussian data, the estimate is still the best linear unbiased estimate (BLUE). Accordingly, the estimate of the variance of the mean was obtained by

$$(X^TV^{-1}X)^{-1}=1/\sum_m \frac{1}{\sigma_m^2}. \tag{7}$$

The metrics from Eqs. 2 and 3 were used to study the uncertainties in the P balance calculations.

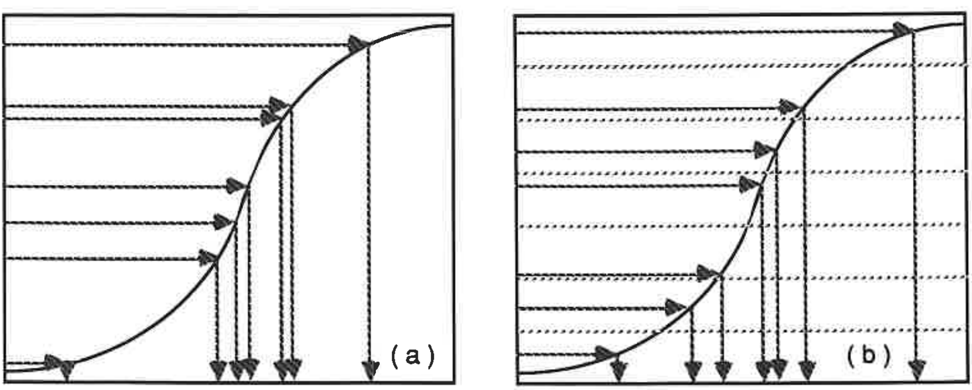


Fig. 4. Principle of (a) Monte Carlo, and (b) Latin Hypercube sampling.

Table 2
The two principles, $A$ and $B$, both consisting of three simulations, $a$, $b$, and $c$, of perturbing the coefficient of variation $cv_i$ of each uncertain component $i$. $cv_i^*$ stands for the perturbed $cv_i$ value

| | $cv_i^*$ | | | $cv_j$, $j \neq i$ |
|---|---|---|---|---|
| | $a$ | $b$ | $c$ | |
| $A$ | $0.5\ cv_i$ | $cv_i$ | $2\ cv_i$ | 0 |
| $B$ | 0 | $0.5\ cv_i$ | $2\ cv_i$ | $cv_j$ |

The sampling in probabilistic simulations was done using the Latin Hypercube principle. In $n$ iterations, the probability distribution was divided into $n$ sections equal in probability mass. Thereafter a random value was generated for each section, which are in random order. In comparison to Monte Carlo sampling, this approach gives typically a better view of the distribution with a modest number of iterations (McKay et al., 1979). The difference is particularly notable when tails of distributions are under concern. Fig. 4 illustrates the difference between Monte Carlo and Latin Hypercube sampling.

The role of each uncertain component $i$ was thereafter analyzed in two ways (Table 2). Principle $A$ analyzes the actual phosphorus balance calculation better, especially in simulations $a$ and $b$, because uncertainty is involved at one component at each simulation. The uncertainties of other components $cv_j$ were set to 0. Principle $B$ was used to simulate the real situation in nature with uncertainties at each component. 1000 Latin Hypercube iterations were performed at each phase.

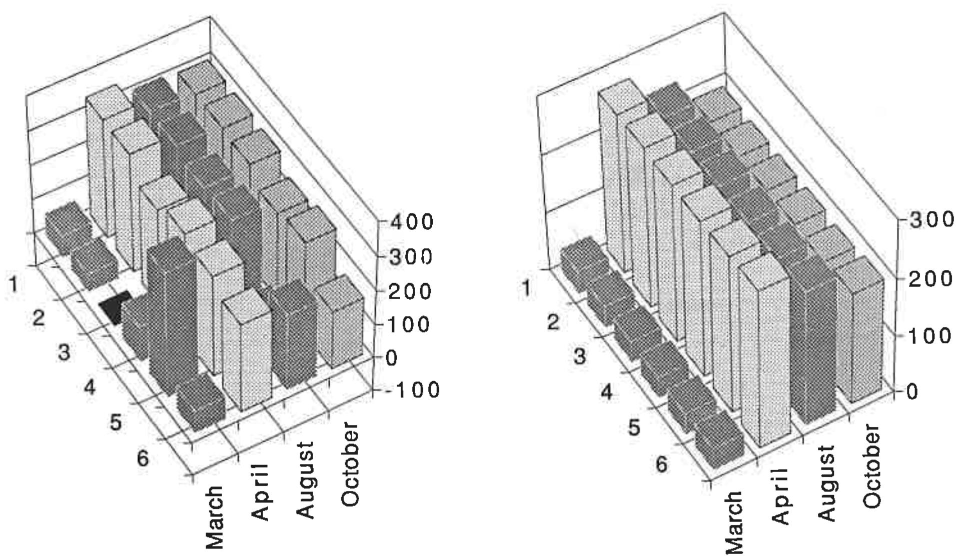


Fig. 5. The average values of the three runs $a$, $b$, and $c$ (cf. Table 2) for the means of the residual term $Y$ [kg P month$^{-1}$] with the perturbation principle $A$ (left) and $B$ (right). Mass balance components are denoted as: $1 = c_u$, $2 = Q_u$, $3 = P_{wtp}$, $4 = P_{ffp_1}$, $5 = P_{ffp_2}$, and $6 = c_l$.

*Perturbation analysis: Results*

The results of the perturbation study (Figs. 5 to 7) can be divided into two parts according to the two strategies *A* and *B* as defined in Table 2. In comparison to variances, the three simulations *a*, *b*, and *c* resulted in very small differences in the sample means and therefore only their averages are presented (Fig. 5). Strategy *A* indicates the sensitivity of the mass balances to levels of each component. Strategy *B* simulates the real situation and indicates the proportional importance of the uncertainty of each component. When comparing these results, the variances are essentially lower with strategy *A* than with strategy *B*, but the impacts on the level of the sample mean are more striking.

With strategy *A*, in March, the uncertainty due to the upstream P flux ($c_u Q_u$) was dominant, while in other months, the downstream P flux ($c_l Q_l$) had the greatest contribution to the total uncertainty. The uncertainties of the other mass balance components were minor in comparison to these two. The perturbation

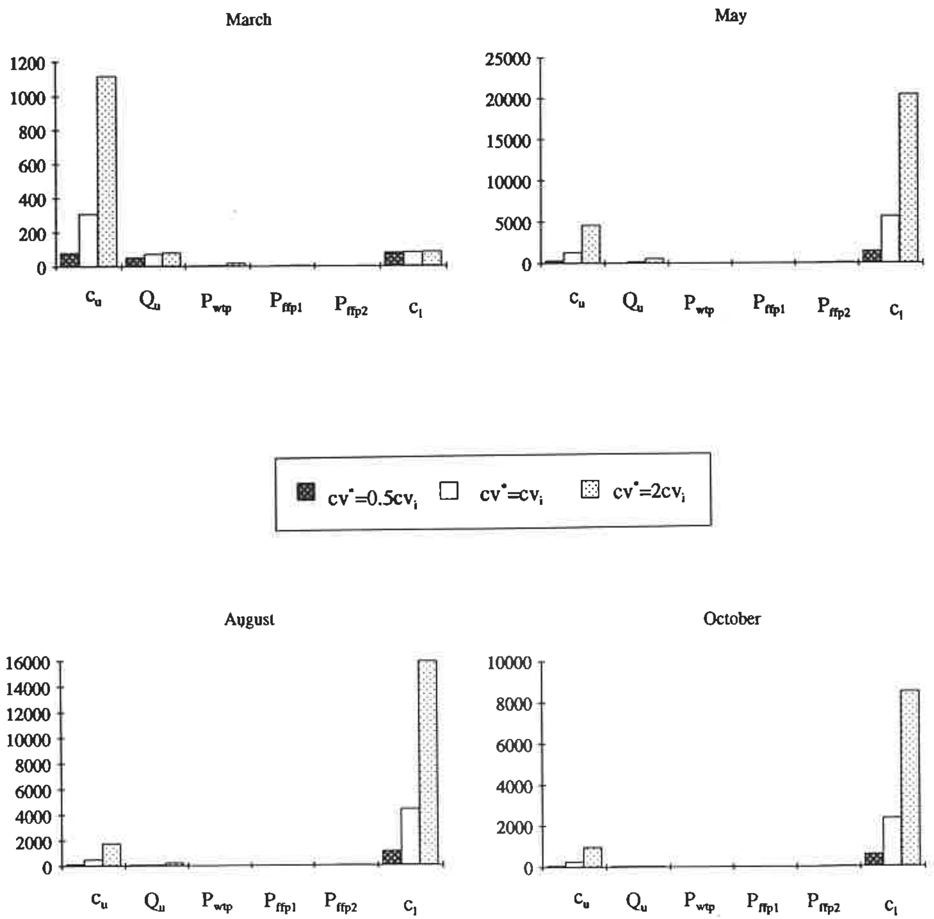


Fig. 6. The variances of the residual term *Y* [(kg P month$^{-1}$)$^2$] with the perturbation principle *A*.

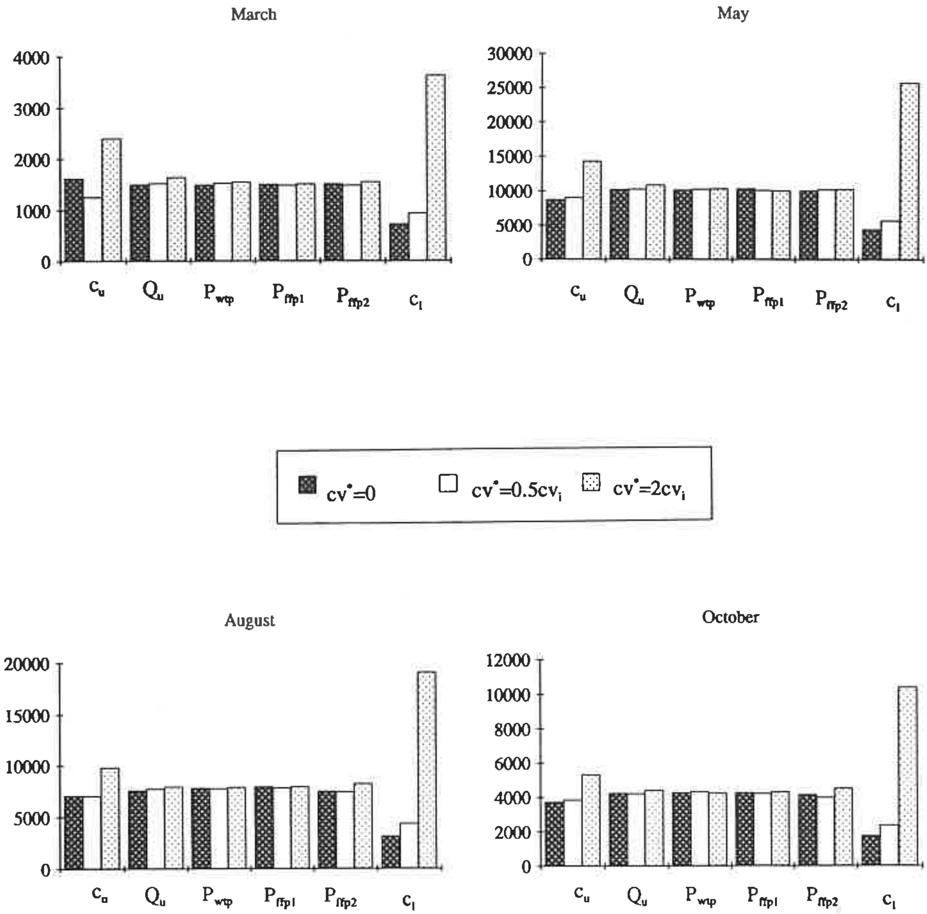


Fig. 7. The variances of the residual term $Y$ [(kg P month$^{-1}$)$^2$] with the perturbation principle $B$.

study $B$ indicated largely the same features as the study $A$. However, the role of the downstream P flux was also the most dominant component in March.

## 4. Discussion and conclusions

In water quality management, the role of uncertainty is often very large. More attention should be paid on uncertainties involved in basic structures of water quality models such as mass balances. In the present case, the analysis would be strongly impaired if only the mean behavior of the system were studied, as is often done. Very dangerous would be to use complex, deterministic mass balance-based water quality models in cases like this, without paying attention on the uncertainty in data and in calculations.

The approach presented appeared suitable both in research and in administrative, management use. The computational implementation – using spreadsheets and a risk analysis add-in (Anonymous, 1990) – provided a feasible tool for performing probabilistic studies which can be linked very closely to environmental data bases. This experience is in accordance with the one by Kuikka and Varis (1991) and Koivusalo et al. (1992). Being flexible, the approach easily allows also the augmentation of other river basin objects to the analysis as shown by Taskinen (1992). He has extended the model to, e.g., give probabilistic forecasts on total phosphorus and chlorophyll-*a* concentrations in the Lower Lake using nonlinear multiple regression models. Furthermore, the analysis could be extended in the management direction to include features such as risk attitude and optimization (see Varis et al., 1994).

In the management of the case study presented, it is essential to know well the throughflowing amount of P. The major monitoring effort should be allocated to observations of the nutrient concentrations at both ends of the river Tainionvirta. If these quantities are not known more precisely, the assessment of the impacts of the two fish farming plants ($ffp_1$ and $ffp_2$) is bound to remain highly uncertain, even if the information of their nutrient loads were more accurate than at present. Additional resources could be raised by deleting the two sampling points within the reach. Also data allowing the consideration of the N balances would be highly useful in further management of the case. The results were not very sensitive to the perturbation principle. The high importance of the upstream P flux in March is most evidently caused by the high proportion of throughflowing P in comparison to non-point load due to ice and snow cover, and off-season low activity in fish farming.

## Acknowledgements

We are greatly indebted to J. Kettunen, H. Teräsvirta, and all our other colleagues who have contributed to the realization of this study, which was realized as a co-project of Helsinki University of Technology and Mikkeli Water and Environmental District Office, National Board of Waters and the Environment.

## References

Anonymous, 1990. @Risk: Risk Analysis and Simulation Add-In. Palisade, Newfield, New York.

Hertz, D.B. and Thomas, H., 1983. Risk Analysis and its Applications. Wiley, New York.

Koivusalo, H., Varis, O. and Somlyódy, L., 1992. Water quality of the Nitra River, Slovakia – analysis of organic material pollution. International Institute for Applied Systems Analysis, WP-92-084, Laxenburg.

Kuikka, S. and Varis, O., 1991. Probabilistic assessment of TAC based fisheries management of Baltic Salmon. International Council for Exploitation of the Sea, C.M. M:30 (Mimeo).

McCormick, N.J., 1981. Reliability and Risk Analysis. Academic Press, New York.

McKay, M.D., Conover, W.J. and Beckman, R.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 211: 239–245.

Megill, R.E., 1985. An Introduction to Risk Analysis, 2nd ed. PennWell Books, Tulsa, OK.

Seber, G.A.F., 1977. Linear Regression Analysis. Wiley, New York.

Taskinen, A., 1992. MSc Thesis (In Finnish, with English Abstract). Helsinki University of Technology, Laboratory of Hydrology & Water Resources Management, Espoo, Finland.

Varis, O., Kuikka, S. and Taskinen, A., 1994. Modeling for water quality decisions: uncertainty and subjectivity in information, in objectives, and in model structure. Ecol. Modelling.