Author(s): Helske, Jouni; Nyblom, Jukka; Ekholm, Petri; Meissner, Kristian

Title: Estimating aggregated nutrient fluxes in four Finnish rivers via Gaussian state space models

Year: 2013

Version:

Please cite the original version:

Helske, J., Nyblom, J., Ekholm, P., & Meissner, K. (2013). Estimating aggregated nutrient fluxes in four Finnish rivers via Gaussian state space models. Environmetrics, 24(4), 237-247. https://doi.org/10.1002/env.2204

Estimating aggregated nutrient fluxes in four Finnish rivers via Gaussian state space models

Jouni Helske *[1], Jukka Nyblom[1], Petri Ekholm[2], and Kristian Meissner[2]

[1]University of Jyväskylä, Department of Mathematics and Statistics

[2]Finnish Environment Institute

Abstract

Reliable estimates of the nutrient fluxes carried by rivers from land-based sources to the sea are needed for efficient abatement of marine eutrophication. While nutrient concentrations in rivers generally display large temporal variation, sampling and analysis for nutrients, unlike flow measurements, are rarely performed on a daily basis. The infrequent data calls for ways to reliably estimate the nutrient concentrations of the missing days. Here we use the Gaussian state space

*Corresponding author: Jouni Helske, jouni.helske@jyu.fi, University of Jyväskylä, Department of Mathematics and Statistics, P.O.Box 35 (MaD) Jyväskylä, FI 40014

models with daily water flow as a predictor variable to predict missing nutrient concentrations for four agriculturally impacted Finnish rivers. Via simulation of Gaussian state space models, we are able to estimate aggregated yearly phosphorus and nitrogen fluxes, and their confidence intervals.

The effect of model uncertainty is evaluated through a Monte Carlo experiment, where randomly selected sets of nutrient measurements are removed and then predicted by the remaining values together with re-estimated parameters. Results show that our model performs well for rivers with long-term records of flow. Finally, despite the drastic decreases in nutrient loads on the agricultural catchments of the rivers over the last 25 years, we observe no corresponding trends in riverine nutrient fluxes.

Keywords: simulation, sparse data, interpolation, Kalman filter, Kalman smoother

1 Introduction

Abatement of marine eutrophication calls for reliable estimates of the nutrient fluxes carried by rivers from land-based sources to the sea. Monitoring programs of many important rivers in Finland, and elsewhere, typically involve daily measurements of water flow, but due to the costs, much more infrequent sampling and analysis of phosphorus and nitrogen concentrations. Yet, the concentrations of nutrients often show large temporal variation, especially in rivers receiving loading from diffuse sources (Kauppila and Koskiaho, 2003). The more infrequent the water quality data are, the more sensitive the flux estimates are to the method used to estimate the concentrations for

the unsampled days. Several inter- and extrapolation methods have been suggested to estimate missing monitoring data (Young et al., 1988; Rekolainen et al., 1991; Kronvang and Bruhn, 1996; Quilbé et al., 2006). While many of the methods simply assume that the observation made on a specific day represents the concentration level for a longer period (e.g. between the midpoints of the preceding, current and next observation), other approaches make use of the relationship between the concentration and some other variable, usually the flow.

Our aim is to develop a method for estimating fluxes of total phosphorus and total nitrogen for rivers mainly impacted by diffuse loading from agriculture for a given time period, commonly a year. For prediction of the missing nutrient concentration measurements we use a time varying regression model with an additional autoregressive component using the water flow measurements as a predictor variables. Various simulation techniques are employed for evaluating our results. As a general framework we use Gaussian state space models together with Kalman filter and smoother.

2 Methods

2.1 Interpolation via state space models and simulation

Our approach to modelling nutrient concentrations and fluxes is based on state space modelling with Kalman filtering, smoothing and interpolation. The form of the Gaussian state space model, sufficient for our purposes, is

$$y_t = X_t\beta_t + \epsilon_t, \quad \epsilon_t \sim NID(0, H) \tag{1}$$

$$\beta_{t+1} = T\beta_t + \eta_t, \quad \eta_t \sim NID(0, Q), \quad t = 1, 2, \ldots, T, \tag{2}$$

where NID stands for "normally and independently distributed". The first row (1) is called an observation equation and the second row (2) a state equation. The observed process $\{y_t\}$ may be a scalar or vector valued. The unobserved state process $\{\beta_t\}$ is often a vector process. The process starts with $\beta_1 \sim N(b_1, P_1)$ independently of error processes $\{\epsilon_t\}, \{\eta_t\}$. In our application the system matrix T is a time invariant diagonal matrix, whereas the system matrices X_t contain time varying predictor values. The state process $\{\beta_t\}$ is a latent process of time varying levels and regression coefficients. The model is defined in more detail in sections 4.1 and 4.2. Further, the covariance matrices H and Q are time invariant.

In our application the interpolation problem arises because there are missing observations. Let Y comprise all the non-missing observations. If the value y_t at time t is missing, then the Kalman smoother provides its estimate as the conditional mean $\hat{y}_t = X_t\hat{\beta}_t$ together with $\hat{\beta}_t = E(\beta_t \,|\, Y)$ and the conditional covariance matrix $\mathrm{Var}(y_t \,|\, Y) = S_t$. The Gaussian assumption then yields

$$y_t \,|\, Y \sim N(\hat{y}_t, S_t), \tag{3}$$

which can be used for obtaining prediction error limits. Plainly, the interpolated value is unbiased in the sense that $E(y_t - \hat{y}_t) = 0$.

Formula (3) is useful for single missing values. However, our primary interest is a nonlinear compound measure over a time span $t+1, \ldots, t+s$ of length s (e.g. a calendar year), denoted by

$$m_{t,s} = \sum_{i=1}^{s} q_{t+i} e^{y_{t+i}},$$

where q_t is the water flow on the day t and e^{y_t} is the daily nutrient concentration. If we had the values q_t and y_t measured on each day, then we

would have correct nutrient fluxes. Admittedly this is not exactly true due to the measurement errors, but it would satisfy the practical needs of evaluating the yearly fluxes. In the subsequent analysis we focus on the effects of missing nutrient measurements compared to the ideal case of having all measurements.

In section 4 we define our model. Under the specified model we replace the missing values with the estimates which are simply their conditional expectations. Furthermore, in order to assess their accuracy we need the conditional variances as well. Formally, we need to determine

$$m_{t,s} = E\left[\sum_{i=1}^{s} q_{t+i} e^{y_{t+i}} \,\middle|\, Y\right], \tag{4}$$

$$V_{t,s} = \mathrm{Var}\left[\sum_{i=1}^{s} q_{t+i} e^{y_{t+i}} \,\middle|\, Y\right]. \tag{5}$$

Although the conditional means are easily estimated by using known results of log-normal variables, the variances are more complicated due to correlations between the smoothed state variables (see Durbin and Koopman (2002, section 4.5)). Therefore, we rely on simulations (see Durbin and Koopman (2002)). Additionally, these simulations allow easy constructions for the prediction intervals which are analytically intractable because the distribution of the sum of the log-normal variables cannot be given in a closed form.

For simulating the missing observations conditionally on Y, we simulate realizations $(\tilde{\beta}, \tilde{\epsilon})$ from their joint conditional distribution $p(\beta, \epsilon|Y)$. Then, simulated observations are obtained from $\tilde{y}_t = X_t\tilde{\beta}_t + \tilde{\epsilon}_t,\ t = 1, \ldots, n$. As we are simulating conditionally on Y, $\tilde{y}_t = y_t$ if y_t is observed, as y_t belongs to Y. The simulation from $p(\beta, \epsilon|Y)$ can be done by augmenting state vector β_t with disturbance ϵ_t, similarly as in Durbin and Koopman (2001, p. 131), and

using the simulation smoothing algorithm of Durbin and Koopman (2002) for the augmented state vector. With a large number of replications the conditional mean (**??**) and variance (**??**) are computed naturally as averages. More specifically let $\tilde{y}_{1j}, \ldots, \tilde{y}_{nj}$ be the j^{th} simulated series, and

$$\tilde{m}_{s,t,j} = \sum_{i=1}^{s} q_{t+i} e^{\tilde{y}_{t+i,j}}.$$

Then, with N replicates, the conditional expectations and variances are obtained respectively as

$$m_{t,s} = \frac{1}{N} \sum_{j=1}^{N} \tilde{m}_{s,t,j},$$

$$V_{t,s} = \frac{1}{N} \sum_{j=1}^{N} (\tilde{m}_{s,t,j} - m_{t,s})^2.$$

Assuming that the estimated model is true, the accuracy of the yearly total nutrient fluxes can be computed in terms of prediction intervals. The prediction interval with coverage probability $1-2\alpha$ is found by taking the r^{th} smallest and the r^{th} largest value among $\{\tilde{m}_{s,t,j}\}, j = 1, \ldots, N$ with $r = N\alpha$; denoted as $\tilde{m}_{s,t,\text{low}}$ and $\tilde{m}_{s,t,\text{up}}$. Assuming the estimated parameters true, the required prediction interval is

$$[\tilde{m}_{s,t,\text{low}}, \tilde{m}_{s,t,\text{up}}].$$

The other measure of accuracy is the coefficient of variation

$$\frac{\sqrt{V_{t,s}}}{m_{t,s}}.$$

All the computations in this paper have been done in R (R Development Core Team, 2012), using the state space modelling package KFAS (Helske, 2012), where the simulation of the state vector is done using the simulation smoothing with two antithetic variables in order to reduce the error due to the simulation (Durbin and Koopman, 2002).

2.2 Model fitting and evaluation

The unknown parameters of the nutrient concentration model can be estimated by maximum likelihood method, using the Kalman filter for computing the log-likelihood of the model. The Kalman filter updating formulas yield us the predicted state $b_{t+1} = E(\beta_{t+1} \,|\, y_1, \ldots, y_t)$, the prediction $X_{t+1}b_{t+1}$ for y_{t+1}, the prediction error $v_{t+1} = y_{t+1} - X_{t+1}b_{t+1}$, and the prediction error variance (or the covariance matrix in multivariate case) $\mathrm{Var}(v_t) = F_t$.

The log-likelihood of a linear Gaussian state space model can be written in terms of prediction errors and their covariance matrices which in applications depend on unknown parameters. Let us denote the parameter vector by ψ, and let $v_{t,\psi}$ and $F_{t,\psi}$ be the prediction errors and their covariance matrices under ψ. Then the likelihood is given by

$$\log L(\psi) = -\frac{np}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{n}(\log|F_{t,\psi}| + v'_{t,\psi}F_{t,\psi}^{-1}v_{t,\psi}),$$

where p is the the dimension of y_t.

The non-stationary part of the state vector is initialized by the diffuse method suggested by (Durbin and Koopman, 2001), whereas the stationary components are assumed to have a stationary distribution at start. When the series $\{y_t\}$ is multivariate, we transform it into a univariate form as in Durbin and Koopman (2001). This enables us to treat totally and partially missing values automatically as well as automatically adjust the likelihood correctly.

The effect of model uncertainty, comprising parameter uncertainty and the uncertainty due to model choice, is evaluated by removing k nutrient measurement vectors from the dataset. The model is then fitted to the thinned data. Let f_k be the total nutrient flux of the the removed days,

and $\widehat{f}_k$ is the corresponding figure estimated using the thinned dataset. The relative error due to thinning is then $(\widehat{f}_k - f_k)/f_k$. Assuming the model is true and ignoring the parameter estimation error, the difference $e_k = \widehat{f}_k - f_k$ has mean zero. Furthermore, with larger k the average error per day e_k/k tends to be smaller. The same is true also for the relative error $e_k/f_k = (e_k/k)/(f_k/k)$. Therefore, if we plot the absolute relative errors $|e_k|/f_k$ on the thinning size k, we expect to see a decreasing curve, given our model is true. However, if these values remain more or less constant or are increasing, then our model is severely biased.

The overall effect of thinning is assessed through a Monte Carlo experiment. We remove randomly k nutrient measurement vectors and compute the mean relative error

$$\mathrm{MRE}_k = \frac{1}{B} \sum_{i=1}^{B} \frac{|\widehat{f}_{i,k} - f_{i,k}|}{f_{i,k}}, \tag{6}$$

where B denotes the number of random replicates and i refers to i-th replicate.

3 Data

Our data consist of the concentrations of total phosphorus and total nitrogen, and daily water flow measurements from four rivers located in southern Finland, Paimionjoki, Aurajoki, Porvoonjoki and Vantaanjoki, during 1985–2010. The nutrient data are taken from the databases of the Finnish Environment Institute. Total phosphorus and total nitrogen concentrations have been determined spectrometrically from water samples after digestion with peroxodisulphate.

The catchments of these rivers all have a high proportion of the agricultural land (24–43%, Table 1) and the soil is dominated by clay, which renders the water turbid. Much of the phosphorus in these rivers is transported in association with eroded soil particles. In addition, the catchments contain only few lakes (lake percentage 0.3-2.6), which results in high day to day variation in flow. In all the rivers, agriculture is the major source of nutrients.

At the beginning of our observation period, the Porvoonjoki has received substantialwastewater loading from the city of Lahti, but due to improved treatment the share of wastewater to total loading has decreased with time, to an average 12% of the anthropogenic loading. In the Vantaanjoki, the respective proportion of wastewater loading is 6.3%, while in the other two rivers it is below 1%.

[Table 1 about here]

Daily measurements on nutrient concentrations are available for only 5–10% of the time period while flow measurements are usually available for each day. A few flow measurements are missing in the Paimionjoki and Aurajoki series. For the Paimionjoki, flow measurements are missing from mid-October to mid-November for 2004, whereas for the Aurajoki, flow values are missing on a single day in 1985, and on a total of 99 days between 2004–2010. The missing flow measurements in Paimionjoki and Aurajoki are estimated from an auxiliary four-variate state space model defined as in (1) and (2) with all matrices X_t and T being identity matrices. The model is called a local level model, e.g. see Harvey (1989). Amisigo and van de Giesen (2005) have used a similar model to patch gaps in daily riverflow series.

4 Results

4.1 Relating nutrient concentration and river flow

It can be argued, as has been done by Wartiovaara (1975) and Rankinen et al. (2010), that the high water flow due to the precipitation has two opposite effects on the nutrient loadings. Precipitation increases the diffuse loading from the agriculture while simultaneously diluting wastewater loading. We have tried to take both these aspects into account. In Figure 1 we have plotted the concentrations on the flow, both in logarithms, but due to zero values we have first added one to the flow values. To address both of the mutually opposing effects caused by precipitation induced high flows, we have decided to regress the log-concentration y_t on both $\log(1+q_t)$ and $1/\log(2+q_t)$. In the latter we have added 2 to ensure a finite value. Figure 1 includes also some regression curves: the loess curves of first degree (Cleveland and Devlin, 1988), and the ordinary least squares regression of y_t on $\beta_0+\beta_1 \log(1+q_t)$ and on $\beta_0 + \beta_1 \log(1+q_t) + \beta_2/\log(2+q_t)$.

[Figure 1 about here]

By visual inspection the relation between the the concentration and the flow seems to be linear or slightly curved in a log-scale. Moreover, the loess curve and the regression curve from model with two predictor variables are quite close to each other, whereas the regression line from model with one predictor lies apart, especially for nitrogen measurements. Therefore, in some cases it seems clearly beneficial to include both $x_{1,t} = \log(1+q_t)$ and $x_{2,t} = 1/\log(2+q_t) = 1/\log(1+e^{x_{1t}})$ as the predictor variables in the model. In order to treat all series equally both predictors are present in each model. Note

that this visual inspection with regression and loess curves is about finding the proper relationship between concentration and flow, and it ignores the time aspect of the problem which, as we will later see, is an important part of the modelling.

4.2 State space specification

As the phosphorus and nitrogen concentration measurements are correlated, we model them together but separately for each river. The model applied to each river is of the form

$$\begin{aligned} y_t^P &= \mu^P + \alpha_t^P + \beta_{1,t}^P x_{1,t} + \beta_{2,t}^P x_{2,t} + \epsilon_t^P, \\ y_t^N &= \mu^N + \alpha_t^P + \beta_{1,t}^N x_{1,t} + \beta_{2,t}^N x_{2,t} + \epsilon_t^N, \\ \alpha_{t+1} &= T\alpha_t + \xi_t, \\ \beta_{t+1} &= \beta_t + \eta_t, \end{aligned} \tag{7}$$

where (y_t^P, y_t^N) is a bivariate processes of the logarithms of phosphorus and nitrogen concentrations, respectively, β_t consists of all coefficients $\beta_{j,t}^i$, $i = P, N, j = 1, 2$, and α_t consists of zero-mean first order autoregressive process α_t^P and α_t^N with $T = \text{diag}[\phi^P, \phi^N]$ containing the corresponding autoregressive parameters. The disturbance processes $\epsilon_t \sim N(0, \Sigma_\epsilon)$, $\eta_t \sim N(0, \Sigma_\eta)$ and $\xi_t \sim N(0, \Sigma_\xi)$ independently of each other. For simplicity, Σ_η is assumed to be a diagonal matrix. When the diagonal elements are positive the regression coefficients vary according to a random walk allowing the dependence between the flow and the nutrient concentration to change in time.

Note that the model collapses to an ordinary regression model when $\Sigma_\eta = 0$ and $T = 0$ (i.e. $\phi^P = \phi^N = 0$). The first restriction means that the regression coefficients β_t are constants. The second one implies that level

processes α_t^P, α_t^N are white noise processes merged into the errors ϵ_t^P and ϵ_t^N, respectively.

Zero variances for the components of coefficient process β_t are sometimes obtained. The state space modeling automatically handles the zero variances in the covariance matrices, so that the time invariant regression coefficients coincide with the appropriate generalized least squares estimates. Also the simulation algorithm is capable of handling the constant states without modifications.

The long-term seasonal weather conditions such as the starting times of snow-melt and autumn rains as well as the short-term weather conditions such as daily temperature or precipitation also affect concentrations. We assume here, that their effects come mainly through flow. In addition we assume that other environmental effects are mostly captured by the latent autoregressive level processes and coefficient processes of the flow series. We deliberately aim at a parsimonious model with practical formulas for the interpolation of the nutrient fluxes, although the true phenomena behind the variation of nutrient concentrations are obviously more complicated than our model suggests.

4.3 Estimated nutrient fluxes and model parameters

The yearly estimates of the nutrient fluxes obtained by simulating the model are given in Table 2 in an Appendix. Yearly estimates of nutrientï¿½fluxes with their simulated 95% prediction intervals are also shown in the Figure 2. Each river exhibits a similar fluctuating patterns without a clear trend. Especially yearly phosphorus fluxes, but also nitrogen fluxes clearly peak

in 2008, followed by an even larger drop in 2009. Overall, fewer nutrient measurements result in somewhat wider prediction intervals for Porvoonjoki and Vantaanjoki than for Paimionjoki and Aurajoki.

[Figure 2 about here]

The estimated values of the unknown variance and autoregressive parameters are shown in Table 3.

[Table 3 about here]

Occasionally the estimation process yields the variance estimates close to zero (i.e. values less than 10^{-8}). In such cases these are replaced with fixed zeros and estimation process for other parameters is repeated. In all cases the likelihood remained practically unchanged.

Standard errors of the estimates are computed by inverting the Hessian matrix given by the optimization function optim in R. The variance parameters are estimated in logarithmic scale. Using their standard errors confidence intervals for the log-variances are obtained. Then the confidence intervals for the variances themselves are easily derived.

In all models, the values of autoregressive parameters are close to one $(0.95 - 0.98)$, and therefore the standard errors might not be very useful, as the sampling distributions are far from normal distribution. The correlations ρ_{ξ^P,ξ^N} between the disturbances of the autoregressive processes are around 0.5 for all rivers. This indicates moderate long-term correlation between the underlying phosphorus and nitrogen concentration processes at a given flow level. The instantaneous correlations $\rho_{\epsilon^P,\epsilon^N}$, again given the flow, are smaller and more variable: 0.2 or slightly higher in the Paimionjoki and Porvoonjoki, and negligible in the Aurajoki and Vantaanjoki.

The coefficient processes are shown in Figure 3. Somewhat larger regression coefficients of the reciprocal log-flow of the Porvoonjoki and Vantaanjoki compared to those of Paimionjoki and Aurajoki are in concordance with the fact that the former rivers are subject to higher wastewater loads. Otherwise, the interpretation of the regression coefficient processes is difficult. Nevertheless, as predictive tools, individual river-specific models appear to be highly useful.

[Figure 3 about here]

4.4 Model criticism

We have also tested models where the autoregressive processes have been replaced with random walks (ie. $\phi^P = \phi^N = 1$), and a multivariate local level model without regressors but where the concentration processes are augmented with water flow. In addition, we also tested the ordinary multivariate regression model. All these models yield large autocorrelations of the standardized residuals and in case of time varying models there is a clear inverse relationship between the size of residuals and observed concentration. These apparent violations are avoided by using the model (5). However, even despite obvious violations of model assumptions yearly estimates of the nutrient fluxes from different time varying models have very similar coefficients of variation with deviations being usually less than one percentage point.

In the case of the ordinary regression model the coefficients of variation are often substantially smaller. In Figure 4 the coefficients of variation are plotted against the yearly sample sizes of the concentration measurements. The coefficients of variation from the model (5) depend on the yearly sample

sizes, while results from the ordinary regression model are overoptimistic and counterintuitive: uncertainty in the yearly flux estimate is independent from the amount of measurements in a given year. Both models use the daily water flow for the prediction of the missing concentration measurements, but the ordinary regression is immune to the time order of the measurements and only the total number of measurements is important. However, we acknowledge that since yearly flux estimates are always conditioned on the model, all models underestimate the true errors of yearly flux estimates and none of the models considered is "true".

[Figure 4 about here]

The quantile-to-quantile plots of the standardized residuals of the models reveal heavier tails compared to the normal distribution. This would be problematic if the interest is on the daily values, but because we are interested in yearly values we believe that the possible non-normality is not critical here. This is because the yearly measure of nutrient fluxes is a sum, which tends to be more normal than its components by the central limit theorem. For evaluating the effects of non-normality, we have made a simulation experiment, where the errors ϵ_t are a random sample from a heavy tailed bivariate t-distribution with 3 degrees of freedom scaled to have $Var(\epsilon_t) = \Sigma_\epsilon$. New values representing the concentration measurements, on the same days as the true ones, are then simulated from the model with the estimated parameters. Using these simulated measurements our proposed model is fitted (under Gaussian assumptions) as well as the coefficients of variation computed for the yearly fluxes. The coefficients of variation from the simulation are, on the average, within one percentage point of those obtained from the

actual dataset thus displaying the negligible effect of non-normality.

The main purpose of our model is to estimate the yearly nutrient flux. To this end we developed the thinning experiment explained at the end of section 2.2. We have made five experiments by randomly removing 10%, 20%, 30%, 40% and 50% from the concentration values. The resulting relative errors (4) are reported in Table 4 in the Appendix. The number of simulations is $B = 2000$, and each time parameters are re-estimated. If the model is correct we expect a decreasing trend, and this is mostly what we observe. The loss of relative accuracy with 30% thinning is about 5% or less. However, the mean absolute errors $\mathrm{MAE}_k = \sum_{i=1}^{B} |\widehat{f}_{i,k} - f_{i,k}|/B$ increase rapidly, as expected, when thinning is increased (Table 5 in the Appendix). When measuring bias using average errors $\mathrm{AE}_k = \sum_{i=1}^{B} (\widehat{f}_{i,k} - f_{i,k})/B$ (Table 6 in the Appendix), the total phosphorus flux is underestimated in all rivers, whereas the total nitrogen flux is usually overestimated, except for Aurajoki, where the nitrogen flux is underestimated. Overall, the results suggest that our model performs well enough for practical purposes. For the ordinary regression model the mean relative and absolute errors are always larger, and prominently so for nitrogen fluxes. The average errors show that the ordinary regression model overestimates the nitrogen fluxes more than our model, whereas the bias of phosphorus fluxes is slightly smaller. Finally, we note that removing predictor $1/\log(2 + q_t)$ from the final model (5), always worsens model performance compared to including it.

5 Discussion

We have used Gaussian state space models with partially sparse data for modelling the yearly nutrient fluxes of four rivers running through catchments dominated by agricultural land use. The large proportion of "missing" daily nutrient concentration measurements for corresponding daily flow measurements increased the uncertainty regarding the model selection, parameter estimation and prediction, thus encouraging the use of models with simple structure and large flexibility.

During the observational period covered by this study Finnish agricultural farmlands experienced a substantial decrease in phosphorus and nitrogen balance (Aakkula et al., 2011). Despite this drastic decrease in nutrient balance we did not observe any corresponding trends in nutrient fluxes over the last 25 years for any of the four rivers examined here. Greatly reduced nutrient balances do not always lead to concurrent reduction in riverine nutrient fluxes, for example, due to high nutrient reserves in soil and groundwater (e.g. Stålnacke et al. (2004)). Moreover, although nutrient balances form a crucial indicator of the risk of nutrient losses from agriculture, changes in other agricultural practices, or in climate, may have had an opposite effect on the load (Ekholm et al., 2007).

While we have reported results when the daily water flow is only predictor variable, we have also augmented the model with locally important variables such as daily air temperature, precipitation and several functions of these. To examine the possible effect of large scale climate patterns we have also used the North Atlantic Oscillation (NAO) and Arctic Oscillation (AO) indices in combination with flow. Additions of variables operating at either small

(temperature, precipitation) or large scales (NAO or AO) did not improve results for any of the models we used.

Many studies examining nutrient dynamics of rivers have stated the need for extensive datasets to be able to make precise statements on the nutrient flux (e.g. Rekolainen et al. (1991). While we are conscious that the thinning of an originally sparse data by half can include possible computational caveats and thus may lead to artifacts, our results seem to indicate that when daily flow data are available, relatively sparse data on nutrient concentrations can be used to estimate yearly fluxes. If the aim of monitoring is to assess yearly fluxes of principal nutrients from agriculturally dominated watersheds to receiving downstream locations (e.g. the sea), our findings imply the potential to lower the frequency of water quality (i.e. nutrient) sampling intensities for rivers with permanent gauging stations and long-term records of flow. It should be noted that these concentration measurements could be used for other types of analysis as well, where the number of samples cannot not be reduced.

6 Acknowledgements

The authors thank two anonymous referees for their comments which have led to considerable improvements. Support by grant 110045 from Emil Aaltonen Foundation for J. Helske is gratefully acknowledged.

References

Aakkula J, Kuussaari M, Rankinen K, Ekholm P, Heliölä J, Hyvönen T, Kitti L, Salo T. 2011. Follow-up study on the impacts of agri-environmental measures in Finland. OECD workshop on the evaluation of agri-environmental policies. The Johann Heinrich von Thünen Institute Germany. URL `http://www.oecd.org/dataoecd/31/56/48143799.pdf`.

Amisigo BA, van de Giesen NC. 2005. Using a spatio-temporal dynamic state-space model with the em algorithm to patch gaps in daily riverflow series. *Hydrology and Earth System Sciences* **9**(3):209–224. doi:10.5194/hess-9-209-2005. URL `http://www.hydrol-earth-syst-sci.net/9/209/2005/`.

Cleveland WS, Devlin SJ. 1988. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* **83**(403):596–610. doi:10.2307/2289282.

Durbin J, Koopman SJ. 2001. *Time Series Analysis by State Space Methods*. Oxford University Press, New York.

Durbin J, Koopman SJ. 2002. A simple and efficient simulation smoother for state space time series analysis. *Biometrika* **89**:603–615. doi:10.1093/biomet/89.3.603.

Ekholm P, Granlund K, Kauppila P, Mitikka S, Niemi J, Rankinen K, Räike A, Räsänen J. 2007. Influence of EU policy on agricultural nutrient losses and the state of receiving surface waters in Finland. *Agricultural and Food Science* **16**(4):282–300.

Harvey AC. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

Helske J. 2012. *KFAS: Kalman Filter and Smoother for Exponential Family State Space Models*. R package version 0.9.11.

Kauppila P, Koskiaho J. 2003. Evaluation of annual loads of nutrients and suspended solids in baltic rivers. *Nordic Hydrology* **34**(3):203–220. doi: 10.2166/nh.2003.013.

Kronvang B, Bruhn A. 1996. Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams. *Hydrological processes* **10**:1483–1501. doi:10.1002/(SICI)1099-1085(199611)10:11<1483::AID-HYP386>3.0.CO;2-Y.

Quilbé R, Rousseau AN, Duchemin M, Poulin A, Gangbazo G, Villeneuve JP. 2006. Selecting a calculation method to estimate sediment and nutrient loads in streams: Application to the beaurivage river (Québec, Canada). *Journal of Hydrology* **326**:295–310. doi:10.1016/j.jhydrol.2005.11.008.

R Development Core Team. 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

Rankinen K, Ekholm P, Sjöblom H, Rita H. 2010. Nutrient losses from catchments and the governing factors. In J Aakkula, T Manninen, M Nurro, editors, *Follow-up study on the impacts of agri-environment measures (MYT-VAS3) - Mid-term report*, Reports of Finland Ministry of Agriculture and Forestry 1, 122–131. In Finnish.

Rekolainen S, Posch M, Kämäri J, Ekholm P. 1991. Evaluation of the accuracy and precision of annual phosphorus load estimates from two agricultural basins in Finland. *Journal of Hydrology* **128**:237–255. doi: 10.1016/0022-1694(91)90140-D.

Stålnacke P, Vandsemb S, Grimvall A, Jolankai G. 2004. Changes in nutrient levels in some Eastern European rivers in response to large-scale changes in agriculture. *Water Science & Technology* **49**(3):29–36.

Wartiovaara J. 1975. Amounts of substances discharged by rivers off the coast of Finland. *Publications of the Water Research Institute* **13**. In Finnish.

Young TC, DePinto JV, Heidtke TM. 1988. Factors affecting the efficiency of some estimators of fluvial total phosphorus load. *Water Resources Research* **24**:1535–1540. doi:10.1029/WR024i009p01535.

7 Appendix

[Table 2 about here]

[Tables 4, 5 and 6 about here]

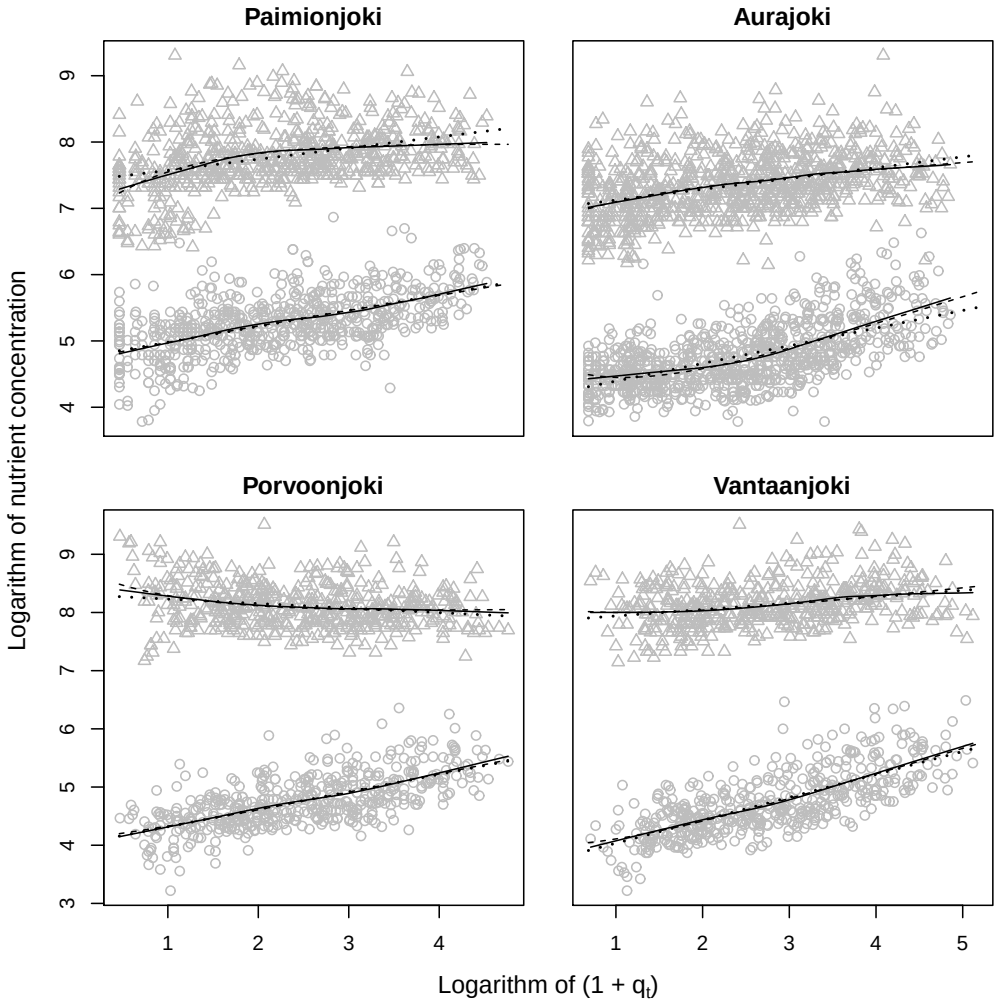

Figure 1: Scatter plots of log-concentrations of nutrients and $\log(1 + q_t)$, with loess curves of first degree (solid line), the regression curves with one explanatory variable (dotted line) and with two explanatory variables (dashed line). The circles correspond to the phosphorus measurements and triangles to the nitrogen measurements.

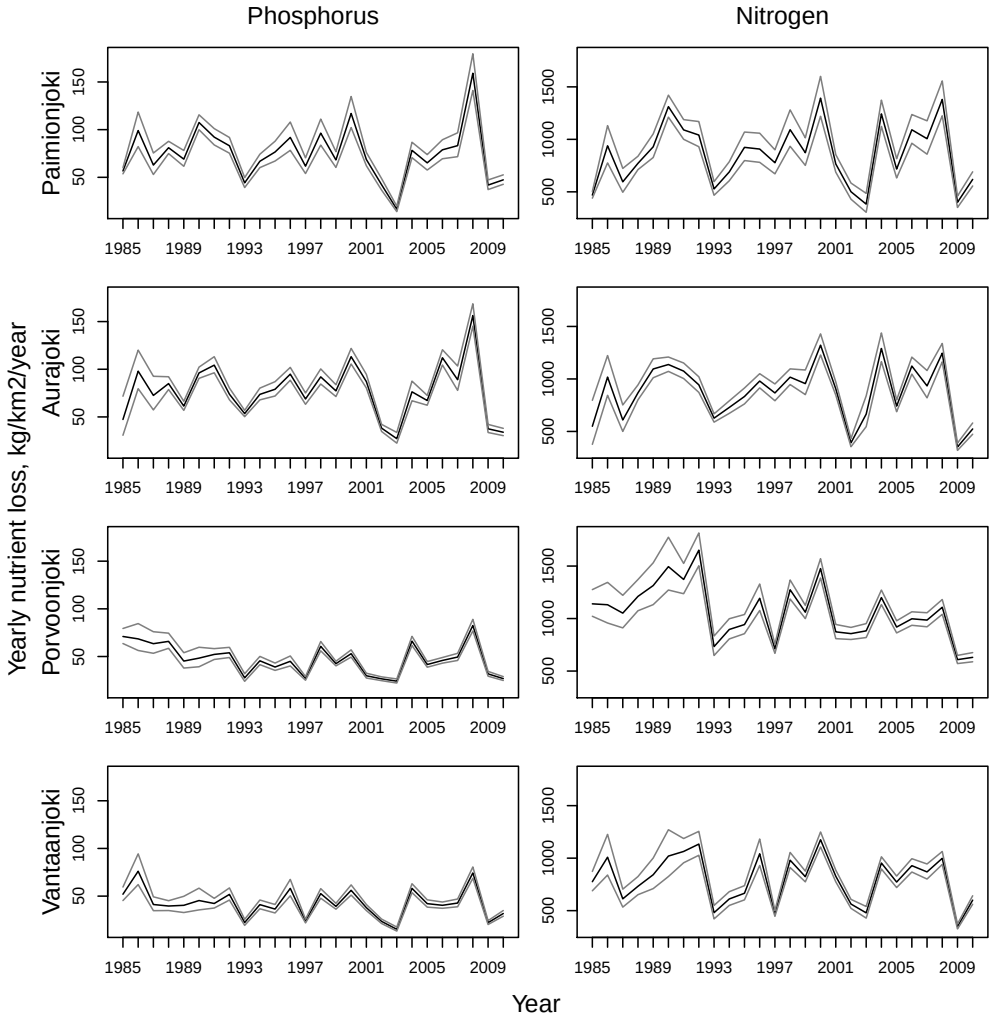

Figure 2: The estimated values and the simulated 95% prediction intervals of the yearly phosphorus and nitrogen fluxex.

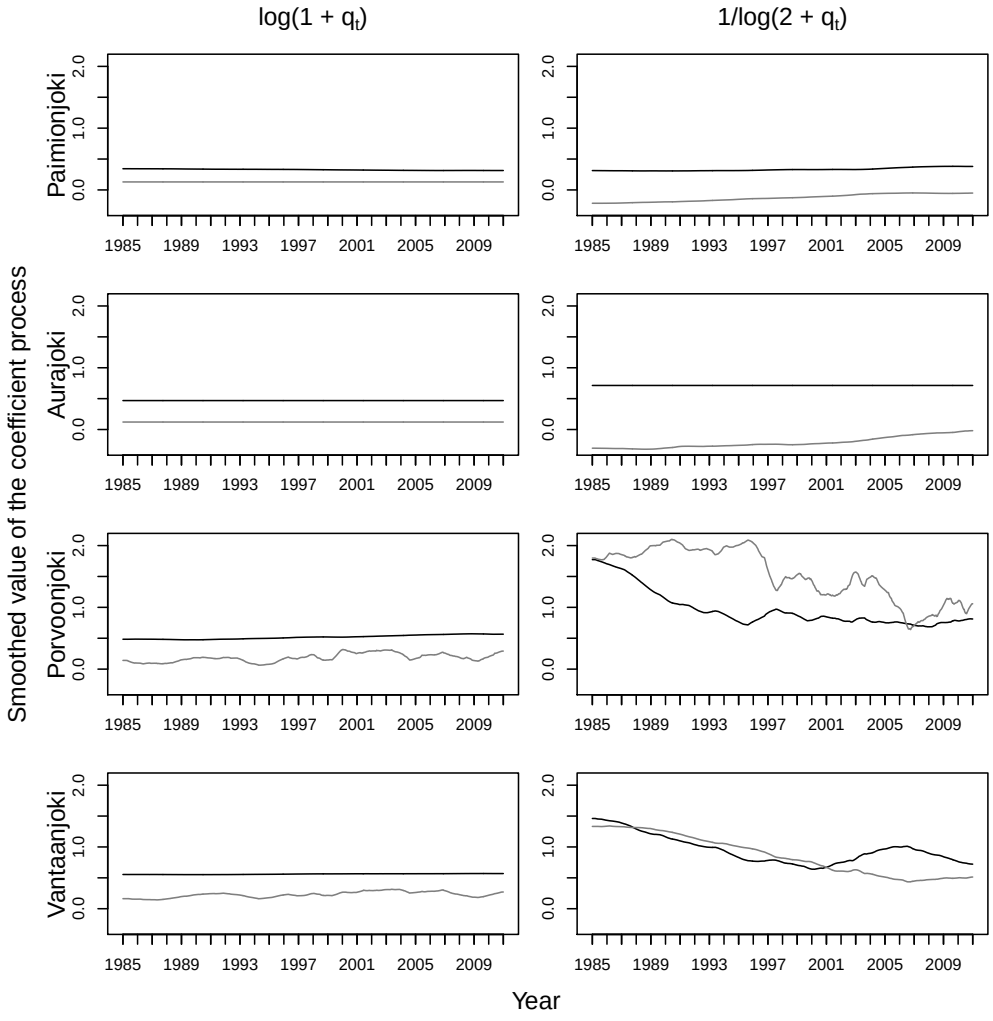

Figure 3: The smoothed coefficient processes corresponding to the predictor variables $\log(1 + q_t)$ (left) and $1/\log(2 + q_t)$ (right) for all four rivers. The black lines represent the processes corresponding to phosphorus observations and the grey lines correspond to the nitrogen observations. Constant horizontal line corresponds to the null variance of the coefficient process.

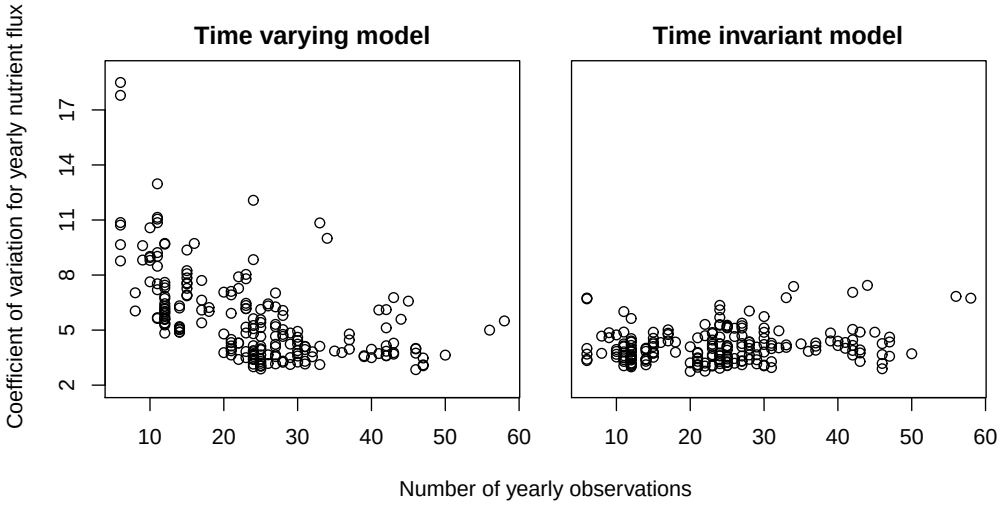

Figure 4: The relationship between coefficients of variation for yearly nutrient flux and the number of yearly nutrient concentration observations. Figure on left corresponds to final model (5), and the figure on right to the ordinary regression model.

Table 1: Catchment characteristics of the rivers studied.

	Paimionjoki	Aurajoki	Porvoonjoki	Vantaanjoki
Catchment area, km^2	1088	874	1273	1686
Lakes, %	1.6	0.3	1.3	2.3
Agricultural land, %	42.8	36.8	31.2	23.8
Constructed area, %	2.5	4.8	4.1	9.2
Mean flow, $m^3 s^{-1}$	9.5	8.5	12.7	15.9
Wastewater load, % of total flux	0.5	0.7	12.3	6.3

Table 2: Annual fluxes (kg/km^2/year) and the coefficient of variation in percentages for each river and nutrient.

	Paimionjoki		Aurajoki		Porvoonjoki		Vantaanjoki	
	P	N	P	N	P	N	P	N
1985	57 (2.8)	469 (3.5)	53 (17.8)	547 (18.5)	71 (5.7)	1140 (5.6)	52 (7.0)	776 (6.0)
1986	99 (9.2)	939 (9.7)	100 (8.8)	1020 (9.6)	68 (10.7)	1132 (8.8)	76 (10.9)	1009 (9.7)
1987	63 (9.4)	596 (9.7)	72 (11.1)	609 (9.7)	63 (9.0)	1052 (7.5)	41 (9.0)	613 (7.2)
1988	81 (4.0)	772 (4.3)	85 (3.8)	867 (4.1)	66 (6.0)	1211 (6.3)	40 (6.7)	732 (6.0)
1989	69 (6.1)	929 (6.1)	62 (3.8)	1094 (4.1)	45 (9.0)	1314 (7.6)	40 (10.6)	842 (8.8)
1990	107 (3.6)	1312 (4.0)	96 (3.1)	1137 (3.1)	48 (10.9)	1494 (8.5)	45 (13.0)	1020 (11.0)
1991	92 (4.8)	1090 (4.5)	101 (3.6)	1076 (3.6)	52 (5.6)	1372 (5.3)	42 (6.2)	1064 (5.6)
1992	83 (5.2)	1041 (5.8)	72 (3.9)	945 (4.1)	54 (5.0)	1651 (4.9)	52 (6.2)	1134 (5.1)
1993	44 (6.0)	528 (6.2)	54 (3.1)	627 (3.3)	28 (7.4)	733 (6.4)	22 (7.3)	481 (6.8)
1994	67 (5.4)	690 (6.6)	73 (4.2)	729 (4.2)	46 (4.8)	896 (5.6)	41 (5.7)	611 (5.4)
1995	77 (6.9)	924 (7.6)	78 (4.6)	832 (4.7)	39 (5.1)	943 (4.9)	36 (6.3)	666 (5.2)
1996	92 (8.1)	907 (7.8)	94 (3.5)	980 (3.6)	45 (5.9)	1193 (5.3)	58 (7.6)	1042 (6.1)
1997	62 (6.9)	777 (7.5)	68 (4.4)	867 (4.8)	27 (3.7)	711 (3.1)	24 (4.1)	473 (3.0)
1998	96 (7.3)	1092 (8.2)	92 (4.0)	1017 (3.9)	61 (4.1)	1274 (3.6)	53 (4.8)	980 (3.8)
1999	68 (6.1)	872 (7.7)	79 (4.1)	950 (6.1)	43 (2.9)	1060 (3.1)	39 (4.0)	823 (3.2)
2000	117 (6.9)	1392 (7.1)	115 (3.7)	1317 (3.8)	53 (3.7)	1476 (3.2)	56 (4.8)	1176 (3.1)
2001	69 (5.1)	766 (5.6)	88 (4.0)	905 (3.6)	30 (4.3)	873 (4.0)	38 (4.3)	828 (3.4)
2002	42 (7.3)	501 (7.9)	39 (4.9)	391 (5.0)	27 (3.7)	856 (3.5)	23 (5.1)	563 (3.9)
2003	17 (8.8)	383 (12.1)	30 (10.0)	662 (10.8)	24 (4.5)	883 (3.9)	15 (7.1)	479 (5.9)
2004	78 (5.2)	1243 (5.2)	76 (6.1)	1251 (5.3)	66 (3.6)	1200 (3.0)	58 (4.1)	953 (3.1)
2005	65 (6.3)	717 (6.5)	65 (3.8)	722 (3.9)	42 (3.6)	919 (3.2)	42 (4.7)	772 (3.7)
2006	79 (6.4)	1091 (6.3)	106 (3.5)	1075 (3.6)	46 (3.2)	998 (3.3)	40 (4.2)	928 (3.5)
2007	83 (7.8)	1006 (8.0)	81 (6.6)	894 (6.8)	49 (3.8)	985 (3.5)	43 (4.8)	869 (4.3)
2008	159 (6.2)	1380 (6.2)	153 (3.7)	1245 (3.7)	83 (3.7)	1108 (3.3)	74 (4.1)	998 (3.2)
2009	42 (6.3)	400 (7.0)	35 (5.5)	335 (5.0)	32 (3.9)	609 (3.3)	22 (4.7)	348 (3.4)
2010	47 (5.4)	619 (5.6)	33 (5.6)	501 (5.1)	27 (3.8)	630 (3.6)	32 (4.8)	600 (3.4)

Table 3: Estimates of the unknown parameters and their standard errors in parenthesis.

	Paimionjoki	Aurajoki	Porvoonjoki	Vantaanjoki
$\sigma^2_{\eta_1^P}$	1.3×10^{-7}	0	6.4×10^{-7}	9.0×10^{-8}
$\log(\sigma^2_{\eta_{11}})$	-15.83 (2.07)	—	-14.27 (1.05)	-16.22 (3.75)
$\sigma^2_{\eta_2^P}$	1.4×10^{-6}	0	7.8×10^{-5}	5.5×10^{-5}
$\log(\sigma^2_{\eta_{12}})$	-13.46 (3.39)	—	-9.46 (0.89)	-9.81 (0.88)
$\sigma^2_{\eta_1^N}$	0	0	2.1×10^{-5}	7.7×10^{-6}
$\log(\sigma^2_{\eta_{21}})$	—	—	-10.78 (0.64)	-11.78 (0.98)
$\sigma^2_{\eta_2^N}$	2.9×10^{-6}	5.8×10^{-6}	3.1×10^{-4}	3.9×10^{-5}
$\log(\sigma^2_{\eta_{22}})$	-12.74 (1.52)	-12.06 (1.00)	-8.08 (1.03)	-10.15 (0.82)
$\sigma^2_{\xi^P}$	5.4×10^{-3}	1.0×10^{-2}	1.1×10^{-2}	7.0×10^{-3}
$\log(\sigma^2_{\xi_1})$	-5.23 (0.16)	-4.58 (0.14)	-4.51 (0.19)	-4.96 (0.25)
$\sigma^2_{\xi^N}$	8.3×10^{-3}	1.1×10^{-2}	4.0×10^{-3}	4.0×10^{-3}
$\log(\sigma^2_{\xi_2})$	-4.80 (0.13)	-4.47 (0.13)	-5.51 (0.24)	-5.51 (0.25)
$\sigma^2_{\epsilon^P}$	3.1×10^{-2}	1.8×10^{-2}	6.5×10^{-3}	2.2×10^{-2}
$\log(\sigma^2_{\epsilon_1})$	-3.47 (0.13)	-4.02 (0.19)	-5.04 (0.89)	-3.81 (0.33)
$\sigma^2_{\epsilon^N}$	3.0×10^{-2}	1.7×10^{-2}	1.5×10^{-2}	1.3×10^{-2}
$\log(\sigma^2_{\epsilon_2})$	-3.49 (0.14)	-4.09 (0.22)	-4.21 (0.30)	-4.32 (0.35)
ρ_{ξ^P,ξ^N}	0.58 (0.04)	0.46 (0.04)	0.53 (0.06)	0.48 (0.06)
$\rho_{\epsilon^P,\epsilon^N}$	0.26 (0.08)	0.07 (0.12)	0.20 (0.29)	0.02 (0.24)
μ^P	4.47 (0.13)	3.83 (0.10)	3.11 (0.25)	2.72 (0.29)
μ^N	7.63 (0.15)	7.62 (0.11)	7.04 (0.23)	6.83 (0.24)
ϕ^P	0.98 (0.004)	0.95 (0.006)	0.95 (0.009)	0.96 (0.007)
ϕ^N	0.98 (0.003)	0.96 (0.005)	0.98 (0.006)	0.98 (0.005)

Table 4: The mean relative error percentages and their standard errors for the final model (5) and for the least squares regression model with two predictors (marked by †).

	Paimionjoki		Aurajoki		Porvoonjoki		Vantaanjoki	
	P	N	P	N	P	N	P	N
MRE_{10}	5.8 (0.09)	4.8 (0.08)	6.5 (0.11)	5.3 (0.11)	6.4 (0.11)	3.9 (0.07)	7.1 (0.13)	4.7 (0.08)
$MRE_{10}^{\dagger}$	7.0 (0.12)	7.4 (0.13)	8.9 (0.15)	8.0 (0.14)	7.6 (0.13)	6.7 (0.11)	8.9 (0.15)	7.6 (0.13)
MRE_{20}	4.5 (0.07)	3.6 (0.06)	5.4 (0.09)	4.6 (0.08)	5.0 (0.08)	3.0 (0.05)	5.8 (0.10)	3.7 (0.06)
$MRE_{20}^{\dagger}$	5.0 (0.08)	5.9 (0.10)	6.8 (0.11)	6.3 (0.10)	6.0 (0.10)	5.1 (0.09)	7.0 (0.11)	6.0 (0.10)
MRE_{30}	4.0 (0.06)	3.3 (0.06)	5.1 (0.08)	4.1 (0.07)	4.5 (0.07)	2.7 (0.05)	5.3 (0.09)	3.3 (0.06)
$MRE_{30}^{\dagger}$	4.5 (0.07)	5.1 (0.09)	6.2 (0.10)	5.3 (0.09)	4.9 (0.08)	4.5 (0.07)	5.9 (0.10)	5.1 (0.09)
MRE_{40}	3.6 (0.06)	3.2 (0.05)	4.9 (0.08)	3.8 (0.06)	4.3 (0.07)	2.7 (0.05)	5.2 (0.08)	3.1 (0.06)
$MRE_{40}^{\dagger}$	4.1 (0.07)	5.0 (0.08)	5.7 (0.09)	4.9 (0.08)	4.7 (0.08)	4.2 (0.07)	5.7 (0.09)	4.7 (0.09)
MRE_{50}	3.5 (0.06)	3.1 (0.05)	4.9 (0.08)	3.9 (0.06)	4.3 (0.07)	2.8 (0.05)	4.9 (0.08)	3.3 (0.06)
$MRE_{50}^{\dagger}$	3.9 (0.07)	4.6 (0.08)	5.5 (0.09)	4.9 (0.08)	4.6 (0.07)	4.1 (0.07)	5.4 (0.09)	4.8 (0.08)

Table 5: The mean absolute errors (metric tons) and their standard errors for the final model (5) and for the ordinary least squares regression model with two predictors (marked by †).

	Paimionjoki		Aurajoki		Porvoonjoki		Vantaanjoki	
	P	N	P	N	P	N	P	N
MAE_{10}	0.9 (0.02)	7.8 (0.14)	1.7 (0.03)	14.0 (0.32)	0.8 (0.02)	9.9 (0.17)	1.3 (0.03)	13.5 (0.24)
$MAE_{10}^{\dagger}$	1.1 (0.02)	11.8 (0.20)	2.2 (0.04)	20.3 (0.35)	1.0 (0.02)	16.7 (0.28)	1.5 (0.03)	21.7 (0.38)
MAE_{20}	1.4 (0.02)	11.8 (0.21)	2.7 (0.05)	23.7 (0.46)	1.3 (0.02)	15.1 (0.26)	2.0 (0.04)	20.9 (0.37)
$MAE_{20}^{\dagger}$	1.6 (0.03)	19.0 (0.31)	3.4 (0.06)	31.4 (0.52)	1.6 (0.03)	25.4 (0.43)	2.4 (0.04)	33.8 (0.59)
MAE_{30}	1.9 (0.03)	16.0 (0.28)	3.9 (0.07)	31.7 (0.55)	1.8 (0.03)	20.0 (0.35)	2.8 (0.05)	28.1 (0.50)
$MAE_{30}^{\dagger}$	2.1 (0.04)	24.8 (0.42)	4.7 (0.08)	40.3 (0.66)	1.9 (0.03)	33.6 (0.55)	3.1 (0.05)	43.6 (0.77)
MAE_{40}	2.3 (0.04)	20.9 (0.36)	5.0 (0.08)	39.1 (0.66)	2.3 (0.04)	26.9 (0.50)	3.6 (0.06)	35.3 (0.64)
$MAE_{40}^{\dagger}$	2.6 (0.04)	32.2 (0.53)	5.8 (0.10)	48.9 (0.83)	2.5 (0.04)	41.5 (0.71)	3.9 (0.07)	53.4 (0.96)
MAE_{50}	2.7 (0.05)	25.7 (0.44)	6.1 (0.10)	49.9 (0.79)	2.8 (0.05)	35.8 (0.65)	4.3 (0.07)	46.9 (0.82)
$MAE_{50}^{\dagger}$	3.0 (0.05)	37.8 (0.64)	6.9 (0.11)	61.2 (1.05)	3.0 (0.05)	51.3 (0.88)	4.7 (0.08)	68.7 (1.20)

Table 6: The mean errors (metric tons) and their standard errors for the final model (5) and for the ordinary least squares regression model with two predictors (marked by †).

	Paimionjoki		Aurajoki		Porvoonjoki		Vantaanjoki	
	P	N	P	N	P	N	P	N
ME_{10}	-0.3 (0.02)	-0.1 (0.22)	-0.9 (0.05)	-4.6 (0.43)	-0.3 (0.02)	-0.1 (0.28)	-0.4 (0.04)	0.7 (0.39)
$ME_{10}^{\dagger}$	-0.2 (0.03)	5.0 (0.31)	-0.8 (0.06)	4.7 (0.57)	-0.2 (0.03)	3.7 (0.46)	-0.3 (0.04)	5.0 (0.61)
ME_{20}	-0.6 (0.04)	0.6 (0.33)	-1.8 (0.07)	-9.0 (0.67)	-0.6 (0.04)	-0.5 (0.43)	-0.9 (0.06)	1.2 (0.59)
$ME_{20}^{\dagger}$	-0.2 (0.04)	10.7 (0.47)	-1.5 (0.09)	9.4 (0.85)	-0.4 (0.04)	8.8 (0.68)	-0.6 (0.07)	11.5 (0.92)
ME_{30}	-0.8 (0.05)	6.1 (0.43)	-2.8 (0.09)	-9.9 (0.87)	-0.9 (0.05)	1.8 (0.57)	-1.3 (0.07)	3.8 (0.80)
$ME_{30}^{\dagger}$	-0.3 (0.06)	15.2 (0.60)	-2.6 (0.12)	12.2 (1.09)	-0.4 (0.05)	13.8 (0.88)	-1.0 (0.08)	16.8 (1.19)
ME_{40}	-0.9 (0.06)	9.1 (0.55)	-3.7 (0.11)	-11.4 (1.07)	-1.3 (0.06)	2.6 (0.78)	-1.9 (0.09)	5.5 (1.01)
$ME_{40}^{\dagger}$	-0.2 (0.07)	22.2 (0.74)	-3.2 (0.14)	18.4 (1.31)	-0.7 (0.07)	19.0 (1.09)	-1.5 (0.11)	22.4 (1.45)
ME_{50}	-0.9 (0.07)	11.6 (0.68)	-4.6 (0.13)	-11.2 (1.34)	-1.5 (0.07)	3.9 (1.03)	-2.3 (0.11)	6.3 (1.33)
$ME_{50}^{\dagger}$	-0.2 (0.09)	25.4 (0.90)	-4.0 (0.17)	23.3 (1.64)	-0.9 (0.08)	21.8 (1.36)	-1.7 (0.13)	28.5 (1.84)