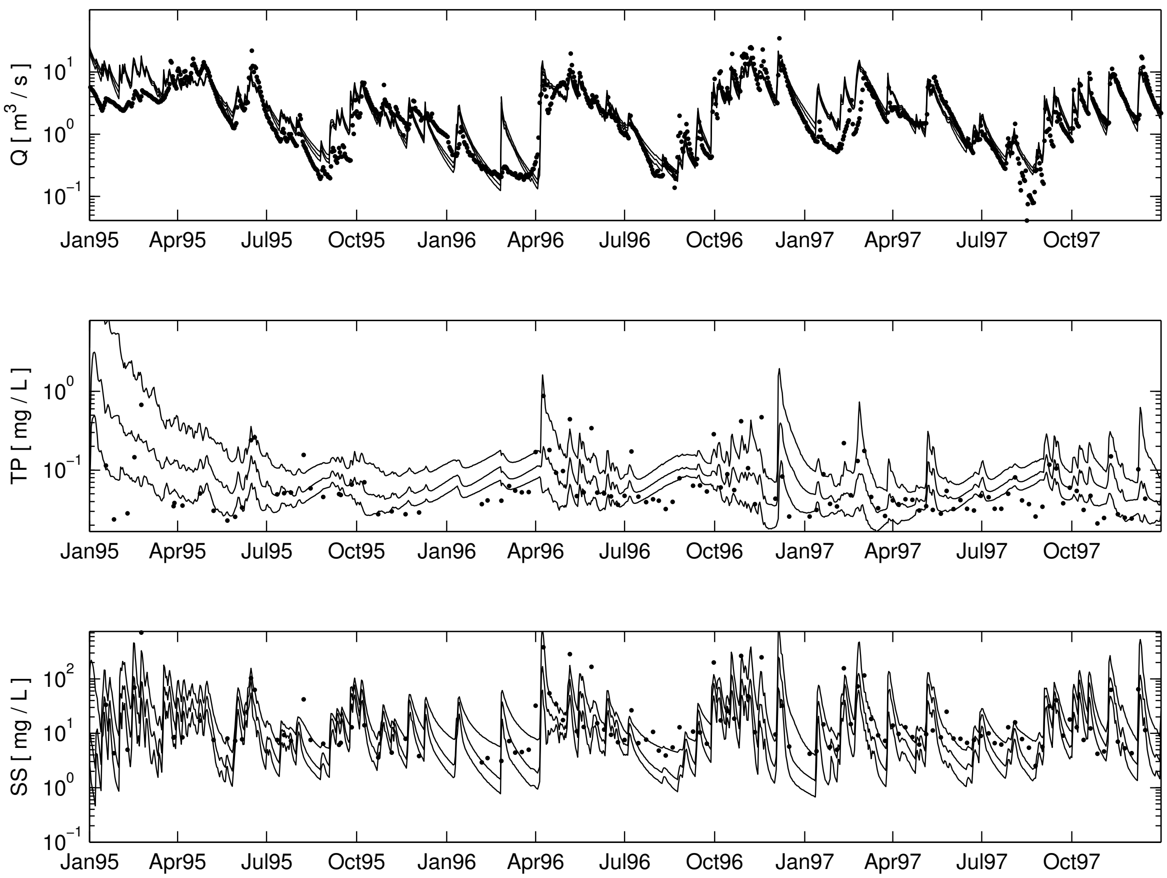


Bayesian parameter estimation on INCA-P highlights the importance of parameter uncertainty in simulating future scenarios

Environmental impact statement for

Bayesian uncertainty assessment of a semi-distributed integrated catchment model of phosphorus transport.

Authors: Starrfelt, J. and Kaste, Ø.

The article *Bayesian uncertainty assessment of a semi-distributed integrated catchment model of phosphorus transport* details the application of a Bayesian scheme for uncertainty assessment of the Integrated Catchment model of Phosphorus (INCA-P). The scheme includes an autocalibration procedure for arriving at posterior distributions of selected parameters and uses these distributions to generate predictions of phosphorus transport under changed land uses, while including the uncertainty surrounding the parameters. This generates distributions of simulated outputs, i.e. probabilistic statements of predictions and can serve as a more solid foundation for management decisions under uncertainty.

Title:

Bayesian uncertainty assessment of a semi-distributed integrated catchment model of phosphorus transport.

Authors:

Jostein Starrfelt[a*] and Øyvind Kaste[b]

[a]Norwegian Institute for Water Research (NIVA), Gaustadalléen 21, N-0349 Oslo, Norway, Phone: + 47 22 18 51 00, Fax: + 47 22 18 52 00, jos@niva.no

[b]Norwegian Institute for Water Research (NIVA), Branch Office South, Jon Lilletuns vei. 3, N-4879 Grimstad, Norway.

[*]Corresponding author.

ABSTRACT

Process-based models of nutrient transport are often used as tools for management of eutrophic waters, as decision makers need to judge the potential effects of alternative remediation measures, under current conditions and with future land use and climate change. All modelling exercises entail uncertainty arising from various sources, such as the input data, selection of parameter values and the choice of model itself. Here we perform Bayesian uncertainty assessment of an integrated catchment model of phosphorus (INCA-P). We use an auto-calibration procedure and an algorithm for including parametric uncertainty to simulate phosphorus transport in a Norwegian lowland river basin. Two future scenarios were defined to exemplify the importance of parametric uncertainty in generating predictions. While a worst case scenario yielded a robust prediction of increased loading of phosphorus, a best case scenario only gave rise to a reduction in load with probability 0.78, highlighting the importance of taking parametric uncertainty into account in process-based catchment scale modelling of possible remediation scenarios. Estimates of uncertainty can be included in information provided to decision makers, thus making a stronger scientific basis for sound decisions to manage water resources.

Keywords: river basin, modelling, INCA-P, phosphorus, Bayesian inference, uncertainty

1. INTRODUCTION

Eutrophication, often caused by excessive inputs of phosphorus (P) and nitrogen (N) compounds from agriculture, urban areas and scattered dwellings, is one of the main environmental concerns for rivers, lakes and coastal waters around the world [1]. Nutrients originate from a variety of sources, and one of the challenging tasks in combatting eutrophication is to identify the main sources and quantify the fluxes of nutrients promoting excessive algal growth. A good understanding of the predominant nutrient sources and pathways is essential to be able to design and effectuate the most cost-effective measures to reduce over-fertilisation of aquatic ecosystems. Process-based, integrated catchment models can provide a tool by which the relative importance of various sources can be quantified, also allowing for manipulative studies with simulation of ecosystem responses to various scenarios of changed policies, land use or climate forcing [2].

Catchment-scale nutrient loss models have been developed for different purposes and thus cover a wide range of complexity, level of process representation, input data requirements, and temporal and spatial resolution [3] . The models simulate water, sediment and P transport from point and non-point sources [3, 4]. Regardless of the span in complexity, all models have inherent uncertainties related to their input data, parameter values and process representation[5-8]. Parts of this uncertainty can be quantified; structural uncertainty (process representation) can be estimated by using different kinds of models of the same phenomena. Uncertainty related to the input data (e.g. forcing data such as meteorological parameters) can be treated by setting up models of observation errors. To address parameter uncertainty various inverse modelling and automatic calibration techniques can be used.

Traditional use of catchment-scale process-based models of nutrient transport often does not take full account of such uncertainties. Modellers have long grappled with the challenge to quantify the uncertainty bounds on simulations generated in model applications. The use of manual calibration techniques has been discussed in the hydrological literature for decades see e.g.[9 and references therein, 10, 11]. Manual calibrations are often dependent on the modellers' subjective opinion of when a fit is good enough. Manual calibration techniques often ignore the problem of equifinality see e.g. [12, 13]; in complex process-based models there are often several parameter sets which will yield the exact same degree of fit, and choosing between them is therefore (in terms of degree of fit) arbitrary. This is particularly problematic in scenario analyses, whereby a parameter can be rather insensitive or insignificant within the calibration or validation period, but have a much larger impact on model results under different scenario conditions. If uncertainty is included in the modelling process, then the outcomes are not as categorical. Addition of uncertainty estimates to the model outcomes means that the scientific basis for decision making becomes stronger for debate on these issues see e.g. [14, 15 and references therein]. Auto-calibration e.g. PEST [16] is a step in the right direction, yet often the goal of such exercises is to arrive at *one* best set of parameters, in which equifinality and parameter uncertainty is not fully addressed.

In recent years different sensitivity and uncertainty analyses have been applied to distributed, or semi-distributed, hydrological and nutrient leaching models as INCA-N [17], INCA-P[18], and INCA-C / INCA-Hg[19, 20]. A General sensitivity analysis (GSA) was adapted to the INCA-type of models by Wade et al. [21] and later adopted for INCA-C[22]. The GSA is performed to identify the model parameters that are most influential in determining system behaviour[23] and is based on a comparison of prior and posterior distributions of model parameter values. Additionally, uncertainty analyses has been carried out within the generalised likelihood uncertainty estimation (GLUE) framework – for INCA-N [24] and INCA-P [25] – and other attempts using Monte Carlo sampling of parameter space [26] or parameter optimisation algorithms[27].

One way to quantify the uncertainty surrounding parameters in catchment-scale nutrient loss models and the resulting uncertainty in modelled outcome is to use Bayesian analysis. Using prior distributions of parameter values and a formal likelihood, the procedure simulates a large number of outcomes and the technique arrives at posterior distributions of parameter values in which we are more confident in than our initial distribution. By using these posterior distributions to generate a range of simulated outputs we thereby quantify the effect of parameter uncertainty in our predictions, including scenario analyses. By using Bayesian methods for auto-calibration many of the shortcomings of manual calibration can be circumvented: the method should in principle arrive at the same posterior distributions for the same priors and input data (i.e. it is reproducible); the problem of equifinality is taken into account by allowing for a multitude of parameter values even though they do not affect the simulations in the calibration period; and the modelled outputs are presented as probability distributions making the uncertainty in the parameter estimates in the model explicitly and visually clear.

Though the principles of Bayesian analysis are fairly simple, in some cases the posterior distributions are complex and hard to fully explore with traditional techniques. Much effort has been put into improving these techniques e.g. [9, 28-33]. Improvements of Markov Chain Monte Carlo methods of posterior exploration include the use of several chains to better sample the full posterior distribution and evolutionary algorithms that include some degree of selection among chains e.g. [32, 33, 34]. MCMC-DiffeRential Evolution Adaptive Metropolis (DREAM) [34] is a recently developed algorithm that includes both several chains and evolutionary aspects and has been shown to be successful in analysing complex hydrological models e.g. [35, 36].

Here our goal was to introduce Bayesian parameter uncertainty on one commonly used catchment scale model of nutrient transport, the Integrated Catchment model for Phosphorus INCA-P; [18, 37] by applying the MCMC-DREAM algorithm. INCA-P was set up for simulating water flow, suspended sediment and phosphorus (P) loads in Hobøl River, the main tributary to Lake Vansjø, SE Norway. The catchment is characterised by high nutrient loads and recurrent blooms of toxin-producing cyanobacteria, and an improved understanding of the sources of P and the uncertainties surrounding them is instrumental in the improvement of the conditions of the lake. Following auto-calibration of the model, two scenarios of land use changes were simulated and the uncertainty associated with model parameters were taken into account through sampling of the posterior distributions.

2. MATERIAL AND METHODS

2.1 Study catchment

The Vansjø-Hobøl catchment comprises several small rivers and lakes, and one large lake (Vansjø) (Figure 1). The catchment area is 690 km^2, and land use is dominated by agriculture (16%) and forestry (80%) [38, 39]. The agricultural production in the area consists mostly of grain production, with a smaller fraction grass production (Table 1). Most of the catchment lies below 200 m elevation and is covered by marine sediments deposited during the last glacial period. Mean annual rainfall is 810 mm and the specific runoff is 14.4 L s-1 km^2. Lake Vansjø is 35 km^2 in surface area and consists of two major basins, Storefjorden and Vanemfjorden, with mean depths of 9 and 4 m, respectively. The main inlet river, Hobøl, has a catchment area of 337 km^2 and discharges into the Storefjorden basin.

<<Figure 1 here>>

<<Table 1 here>>

2.2 Description of the INCA-P model

INCA-P is a process-based mass balance model designed for simulating the P dynamics in catchments[18, 40], and was developed based on an integrated catchment model of nitrogen[41, 42]. INCA-P simulates the flow of water and addition/removal of P in the plant/soil system in different land use types. The water containing both P and suspended particles is then routed downstream in the catchment after accounting for direct effluent discharges and in-stream biological and sediment processes. Effluent discharges, inorganic-P fertilisers and farmyard manure, slurry applications, livestock wastes, and atmospheric deposition can be applied as input fluxes. The input fluxes and P addition/removal processes are differentiated by land use type and varied according to environmental conditions (e.g. soil moisture and temperature). The model also accounts for accumulated pools of inorganic and organic P in the soil (in readily available and firmly bound forms), in groundwater and in the stream reaches.

Since INCA-P is semi-distributed rather than fully-distributed, the catchment is decomposed into three spatial levels: At level 1, the catchment is decomposed into sub-catchments. At level 2, each sub-catchment is further decomposed into a maximum of six land use classes. At the third level, a generic cell is then applied to each land-use type within each sub-catchment. Generalised equations define the P transformations and stores within the cell, and six user-defined parameter sets derived through calibration are used to simulate the differences between the land-use types. The numerical method for solving the equations is based on the fourth-order Runge-Kutta technique, which allows a simultaneous solution of the model equations, thereby ensuring that no single process represented by the equations takes precedence over another[18].

Being a model for river transport of nutrients, INCA-P requires hydrological forcing on daily time steps. The HBV model [43] was used to produce the hydrological input time series. HBV is a semi-distributed conceptual model with subdivision in altitude zones and distributed snow and soil moisture descriptions. For Norwegian conditions, a version of the model developed by Killingtveit and Sælthun [44] and Sælthun [45] is most suitable. The general model structure consists of four main components: a snow module, a soil moisture zone module, a dynamic module comprising the upper and lower soil zone, and a routing module. The HBV model parameters can be grouped into two main categories, free and confined parameters. The confined parameters are based on physical measurements and not subject to calibration, for instance catchment area, area elevation curve and lake percentage. The free parameters must be determined by calibration. The external forcings for HBV are time series of precipitation and air temperature. The areal precipitation is based on point correction for rainfall and snowfall measurement errors, fixed station weights and linear altitude increase of precipitation. HBV produces daily hydrological input data for INCA-P -- hydrological effective rainfall (HER, the part of the precipitation/snowmelt that contributes directly to runoff) and the soil moisture deficit (SMD). The HBV model was not subject to uncertainty estimation, and was calibrated using the PEST procedure [46].

Among its outputs, INCA-P produces daily estimates of discharge *Q* (i.e., water flow), concentrations of suspended solids (*SS*) and P at discrete points along a river's main channel. The different P fractions simulated in the model are total dissolved P (*TDP*), particulate P (*PP*), and soluble reactive P (*SRP*). The *TDP* and *PP* sum up to total P (*TP*). Because the model is semi-distributed, the hydrological and nutrient fluxes from different land use classes and sub-catchment boundaries are modelled simultaneously, and information is fed sequentially into a multi-reach river model. Therefore, spatial variations in land use and farming practises can be taken into account, although the hydrological connectivity of different land use patches is not modelled in the same manner as in a fully distributed modelling approach.

2.3 Reach structure and model parameters.

The Hobøl River was divided into 5 sub-catchments (reaches, Figure 1) and five land use types were defined (see Table 1). For some of the parameters, individual values can be given for each simulated land use type or reach/subcatchment. The model parameter set can be roughly divided into four main categories of parameters:

- land phase parameters and initial values (53 parameters per each land use class);
- in-stream parameters and initial values (8 parameters);
- reach parameters and initial values (46 parameters per each reach);
- subcatchment parameters and initial values (31 parameters per each subcatchment)

The main analysis was performed with the first 4 reaches, as the observations are from the end of reach 4 (Kure). Parameters for reach 5 and its subcatchment were assumed to be identical to those of reach 4 (except for land use proportions, length and area) for the posterior predictive simulations.

A total of 94 parameters involved in all phases of P transport were estimated. In addition, the effluent inputs to the model were deemed to be uncertain and one parameter scaling these inputs was also estimated. Some parameters of the model were set to be identical for specific land use classes, subcatchments or reaches; this reduced the effective number of parameters to 49. The supplementary material lists all parameters estimated, their priors and posteriors. The parameters not varied in this exercise were based on an application of INCA-P in a smaller but similar catchment in the same area [47]

2.4 Formal likelihood.

For application of Bayesian methods a formal likelihood relating the simulated variables and observations needs to be defined. Following [34, 48, 49] we here develop a formal likelihood used for analysing INCA-P in a Bayesian framework. Note that in a Bayesian framework the prior and posterior are linked through a formal likelihood, as opposed to the informal metrics used in a GSA approach[23]. The output from INCA-P for which relevant observations are available are flow (Q, [m^{3}s^{-1}]), suspended sediments in the water column (SS [mg L^{-1}]) and total P in the water column (TP [mg L^{-1}]).

Treating INCA-P as a model (h) yielding a set of outputs $\mathbf{Y}_{\boldsymbol{o}} = \{y_{o,1}, \dots y_{o,n_o}\}$ given a set of forcing data ($\mathbf{X}$) and a set of parameters ($\boldsymbol{\theta}$)

$$\mathbf{Y}_{\boldsymbol{i}} = h(\mathbf{X}, \boldsymbol{\theta}),$$

we get a set of residuals for each type of observations ($\hat{y}_{o,t}$)

$$\varepsilon_{o,t} = y_{o,t}(\mathbf{X}, \boldsymbol{\theta}) - \hat{y}_{o,t} \, , o = \{Q, SS, TP\}, t = 1, \dots n_o \, .$$

We perform logarithmic transformations of our observed variables. The errors of the transformed variables (i.e. the error model) are assumed normally distributed, and we use Gibbs sampling of the error variances during the MCMC simulations. We thus assume that these residuals are mutually independent (uncorrelated) and normally distributed with a variance associated with each type of observation (${\sigma_o}^2$, $o = \{Q, SS, TP\}$) and get an expression for the posterior probability density function (pdf) [48]

$$p\left(\theta | \widehat{\mathbf{Y}}, \mathbf{X}, \boldsymbol{\sigma}^{\mathbf{2}}\right) = c \cdot p(\theta) \prod_{o} \prod_{t=1}^{n_o} \frac{1}{\sqrt{2\pi{\sigma_o}^2}} \times \exp\left(-\frac{\left(y_{o,t}(\mathbf{X}, \boldsymbol{\theta}) - \hat{y}_{o,t}\right)^2}{2{\sigma_o}^2}\right),$$

where c is a normalizing constant, $p(\theta)$ is the prior probability of a set of parameters, combining the data likelihood (the multiplicative part) and with a prior distribution using Bayes theorem. The posterior ($p\left(\theta | \widehat{\mathbf{Y}}, \mathbf{X}, \boldsymbol{\sigma}^{\mathbf{2}}\right)$) is thus the distribution of parameters given the model, input data and observations. Working with the logarithm of likelihoods $(\mathcal{L})$ is often preferred both for simplicity and stability of calculations;

$$\mathcal{L}(\theta|\hat{\mathbf{Y}},\mathbf{X},\boldsymbol{\sigma}^{\mathbf{2}}) = \sum_{o}\left(-\frac{n_o}{2}\ln(2\pi) - \frac{n_o}{2}\ln({\sigma_o}^2) - \frac{1}{2{\sigma_o}^2}\sum_{t=1}^{n_o}\left(y_{o,t}(\mathbf{X},\boldsymbol{\theta}) - \hat{y}_{o,t}\right)^2\right)$$

2.5 MCMC-DREAM algorithm

Estimating posterior probability distributions (as $(\theta|\hat{\mathbf{Y}},\mathbf{X},\boldsymbol{\sigma}^{\mathbf{2}})$) can be a fairly complicated exercise. We proceed by using Markov Chain Monte Carlo techniques; instead of trying to get a function describing the parameter distributions, an algorithm samples from it. MCMC algorithms start with a given parameter set, runs the model and calculates a likelihood for this specific parameter combination. A new (and usually fairly similar) *proposed* parameter set is then put into the model and a new likelihood is calculated using this proposed parameter set. If the new parameter set yields a better fit in terms of the likelihood (and the prior) it is then kept. If the new parameter set yields a slightly worse fit it is sometimes kept (according to the Metropolis Hastings method), in all other cases the *old* parameter set is kept. When run iteratively for a long period of time, this yields a *chain* of parameter sets that have been kept (a Markov Chain). When a histogram is generated from these kept parameter values, an estimate of the posterior distribution is achieved. Usually the first portion of the chain is discarded, to reduce the impact of badly chosen initial parameter guesses.

To estimate the posterior probability density ($p(\theta|\hat{\mathbf{Y}},\mathbf{X},\boldsymbol{\sigma}^{\mathbf{2}})$) we utilize the MCMC-DREAM algorithm[34]. Essentially this algorithm works by simulating several Markov Chains at the same time, which sample parameter proposals from distributions that are automatically tuned in both magnitude and direction during the evolution of the chains. The likelihood of these parameter proposals are then evaluated in a traditional Metropolis Hastings algorithm. The algorithm is succinctly described in[34]. Delayed rejection of parameter proposals was originally included in the DREAM algorithm, but we have not included this feature in our application.

Included in the DREAM algorithm are checks for convergence of chains through the calculation of Gelman-Rubin statistics[34, 50] after which the chains are runs for a given number of iterations to sample the posterior distributions. These chains were then stored and resampled to simulate all 5 reaches for a longer period (1995-2005), a form of posterior predictive modelling [50].

2.6 Routines for model calibration and evaluation of outputs

Model code for MCMC-DREAM evaluation of INCA-P was coded in Matlab [51] and utilized a command-line version of INCA-P. Matlab code was used to generate the proposal values and store the parameter and input files for each chain, after which INCA-P was called and output stored. The INCA-P output was then read by the code and evaluated according to the algorithm described above. Due to the computational cost we simulated INCA-P for years 1995-1997 and used observations from 1996 and 1997 to calculate the likelihoods used in parameter estimation. After convergence of the algorithm we then re-ran INCA-P using the estimated parameters sampled from the chains for a baseline run to estimate yearly loads from the river (including all 5 reaches) as well as for the scenarios for the period 1995-2005.

2.7 Posterior predictive simulations and scenario definitions.

After distributions of parameters were estimated posterior predictive simulations were performed, i.e. parameter values were sampled from the converged chains and INCA-P was rerun with these sets of parameters. This was performed for an extended period (1995-2005) for both a baseline (i.e. with current land use and management) and two scenarios. These future scenarios were developed together with stakeholders at a workshop discussing the future of Vansjø catchment. Two possible futures were envisioned and further limited to changes easily implemented in INCA-P. In the worst case scenario, greater demand for agricultural products was expected and we parameterized this as 25 % of all grassland was put into vegetable production (with a corresponding fertilizer regime), 10 % of the forest was made into grass production and the amount of fertilizer application was overall increased by 25% for all relevant land uses. Best case scenario included a 90 % reduction in effluent inputs (from scattered dwellings and waste water treatment plants), 25% of land allocated to vegetables and crops were changed to grassland production and 50 % direct decreases in fertilizer amounts were implemented. The reduction in effluent inputs was implemented as a second scaling factor in addition to the scaling factor estimated in the parameter estimation, thus still including the uncertainty of these inputs.

2.8 Data sources

Meteorology: Daily data on temperature, precipitation, and snow cover are obtained from three stations (1715 Rygge; 1750 Fløter; 378 Igsi) operated by the Norwegian Meteorological Institute (met.no).

Hydrology: Daily flow at the gauging station 3.22.0.1000.1 Høgfoss are obtained from the Norwegian Water Resources and Energy Directorate (NVE). The HBV model was calibrated using the PEST procedure [46] for the period 1.9.1990 to 31.8.2000, achieving a Nash Sutcliffe value of 0.78 and 0% volume error. The HBV was then run to simulate SMD and HER up to 31.12.2005.

Water chemistry: Water chemistry data come from the MORSA monitoring programme, conducted by Norwegian Institute for Agricultural and Environmental Research (Bioforsk) and Norwegian Institute for Water Research (NIVA). Data from the monitoring station Kure (reach 4) were used for calibration of INCA-P.

Land-cover: General land cover data are obtained from the Norwegian Forest and Landscape Research Institute, whereas more detailed information about land use fertilisation regimes on agricultural fields are provided by Bioforsk.

Municipal wastewater: Nutrient outputs from sewage treatment plants are obtained from Statistics Norway and the database KOSTRA. Outputs from scattered dwellings are provided by Bioforsk and their information system “GISavløp”.

3. RESULTS

The MCMC-DREAM algorithm successfully managed to quantify uncertainty associated with the selected parameters in the INCA-P model. The 40 chains used in the analysis achieved a Gelman

Rubin diagnostic of < 1.2 for all parameters after about 3350 iterations, indicating that the algorithm managed to converge and sample the posteriors [50]. The chains were run for 10 000 iterations and the last 2500 iterations were considered to satisfactorily sample the posterior distributions, with an acceptance rate of 0.288, indicating well-mixed sampling. INCA-P is a computation-time demanding model with each simulation in the parameter estimation phase requiring about 8 minutes to run, leading to a CPU demand of approximately 54 000 hrs. This was drastically reduced by using a multicore computer and the 10 000 iterations were completed in about two weeks.

Several of the marginal posterior distributions were dramatically smaller than the priors, but for others a wide range was still evident in the posteriors, indicating that the parameters did not have clear non-interactive effects on the likelihood. Several of the parameters, however, exhibit strong covariances, which are not evident in marginal distributions. Marginal distributions of all parameters and an example of the importance of covariance of parameters are presented in the supplementary material.

Figure 2 presents the simulated flow, total P and suspended sediment for the fourth reach in the calibration period 1995-1997 as well as observed values for 1996 and 1997, clearly showing the parametric uncertainty has low impacts on the flow (top panel), while giving rise to considerable uncertainty in P and particle matter predictions. For P the simulations span up to a factor of magnitude, particularly so in cases of heavy rain, a pattern opposite to the one exhibited for sediments in which the variation in predictions seem to narrow in the case of rain events but increase in periods of lower flow. While the confidence bands of both P and sediments are spanning up to an order of magnitude, several of the observations fall outside the prediction interval. This is partly due to the exclusion of the impact of the error variance estimated (see supplementary material), which would increase the uncertainty in the predictions considerably. It also underlines the importance of additional data potentially narrowing the priors for the parameters to get at both more constrained and more accurate predictions.

<<Figure 2 here>>

In the posterior predictive simulations for the period 1995-2005 (Figure 3) the flood in the end of 2000 and early 2001 is clearly evident, resulting in a dramatic increase in the uncertainty of the predictions. The winter 2000/2001 was mild with heavy rain, increasing the transport of sediments and particles to the lake (unpublished data). In our simulations, the effect of the flooding on P transport is noticed as an increase in the uncertainty in predictions of daily concentrations (Figure 3) and yearly loads (Figure 4) of P.

<<Figure 3 here>>

<<Figure 4 here>>

The scenarios were developed with the aim of spanning the potential futures of the catchment by positing best case and worst case developments, and the majority of the simulations predict a corresponding decrease or increase in the yearly loads (Table 2). The median reduction in P loads under the best case scenario was about 10%. Quite a few simulations (22%) actually showed an increase in annual loads, and conversely, taking parameter uncertainty into account yields a 78% probability of a decrease in loading. The parameter sets yielding an increase of loading under the best case scenario were the initial values for P storage in the changes land use classes and the parameter scaling the effluent inputs. Essentially, with the degree of uncertainty surrounding the parameters for the grass and crop areas from our calibration, there is a corresponding degree of uncertainty with the effect of the scenario.

<<Table 2 here>>

4. DISCUSSION

Our analysis of the parametric uncertainty in the INCA-P model is the first example of the use of a Bayesian framework on any model in the INCA-family. Earlier attempts at analysing uncertainty on

INCA models are limited to one application of the generalized likelihood uncertainty estimation GLUE, see [13, 52] technique on INCA-P [25] and other attempts using Monte Carlo sampling of parameter space of INCA-N, the nitrogen cousin of INCA-P [24, 26]. Even though GLUE and Bayesian approaches formally have different fundamental philosophies, predicted outcomes can be fairly similar, see e.g. [48].

Though our analysis shows that it is possible to both estimate parameters and perform scenario simulations using INCA-P while taking the parameter uncertainty into account, several assumptions were needed. Firstly, as opposed to the GLUE framework, a proper Bayesian analysis requires a formal definition of the error model (i.e. how predictions relate to observations). As a first step we assumed that these errors were uncorrelated, an assumption which is clearly invalid for most hydrological data. Though we did not include them here, there are more sophisticated ways to define the error model which could take into account the auto-correlated nature of hydrological data e.g. [53]. Secondly, the INCA-P model can only simulate nutrient transport using hydrological input which itself is output from an external model (HBV), and by using a single parameterization of HBV much of the uncertainty associated with parameters in the hydrological model was ignored. Ideally, the whole model chain HBV→ INCA-P should be treated as one model, and analysed in a Bayesian framework, but this has to be left for a future exercise. Lastly, as with any exercise where model parameters are varied, several parameters of the model are still left fixed, even though they entail their own uncertainty. A final improvement to the analysis would be to better select which parameters to include in the Bayesian estimation, and also gather more information for defining the priors.

The outcome of the posterior predictive simulations (Figure 3, 4 and Table 2) of the scenarios underlines the importance of taking parameter uncertainty into account particularly when models such as INCA-P are used for policy purposes. When the parameter values that gave the highest degree of fit (in terms of the Nash Sutcliffe criterion) are used alone, the model predicts a 19% decrease in the yearly loads of P to Lake Vansjø under the best case scenario. Though the scenarios here are not meant as real policy alternatives, a simulated decrease in loading of 19% could be seen as an argument for management of the catchment in the direction of this scenario (e.g. implementing incentives for farmers to switch from cereal crops to grass production). When the full uncertainty of the parameter values are taken into account, however, the median prediction is a much lower reduction in load (11 %) and about 22% of the simulations actually showed an increase in loads given the best case scenario. In summary, with the current knowledge and quantification of our parametric uncertainty there's only about 80 % chance that such mitigation measures would actually decrease the loading from the river. An inspection of which parameter sets that yield an increase in loading under the best case scenario (not shown) reveals that these predictions arise from parameter combinations of high and low initial P values for the grass and crop land use classes respectively. In addition this increased loading under a proposed better future are predicted when the parameter scaling the effluent inputs are in the lower range, reducing the impact of effluent reduction (see supplementary information for all parameters estimated). As the observations against which the model is calibrated are from the river itself, calibration can not distinguish between exports from different land use classes (unless they change over time, which they do not in our case). This is a good example of equifinality; parameters detailing the mobilization and initial storage of P in different land use classes will covary negatively and can result in the same degree of fit for a wide range of parameter settings. Instead of choosing one single parameterization for this process, our approach captures the uncertainty surrounding this process through our scenario simulations.

Information regarding the probability of success of mitigation measures is highly useful for policymakers, and knowing that a particular management decision has a 1 in 5 chance of having the opposite effect would be beneficial. This underlines the importance of communicating uncertainty in the interaction between science and policy, and not to obscure possible risks of non-optimal effects or even failure when implementing a mitigation measure for a perspective see [14]. It also underlines the importance of equifinality issues in models such as INCA-P; even though the best calibration (based on classical NS values) predicts a certain reduction in load, incorporating a wider range of both plausible and probable parameter ranges may have a substantial effect on the probability of modelled outcomes under scenario analysis.

ACKNOWLEDGEMENTS

This work was supported by the EU project REFRESH (Adaptive Strategies to Mitigate the Impacts of Climate Change on European Freshwater Ecosystems), the SEALINK and EUTROPIA projects funded by Norwegian Research Council, and Strategic Internal funding at NIVA. The HBV model runs were performed by Stein Beldring and Deborah Lawrence at Norwegian Water Resources and Energy Directorate, and their efforts are greatly appreciated. Anne Bjørkenes Christiansen's help with the Figures is acknowledged, and Richard Wright streamlined the manuscript. Fakutsi for inspiration.

TABLES

Table 1. Reaches/subcatchments defined within the Vansjø-Hobøl catchment, including land use information.

Reach no.	**Name**	**Size** *km^2*	Wetland *km^2*	Forest *km^2*	Grass *km^2*	Crops *km^2*	Vegetables *km^2*
1	Mjær	146.32	13.2	125.8	1.5	5.9	0.0
2	Tomter	88.16	0.9	66.1	1.8	19.4	0.0
3	Kråkstad	50.6	0.5	33.9	1.0	14.7	0.5
4	Kure	19.32	0.4	12.6	0.6	5.8	0.0
5	Våler	32.41	1.0	25.0	0.6	5.8	0.0

Table 2. Summary results of scenario runs including parameter uncertainty. Scenarios are run by sampling parameter sets from the Markov Chains, and for each sampled parameter set we run all scenarios. The ratio of change is the ratio of individual yearly loads for each specific parameter set under all scenarios (i.e. the only difference is the scenario parameters), i.e.$r = Load_{year,\ scenario} / Load_{year,\ baseline}$. First year of the simulation (1995) is excluded. The single INCA-P parameterization with the highest Nash-Sutcliffe value for total P in the calibration period 1996-1997 (NS_{TP} = 0.4616) gives the following median loads in the period 1996-2005; 15.00 T/y (Baseline), 18.21 T/y (Worst Case) and 11.96 T/y (Best Cse), withmedian ratios of change at 1.22 and 0.81 respectively. Please note that the ratio of change to the right are calculated on the distribution of yearly values and not on the median values reported to the left.

	Percentiles and mean yearly loads 1996-2005 (ton / year)			Percentiles and mean ratio of change.		
	2.5%	50% (mean)	97.5%	2.5%	50% (mean)	97.5%
Baseline	5.18	11.96 (24.24)	108.13	1	1	1
Worst Case	7.18	16.73 (30.48)	131.61	1.02	1.30 (1.41)	2.39
Best Case	4.07	11.10 (23.03)	103.41	0.69	0.89 (0.92)	1.39

FIGURE CAPTIONS

Figure 1. The Vansjø-Hobøl catchment. The 5 subcatchments that constitute Vansjø's main tributary river, Hobøl flowing from North to south and a map of Norway showing location of the catchment. The observations against which the model was calibrated was collected at the southern end of the Kure subcatchment. Lake Vansjø, with its basins Vanemfjorden to the west and Storefjorden to the east, can be seen at the lower part of the catchment.

Figure 2. Confidence intervals for simulated flow, P and sediment 1995-1997. Simulated flow, total P and suspended sediment (log scale) for the burn-in (1995) and calibration period (1996-1997) for the 4^{th} reach (Kure). The lines represent the daily 2.5, 50 and 97.5 percentiles of the output from the posterior predictive simulations performed with the converged chains sampled 100 times (i.e. from 4000 simulations). Observations are represented by dots. Error variance is not included in these plots, so the uncertainty in predictions arises solely from parameter uncertainty.

Figure 3. Confidence intervals for simulated flow, P and sediment 1995-2005. Simulated flow, total P concentrations and suspended sediment (log scale) for the period 1995-2005, with lines representing the 2.5, 50 and 97.5 percentiles of the posterior predictive simulations for the 4^{th} reach. Observations are represented by dots. When running the posterior predictive modelling for the whole period 196 of the parameter sets resulted in crashes of INCA-P and confidence intervals are calculated from the 3804 successful runs. Note the flood occurring Oct-Nov 2000, and how it affects the uncertainty in the predicted total P concentrations. Error variance is not included in these plots, so the uncertainty in predictions arises solely from parameter uncertainty.

Figure 4. Yearly loading from Hobøl River. Simulated yearly loading of P (log scale) from Hobøl River into Lake Vansjø in metric ton per year, i.e. from the 5^{th} reach. The lines represent the 2.5, 50 and 97.5 percentiles. The high degree of uncertainty in 1995 is due to initial conditions that have reduced impact outside the burn-in year. Also note how the flooding in Autumn 2000 affects both the

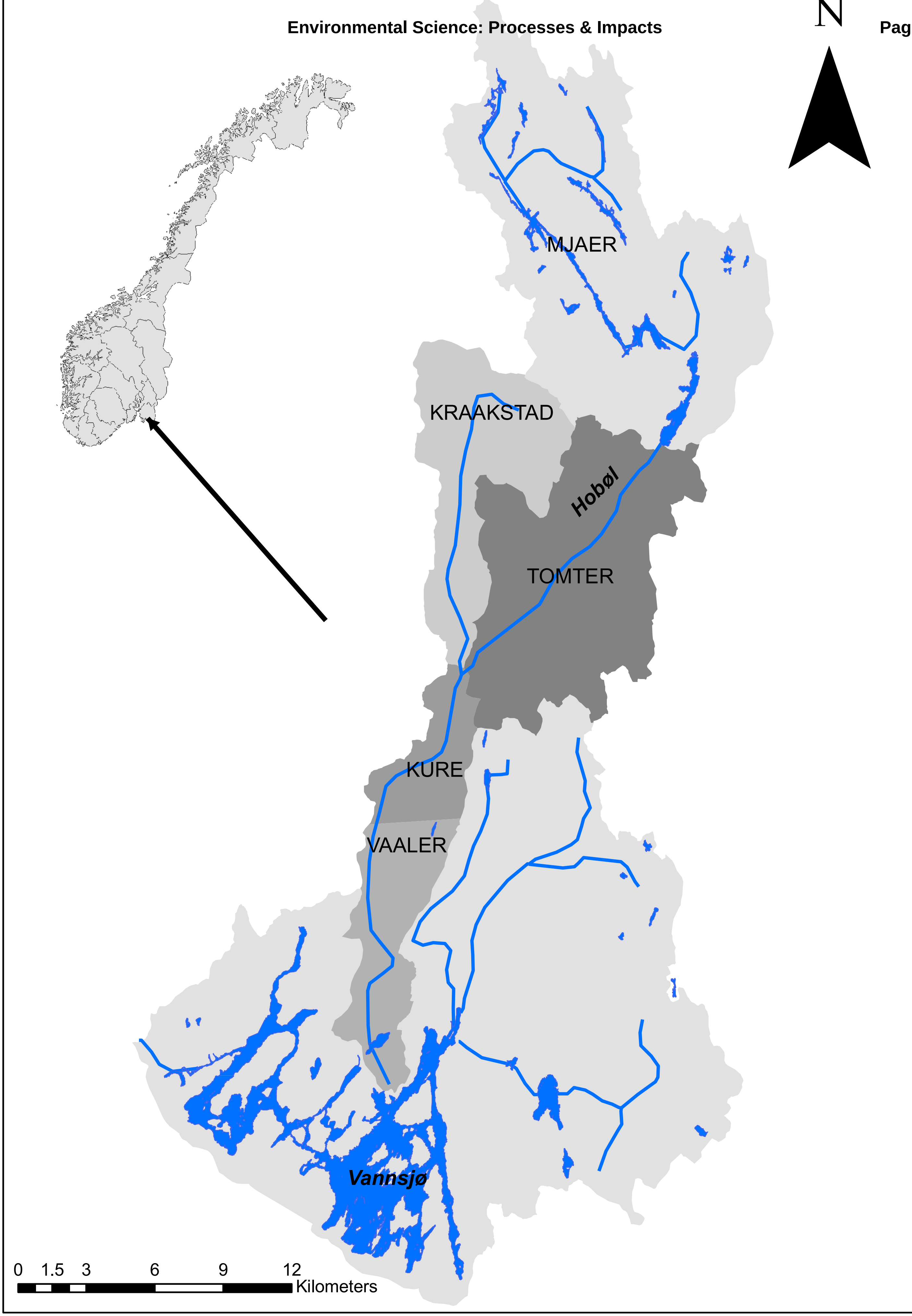
median and the surrounding uncertainty. The median loading for the whole period is approximately 12 tons/year (see Table 2).

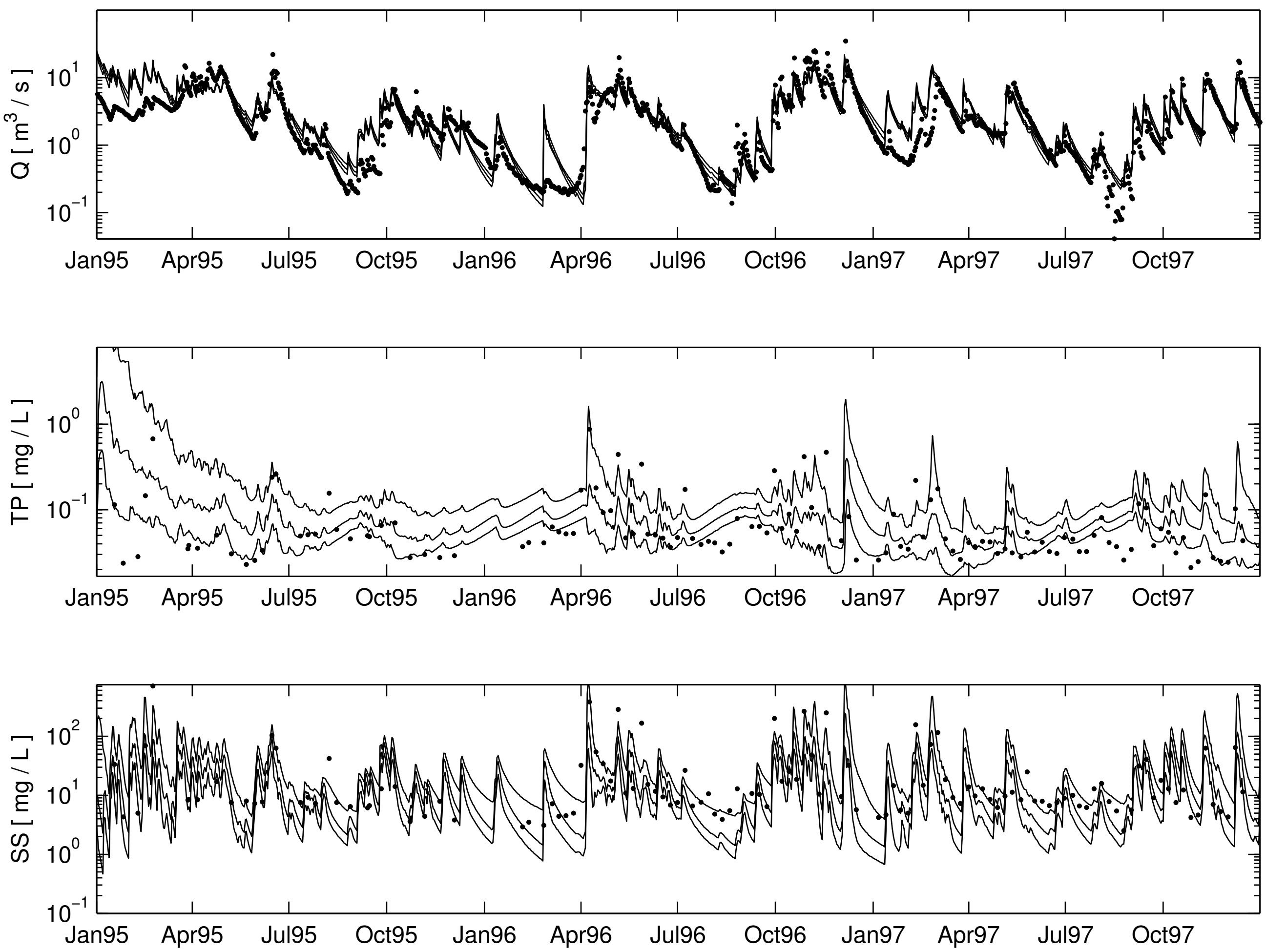
References

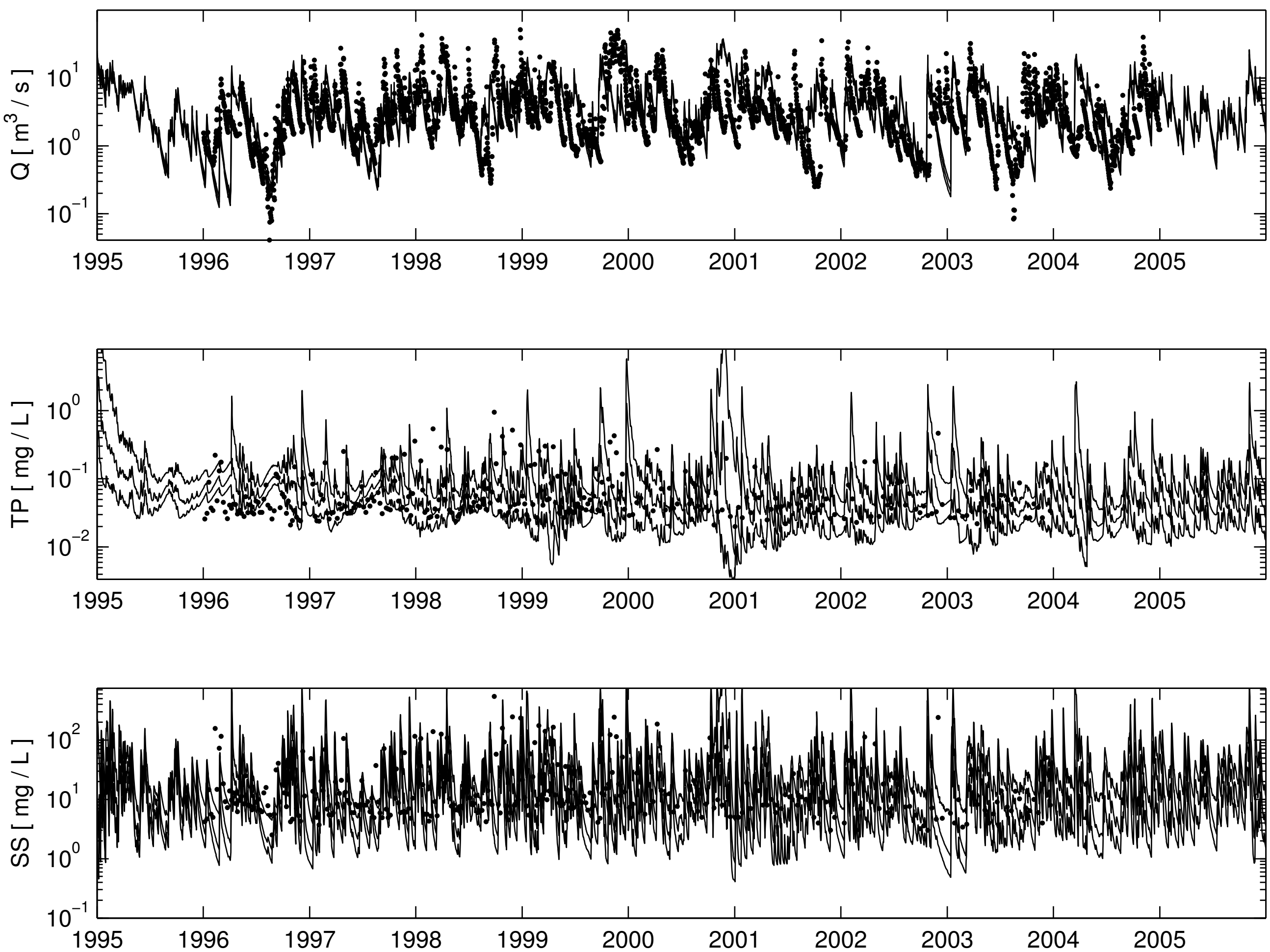
1. D. J. Conley, H. W. Paerl, R. W. Howarth, D. F. Boesch, S. P. Seitzinger, K. E. Havens, C. Lancelot and G. E. Likens, *Science (Washington)*, 2009, **323**.
2. T. M. Saloranta, *Ecological Modelling*, 2006, **194**, 316-327.
3. O. F. Schoumans, M. Silgram, P. Groenendijk, F. Bouraoui, H. E. Andersen, B. Kronvang, H. Behrendt, B. Arheimer, H. Johnsson, Y. Panagopoulos, M. Mimikou, A. Lo Porto, H. Reisser, G. Le Gall, A. Barr and S. G. Anthony, *Journal Of Environmental Monitoring*, 2009, **11**, 506-514.
4. D. Radcliffe, J. Freer and O. Schoumans, *Journal of environmental quality*, 2009, **38**, 1956-1967.
5. T. Page, A. L. Heathwaite, B. Moss, C. Reynolds, K. J. Beven, L. Pope and R. Willows, *Freshwater Biology*, 2012, **57**, 108-123.
6. T. Wagener and H. V. Gupta, *Stoch Env Res Risk A*, 2005, **19**, 378-387.
7. J. C. Refsgaard, J. P. van der Sluijs, A. L. Hojberg and P. A. Vanrolleghem, *Environ Modell Softw*, 2007, **22**, 1543-1556.
8. W. E. Walker, P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen and M. P. Krayer von Krauss, *Integrated Assessment*, 2003, **4**, 5-17.
9. H. V. Gupta, S. Sorooshian and P. O. Yapo, *Water Resources Research*, 1998, **34**, 751-763.
10. D. P. Boyle, H. V. Gupta and S. Sorooshian, *Water Resour. Res*, 2000, **36**, 3663-3674.
11. K. Beven, *Journal of Hydrology*, 1989, **105**, 157-172.
12. K. Beven, *Journal of Hydrology*, 2006, **320**, 18-36.
13. K. Beven and J. Freer, *Journal of Hydrology*, 2001, **249**, 11-29.
14. K. Beven, *Hydrological Processes*, 2006, **20**, 3141-3146.
15. K. Beven, *Hydrological Processes*, 2008, **22**, 3549-3553.
16. J. Doherty, *PEST: Model Independent Parameter Estimation.*, Watermark Numerical Computing, Brisbane, Australia, 2004.
17. A. Wade, P. Durand, V. Beaujouan, W. Wessel, K. Raat, P. Whitehead, D. Butterfield, K. Rankinen and A. Lepisto, *Hydrol Earth Syst Sc*, 2002, **6**, 559-582.
18. A. J. Wade, P. G. Whitehead and D. Butterfield, *Hydrol Earth Syst Sc*, 2002, **6**, 583-606.
19. M. Futter, D. Butterfield, B. Cosby, P. Dillon, A. Wade and P. Whitehead, *Water Resources Research*, 2007, **43**, 2424.
20. M. Futter, A. Poste, D. Butterfield, P. Dillon, P. Whitehead, A. Dastoor and D. Lean, *Science of the Total Environment*, 2012.
21. A. Wade, G. M. Hornberger, P. Whitehead, H. Jarvie and N. Flynn, *Water Resources Research*, 2001, **37**, 2777-2792.
22. M. N. Futter, D. Butterfield, B. J. Cosby, P. J. Dillon, A. J. Wade and P. G. Whitehead, *Water Resources Research*, 2007, **43**.
23. R. Spear and G. Hornberger, *Water Research*, 1980, **14**, 43-49.
24. K. Rankinen, T. Karvonen and D. Butterfield, *Science of the Total Environment*, 2006, **365**, 123-139.
25. S. Dean, J. Freer, K. Beven, A. J. Wade and D. Butterfield, *Stoch Env Res Risk A*, 2009, **23**, 991-1010.

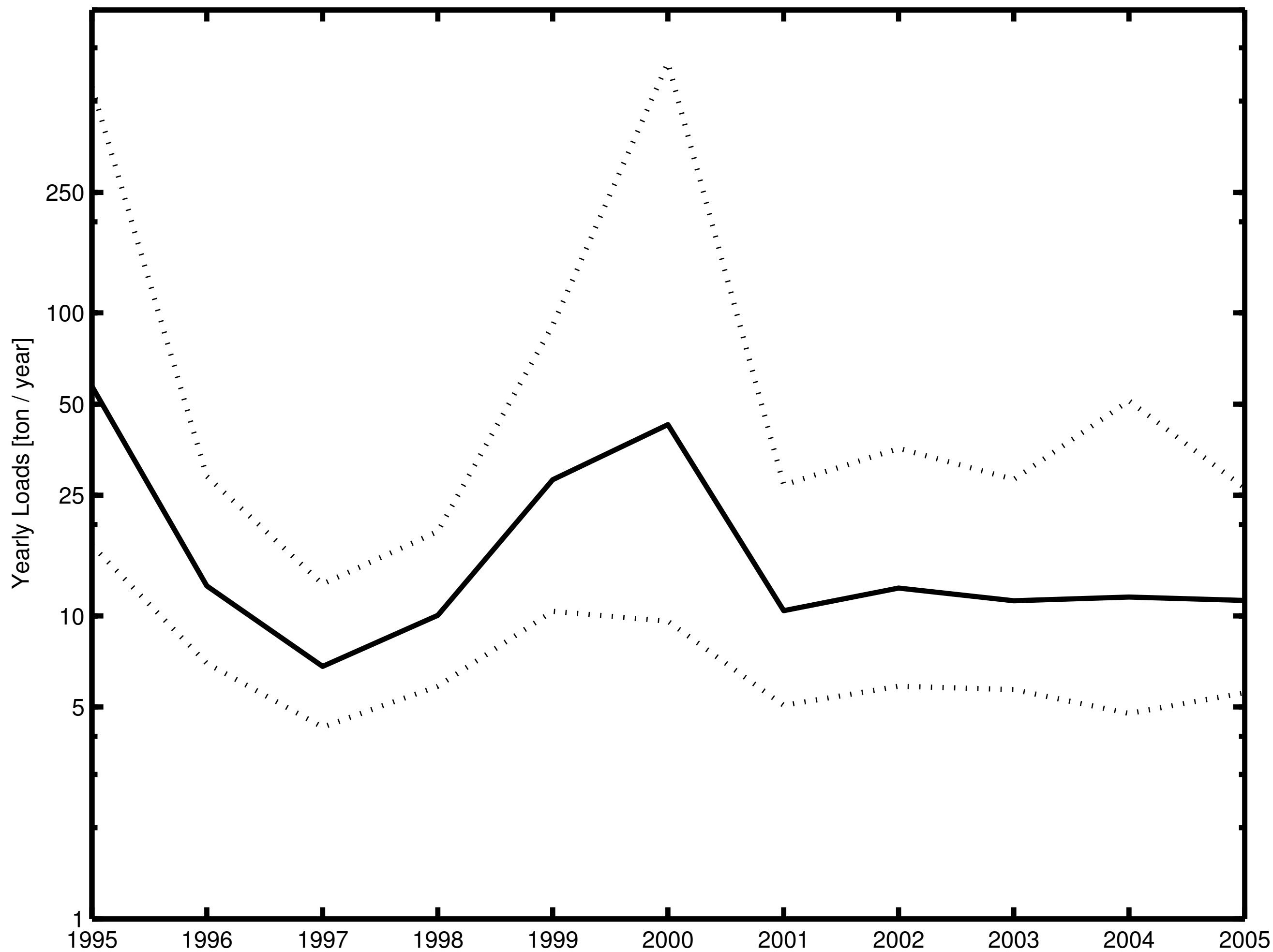
26. N. McIntyre, B. Jackson, A. Wade, D. Butterfield and H. Wheater, *Journal of Hydrology*, 2005, **315**, 71-92.
27. K. J. Raat, J. A. Vrugt, W. Bouten and A. Tietema, *Hydrol Earth Syst Sc*, 2004, **8**, 751-763.
28. G. Kuczera and E. Parent, *Journal of Hydrology*, 1998, **211**, 69-85.
29. J. C. Refsgaard and J. Knudsen, *Water Resources Research*, 1996, **32**, 2189-2202.
30. H. Haario, M. Laine, A. Mira and E. Saksman, *Stat. Comput.*, 2006, **16**, 339-354.
31. J. A. Vrugt, H. V. Gupta, L. A. Bastidas, W. Bouten and S. Sorooshian, *Water Resources Research*, 2003, **39**.
32. J. A. Vrugt, H. V. Gupta, W. Bouten and S. Sorooshian, *Water Resources Research*, 2003, **39**.
33. J. A. Vrugt and B. A. Robinson, *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 708-711.
34. J. A. Vrugt, C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman and D. Higdon, *International Journal of Nonlinear Sciences and Numerical Simulation*, 2009, **10**, 273-290.
35. E. Laloy and J. A. Vrugt, *Water Resources Research*, 2012, **48**, W01526.
36. J. Huisman, J. Rings, J. Vrugt, J. Sorg and H. Vereecken, *Journal of Hydrology*, 2010, **380**, 62-73.
37. P. G. Whitehead, L. Jin, H. M. Baulch, D. A. Butterfield, S. K. Oni, P. J. Dillon, M. Futter, A. J. Wade, R. North, E. M. O'Connor and H. P. Jarvie, *Science of the Total Environment*, 2011, **412**, 315-323.
38. A. L. Solheim, N. Vagstad, P. Kraft, S. Turtumøygard, S. Skoglund, Ø. Løvstad and J. R. Selvik, *Analysis of abatemant measures for the Vansjø-Hobøl-watercourse (Norwegian).* 4377-2001, Norwegian Institute for Water Research, Norway, 2001.
39. F. Bouraoui, B. Grizzetti, G. Adelskold, H. Behrendt, I. de Miguel, M. Silgram, S. Gomez, K. Granlund, L. Hoffmann, B. Kronvang, S. Kvaerno, A. Lazar, M. Mimikou, G. Passarella, P. Panagos, H. Reisser, B. Schwarzl, C. Siderius, A. S. Sileika, A. Smit, R. Sugrue, M. VanLiedekerke and J. Zaloudik, *Journal Of Environmental Monitoring*, 2009, **11**, 515-525.
40. A. J. Wade, D. Butterfield, D. S. Lawrence, I. Bärlund, P. Ekholm, A. Lepistö, M. Yli-Halla, K. Rankinen, K. Granlund, P. Durand and Ø. Kaste, *The integrated catchment model of phosphorus (INCA-P), a new structure to simulate particulate and soluble phosphorus transport in European catchments.* Deliverable 185 to the EU Euro-limpacs project, UCL, London., 2009.
41. P. Whitehead, E. Wilson and D. Butterfield, *Science of the total environment*, 1998, **210**, 547-558.
42. P. Whitehead, E. Wilson, D. Butterfield and K. Seed, *Science of the total environment*, 1998, **210**, 559-583.
43. S. Bergström, *Development and application of a conceptual runoff model for Scandinavian catchments.*, SMHI, 1976.
44. Å. Killingtveit and N. R. Sælthun, *Hydrology*, Norwegian Institute of Technology, Division of Hydraulic Engineering, Trondheim, Norway, 1995.
45. N. R. Sælthun, *The "Nordic" HBV model*, Norwegian Water Resources and Energy Administration., Oslo, Norway, 1996.
46. D. Lawrence, I. Haddeland and E. Langsholt, *Calibration of HBV hydrological models using PEST parameter estimation.* 1-2009, Norwegian Water Resources and Energy Directorate, Oslo, Norway, 2009.
47. C. Farkas, S. Beldring, M. Bechmann and J. Deelstra, *Soil Use and Management*, in press.

48. J. A. Vrugt, C. J. F. ter Braak, H. V. Gupta and B. A. Robinson, *Stoch Env Res Risk A*, 2009, **23**, 1011-1026.
49. D. Kavetski, G. Kuczera and S. W. Franks, *Water Resources Research*, 2006, **42**.
50. A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin, *Bayesian data analysis*, 2nd edn., Chapman & Hall/CRC, Boca Raton, Fla., 2004.
51. , The MathWorks Inc., Natick, Massachusetts, Editon edn., 2011.
52. K. Beven and A. Binley, *Hydrological Processes*, 1992, **6**, 279-298.
53. G. Schoups and J. A. Vrugt, *Water Resources Research*, 2010, **46**.


N
MJAER
KRAAKSTAD
Hobøl
TOMTER
KURE
VAALER
Vannsjø
0 1.5 3 6 9 12
Kilometers


Q [m^3 / s]
10^1
10^0
10^-1
Jan95
Apr95
Jul95
Oct95
Jan96
Apr96
Jul96
Oct96
Jan97
Apr97
Jul97
Oct97
TP [mg / L]
10^0
10^-1
SS [mg / L]
10^2
10^1
10^0
10^-1


Q [m^3 / s]
10^1
10^0
10^−1
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
TP [mg / L]
10^0
10^−1
10^−2
SS [mg / L]
10^2
10^1
10^0
10^−1


Yearly Loads [ton / year]
250
100
50
25
10
5
1
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005