# Forecasting Acidification Effects Using a Bayesian Calibration and Uncertainty Propagation Approach

T H O R J Ø R N   L A R S S E N , * , †
R A G N A R   B .   H U S E B Y , ‡
B E R N A R D   J .   C O S B Y , §   G U D M U N D   H Ø S T , ‡ , ∥
T O R E   H Ø G Å S E N , †   A N D   M A G N E   A L D R I N ‡

*Norwegian Institute for Water Research, Gaustadalleen 21, 0349 Oslo, Norway, Norwegian Computing Center, P.O. Box 114 Blindern, 0314 Oslo, Norway, Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia 22904-4123*

We present a statistical framework for model calibration and uncertainty estimation for complex deterministic models. A Bayesian approach is used to combine data from observations, the deterministic model, and prior parameter distributions to obtain forecast distributions. A case study is presented in which the statistical framework is applied using the hydrogeochemical model (MAGIC) for an assessment of recovery from acidification of soils and surface waters at a long-term study site in Norway under different future acid deposition conditions. The water quality parameters are coupled with a simple dose− response model for trout population health. Uncertainties in model output parameters are estimated and forecast results are presented as probability distributions for future water chemistry and as probability distributions of future healthy trout populations. The forecast results are examined for three different scenarios of future acid deposition corresponding to three different emissions control strategies for Europe. Despite the explicit consideration of uncertainties propagated into the future forecasts, there are clear differences among the scenarios. The case study illustrates how inclusion of uncertainties in model predictions can strengthen the inferences drawn from model results in support of decision making and assessments.

## 1. Introduction

Deterministic models are often used to simulate hydrological, hydrochemical, or other complex environmental systems, and to make predictions about system response under different input regimes. Such deterministic models produce unique output parameters as a function of a unique group of input parameters. Calibration of input parameters through fitting the corresponding output parameters to observed data is a fundamental step in developing dynamic, process-oriented models for use in prediction of future system response. Systematic calibration procedures (instead of ad hoc practices) are necessary to improve the relevance and reliability of model applications. Although such systematic

procedures are becoming more frequent in the literature they are still not regular practice in much of the applied environmental modeling community.

In addition, it is common to present model predictions or forecasts without quantitative estimates of the associated uncertainties, particularly for "real-world" applications in a policy context. In many cases, a prediction is based on one set of input parameters assumed to be the "best guess" for the parameter values. However, in a deterministic model with multiple input parameters, many different combinations of input parameters may give similar output parameters, and hence similar fits to measurements of the output parameters (*1, 2*). Taking into account the uncertainties in input parameters and observations used for calibration, a large ensemble of plausible combinations of input parameters can be found for any model application. The ensemble of output parameters produced can then be presented as probability distributions instead of fixed values.

Several approaches have been developed for estimating uncertainty using Bayesian techniques with deterministic models of water resources. These approaches have been used primarily with conceptual hydrological catchment models (*2−7*). The use of Bayesian techniques with biogeochemical or hydrogeochemical models is relatively rare, probably due to the higher dimensionality of the input and output parameter spaces in water quality models as compared to water quantity (rainfall and runoff) models. In this paper we present a Bayesian framework that can be used with complex, deterministic models to calibrate the models and propagate uncertainty through the model simulations. The analysis framework is demonstrated using the deterministic acid deposition response model MAGIC (*8, 9*). The water quality parameters are coupled with a simple dose−response model for trout population health (*10*).

The goal of this paper is to describe and demonstrate how a statistical framework for calibrating and propagating uncertainty can be used in generating probability distributions for model forecasts of acidification effects in terms of water chemistry and fish status.

## 2. Materials and Methods

Our approach is to combine the deterministic model MAGIC with a stochastic model for the observed output data, and to estimate all unknown parameters by Bayesian computations using Markov Chain Monte Carlo (MCMC) techniques. This involves specifying prior probability distributions for the input parameters and the likelihood functions for the output data. The posterior distributions are calculated by running the deterministic model repeatedly with parameters suggested by an MCMC scheme. This approach allows a formal management of uncertainties in both parameters and data and allows the impact of the uncertainty on model results to be explicitly shown.

**2.1.** The Combined Model Framework. The deterministic model **m** (i.e., MAGIC) relates output parameters at time $t$ to input parameters up to and including $t$. This model may be written

$$\mathbf{m}_t = \mathbf{m}(\mathbf{v}, \mathbf{x}_t, \mathbf{x}_{t-1}, ...),\qquad(1)$$

where **m** is an exact vector function, **v** is a vector of time-independent input parameters, $\mathbf{x}_t$ is a vector of time-dependent input parameters at time $t$, and $\mathbf{m}_t$ is a vector of time-varying output parameters at time $t$.

Assume that there exist observations $\mathbf{y}_t$ corresponding to the output parameters $\mathbf{m}_t$, and that these output data are

---

* Corresponding author phone: +47 22 18 51 00; fax: 47 22 18 52 00; e-mail: thorjorn.larssen@niva.no.
† Norwegian Institute for Water Research.
‡ Norwegian Computing Center.
§ University of Virginia.
∥ Now at Norwegian Research Council, 0131 Oslo.

unbiased measurements of the underlying, "true" output parameters. The output data and output parameters can then be linked through the stochastic model

$$\mathbf{y}_t = \mathbf{m}_t + \epsilon_t \qquad (2)$$

where the error $\epsilon_t$ has zero mean and is the difference between the output parameters $\mathbf{m}_t$ from the model and the output data $\mathbf{y}_t$. If the error $\epsilon_t$ is assumed to follow a specified statistical distribution, a likelihood function for the output data can be specified.

Assume further that there exist a set of optimal input parameters that represents the output data $\mathbf{y}_t$ faithfully, i.e., that minimizes $\epsilon_t$ in some sense, both for a calibration period with known output data and for hypothetical future observations. One of our aims is to estimate the unknown, optimal input parameters with uncertainty, with corresponding predictions (with uncertainty) of the output parameters $\mathbf{m}_t$ given by the model (eq 1). Adding the uncertainty due to $\epsilon_t$ gives the uncertainty in predicting future output data $\mathbf{y}_t$.

The input parameters typically have concrete physical meanings, and for many of the input parameters there exist corresponding observations, here called input data. However, the unknown, optimal input parameters usually differ from the observed input data, since the model (eq 1) is only an approximation to the truth and the input data may have measurement errors. Anyhow, the input data do contain information on the optimal input parameters, and together with subjective expert knowledge, the input data are used to formulate prior probability distributions for the (optimal) input parameters.

We put the input parameters and other unknown parameters into a vector $\boldsymbol{\theta}$, and denote the prior probability distribution of $\theta$ by $\mathbf{p}(\boldsymbol{\theta})$. Further, let $\mathbf{y}$ denote the output data from the calibration period, with likelihood $L(\mathbf{y};\boldsymbol{\theta})$ conditioned on $\boldsymbol{\theta}$. By Bayes' formula, the conditional distribution $\mathbf{p}(\boldsymbol{\theta} \mid \mathbf{y})$, also called the posterior distribution, is

$$\mathbf{p}(\boldsymbol{\theta} \mid \mathbf{y}) \propto L(\mathbf{y};\boldsymbol{\theta}) \; \mathbf{p}(\boldsymbol{\theta}) \qquad (3)$$

This posterior distribution contains all information relevant to the parameter vector $\boldsymbol{\theta}$, given the model $\mathbf{m}$ and the data $\mathbf{y}$. A more thorough discussion of many aspects of Bayesian calibration of deterministic models is found in ref *11*.

**2.2. The MAGIC Model.** The MAGIC model was developed to predict long-term effects of acid deposition on soil and surface water chemistry (*8, 9*). The model has been extensively used at a range of different sites and applications (*12–14*).

MAGIC calculates annual or monthly concentrations of ions in soil solution and surface water using mathematical solutions to simultaneous equations describing sulfate adsorption, cation exchange, dissolution–precipitation–speciation of aluminum, and dissolution–speciation of inorganic and organic carbon. The model accounts for the mass balance of major ions by simulating ionic fluxes from atmospheric inputs, chemical weathering, net uptake in biomass, and loss to runoff. A model simulation is initialized in pre-industrial times (1850 is used in the current application) and is run forward in time to a calibration period using historical changes in the deposition of the major ions to drive the model.

In the traditional procedure for calibrating the model two sets of input parameters are adjusted by trial and error, whereas the remaining input parameters are fixed:

(1) total soil base cation weathering ($BC_w$), which is the sum of the weathering rate of the four base cations $Ca^{2+}$, $Mg^{2+}$, $Na^+$, and $K^+$;

( 2) total initial soil base cation saturation ($BS_{ini}$), which is the sum of the relative amounts of the individual exchangeable base cations $Ca^{2+}$, $Mg^{2+}$, $Na^+$, and $K^+$ in the soil during the year the simulation is initiated.

Calibration is carried out by adjusting values of $BC_w$ and $BS_{ini}$ (and the partitioning of each ion) until output parameters match output data for water chemistry and soil base saturation in the calibration period. After calibration, the model is used for predicting the future soil and water chemistry under different scenarios of changes in atmospheric deposition.

In this paper we disregard the traditional calibration approach and, instead, implement the Bayesian approach outlined above which considers all input and output parameters in the model as being subject to uncertainty. This approach allows for a larger suite of input parameters to be considered during calibration. In all, there are 28 time-independent input parameters ($\mathbf{v}$) and nine time-dependent input parameters ($\mathbf{x}_t$) (Table 1). The output parameters ($\mathbf{m}_t$) from this application of MAGIC consist of 14 time variable parameters related to soil and water chemistry (Table 2).

A particularly relevant output parameter from the model is the combined parameter acid neutralizing capacity (ANC) of the stream water. It is calculated from the equivalent concentration of seven ions in the surface water (Table 2)

$$\text{ANC} = [Ca^{2+}] + [Mg^{2+}] + [Na^+] + [K^+] - [SO_4^{2-}] - [NO_3^-] - [Cl^-]$$

where the brackets indicate concentrations expressed as $\mu$eq/L. ANC is commonly used as indicator for water quality related to long-term fish population survival in connection with acidification (*10, 15*).

**2.3. Data.** The Birkenes catchment is located in southern Norway, about 20 km inland from the coast. The catchment area is 0.41 km² with elevation of 200–300 m. The vegetation is mainly 100-year-old Norway spruce with some pine and birch and an undergrowth of mosses, blueberry, and fern. Podzolic and brown-earth soils have developed from a shallow layer of glacial till on granitic bedrock. Peaty deposits have developed on poorly drained sites in the catchment. On the slopes, well-drained thin organic layers on gravel or bedrock are common. The catchment is drained by three small second-order streams, which converge about 150 m above the lower catchment border. A station for sampling precipitation and air is located about 500 m north of the catchment (*16*). The catchment is substantially influenced by seasalt deposition, which has considerable influence on the stream and soil water chemistry.

Measurements of major anions and cations have been recorded since 1974 for both bulk deposition (daily) and runoff chemistry (weekly) (*16, 17*). In the current application, the chemical composition of the annual deposition and the runoff amount are obtained from volume-weighted daily observations. These input data ($\mathbf{z}_t$) correspond to the nine time-dependent input parameters ($\mathbf{x}_t$) in Table 1.

In addition to these time series, physical and chemical properties of the catchment have been measured, including soil depth, soil porosity, soil density, and cation exchange capacity (CEC) (Table 1). Soil property data are based on samples from four soil profiles, each sampled in five depths, analyzed in 1992 and 2000 (*18, 19*).

The relative changes in the long-term historic deposition data are taken from calculations with a European-scale atmospheric dispersion and transportation model (the EMEP model) (*20*).

**2.4. Prior Distributions and Likelihood.** Prior distributions for the input parameters have been constructed based on corresponding input data and expert knowledge (Table 1). The prior distributions for parameters like soil depth and porosity are based on measurements and are rather narrow.

**TABLE 1. Input Parameters with Prior Distributions**

| input parameters | unit | prior distribution[a] |
|---|---|---|
| **constant parameters $v$** | | |
| fixed soil parameters | | |
|   depth | m | uniform (0.3, 0.5) |
|   porosity | % | uniform (40,60) |
|   bulk density | kg m$^{-3}$ | uniform (696,850) |
|   cation exchange capacity | meq kg$^{-1}$ | uniform (95,117) |
|   sulfate adsorption half saturation | $\mu$eq L$^{-1}$ | uniform (1,500) |
|   sulfate adsorption maximum | meq kg$^{-1}$ | uniform (0,50) |
|   aluminum dissolution constant | Log10 | uniform (6,11) |
|   temperature | °C | fixed (5) |
|   partial $CO_2$ pressure | % atm | uniform (0.50,2) |
|   dissolved organic C concentration | $\mu$mol L$^{-1}$ | uniform (0,250) |
|   nitrification (% of incoming flux) | % | fixed (100) |
|   ammonium uptake (% of incoming flux) | % | uniform (0,100) |
|   nitrate uptake (% of incoming flux) | % | uniform (0,100) |
| soil base cation weathering ($BC_w$) | | |
|   $Ca^{2+}$ weathering | meq m$^{-2}$ yr$^{-1}$ | uniform (0,100) |
|   $Mg^{2+}$ | meq m$^{-2}$ yr$^{-1}$ | uniform (0,100) |
|   $Na^+$ | meq m$^{-2}$ yr$^{-1}$ | uniform (0,100) |
|   $K^+$ | meq m$^{-2}$ yr$^{-1}$ | uniform (0,100) |
| initial soil base cation saturation ($BS_{ini}$)[b] | | |
|   exchangeable $Ca^{2+}$ | % | uniform (0.1,50) |
|   exchangeable $Mg^{2+}$ | % | uniform (0.1,50) |
|   exchangeable $Na^+$ | % | uniform (0.1,50) |
|   exchangeable $K^+$ | % | uniform (0.1,50) |
| surface water characteristics | | |
|   relative area water:catchment | % | fixed (0.01) |
|   water retention time | year | fixed (0) |
|   aluminum dissolution constant | Log10 | uniform (6,11) |
|   temperature | °C | uniform (3,10) |
|   $CO_2$ partial pressure | % of atm | uniform (0.05,0.2) |
|   dissolved organic C concentration | $\mu$mol L$^{-1}$ | uniform (0,100) |
|   nitrification (% of incoming flux) | % | fixed (100) |
| **time-dependent parameters $x_t$** | | |
|   runoff amount[c] | m yr$^{-1}$ | normal ($z_{1t}$, $\gamma_1 z_{1t}$) |
| deposition concentrations[d] | | |
|   $Ca^{2+}$ | $\mu$eq L$^{-1}$ | normal ($z_{2t}$, $\gamma_2 z_{2t}$) |
|   $Mg^{2+}$ | $\mu$eq L$^{-1}$ | normal ($z_{3t}$, $\gamma_2 z_{3t}$) |
|   $Na^+$ | $\mu$eq L$^{-1}$ | normal ($z_{4t}$, $\gamma_2 z_{4t}$) |
|   $K^+$ | $\mu$eq L$^{-1}$ | normal ($z_{5t}$, $\gamma_2 z_{5t}$) |
|   $NH_4^+$ | $\mu$eq L$^{-1}$ | normal ($z_{6t}$, $\gamma_2 z_{6t}$) |
|   $SO_4^2$ | $\mu$eq L$^{-1}$ | normal ($z_{7t}$, $\gamma_2 z_{7t}$) |
|   $Cl^-$ | $\mu$eq L$^{-1}$ | normal ($z_{8t}$, $\gamma_2 z_{8t}$) |
|   $NO_3^-$ | $\mu$eq L$^{-1}$ | normal ($z_{9t}$, $\gamma_2 z_{9t}$) |

[a] Prior distributions are selected from the available knowledge on each parameter. In cases where an expert judgment has suggested upper and lower limits, we have used uniform distributions. [b] The sum of four exchangeable base cations ($Ca^{2+}$, $Mg^{2+}$, $Na^+$, $K^+$) is constrained to be $\leq 100$. [c] The observed runoff amount in year $t$ is $z_{1t}$. (The MAGIC model uses observed runoff and contains no hydrological model.) $\gamma$ denotes hyper parameters that are estimated from data (see details in the Supporting Information). [d] The observed deposition concentration in year $t$ for component $i$ ($i = 2, ..., 9$) is $z_{it}$. The distributions for all components two to nine are exactly correlated at the same time with correlation 1. All other prior distributions are independent.

On the other hand, weathering rates ($BC_w$) and initial base saturation ($BS_{ini}$) are either difficult or impossible to measure, and hence their prior distributions are rather wide and non-informative. Five of the constant input parameters in $\mathbf{v}$ are fixed, and hence 23 constant input parameters remain to be estimated.

The prior distributions for the time-dependent parameter vector $\mathbf{x}_t$ are explicitly based on corresponding time-dependent input data $\mathbf{z}_t$. We expect that $\mathbf{x}_t$ is close to $\mathbf{z}_t$, but we are not willing to specify the magnitude of the difference. The prior distribution for $\mathbf{x}_t$ therefore has a mean equal to $\mathbf{z}_t$ and standard deviation proportional to $\mathbf{z}_t$, where the proportionality factors $\gamma_1$ and $\gamma_2$ are estimated from data.

The error $\epsilon_t$ is assumed to be normally distributed, with mean 0, uncorrelated elements and uncorrelated in time. These assumptions are discussed further later. We denote the vector of standard deviations by $\boldsymbol{\sigma}$. The first five elements of $\boldsymbol{\sigma}$ ($\sigma_1, ..., \sigma_5$) are fixed at plausible values (subjective guesses given by experts), since there is only one observation available for each of these. The remaining elements ($\sigma_6, ..., \sigma_{14}$) are parameters to be estimated from the data.

Let $\mathbf{y}$ denote all available output data from year $t = 1$ (1974) to year $T$ (2002). Further, let $\boldsymbol{\theta}$ denote all parameters to be estimated. These comprise (a) the time-independent input parameters $\mathbf{v}$, (b) the time-dependent input parameters $\mathbf{x}_t$ up to year $T$, (c) the parameters $\gamma_1$ and $\gamma_2$, and (d) the unknown's. The likelihood is then

$$L(\mathbf{y};\boldsymbol{\theta}) \propto \{1/\prod_{t=1}^{T}\prod_{i=1}^{14}\sigma_i\}\exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{14}\sigma_i^{-2}(y_{it}-m_{it})^2\right\}$$

(4)

where the index $i$ denotes the $i$'th element of a vector. Missing observations (see the data section) are just ignored in the likelihood.

Our parameter space is large, but hopefully it is possible to estimate the parameters with reasonable precision, because (*i*) we have informative priors for all input param-

**TABLE 2. Output Parameters, Likelihood, and Number of Observations of Corresponding Output Data**

| output parameters $m_t$ | unit | likelihood of output data $y_t$ | number of observations |
|---|---|---|---|
| soil pH | pH | normal $(\mu_1, \sigma_1 = 0.15)$ | 1 |
| soil base cation saturation | | | 1 |
| exchangeable $Ca^{2+}$ | % | normal $(\mu_2, \sigma_2 = 0.5)$ | 1 |
| exchangeable $Mg^{2+}$ | % | normal $(\mu_3, \sigma_3 = 0.5)$ | 1 |
| exchangeable $Na^+$ | % | normal $(\mu_4, \sigma_4 = 0.5)$ | 1 |
| exchangeable $K^+$ | % | normal $(\mu_5, \sigma_5 = 0.5)$ | 1 |
| surface water concentrations[a] | | | |
| $Ca^{2+}$ | $\mu$eq $L^{-1}$ | normal $(\mu_6 t, \sigma_6)$ | 27 |
| $Mg^{2+}$ | $\mu$eq $L^{-1}$ | normal $(\mu_7 t, \sigma_7)$ | 27 |
| $Na^+$ | $\mu$eq $L^{-1}$ | normal $(\mu_8 t, \sigma_8)$ | 27 |
| $K^+$ | $\mu$eq $L^{-1}$ | normal $(\mu_9 t, \sigma_9)$ | 27 |
| $SO_4^{2-}$ | $\mu$eq $L^{-1}$ | normal $(\mu_{10} t, \sigma_{10})$ | 27 |
| $Cl^-$ | $\mu$eq $L^{-1}$ | normal $(\mu_{11} t, \sigma_{11})$ | 27 |
| $NO_3^-$ | $\mu$eq $L^{-1}$ | normal $(\mu_{12} t, \sigma_{12})$ | 27 |
| $H^+$ | $\mu$eq $L^{-1}$ | normal $(\mu_{13} t, \sigma_{13})$ | 27 |
| Al | $\mu$eq $L^{-1}$ | normal $(\mu_{14} t, \sigma_{14})$ | 18 |

[a] $\sigma_6 - \sigma_{14}$ denote hyper parameters that are estimated from data (see details in the Supporting Information).

eters, (*ii*) there are several output data for calibration, since they are multivariate and from long time period, and (*iii*) the MAGIC model has shown to mimic the true process reasonably well in many applications.

## 3. Results

**3.1. Examination of Distributional Assumptions.** We calibrated the model on the entire period of data (1974−2002). To check some of the distributional assumptions, we calculated the empirical residuals $\hat{\epsilon}_t = y_t - \hat{m}_t$, where $\hat{m}_t$ are the posterior means. Details are found in the Supporting Information. The normality assumption seemed to be reasonable in the calibration period. However, the individual elements of the residuals were correlated with each other within the same time, as already mentioned in the previous section. Furthermore, the residuals were positively autocorrelated over time.

**3.2. Examination of Prediction Ability.** To test the prediction ability of the model, we re-calibrated it to the 1974−1990 data only. In addition, the estimated model was used to predict output parameters (with uncertainty) from 1991 to 2002, conditioned on the observed time-dependent input data in the same period.

For all parameters the estimated (1974−1990) and predicted (1991−2002) output parameters are in reasonable agreement with the corresponding observations (Figure 1). For instance, the observed declines in the concentrations of sulfate and calcium are reproduced, and the falls or rises in 1993 in various parameters are predicted well (1993 had an unusual water chemistry, induced by an extreme storm event (*21*)).

Figure 1 also shows 95% pointwise credible intervals (sometimes also called Bayesian confidence intervals) for the output parameters $m_t$ (inner bands) and 95% prediction intervals for the observations $y_t$ (outer bands), where the latter include the uncertainty of the error $\epsilon_t$. We expect that around 95% of the observations fall inside the prediction intervals (the outer intervals), and this holds for all individual parameters. For ANC, however, the width of the prediction interval seems to be overestimated. The reason is that correlations within the elements of the errors $\epsilon_t$ are ignored in our model.

We did not re-estimate the whole model, but calculated a modified 95% prediction interval, taking into account the empirical covariance matrix of $\epsilon_t$. The spread of the observations seems reasonable compared to this modified interval, shown as the medium interval in the ANC panel in Figure 1. The covariance matrix of $\epsilon_t$ was estimated by combining

the estimated variances (with uncertainties) and the correlation matrix from Table S-1 in the Supporting Information (without uncertainty).

Overall, this test of the calibration routine shows that the last 12 years of output data, excluded from the calibration procedure, are reproduced satisfactorily by the conditional predictions of the output parameters. Therefore, even if the model definitively may be improved according to the results of Section 3.1, it still seems to be useful for predictions ahead in time.

**3.3. Scenario Forecasts.** We now return to the model calibrated on the entire period of data (1974−2002). Future values (until 2050) of the output parameters were forecasted conditional on certain scenarios for the input data in the same period. The scenario for measured runoff amount is constant, taken as the average of all the observations. Scenarios for measured S and N deposition have been estimated at the EMEP grid scale based on European emission scenarios (*20*). The deposition sequences used were averages for the four EMEP grid cells covering most of southern Norway, scaled to current deposition. Three different scenarios were used:

*Current Legislation (CLE).* This scenario is based on the current legislation in Europe, assuming these are followed as legislated.

*Maximum Feasible Reduction (MFR).* This scenario assumes that best available technology will be fully implemented in Europe between 2010 and 2015. This scenario entails considerable reductions of both S and N emissions compared to the CLE scenario.

*No Anthropogenic Deposition (NOD).* This scenario assumes a linear decrease to zero from 2010 to 2015 in all deposition of anthropogenic origin. Although clearly unreasonable, this scenario provides a measure of maximum recovery in the future.

Time trends for the conditional forecasts (mean posterior values) for each year are shown in Figure 2. Under all scenarios a substantial change from the current situation is expected for the different major ions. The changes in $SO_4^{2-}$ and $NO_3^-$ are direct responses to the changes in the deposition input. The concentrations of $Ca^{2+}$, $H^+$, and Al decrease as a result of reduced concentration of mobile anions (in particular $SO_4^{2-}$) and increased pH of the deposition. ANC is projected to increase and there are clear differences among the different scenarios.

Figure 3 shows the posterior distributions for the ANC parameter for the different scenarios in 2020 and 2050. The uncertainty includes the uncertainty of the output parameter
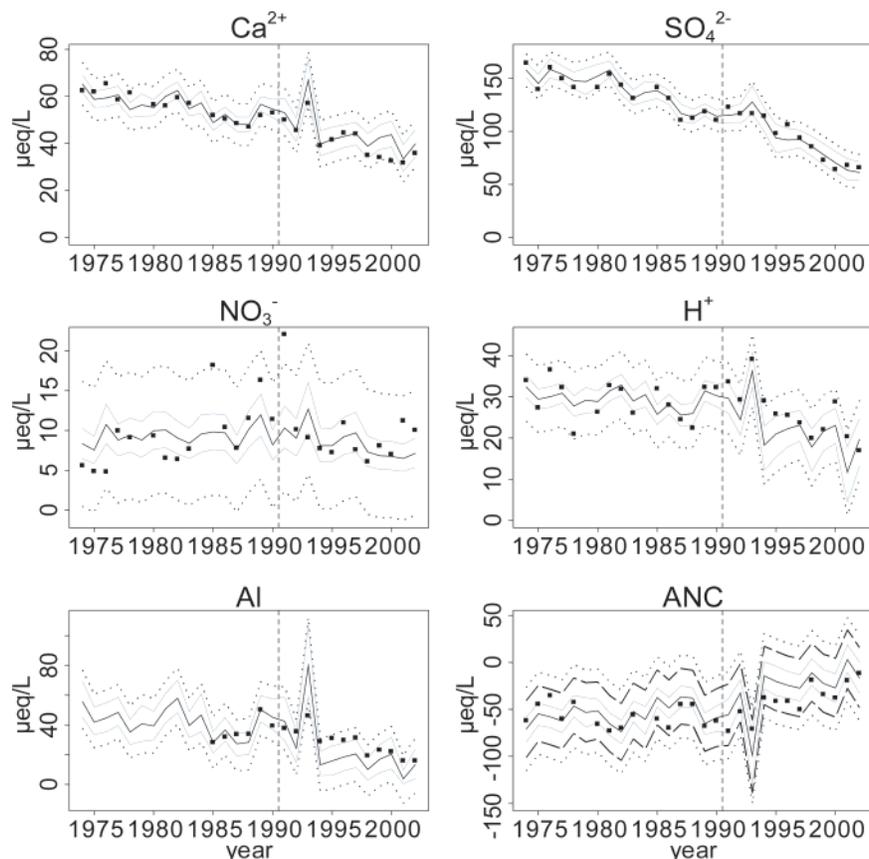
**FIGURE 1. Estimates (1974−1990) and conditional predictions (1991−2002) of selected output parameters (solid lines) with corresponding observations (squares). Estimated and predicted values are calculated as the posterior means. In addition the light blue lines show the 95% credible intervals for the output parameters; the outer, dotted lines show the 95% prediction intervals for the observations. For ANC, the dashed lines show the adjusted 95% prediction intervals. Output data up to 1990 (indicated by the dotted vertical line) were used for calibration; the output data after that were used for comparison with predictions conditioned on the input data after 1990.**

$\mathbf{m}_t$, corresponding to the (inner) credibility intervals in Figure 1. Uncertainty due to $\epsilon_t$ is ignored here, since $\epsilon_t$ only represents short-term variations from year to year.

**3.4. Coupling to a Dose−Response Model for Fish Populations.** The scenario forecasts for ANC were coupled to a simple, static dose−response model for fish survival (*10*, *22*). The model is a logistic regression model with three classes of fish population status related to ANC in water (healthy, damaged, and extinct). The probability for having a healthy trout population as a function of ANC is

$$p = \frac{1}{1 + e^{a+bANC}}, \qquad (5)$$

where *a* and *b* are parameters with estimated values and standard errors given in ref *22*. Based on this, we have constructed corresponding priors: the priors for *a* and *b* are both normal with means −1.01 and −0.13 and standard deviations 0.012 and 0.05, respectively.

Combining this model with the posterior distribution for predicted ANC conditioned on the different scenarios gives a posterior distribution for the probability of achieving a healthy fish population in the stream at a given year. For illustration, we show the predicted situations in 2020 and 2050 in Figure 4. The results show very clear differences for the different scenarios. In 2020, there is an estimated probability of 0.29 (the posterior mean) for a healthy fish population under the CLE scenario, increasing to 0.53 under the MFR scenario and further increasing to 0.89 under the NOD scenario. The difference between the estimated probabilities for the CLE and MFR scenarios in 2020 is 0.24 with 95% credible interval 0.19−0.29 (see Table S-4 in the

Supporting Information). Hence, even though the individual estimated probabilities have rather wide credible intervals as seen in Figure 4, the differences among them are highly significant, with rather small credible intervals, taking into account the correlations between the individual posterior distributions.

The considerable time lag in recovery of the water chemistry is seen from the difference between the years 2020 and 2050; there is considerably larger probability of having a healthy fish population in 2050 than in 2020, even though the deposition is kept constant from 2015. The time lag seen from the onset of a deposition change to the response in the predicted fish population status is only related to the time lag in water chemistry in this approach. The dose−response model does not include any information about the time lag in the fish response to the change in ANC. For natural recovery of a fish population the recovery delay time can be considerable. However, for lakes and streams being stocked with fish, the response time can be assumed immediate.

## 4. Discussion

Our stochastic model (eq 2) can be modified in several ways. We have already pointed out that the errors $\epsilon_t$ were correlated at a single point in time, and autocorrelated over time, so an improved model for $\epsilon_t$ could be a multivariate time series model. Another useful modification could be to take into account that the output data $\mathbf{y}_t$ in our case always are nonnegative, by using multiplicative instead of additive errors. The observant reader may have seen that the (outer) prediction intervals for $NO_3^-$ and Al in Figure 1 are in fact negative. However, this does not have a dramatic effect, since
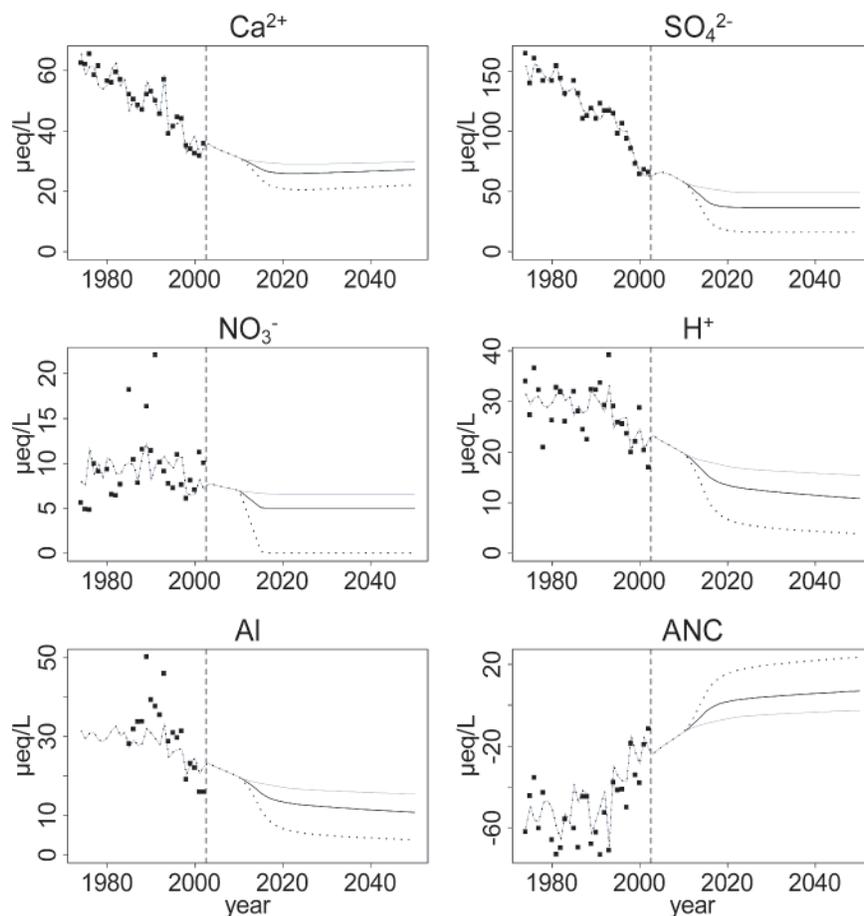
**FIGURE 2. Estimates (dotted line until 2002) and conditional forecasts (lines after 2002) of selected output parameters with corresponding observations (squares). The forecasts are conditioned on each of the three different forecast scenarios. The light blue lines show the CLE (current legislation) scenario, the solid black line shows the MFR (maximum feasible reduction) scenario, and the dotted line shows the NOD (no anthropogenic deposition) scenario. Observations for all years (1974–2002) were used for calibration (indicated by the dotted vertical line).**
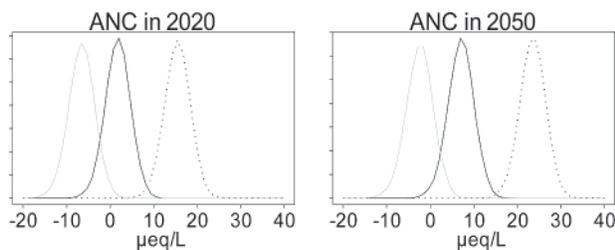


**FIGURE 3. Posterior distributions (corresponding to the inner band in Figure 1) for the ANC parameter in 2020 and 2050 conditioned on each of the three different scenarios. The light blue line shows the CLE (current legislation) scenario, the solid black line shows the MFR (maximum feasible reduction) scenario, and the dotted line shows the NOD (no anthropogenic deposition) scenario.**
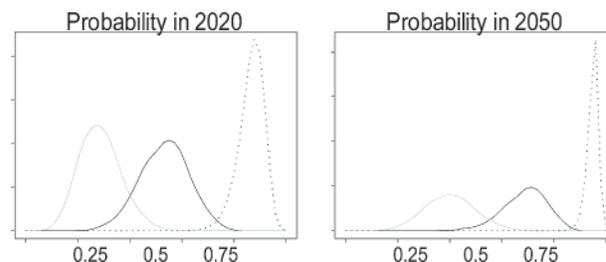


**FIGURE 4. Posterior distributions for probability of having a healthy fish population in 2020 and 2050 conditioned on each of the three different scenarios. The light blue line show the CLE (current legislation) scenario, the solid black line shows the MFR (maximum feasible reduction) scenario, and the dotted line shows the NOD (no anthropogenic deposition) scenario.**

the MAGIC model guarantees the output parameter $m_t$ to always be non-negative. Furthermore, the prior distribution for $x_t$ (Table 1) could also be replaced by a non-negative distribution, for instance log-normal.

We have so far not given any interpretation of the error $\epsilon_t = y_t - m_t$. It may contain both model error, i.e., the difference between $m_t$ and a hypothetical true value $m_t^{\text{TRUE}}$, and measurement error, i.e., the difference between $y_t$ and $m_t^{\text{TRUE}}$, and it can be difficult to separate these two sources of error. One important consequence of model error may be model bias, i.e., a systematic difference between $m_t$ and $m_t^{\text{TRUE}}$. This may be accounted for if the output data $y_t$ are

unbiased measurements of $m_t^{\text{TRUE}}$, by applying the model

$$y_t = \beta_0 + \beta_1 \, m_t + \epsilon_t \qquad (6)$$

instead of the model in eq 2. Here $\beta_0$ is a parameter that accounts for additive bias, whereas $\beta_1$ is a parameter that accounts for multiplicative bias. A further discussion on model inadequacy and various sources of error can be found in ref *11*.

Model predictions will always have uncertainties, although they are commonly ignored in many model applications. We have presented a statistical framework for automatic calibration and uncertainty estimation in deterministic models

using a Bayesian approach based on an MCMC algorithm for combing information from observations with the deterministic model. The approach makes it possible to present model prediction results as probability distributions rather than single values.

In most model applications, in particular those linked to decision support or policy making, uncertainty analyses are commonly ignored. Uncertainty analysis is often considered complicated and may even be seen as undermining the conclusions from the model simulations because one actually reveals that the results are uncertain. In the development of our approach we have been mindful that, although the method itself is technically complex, the final outputs from the analysis should be easy to communicate in a straightforward manner.

Many policy makers and other stake holders have experience in dealing with risks. If results from model applications are presented as probability distributions rather than single numbers, those using the results may be able to combine the probabilities with the consequences and act according to the resultant perceived risk. In this way models can be even more powerful tools in support of decision making and assessments.

## Supporting Information Available

More detailed description of the prior and posterior distributions and some details about the algorithm used; additional results, including an illustration of how the MCMC algorithm works and an examination of the distributional assumptions used; a discussion on how our approach compares to other Bayesian approaches. This material is available free of charge via the Internet at http://pubs.acs.org.

## Literature Cited

(1) Hooper, R. P.; Stone, A.; Christophersen, N.; Degrosbios, E.; Seip, H. M. Assessing the Birkenes model of stream acidification using a multisignal calibration methodology. *Water Resour. Res.* **1988**, *24*, 1308–1316.

(2) Beven, K.; Binley, A. The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process.* **1992**, *6*, 279–298.

(3) Kuczera, G.; Parent, E. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the metropolis algorithm. *J. Hydrol.* **1998**, *211*, 69–85.

(4) Page, T.; Beven, K.; Freer, J.; Jenkins, A. Investigating the uncertainty in predicting responses to atmospheric deposition using the Model of Acidification of Groundwater in Catchments (MAGIC) within a Generalized Likelihood Uncertainty Estimation (GLUE) framework. *Water Air Soil Pollut.* **2003**, *142*, 71–94.

(5) Thiemann, M.; Trosset, M.; Gupta, H.; Sorooshian, S. Bayesian recursive parameter estimation for hydrologic models. *Water Resour. Res.* **2001**, *37*, 2521–2535.

(6) Bates, S.; Cullen, A.; Raftery, A. Bayesian uncertainty assessment in multicompartment deterministic simulation models for environmental risk assessment. *Environmetrics* **2003**, *14*, 355–371.

(7) Beven, K. A manifesto for the equifinality thesis. *J. Hydrol.* **2006**, *320*, 18–36.

(8) Cosby, B. J.; Hornberger, G. M.; Galloway, J. N.; Wright, R. F. Modelling the effects of acid deposition: assessment of a lumped parameter model of soil water and streamwater chemistry. *Water Resour. Res.* **1985**, *21*, 51–63.

(9) Cosby, B. J.; Ferrier, R. C.; Jenkins, A.; Wright, R. F. Modelling the effects of acid deposition: refinements, adjustments and inclusion of nitrogen dynamics in the MAGIC model. *Hydrol. Earth Syst. Sci.* **2001**, *5*, 499–518.

(10) Lien, L.; Raddum, G. G.; Fjellheim, A. *Critical Loads of Acidity to Freshwater-Fish and Invertebrates*; NIVA-report 2732/1992; Norwegian Institute for Water Research: Oslo, 1992.

(11) Kennedy, M. C.; O'Hagan, A., Bayesian calibration of computer models (with discussion). *J. Roy. Stat. Soc.* **2001**, *63*, 425–464.

(12) Wright, R. F.; Larssen, T.; Camarero, L.; Cosby, B. J.; Ferrier, R. C.; Helliwell, R. C.; Forsius, M.; Jenkins, A.; Kopácek, J.; Majer, V.; Moldan, F.; Posch, M.; Rogora, M.; Schöpp, W. Recovery of acidified European surface waters. *Environ. Sci. Technol.* **2005**, *39*, 64A–72A.

(13) Larssen, T., Model Prognoses for Future Acidification Recovery of Surface Waters in Norway Using Long-Term Monitoring Data. *Environ. Sci. Technol.* **2005**, *39*, 7970–7979.

(14) Larssen, T.; Brereton, C. I.; Gunn, J. Dynamic modelling of Recovery from Acidification of Lakes in Killarney Park, Ontario, Canada. *Ambio* **2003**, *32*, 244–248.

(15) Henriksen, A.; Fjeld, E.; Hesthagen, T. Critical load exceedance and damage to fish populations. *Ambio* **1999**, *28*, 583–586.

(16) Aas, W.; Tørseth, K.; Solberg, S.; Berg, T.; Manø, S.; Yttri, K. E. *Monitoring of Long Range Transported Air Pollution and Precipitation - Atmospheric deposition 2002*; The Norwegian Pollution Control Authority (SFT): Oslo, 2003.

(17) SFT. *The Norwegian Monitoring Programme for Long-Range Transported Air Pollutants. Annual Report - Effects 2003*; The Norwegian Pollution Control Authority (SFT): Oslo, 2004.

(18) SFT. *The Norwegian Monitoring Programme for Long-Range Transported Air Pollutants. Annual Report 1992*; The Norwegian Pollution Control Authority (SFT): Oslo, 1993.

(19) SFT. *The Norwegian Monitoring Programme for Long-Range Transported Air Pollutants. Annual Report - Effects 2000*; The Norwegian Pollution Control Authority (SFT): Oslo, 2001.

(20) Schöpp, W.; Posch, M.; Mylona, S.; Johansson, M. Long-term development of acid deposition (1880–2030.) in sensitive freshwater regions in Europe. *Hydrol. Earth Syst. Sci.* **2003**, *7*, 436–446.

(21) Hindar, A.; Henriksen, A.; Tørseth, K.; Semb, A. Acid water and fish death. *Nature* **1994**, *372*, 327–328.

(22) Lydersen, E.; Fjeld, E.; Andersen, T. *Fish Status and Water Chemistry in Norwegian Lakes*; NIVA-report 3002/1994; Norwegian Institute for Water Research: Oslo, NIVA, 1994.

# Supporting Information

**Uncertainties in forecasts of acidification effects using a Bayesian calibration and**

**uncertainty propagation approach**

Thorjørn Larssen[1,*], Ragnar B. Huseby[2], Bernard J. Cosby[3], Gudmund Høst[4], Tore Høgåsen[1],

Magne Aldrin[2]

[1] Norwegian Institute for Water Research, P.O.Box 173 Kjelsaas, 0411 Oslo, Norway;

tla@niva.no, toh@niva.no

[2] Norwegian Computing Center, P.O.Box 114 Blindern, 0314 Oslo, Norway;

ragnar.huseby@nr.no, magne.aldrin@nr.no

[3] Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22904-4123, USA; bjc4@virginia.edu

[4] Norwegian Computing Center, now at Norwegian Research Council, 0131 Oslo;

gho@forskningsradet.no

[*] Corresponding author

This Supporting Information consists of 12 pages of text, 4 tables and 4 figures.

## 1. Model description

21

22 The model framework itself is described in the main paper (2.1). A sketch of model

23 framework showing the interaction between input parameters, input data, the deterministic

24 model, output parameters and output data are shown in Figure S-1.

25

### 1.1. Prior distributions and likelihood

26

27 The prior for the time-dependent input parameter $\mathbf{x}_t$ is normal distributed with mean equal to

28 the corresponding observation $\mathbf{z}_t$ (Table 1 in the main paper). Let $r_{it}=x_{it}-z_{it}$ denote the

29 difference between the i-th output parameter and it's corresponding observation. To keep the

30 number of unknown parameters within reasonable limits and save computer time, we have

31 assumed that $r_{2t}$, …, $r_{9t}$ are exact correlated. Note that this does not mean that there is an exact

32 relationship between the parameter values $x_{2t}$, …, $x_{9t}$, since there is no exact relationship

33 between the observations $z_{2t}$, …, $z_{9t}$. Our approach is still much more flexible than other

34 calibrations of the MAGIC model, including references (1,2) where only 12 time-independent

35 input parameters were calibrated, and all remaining input parameters were fixed.

36

37 The two proportionality factors $\gamma_1$ and $\gamma_2$ are treated as hyper-parameters (i.e. parameters of

38 the prior distribution) to be estimated with vague priors ($1/\gamma_1^2$ and $1/\gamma_2^2$ being gamma

39 distributed with shape and rate equal to 0.01).

40

41 The standard deviations $\sigma_6$, …, $\sigma_{14}$ are parameters to be estimated with vague prior

42 distributions ($1/\sigma_i^2$ being gamma distributed with shape and rate equal to 0.01).

43

44

### *1.2.    Posterior distribution*

46    In the Bayesian framework, the parameter vector $\theta$ is regarded as a multivariate random

47    variable with a prior distribution, which we now will denote $\mathbf{p}(\theta|\mathbf{z})$ to emphasize that the prior

48    distribution is based on the input data $\mathbf{z}$. This prior distribution is defined through the

49    individual prior distributions given in the previous section. The purpose of a Bayesian

50    analysis is to estimate the conditional or posterior distribution of the parameter vector $\theta$ given

51    all the data $\mathbf{y}$ and $\mathbf{z}$. By Bayes' formula, the conditional distribution $\mathbf{p}(\theta \mid \mathbf{y},\mathbf{z})$, also called the

52    posterior distribution, is given by $\mathbf{p}(\theta \mid \mathbf{y},\mathbf{z}) \propto L(\mathbf{y};\theta)\, \mathbf{p}(\theta|\mathbf{z})$ .  (See discussion at eq. 3 in the

53    main paper).

54

55    By the Bayesian paradigm, this posterior distribution contains all information relevant to the

56    input parameters, given the model $\mathbf{m}$ and the data $\mathbf{y}$ and $\mathbf{z}$. We see that the posterior is a

57    combination of the likelihood and the prior. For example, input parameters which provide

58    output parameters that correspond well with the output data are associated with high posterior

59    probability unless these input parameters have low prior probability. We use vague (wide)

60    prior distributions for parameters for which we have no prior information, and the results

61    should not be very sensitive to these vague prior distributions. In classical calibration

62    problems, the main interest concerns the "optimal" value of $\theta$. Such "calibrated" values could

63    be taken as the mean of the posterior distribution. However, the Bayesian formulation

64    provides a complete probability distribution for the input parameters, thereby describing the

65    uncertainty of the parameters. This is of particular interest in decision-making and in

66    subsequent use of the results in assessment of ecological risk. The posterior probabilities of

67    the output parameters from MAGIC may be directly generated by running simulations

68    drawing parameter values from the posterior distribution $\mathbf{p}(\theta \mid \mathbf{y},\mathbf{z})$.

69

70 We are interested in using the estimated model to make forecasts for future output parameters,

71 say up to year T*>T, conditional on certain scenarios for future input parameters $\mathbf{x}_{\rhd T}$ One

72 possibility is to specify exact values for $\mathbf{x}_{\rhd T}$. Another possibility is to use a vague scenario for

73 $\mathbf{x}_{\rhd T}$ by specifying a prior distribution for $\mathbf{x}_{\rhd T}$ based on a given scenario for future input data

74 $\mathbf{z}_{\rhd T}$. We will use this latter approach in the result section. The posterior distribution of $\mathbf{m}_{\rhd T}$,

75 $\mathbf{p}(\mathbf{x}_{\rhd T}|\mathbf{y},\mathbf{z}, \mathbf{z}_{\rhd T})$, is easily found by simulations from the model $\mathbf{m}$ using the prior distribution

76 for $\mathbf{x}_{\rhd T}$ and the posterior distribution for $\mathbf{\theta}$.

77

78 The posterior distribution of $\mathbf{\theta}$ is virtually impossible to specify analytically, due to the

79 complexity of the relationship between the input parameters, the output parameters and the

80 data. However, samples from the posterior may be generated by Monte Carlo simulation,

81 using an MCMC method. A general description of the MCMC algorithm can be found in

82 numerous references, see for example (3). In the current example, we have used the

83 Metropolis-Hastings MCMC algorithm (4). This algorithm generates samples from a random

84 walk, which eventually converges to the true posterior.

85 *1.3.    Algorithm details*

86 We want to generate samples from the posterior distribution  $\mathbf{p}(\mathbf{\theta}|\mathbf{y},\mathbf{z})$. The input parameters

87 consist of 23 time-independent scalars, eight time-series of annual deposition concentrations,

88 one time-series of annual runoff amount,  the unknown $\sigma$'s and the hyper parameters $\gamma_1$ and

89 $\gamma_2$. The Metropolis-Hastings scheme can be described as follows.

90 Assume that we have generated the sample vectors $\mathbf{\theta}^0, \mathbf{\theta}^1, ..., \mathbf{\theta}^K$ .

91     1.  A candidate $\mathbf{\theta}^{proposal}$ for the next sample is determined randomly according to some

92         probability distribution.

93     2.  Calculate the Hastings ratio

94 $$r = \frac{q(\boldsymbol{\theta}^{proposal}, \boldsymbol{\theta}^K) p(\boldsymbol{\theta}^{proposal} \mid \mathbf{y}, \mathbf{z})}{q(\boldsymbol{\theta}^K, \boldsymbol{\theta}^{proposal}) p(\boldsymbol{\theta}^K \mid \mathbf{y}, \mathbf{z})}$$

95    where q(α,β). is the density of the proposed sample vector β given α.

96    If the Hastings ratio is greater than 1, this means that the proposed sample is more

97    likely than the current one.

98    3.  The value of $r$ determines whether the proposed sample vector is accepted   (i.e.

99    $\boldsymbol{\theta}^{K+1} = \boldsymbol{\theta}^{proposal}$  or rejected (i.e. $\boldsymbol{\theta}^{K+1} = \boldsymbol{\theta}^K$). If $r$ is equal to 1 or greater the proposed

100    sample vector is always accepted.  If $r$ is smaller than 1 it is accepted with probability

101    $r$. Thus, we always accept a proposed sample vector that is more likely than the

102    current one but we sometimes accept a proposed sample vector that is less likely.

103    In our implementation, $\boldsymbol{\theta}^{proposal}$ and $\boldsymbol{\theta}^K$ differ in only a subset of the parameters. We start by

104    updating the unknown σ's and the hyper parameters, one parameter at the time, according

105    their posterior distribution given the data and the remaining parameters. This ensures that the

106    Hastings ratio is equal to one. Consequently, the proposal is always accepted. This can be

107    done because their posterior distributions are simple. For instance, the posterior distribution of

108    $\sigma_i^{-2}$ is the gamma distribution with shape $= 0.01 + T/2$ and

109    $\text{rate} = 0.01 + \dfrac{1}{2} \sum_{t=1}^{T} (y_{it} - \mu_{it})^2$ .

110    This is due to the fact that the posterior distribution is proportional to the product of the

111    probability density of the output data, which is Gaussian, and the prior distribution, which is a

112    gamma distribution. For the 23 time-independent scalars and the time-series, their individual

113    posterior distributions given the other parameters and the data are not simple. Therefore, we

114    apply the rejection sampling strategy in the following steps. In general the proposed update of

115    the parameter is drawn from an interval surrounding the current value of the parameter. We

116    first generate proposals where one of the time-independent scalars is changed while all the

117  order parameters are kept fixed. Next we generate proposals where all the wet concentrations

118  for a given year is multiplied with a common scalar (to account for the perfect correlation

119  between the eight components) while the remaining parameters are fixed. We generate one

120  such proposal for each year. Finally, we propose changes in runoff amount  for each year. The

121  loop consisting of updating all the parameters is executed a large number of times.

122

123  **2.  Results**

124  ***2.1.    Illustration of the MCMC algorithm***

125  The Metropolis-Hastings algorithm is iterative. Each iteration produces a set of values for $\theta$.

126  When the algorithm has converged, it will sample values from the posterior distribution of $\theta$.

127  Since the iterations are correlated, many iterations are usually needed to calculate the

128  posterior probability with sufficient precision.

129

130  The convergence of the algorithm was judged manually, by visual inspection of trace plots of

131  various parameters (Figures S-3 and S-4). A trace plot shows the sampled value of a

132  parameter on the y-axis and the iteration number on the x-axis. The left panels of Figure S-3

133  show the trace plots for the lumped parameters $BC_w$ and $BS_{ini}$. Initially, the algorithm

134  suggests values from the full range of the prior distributions, but stabilizes to a much narrower

135  range of values after some thousand iterations. The initially large fluctuations are typical for

136  the "burn-in" of MCMC algorithms.  After the simulated values have stabilized around a

137  fixed level, they still show rather large fluctuations and are highly correlated. Therefore

138  several thousand iterations are necessary to estimate the prior distributions with sufficient

139  precision. Based on these plots and trace plots of other parameters (Figure S-4) we regard the

140     first 10,000 iterations as the burn-in phase, and use the last 70,000 iterations for calculating

141     the posterior distributions.

142

143     The prior and posterior distributions for $BC_w$ and $BS_{ini}$ show that the posterior distributions

144     are comparatively narrow and in the lower part of the prior distributions (right panels of

145     Figure S-3). The rather flat prior for $BC_w$ between 0 and 400 meq m$^{-2}$ yr$^{-1}$ arises as the

146     combination of the four uniform distributions for the four individual input parameters that

147     constitute the total. The posterior distribution is much more peaked with probability mass

148     concentrated around 60 meq m$^{-2}$ yr$^{-1}$. For the $BS_{ini}$ the prior is also rather flat, with somewhat

149     higher density in the higher end of the distribution. Again, this shape of the distribution is

150     related to the fact that the parameter shown is a combination of four distributions with the

151     constraint that the sum of the four should not exceed 100%. The posterior is again rather

152     narrow, with a maximum around 40%.

153

154     Trace plots for selected output parameters show that these also stabilize near the

155     corresponding observed value after rather a few iterations (Figure S-4). The variation after

156     burn-in is about 10 µeq L$^{-1}$ for $SO_4^{2-}$ (corresponding to 15% of the observed value in 2002)

157     and about 7 µeq L$^{-1}$ for $Ca^{2+}$ (corresponding to 20% of the observed value in 2002), with

158     some more extreme peaks every now and then throughout the 80,000 iterations. The variation

159     in modeled ANC is about 20 µeq L$^{-1}$. Remember that observations of ANC are not directly

160     used in the model calibration. Rather observations of the seven individual components of

161     ANC are used. The performance of ANC thus sums up the overall calibration performance of

162     the seven ions.

163

164     Running 80,000 iterations take about 40 hours on a 2 GHz computer.

165

## 2.2.    *Examination of distributional assumptions*

167    To check some distributional assumptions, we first calculated the empirical residuals

168    $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \hat{\mathbf{m}}_t$, where $\hat{\mathbf{m}}_t$ are the posterior means. Remember that we have only one observation

169    for each of the first five elements of $\mathbf{y}_t$, but for the last nine elements we have 27 observations,

170    except for Al where we have 18 observations. The validations will be based on the residuals

171    for the nine outputs with more than one observation.

172

173    Figure S-2 shows quantiles of the residuals versus quantiles of the standard normal

174    distribution. If $\boldsymbol{\varepsilon}_t$ is normal distributed, these plots should roughly show straight lines, which

175    they approximately do. Thus, we find that the normality assumption for $\varepsilon_t$ is reasonable.

176

177    Another assumption is that the individual elements of $\boldsymbol{\varepsilon}_t$ are uncorrelated. The empirical

178    correlation matrix is given in Table S-1. Seven out of 36 correlations are significantly

179    different from 0, which indicates that the assumption of uncorrelated errors is violated. A

180    possible future modification of the statistical model is therefore to include correlations. The

181    basis for these hypothesis tests is that a correlation estimate based on n observation has a

182    standard deviation of roughly $1/\sqrt{n}$ , where n is number of observations, such that empirical

183    correlations with absolute values larger than $\pm 1/\sqrt{n}$ are significant.

184

185    We have also assumed that $\boldsymbol{\varepsilon}_t$ is uncorrelated in time. Table S-2 shows the empirical

186    autocorrelations at lag 1. Four out of nine correlations are significantly positive correlated,

187    clearly indicating that $\boldsymbol{\varepsilon}_t$ is dependent over time.

188

189    *2.3.    Scenario forecasts for ANC and the probability for a healthy fish populations*

190    Table S-3 shows the credibility intervals for the ANC parameter in 2020 and 2050

191    conditioned on the various scenarios, corresponding to Figure 3 in the main paper. The

192    predicted ANC for the CLE scenario in 2020 is -6.6 µeq L$^{-1}$ with the 95% credible interval

193    from -12.5 µeq L$^{-1}$ to -0.9 µeq L$^{-1}$. For the MFR scenario the estimate is 1.8 µeq L$^{-1}$ with the

194    95% credible interval from -3.9 µeq L$^{-1}$ to 7.2 µeq L$^{-1}$. The difference between the two

195    estimates is 8.3 µeq L$^{-1}$, with a 95 % credible interval of only ±1 µeq L$^{-1}$ around the mean,

196    thus the two scenarios are significantly different. The credible intervals for the differences

197    take the correlations between the individual posterior distributions into account (pair-wise

198    comparison), which result in small uncertainties.

199

200    Combining this model with the posterior distribution for predicted ANC conditioned on the

201    different scenarios gives a posterior distribution for achieving a healthy fish population in the

202    stream at a given year. Table S-4 supplements Figure 4 in the main paper. The CLE scenario

203    shows an estimated probability of 0.29 in 2020 and 0.40 in 2050 with quite a large span for

204    the 95% credible interval (0.15-0.46 and 0.22-0.60, respectively). The estimated probability is

205    considerably higher for the MFR scenario with an estimated probability of 0.53 and 95%

206    credible interval of 0.35-0.71 in 2020. For the NOD scenario the estimated probability is 0.87

207    in 2020 and 0.95 in 2050 for a healthy fish population to be successfully established. The

208    difference between the estimated probabilities for the CLE and MFR scenarios in 2020 is 0.24

209    with 95 % credible interval 0.19-0.29.

210

211    **3.  Comparison with other Bayesian approaches**

212    The various approaches for modeling complex environmental systems can roughly be divided

213    into two groups depending on the basic model. One group is based on empirical stochastic

214    models that are rather general and may be used for several applications. in this group

215    complexity is often handled by a hierarchical structure (see review by (5)). The other group

216    (ours being one example) is based on deterministic simulation models designed for a specific

217    applications and based on chemical and physical principles. Combinations of the two

218    approaches also exist. Reference (6) present a meteorological application in which the

219    underlying process is modeled by an empirical spatio-temporal model with input data derived

220    from point measurements and from the output of a deterministic model on an aggregated

221    spatial scale. Similar combined approaches have also been used within air pollution modeling

222    (7,8). Many of the approaches are based on Bayesian modeling, because prior information

223    often exists in addition to data and because the combination of Bayesian algorithms with

224    simulation based estimation methods is a convenient way to handle complex systems.

225

226    The strength of deterministic simulation models is that they can take into account complex

227    structures and dependencies that are difficult to handle within pure empirical models.

228    However, since they are deterministic, it is not obvious how they should be fitted to

229    observational data. The so called generalized likelihood uncertainty estimation (GLUE)

230    method (9) has become a popular way to account for uncertainty. The GLUE method has also

231    been applied with the MAGIC model (1,2). In GLUE, the user specifies prior distributions for

232    the input parameters and an objective function that quantifies how well the output parameters

233    (for given input parameters) fit the output data. This objective function is called a

234    "generalized likelihood". The prior information and the objective function are combined

235    following computational rules from Bayesian statistics, but are not based on a statistical

236    model, and usually the objective function is not a data likelihood function. Prediction of

237    uncertainty in GLUE is based on simulations of $\mathbf{m}(\mathbf{x}_t)$, taking into account the uncertainty in

238    $\mathbf{x}_t$, but ignoring $\varepsilon_t$. Therefore, GLUE uncertainty intervals may be too narrow compared to

239    observations, as pointed out by several authors (9-11).

240

241    However, if the GLUE methodology is adjusted to take into account the error $\varepsilon_t$, as in (12) or

242    (13), the difference between GLUE and a method like ours is then only the Monte Carlo

243    sampling strategy. GLUE sample independently from the prior distribution, and weight each

244    sample to get the correct posterior distribution. Our approach uses the Metropolis-Hastings

245    algorithm, which gives dependent samples from the posterior distribution. Both references

246    (12) and (13) compared the GLUE sampling with Metropolis-Hastings, and concluded that

247    the latter should be preferred, at least in situations with high-dimensional parameter space.

248    Nevertheless, Monte Carlo based techniques like GLUE, our own and several others are very

249    efficient tools to handle uncertainty in complex systems.

250

251    In some circumstances one may also have prior information about the output parameters, in

252    addition to the priors that are induced from the priors on the input parameters. Then one can

253    use a method called Bayesian melding (14).

254

255    **References**

256    (1)      Page, T.; Beven, K.; Freer, J.; Jenkins, A., Investigating the uncertainty in predicting
257    responses to atmospheric deposition using the Model of Acidification of Groundwater in
258    Catchments (MAGIC) within a Generalized Likelihood Uncertainty Estimation (GLUE)
259    framework. *Water Air Soil Pollut*. **2003**, *142*, 71-94.

260    (2)      Page, T.; Beven, K.; Whyatt, D., Predictive capability in estimating changes in water
261    quality: Long-term responses to atmospheric deposition. *Water Air Soil Poll.* **2004**, *151*, 215-
262    244.

263     (3)      Besag, J.; Green, P.; Higdon, D.; Mengersen, K., Bayesian Computation and
264    Stochastic Systems. *Stat. Sci.* **1995**, *10*, 3-41.

265    (4)      Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E.,
266    Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087-
267    1091.

268    (5)      Wikle, C., Hierarchical models in environmental science. *Int. Stat Rev.* **2003**, *71*, 181-
269    199.

270    (6)      Gel, Y.; Raftery, A.; Gneiting, T., Calibrated probabilistic mesoscale weather field
271    forecasting: The geostatistical output perturbation method. *J. Am. Stat. Assoc.* **2004**, *99*, 575-
272    583.

273    (7)      Hirst, D.; Storvik, G.; Syversveen, A., A hierarchical modelling approach to
274    combining environmental data at different scales. *J. Roy. Stat. Soc. C-App.* **2003**, *52*, 377-390.

275    (8)      Fuentes, M.; Raftery, A., Model evaluation and spatial interpolation by Bayesian
276    combination of observations with outputs from numerical models. *Biometrics* **2005**, *61*, 36-
277    45.

278    (9)      Beven, K.; Binley, A., The future of distributed models: model calibration and
279    uncertainty prediction. *Hydrol. Process.* **1992**, *6*, 279-298.

280     (10)    Christensen, S., A synthetic groundwater modelling study of the accuracy of GLUE
281    uncertainty intervals. *Nordic Hydrol.* **2004**, *35*, 45-59.

282    (11)     Montanari, A., Large sample behaviors of the generalized likelihood uncertainty
283    estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. *Water Resour.*
284    *Res.* **2005**, *41*, doi:10.1029/2004WR003826.

285    (12)     Kuczera, G.; Parent, E., Monte Carlo assessment of parameter uncertainty in
286    conceptual catchment models: the metropolis algorithm. *J. Hydrol.* **1998**, *211*, 69-85.

287    (13)     Makowski, D.; Wallach, D.; Tremblay, M., Using a Bayesian approach to parameter
288    estimation; comparison of the GLUE and MCMC methods. *Agronomie* **2002**, *22*, 191-203.

289    (14)     Poole, D.; Raftery, A., Inference for deterministic simulation models: The Bayesian
290    melding approach. *J. Am. Stat. Assoc.* **2000**, *95*, 1244-1255.
291
292

292  **Tables**

293

294  Table S-1: Lower triangle of the correlation matrix for empirical residuals from 1974- 2002.

295  Significant correlations are marked with *.

| | $Ca^{2+}$ | $Mg^{2+}$ | $Na^+$ | $K^+$ | $SO_4^{2-}$ | $Cl^-$ | $NO_3^-$ | $H^+$ | $Al$ |
|---|---|---|---|---|---|---|---|---|---|
| $Ca^{2+}$ | 1.00 | | | | | | | | |
| $Mg^{2+}$ | 0.08 | 1.00 | | | | | | | |
| $Na^+$ | -0.42* | -0.58* | 1.00 | | | | | | |
| $K^+$ | 0.11 | 0.38 | -0.3 | 1.00 | | | | | |
| $SO_4^{2-}$ | -0.01 | -0.26 | 0.00 | -0.22 | 1.00 | | | | |
| $Cl^-$ | -0.24 | -0.16 | 0.47* | -0.13 | -0.58* | 1.00 | | | |
| $NO_3^-$ | 0.04 | -0.27 | -0.20 | 0.25 | 0.39* | -0.20 | 1.00 | | |
| $H^+$ | -0.30 | 0.06 | -0.22 | -0.23 | 0.39* | -0.37 | 0.12 | 1.00 | |
| $Al$ | -0.27 | 0.59* | -0.11 | 0.40 | -0.20 | 0.21 | 0.04 | 0.26 | 1.00 |

296

297

298

299

300  Table S-2: Autocorrelation at lag 1 for empirical residuals from 1974- 2002. Significant

301  correlations are marked with *.

| | |
|---|---|
| $Ca^{2+}$ | -0.06 |
| $Mg^{2+}$ | 0.48* |
| $Na^+$ | 0.42* |
| $K^+$ | 0.75* |
| $SO_4^{2-}$ | -0.08 |
| $Cl^-$ | 0.25 |
| $NO_3^-$ | 0.33 |
| $H^+$ | 0.17 |
| $Al$ | 0.62* |

302

303

303    Table S-3: Forecasted (posterior mean) ANC parameter in 2020 and 2050 conditioned on

304    each of the three scenarios: current legislation (CLE), maximum feasible reduction (MFR)

305    and no anthropogenic deposition (NOD). Differences between the different scenarios are also

306    shown.

| Scenario | 2020 | | 2050 | |
|---|---|---|---|---|
| | Forecast | 95% credible interval | Forecast | 95% credible interval |
| CLE | −6.6 | (−12.5, −0.9) | −2.5 | (−8.6, 3.3) |
| MFR | 1.8 | (−3.9, 7.2) | 7.0 | (1.1, 12.6) |
| NOD | 15.5 | (9.7, 20.8) | 23.5 | (17.7, 28.9) |
| MFR − CLE | 8.3 | (7.3, 9.2) | 9.5 | (8.7, 10.4) |
| NOD − CLE | 22.0 | (19.0, 24.6) | 26.0 | (23.5, 28.4) |
| NOD − MFR | 13.7 | (11.7, 15.4) | 16.5 | (14.8, 18.1) |

307

308

309

310    Table S-4: Estimated (posterior mean) probability of having a health trout population in 2020

311    and 2050 conditioned on each of the three scenarios: current legislation (CLE), maximum

312    feasible reduction (MFR) and no anthropogenic deposition (NOD). Differences between the

313    different scenarios are also shown.

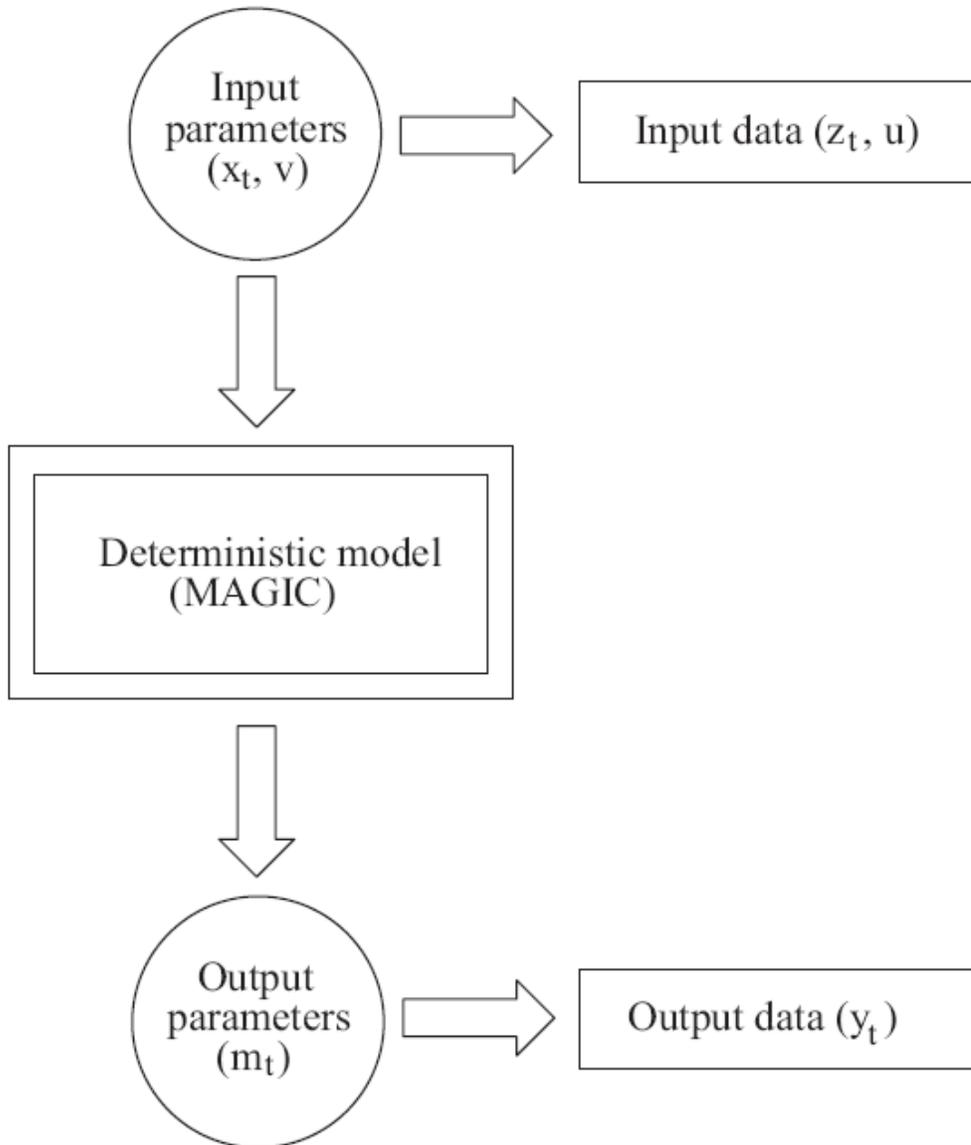| Scenario | 2020 | | 2050 | |
|---|---|---|---|---|
| | ESTIMATE | 95% credible interval | Estimate | 95% credible interval |
| CLE | 0.29 | (0.15, 0.46) | 0.40 | (0.22, 0.60) |
| MFR | 0.53 | (0.35, 0.71) | 0.69 | (0.50, 0.83) |
| NOD | 0.87 | (0.76, 0.94) | 0.95 | (0.90, 0.98) |
| MFR − CLE | 0.24 | (0.19, 0.29) | 0.29 | (0.23, 0.33) |
| NOD − CLE | 0.58 | (0.47, 0.66) | 0.55 | (0.38, 0.68) |
| NOD− MFR | 0.33 | (0.22, 0.43) | 0.26 | (0.14, 0.40) |

314

315

316

316 **Figures**

317



318

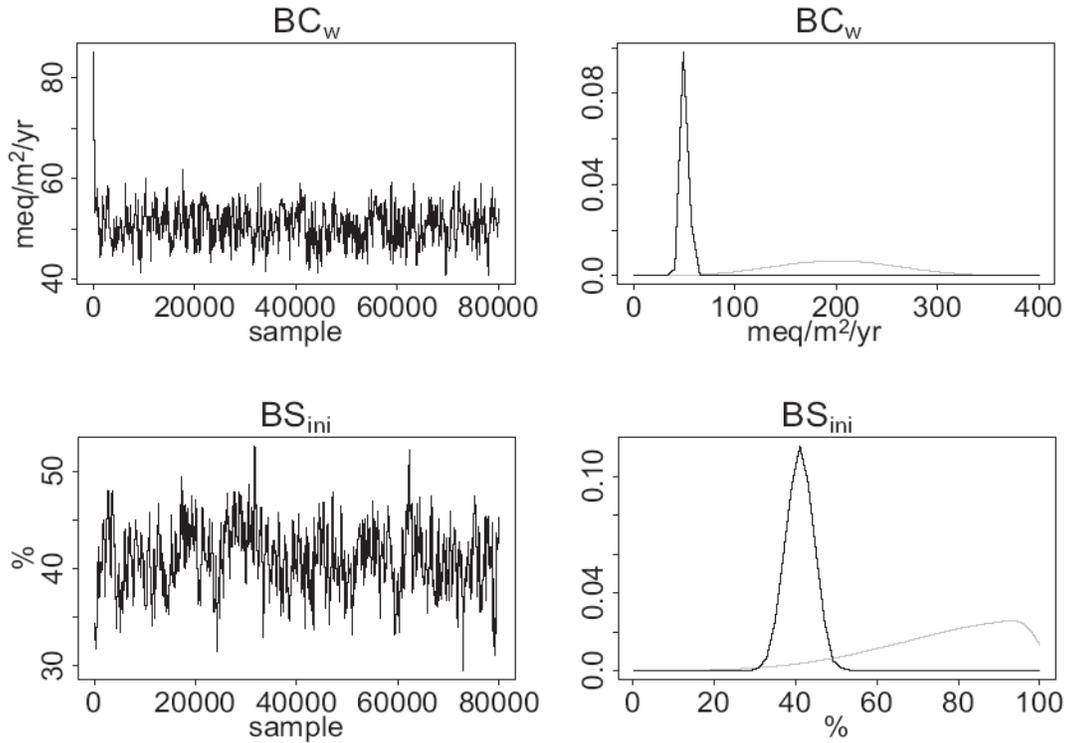319   Figure S-1. Sketch of model framework showing the interaction between input parameters,

320   input data, the deterministic model, output parameters and output data.

321

322
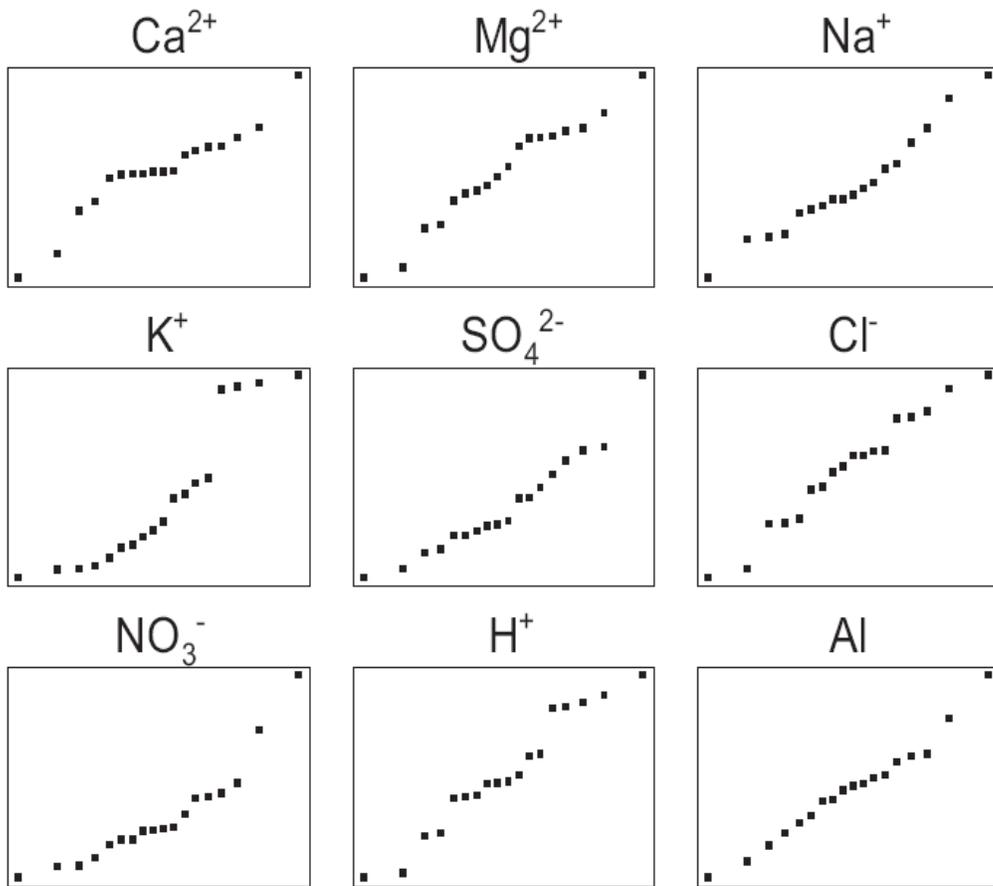
Figure S-2. Quantiles of empirical residuals versus quantiles of the standard normal distribution.

327

328    Figure S-3: Left panels: Trace plots for selected input parameters (base cation weathering,

329    $BC_w$ and initial base saturation, $BS_{ini}$) for 80 000 iterations of the algorithm. For clarity in the

330    figure, every 100th sample is shown. Right panels: Corresponding prior (grey) and posterior

331    (black) distributions. The latter is based on the last 70 000 iterations.

332
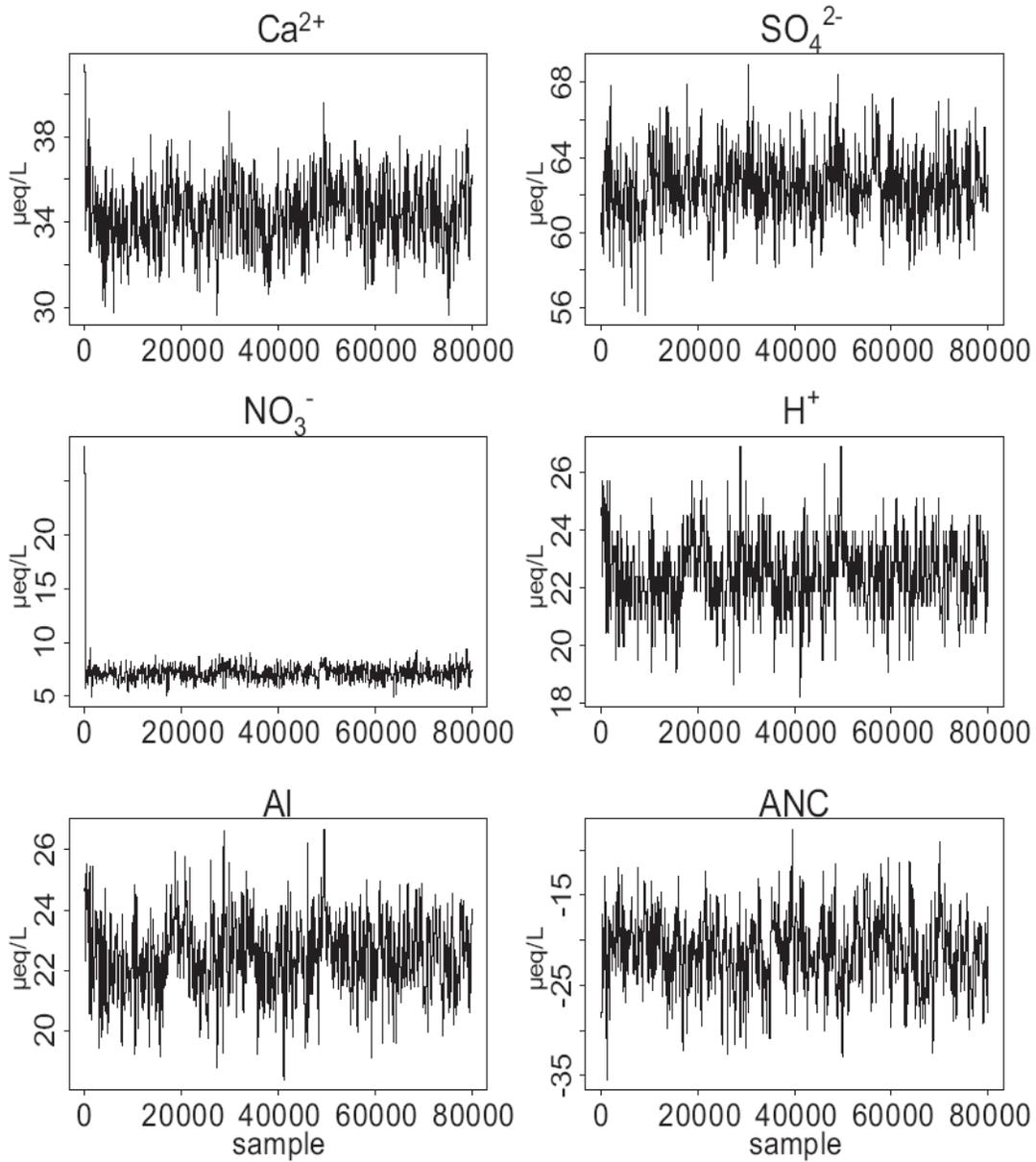
Figure S-4. Trace plots for selected output parameters for 2002. For clarity in the figure, every 100th sample is shown.