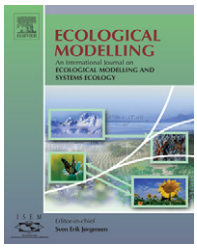
available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ecolmodel



Impact of time series data on calibration and prediction uncertainty for a deterministic hydrogeochemical model

Thorjørn Larssen[a,*], *Tore Høgåsen*[a], *B. Jack Cosby*[b]

[a] *Norwegian Institute for Water Research, Gaustadallèen 21, 0349 Oslo, Norway*
[b] *University of Virginia, Charlottesville, VA 22904-4123, USA*

ARTICLE INFO

Article history:
Published on line 20 April 2007

Keywords:
MAGIC
Acidification modelling
Recovery
Markov Chain Monte Carlo
Model calibration
Data availability
Time series data

ABSTRACT

Model calibration is fundamental in applications of deterministic process-based models. Uncertainty in model predictions depends much on the input data and observations available for model calibration. Here we explored how model predictions (forecasts) and their uncertainties vary with the length of time series data used in calibration. As an example we used the hydrogeochemical model MAGIC and data from Birkenes, a small catchment in southern Norway, to simulate future water chemistry under a scenario of reduced acid deposition. A Bayesian approach with a Markov Chain Monte Carlo (MCMC) technique was used to calibrate the model to different lengths of observed data (4–29 years) and to estimate the prediction uncertainty each calibration. The results show that the difference between modelled and observed water chemistry (calibration goodness of fit) in general decreases with increasing length of the time series used in calibration. However, there are considerable differences for different time series of the same length. The results also show that the uncertainties in predicted future acid neutralizing capacity were lowest (i.e. the distribution peak narrowest) when using the longest time series for calibration. As for calibration success, there were considerable differences between the future distributions (prediction uncertainty) for the different calibrations.

1. Introduction

Deterministic simulation models are often used to mimic hydrological, hydrogeochemical or other complex environmental systems and to make predictions under different scenarios. Such models produce certain output parameters as a function of a group of input parameters.

There are several sources of uncertainty in predictions from such models (Funtowicz and Ravetz, 1990; Saloranta et al., 2003). One important source of uncertainty is related to the quality, quantity and representativity of input data and observations used in calibration. Uncertainty of this origin is often considered (by the modellers, at least) to be the largest contributor to the overall uncertainty (Blöschl and Grayson, 2000; Uhlenbrook and Sieber, 2005). Another type of uncertainty is related to the model structure and the extent to which the model properly describes the system to be modelled. Models are by necessity simplified descriptions of the natural system, and the challenge when selecting or building a model is to include the important drivers for the problem to be investigated, but yet not over parameterize (Perrin et al., 2001; Snowling and Kramer, 2001). The uncertainties related to input data and model structure are closely related in model applications. Additional uncertainty, usually impossible to quantify, is of an epistemological nature, such as sudden dramatic changes in the system, or system processes not included in

* *Corresponding author*. Tel.: +47 22185100; fax: +47 22185200.
E-mail address: thorjorn.larssen@niva.no (T. Larssen).
0304-3800/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2007.03.016

the model (for example wildfires, landslides, volcanic eruptions) (Funtowicz and Ravetz, 1990). The focus in this paper is the prediction uncertainty related to the length of the time series data available for model calibration.

Calibration of input parameters through fitting the corresponding output parameters to observational data is fundamental in dynamic, process-oriented models used for prediction. Model predictions depend strongly on model calibration, and calibration depends on available observational data. Some recent studies discuss the relationship between uncertainty in predictions and length of calibration data time series in hydrological modelling (Anctil et al., 2004; Brath et al., 2004; Xia et al., 2004). Much less work has been done on uncertainty assessment and time series data with hydrogeochemical models, although it has been demonstrated that increasing numbers of observational data at different points in time will significantly reduce uncertainties in predictions (Larssen et al., 2004).

A principal approach to quantifying uncertainty in model predictions is to give some sort of probability distribution to the input parameters to the deterministic model and thereby obtain probability distributions rather than single values as model outputs. Several approaches have been developed for uncertainty assessment in which Bayesian techniques are combined with conceptual hydrological catchment models (Bates et al., 2003; Kuczera and Parent, 1998; Makowski et al., 2002; Thiemann et al., 2001). For biogeochemical or hydrogeochemical models as well as other types of integrated water resources modelling, such approaches are less developed, but are now becoming increasingly available (Larssen et al., 2006).

The hydrogeochemical case study presented in this paper is recovery from acidification in response to reduced acid deposition. The case is particularly suitable for illustrating the variation in uncertainty with the observation data time series because long-term (30+ year) monitoring data exist, and widely tested and well-documented models are available. Data from the Birkenes catchment in southern Norway are used. The deterministic model is the widely used hydrogeochemical model MAGIC (Model for Acidification of Groundwater in Catchments) (Cosby et al., 1985, 2001). MAGIC is mostly used for predicting future development in water chemistry as a result of changing inputs of sulphur and nitrogen from the atmosphere (acid rain). Here we present predictions of future water chemistry based on model calibrations made with different time series of model calibration data.

The objectives are to illustrate how (1) the calibration success and (2) the uncertainties in model predictions may vary with different lengths of time series data used for model calibration. In order to utilize the information in the time series observations, we combine the deterministic model (MAGIC) with a stochastic model for the observed output data and estimate all unknown parameters by Bayesian computations. We used a Markov Chain Monte Carlo (MCMC) technique for calibrating the MAGIC model to different sub-sets of the Birkenes stream chemistry data. We then compare (1) the agreement between simulated and observed data for each calibration and (2) the probability distributions of predicted future water chemistry under a future deposition scenario for each calibration. Different sub-sets of 4–5 years length and 11 years length as well as the full 29-year series were used for calibration to examine the effects of different time series of observations on calibration success and uncertainty in predictions.

2. Material and methods

2.1. *The MAGIC model and the calibration routine*

The MAGIC model is used to predict long-term effects of acid deposition on soils and surface water chemistry (Cosby et al., 1985, 2001). The model has been extensively used at a range of different sites and applications (see appendix in Cosby et al., 2001).

The predictions are largely driven by atmospheric deposition of the major ions sulphate (SO_4^{2-}), nitrate (NO_3^-), chloride (Cl^-), ammonium (NH_4^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), sodium (Na^+), potassium (K^+) and hydrogen (H^+). MAGIC calculates for each year the concentrations of ions in soil solution and surface water under the assumption of simultaneous reactions involving sulphate adsorption, cation exchange, dissolution–precipitation–speciation of aluminium, and dissolution–speciation of inorganic and organic carbon. MAGIC accounts for the mass balance of major ions in the soil by book-keeping the fluxes from atmospheric inputs, chemical weathering, net uptake in biomass, and loss to runoff. A model run is initialized in pre-industrial times (1850 is used in the current application) and driven forward in time by changing input of ions from deposition. During calibration, the model is run from the initial year to the period in time for which observations are available for surface water chemistry and soil chemistry. In traditional calibrations of MAGIC, where the input parameter uncertainty is not estimated, two sets of input parameters are adjusted to produce a match between simulated and observed soil and water variables:

- The weathering rate of the four base cations Ca^{2+}, Mg^{2+}, Na^+, and K^+, the sum of which is the total soil base cation weathering (BC_w).
- The relative amounts of the individual exchangeable base cations Ca^{2+}, Mg^{2+}, Na^+, and K^+ in the soil, the sum of which is the total initial soil base cation saturation (BS_0). The subscript refers to the year the model is initiated, i.e. 1850.

Calibration is then done by adjusting BC_w and BS_0 until output parameters match observed data for water chemistry and soil base saturation in the calibration year or years.

In the traditional calibration of the MAGIC model, the values for all input parameters, except BC_w and BS_0, are treated as known. In the Bayesian approach, however, all input parameters and output parameters are technically interpreted as unknown, ideal or effective values which are representative of the whole catchment and a whole year. They are to some degree measurable, but cannot be specified exactly in practice, since the corresponding measurements (input data and output data) typically will be point measurements in space and often also in time.

The input parameters required for calibration of MAGIC comprise 28 parameters assumed independent of time, including lake and catchment characteristics and soil chemical and physical characteristics (see Table 1 for a detailed