Impacts of acid deposition at Plastic Lake: forecasting chemical recovery using a Bayesian calibration and uncertainty propagation approach

George MacDougall, Julian Aherne and Shaun Watmough

ABSTRACT

Given the importance of models in the development of environmental polices it is necessary to assess the uncertainty introduced by model parameterisation and its impact on predictions. In the current study, an uncertainty framework designed to perform automated calibrations and developed for use with the Model of Acidification of Groundwater in Catchments (MAGIC) was applied to Plastic Lake, a long-term study site in Southern Ontario, Canada. The primary objectives were to investigate the chemical response of soil and surface water at Plastic Lake to proposed acid (sulfur and nitrogen) emissions and assess the use of the framework at a regional level. Despite the relatively high amount of uncertainty associated with many of the model parameters, calibration resulted in relatively narrow parameter convergence. The importance of time-series stream data was clearly evident, with uncertainty decreasing with more observation years. The forecast improvements in stream Acid Neutralizing Capacity at Plastic Lake from −40 μeq/L in 1988 to 14 μeq/L in 2060 had 5 and 95% confidence bounds of −3 and 29 μeq/L, respectively. Despite the limited availability of soil chemical data in Ontario, the approach applied at Plastic Lake is viable on a regional basis given the abundance of water chemistry data.

Key words | acidification recovery, Monte Carlo Markov chain, uncertainty assessment

George MacDougall (corresponding author)
Julian Aherne
Shaun Watmough
Environmental and Life Sciences,
Trent University,
1600 West Bank Drive,
Peterborough,
ON K9J 7B8,
Canada
Tel.: +1 705 748 1011
Fax: +1 705 748 1569
E-mail: *gmacdoug@trentu.ca*

INTRODUCTION

The Model of Acidification of Groundwater in Catchments (MAGIC) has been used since 1985 to predict the long-term effects of acid deposition on soil and surface water chemistry at a range of sites around the globe (see the appendix in Cosby *et al.* (2001) and Wright *et al.* (2005)). Overall, the model has proven to be robust, reliable and reproducible. Nonetheless, similar to other deterministic process-oriented models, MAGIC is subject to uncertainty from several sources (Saloranta *et al.* 2003; van der Sluijs *et al.* 2005) but the availability (quantity and quality) of site-specific data is usually what determines the success (reproducibility) of an individual application. In this study uncertainties in model parameters (physico-chemical characteristics such as soil data) were assessed. MAGIC is a lumped parameter model; site-specific parameters represent "average" values for the various soil and water model compartments, regardless of the heterogeneity of the actual catchment, i.e. soil depth observations collected at numerous locations at study sites are averaged to a single effective value for each soil compartment. At intensively studied sites, the variability of model parameters such as soil chemistry becomes evident as the number of soil pits (observations) increases. Accordingly, regional studies are often limited by the availability of multiple soil pits; a framework that utilises a distribution based on field data offers advantages over model configurations that use a single (or a small number of) observations to estimate average values for an entire catchment (or region). Given the importance of models in the development of environmental polices it is necessary to present model forecasts along with an estimate of uncertainties.

doi: 10.2166/nh.2009.084

The uncertainty involved in parameterising and calibrating dynamic models has been studied since their development (Hornberger *et al.* 1986). More recently techniques for assessing this uncertainty have become more complex and versatile (Gallagher & Doherty 2007). Page *et al.* (2003) indicated that, when the number of model parameters is high and the number of observations low, it is possible to calibrate a model equally well with more than one set of parameters. Automated calibration techniques have been developed to replace *ad hoc* techniques and to address the problem of assessing and tracking uncertainties in model parameters (Kennedy & O'Hagan 2001; Larssen *et al.* 2006).

A statistical framework for model calibration and uncertainty estimation developed for use with MAGIC (Larssen *et al.* 2006) was used in this study. The uncertainty framework uses Markov chain Monte Carlo (MCMC) input sampling and likelihood functions determined from available observations of water and soil chemistry. The framework was developed by Larssen *et al.* (2006) and applied to Birkenes, Norway. In the current study the framework was applied to a sub-catchment at Plastic Lake (PC1-08) in Southern Ontario. Acid (sulfur (S)) deposition has declined by about 50% since the early 1980s in parts of Ontario (including Plastic Lake) but the chemical recovery of lakes has been limited (Jeffries *et al.* 2003). Under current legislation emissions will decline further; the impact of these reductions on water chemistry will influence future emission reduction policies. The primary objective of this study was to apply the framework to predict future soil and surface water chemistry at Plastic Lake while assessing the applicability of the framework at a regional level. In concert, the impact of calibrating the model to truncated (observed) stream chemistry data was investigated to evaluate the influence on model predictions when the framework is used at less intensively studied sites.

METHODS

Study site

Plastic Lake (32.1 ha headwater lake) is located on the Precambrian Shield in Haliburton County, south-central Ontario (45°11′N, 78°50′W). Plastic Lake's catchment (95.5 ha) is drained by seven small streams, six of which are ephemeral. The largest of the seven sub-catchments (PC1, 23.3 ha) drains a wetland (sphagnum – conifer swamp, 2.2 ha) and flows year round except in very dry years. The upland soils of PC1 (and its gauged subcatchment PC1-08, 3.5 ha) are sandy, shallow, acidic humo-ferric or ferro-humic podzols (Lozano 1987). The PC1 upland forest cover is primarily coniferous, dominated (80% of basal area in 1999: Watmough & Dillon 2001) by white pine (*Pinus strobus* L.), red maple (*Acer rubrum* L.), eastern hemlock (*Tsuga canadensis* (L.) Carrière) and red oak (*Quercus rubra* L.). The average age of the stand is ~87 yr, with a maximum of ~200 yr (baseline 2000). The catchment altitude ranges from 380 – 420 m a.s.l., annual rainfall averages ~1,000 mm and the mean January and July air temperatures are – 9.4°C and 18.6°C, respectively.

The chemistry and volume of bulk precipitation have been measured since 1976 at a clearing adjacent to PC1 (Dillon *et al.* 1988). Precipitation samples were removed from collectors when there was sufficient volume for chemical analyses, typically weekly. On the rare occasions when data were missing the average precipitation from three nearby (within 50 km) bulk collectors were used. Between 1986 and 1995 stream discharge and chemistry were monitored at PC1-08. A V-notched weir was installed at the catchment outflow and water samples for chemical analyses were collected regularly (at least biweekly) when there was flow. The field, hydrologic and analytical methods are described in detail elsewhere (OME 1983; Scheider *et al.* 1983; Locke & Scott 1986; methods have remained unchanged). During 1983, soil samples were collected from 15 pits located in the upland part of the catchment (Lozano 1987); sampling was repeated in 1999 using adjacent soil pits (Watmough & Dillon 2004). Bulk density and exchangeable cations were determined for both surveys using standard methods (OME 1983). The framework was applied to the PC1-08 sub-catchment to avoid complexities associated with the response of the wetland to drought that has been observed at PC1 (LaZerte 1993).

THE MODEL OF ACIDIFICATION OF GROUNDWATER IN CATCHMENTS

MAGIC is a lumped-parameter model of intermediate complexity, developed to predict the long-term effects of

acidic deposition on soils and surface water chemistry. The model was first described by Cosby *et al.* (1985) and developments are reviewed by Cosby *et al.* (2001). In brief, the model predicts the monthly and annual average concentrations of the major ions for soil solution and surface water chemistry. MAGIC represents the catchment with aggregated, uniform soil compartments (one to three) and a surface water compartment that can be either a lake or a stream. Time-series (forcing) inputs to the model include annual or monthly estimates of: deposition of ions from the atmosphere (wet plus dry deposition); discharge volumes and flow routing within the catchment; biological production, removal and transformation of ions; internal sources and sinks of ions from weathering or precipitation reactions; and climate data. Constant parameters in the model (i.e. model parameters) include physical and chemical characteristics of the soils and surface waters, and thermodynamic constants. The model is calibrated using observed values of surface water and soil chemistry for a specified period. In the current study the PC1-08 subcatchment was simulated using one soil and one stream compartment on an annual time step.

The uncertainty framework

Larssen *et al.* (2006) presented an approach for automatic calibration and uncertainty propagation for forecasting acidification effects at a long-term study site at Birkenes, Norway. The approach combines data from observations, the deterministic model and prior distributions of model parameters to obtain forecast distributions. The implementation requires prior probability distributions to be specified for the model parameters and likelihood functions to be specified for the output data (observed values of surface water (annual time series) and soil chemistry (single observations of exchangeable cations)). Posterior distributions are generated by running the model (MAGIC) repeatedly using a MCMC scheme to sample model parameters from prior distributions. The likelihood functions are determined by available observations of the model output data for a calibration period. The calibration is Bayesian in that unknown parameters are estimated using prior knowledge of model parameters and output data. The approach provides an assessment of uncertainties in the model parameters and allows the impact of the uncertainty on model outputs to be explicitly shown. Other sources of uncertainty, such as model structure and epistemological concerns (ignorance), certainly exist but are not addressed in this study. Uncertainties in forcing data (deposition chemistry and discharge volumes) were incorporated by stochastic sampling from narrow distributions around the time-series observations (representing measurement error).

Parameter sampling

Successful configuration of the framework requires a balance between specifying suitable constraints on the model parameters so that it converges within a practical number of iterations, and allowing the framework sufficient freedom so that the algorithm can effectively seek suitable values throughout the entire available parameter space. Only a subset of the model parameters were subject to uncertainty assessment; the remaining parameters, which have minimal influence on model outputs, were specified as single values. The parameters specified with prior distributions include physico-chemical characteristics such as soil depth and the aluminum dissolution constant. Parameters that were excluded from the uncertainty assessment included those related to simplified nitrogen dynamics, e.g., soil and lake nitrification and soil ammonium uptake (Table 1). The prior distributions were specified as uniform with a minimum to maximum range, e.g. soil depth was specified as a uniform distribution between 0.3 and 0.5 m (Table 1).

While the framework supports input distributions other than uniform, the MCMC sampling method uses its own likelihood function based on observations to choose optimum values. Interpreting the framework results is less complicated with uniform prior distributions as selection of the inputs will depend solely on the weightings assigned by the likelihood function used in the MCMC scheme. High confidence in a model parameter can be implemented by specifying a narrower range.

Several inputs to MAGIC are specified as time series, e.g. deposition chemistry and discharge volumes. Uncertainty was included in these observed values by specifying them as normally distributed with fixed standard deviations of 10%. There is uncertainty associated with these observations but

Table 1 | MAGIC input parameters with prior distributions

Constant parameters	Distributions (min, max)
Soil parameters	
Depth (m)	Uniform (0.3,0.5)
Porosity (%)	Uniform (20,40)
Bulk density (kg/m^3)	Uniform (500,1200)
Cation exchange capacity (meq/kg)	Uniform (65,95)
Sulfate adsorption half saturation (μeq/L)	Uniform (0,500)
Sulfate adsorption maximum (meq/kg)	Uniform (0,20)
Aluminum dissolution constant ($\log_{10}$)	Uniform (6,11)
Temperature (°C)	Fixed (6.7)
Partial CO_2 pressure (% atm)	Uniform (0.2,2)
Dissolved organic carbon (μmol/L)	Uniform (0,100)
Nitrification (%)	Fixed (100)
Ammonium immobilisation (%)	Fixed (100)
Nitrate immobilisation (%)	Fixed (99.8)
Ca weathering (meq/m^2/yr)	Uniform (0,50)
Mg weathering (meq/m^2/yr)	Uniform (0,30)
Na weathering (meq/m^2/yr)	Uniform (0,20)
K weathering (meq/m^2/yr)	Uniform (0,5)
Sulfate weathering (meq/m^2/yr)	Uniform (0,20)
Initial exchangeable Ca (%)	Uniform (0.1,50)
Initial exchangeable Mg (%)	Uniform (0.1,50)
Initial exchangeable Na (%)	Uniform (0.1,50)
Initial exchangeable K (%)	Uniform (0.1,50)
Surface water parameters	
Relative area water:catchment (ratio)	Fixed (0.01)
Water retention time (yr)	Fixed (0)
Aluminum dissolution constant ($\log_{10}$)	Uniform (6,11)
Temperature (°C)	Uniform (3,10)
Partial CO_2 pressure (% atm)	Uniform (0.03,0.2)
Dissolved organic carbon (μmol/L)	Uniform (15,85)
Nitrification (%)	Fixed (100)

it was assumed that the observations were unbiased, and the narrow distributions represent measurement error. This was chosen as the simplest approach (supported by the framework) that still included some estimation of uncertainty.

The uncertainty framework uses a Metropolis–Hastings approach which has advantages over other approaches such as the generalised likelihood uncertainty estimation (GLUE) method (Beven & Binley 1992) when used in similar MCMC applications (Kuczera & Parent 1998). The likelihood ratios are calculated by evaluating each candidate set of parameter values against the current (most recently accepted) set of parameter values. The ratio is driven by matching model output values with observations using a least-sum-of-squares-based function such that the ratio is greater than 1 if the candidate set of parameters produces a better match to observations than the current set. Parameter sets that generate more likely outputs are always accepted. Parameter sets that generate less likely outputs are either rejected if the ratio is below a specific threshold or, if above the threshold, they are accepted with a weighting equal to the ratio (always less than 1). The process repeats until the specified number of accepted parameter sets are generated.

Automated calibration

Calibrating MAGIC typically involves some manual processes that optimise a few parameters (weathering and initial base saturation levels) against a single year's observations (surface water chemistry and current base saturation levels). The Bayesian calibration process optimises many parameters against many years of (water chemistry) observations simultaneously. By specifying each parameter distribution carefully the model calibration can be reasonably constrained based on site characteristics, available data and experience.

In the current study, soil physical and chemical characteristics, such as soil depth, density, porosity, cation exchange capacity (CEC), sulfate adsorption parameters and aluminum dissolution constant were specified as narrow distributions based on site knowledge and observations. While these parameters are included in the calibration process, the narrow distributions are a reflection of the uncertainty being assessed. Parameters such as base cation weathering rates, which are notoriously difficult to measure, and initial base saturation values were specified with the widest reasonable distributions (Table 1). These parameters are key to calibrating MAGIC, and by not providing the framework with best estimates of these values or even narrow distributions it was ensured that only the algorithms within the MCMC framework complete the calibration. Additionally the framework is supplied with site

observations to which the model outputs are calibrated. The model was calibrated to a single soil (base saturation) observation year (set to 1997) and 11 years of water chemistry (1984–1994) under 18 years of forcing data: observed deposition of major ions and discharge volumes (1980–1997). For a broader discussion on Bayesian calibration of deterministic models see Kennedy & O'Hagan (2001). A total of 50,000 calibrated parameter sets were generated for PC1-08 and subsequently used to simulate historic and future soil and water chemistry. MAGIC was run with each parameter set using a single deposition scenario (1850–2100) which was constructed from historical emission inventories, recent observed data (1980–1997) and current legislated emissions under the Canada–US Air Quality Agreement (Aherne *et al.* 2003).

MCMC framework and MAGIC simplifications

The framework was also used to address a number of questions towards regionalisation, i.e. application to many more sites, but with less site-specific data. While a number of possible variations can be explored, in this study it was asked: What is the impact of varying the model parameters' initial starting points, step size (also known as proposal or jump distribution) and number of iterations on the behaviour of the model forecasts? How do predictions change when less water chemistry data are available (does the framework remain viable)? The starting point for each parameter was the initial value chosen within the input distribution. Choosing the mid-point of uniform distributions is reasonable and facilitates automation. Step size is the range, within the specified input distribution, over which the Monte Carlo random sampling algorithm will choose the next candidate value. This parameter determines how quickly the MCMC framework moves through a distribution. It specifies the per iteration range, centered on the current value, over which the next candidate value will be randomly sampled. The step size is known to impact Metropolis–Hastings algorithms (Gelman *et al.* 1995). The number of iterations refers to the number of model runs (or evaluations) required for convergence (i.e. iterations required for model parameters to reach a stationary distribution). Iterations (or parameter sets) prior to convergence (the burn-in phase) are typically discarded.

Few study sites in Ontario have decades of continuous water chemistry data but there are an increasing number of sites with multi-year survey data. Larssen *et al.* (2004) presented some techniques for calibrating dynamic models effectively using variable time series of data. Generating results calibrated with two years of data separated by ten years in time, for example, yields superior long-term forecast results over using two adjacent years. Under this guidance the number of years of observations was reduced: from 11 years of data 2- and 5-year subsets separated as widely as possible were selected, e.g. the 5-year subset used data from 1984, 1986, 1989, 1992 and 1994. The impact of reduced data on surface water ANC in the year 2060 was assessed.

Table 2 | Input parameters derived from manual and Bayesian calibration of MAGIC

Parameter	Manually calibrated*	Automatically calibrated (percentiles)		
		5	50	95
Ca weathering (meq/m^2/yr)	19.5	12.9	21.0	27.5
Mg weathering (meq/m^2/yr)	13.5	11.2	14.1	15.6
K weathering (meq/m^2/yr)	1.0	0.0	0.1	0.5
Na weathering (meq/m^2/yr)	10.5	7.0	9.4	11.4
Initial exchangeable Ca (%)	17.6	14.2	22.5	31.0
Initial exchangeable Mg (%)	4.7	3.0	6.2	9.4
Initial exchangeable K (%)	2.3	0.4	1.5	2.4
Initial exchangeable Na (%)	2.0	2.4	3.6	4.7
Aluminum dissolution constant ($\log_{10}$)	9.2	8.7	9.6	10.2
Dissolved organic carbon (μmol/L)	55.0	25.0	69.0	94.0

*Data taken from Aherne *et al.* (2006) under similar but not identical MAGIC configurations.

RESULTS AND DISCUSSION

The uncertainty framework at Plastic Lake

The application of the framework on PC1-08 was consistent with earlier applications of MAGIC (see baseline results in Aherne *et al.* (2007) and calibration details in Aherne *et al.* (2004, 2006)) (Table 2). While not a surprising result given that data and model were similar to previous applications, it does validate the MCMC framework's objective calibration abilities, and further provides a basis for uncertainty assessment at Plastic Lake.

The Bayesian calibration resulted in close correspondence between observed and simulated (mean) stream chemistry (e.g. normalised absolute mean error for stream sulfate (SO_4^{2-}) was 0.01; zero indicates perfect correspondence) during the calibration period (Figure 1). The 5 and 95% percentile values for simulated chemistry during the calibration period (1984–1994) are narrow, e.g. 1985 stream SO_4^{2-} 5 and 95 percentile values were 170 and 177 μeq/L, respectively (Figure 1). In contrast, the simulated 5 and 95 percentile values for stream SO_4^{2-} in 1975 were approximately 200 and 250 μeq/L; Larssen *et al.* (2007) showed similar behaviour for Birkenes. In the current study, deposition data (1980–1997) extended beyond the calibration period for surface water observations (Figure 1). The variability in the deposition data contributed to variability outside the calibration period but the pattern of increasing variability in simulated values further outside the calibration period is a common characteristic of MCMC applications. It reflects that the MCMC algorithm samples randomly from a very large and complex parameter space and the only constraint on the thousands of candidate parameter sets is they match the observed data during the calibration period.

The MCMC framework calibration process converged on reasonable values for all the key parameters despite the fact that some were specified as unconstrained distributions,

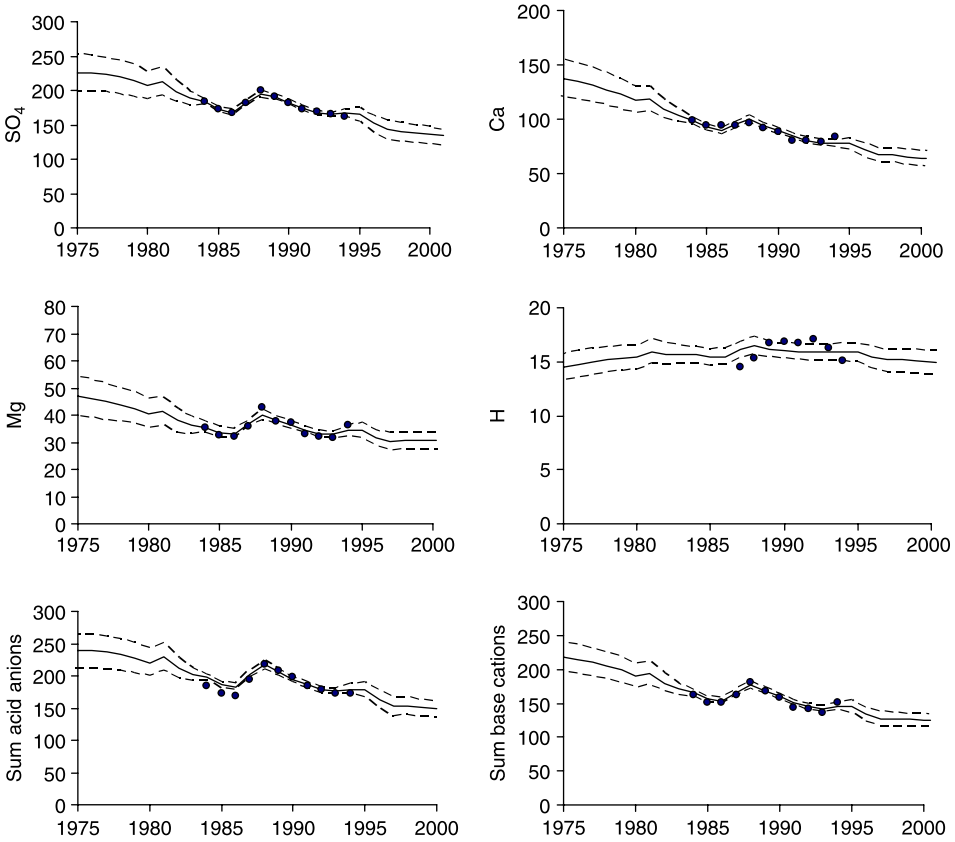

Figure 1 | Simulated and observed stream concentrations (μeq/L) for sulfate, calcium, magnesium, hydrogen, sum of acid anions and sum of base cations for the period 1975–2000. Thin lines are the simulated values and the filled circles are observed values during the calibration period (1984–1994). The dashed lines are 5 and 95 percentiles. The results are from 50,000 framework iterations.

e.g. the prior distributions for weathering rates and initial exchangeable base cations. All four base cation weathering rates converged to relatively narrow ranges; potassium appears to be constrained by the lower limit of the specified distribution but MAGIC reasonably only accepts positive values for weathering rates (see Figure 2). The annual weathering rate for calcium was specified as a uniform distribution between 0 and 50 meq/m^2 and the MCMC framework converged to values near 21 meq/m^2 (Figure 2). Despite some differences in the calibration data and procedures, this mean value compares well with previous site assessments, i.e. Aherne *et al.* (2004) estimated calcium weathering at 19.5 meq/m^2/year using MAGIC (Table 2). Watmough & Dillon (2004) used several techniques including the PROFILE model (Sverdrup & Warfvinge 1993) to estimate calcium weathering at Plastic Lake and their results ranged from approximately 10–13 meq/m^2/year. These low calcium weathering rates are consistent with sites that have shallow soils on the Canadian Shield (Ouimet & Duchesne 2005; Watmough *et al.* 2005). The four initial base cation saturation parameters were all specified with identical unrestrained input distributions between 0–50% of CEC. Each converged to separate and relatively narrow ranges which are comparable with Aherne *et al.* (2004); except for the low value for sodium, the manually calibrated values fell within the MCMC framework 5 and 95 percentiles (Table 2). The mean initial exchangeable magnesium was calibrated to be 6.2% of CEC with 5 and 95 percentile values

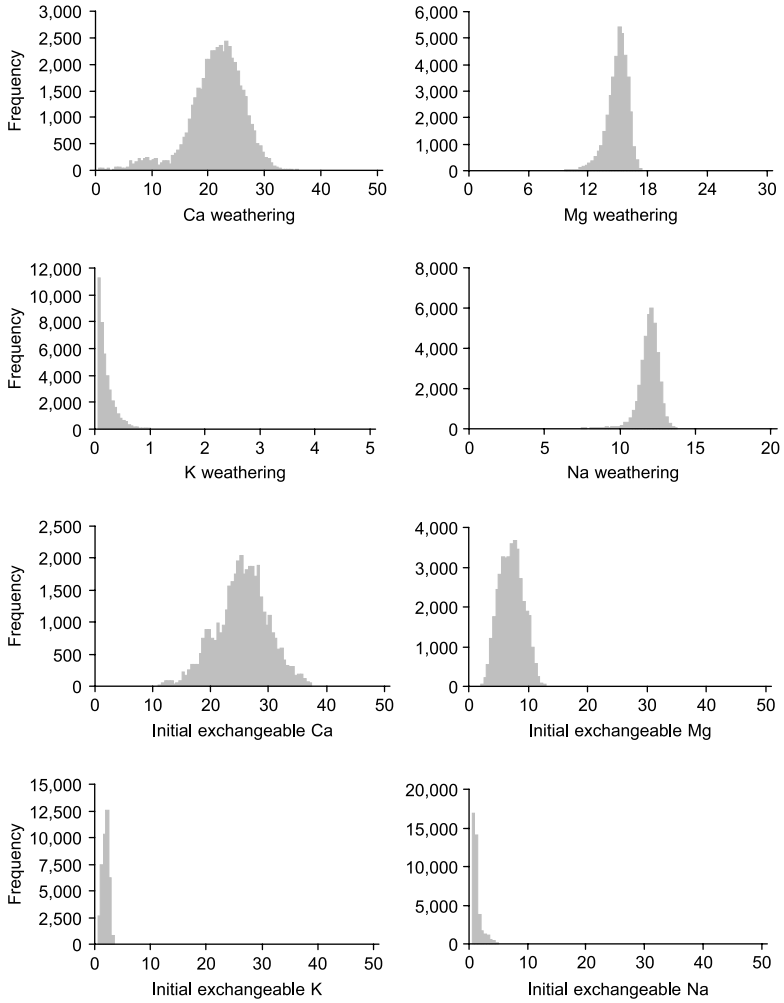

Figure 2 | Input parameter distributions for base cation weathering (meq/m^2/yr) and initial base cation fraction (%). The *x* axis of each chart is the minimum to maximum range specified (prior distributions: see Table 1). The results are from 50,000 framework iterations.

of 3 and 9.4% of CEC. This parameter was calibrated by Aherne *et al.* (2004) to 4.7% of CEC.

Monte Carlo techniques can usually converge to optimum values (assuming such values exist), if they are given adequate configuration parameters and if they are run for enough iterations. Starting with identical seeds and configuration values, there was improved visual evidence of convergence as the number of iterations was increased from 5,000 to 10,000 to 25,000 to 50,000 (Figure 3). Studies into objective convergence assessment techniques for MCMC applications have found some algorithms that appear promising but they tend to be non-universal, dependent on the characteristics of the sample space, and computationally intensive (Rosenthal 1994; Brooks & Roberts 1998). When relying on subjective visual inspection of parameter trace plots the improvements from 25,000 to 50,000 are arguably cosmetic if one is solely interested in getting an estimate of the parameter means. It is important to note that, if too few iterations are specified, parameter sample distributions have a greater variability and may have shifted means. For example, calcium weathering did not stabilise until 25,000 iterations (slope of trend line (not shown) for 25,000 to 50,000 iterations is zero; Figure 3). The trace plots often have common characteristics that become important when making assessments on convergence. Initially, MCMC algorithms wander widely over the parameter space. The values selected during this "burn-in" period are often discarded because they can potentially distort the means of the parameters. As an MCMC algorithm converges the trace typically continues to fluctuate near the optimum parameter space which can make assessment of convergence difficult. This problem is aggravated when there are correlated parameters.

There are model parameters that can be further constrained based on knowledge of the process descriptions included in MAGIC. Several of the parameters in MAGIC are aggregated, e.g. soil depth, soil bulk density and CEC. They are used in combination with exchangeable (base cation) fractions to set the size of the individual base cation exchange pools. The size of the cation exchange pool is not specified directly; rather, it is specified as separate measurable parameters. While it is clear that no optimum values for soil depth, bulk density or CEC were preferred by the framework, there are values for the total cation exchangeable pool that are preferred by the calibration process (Figure 4). It is

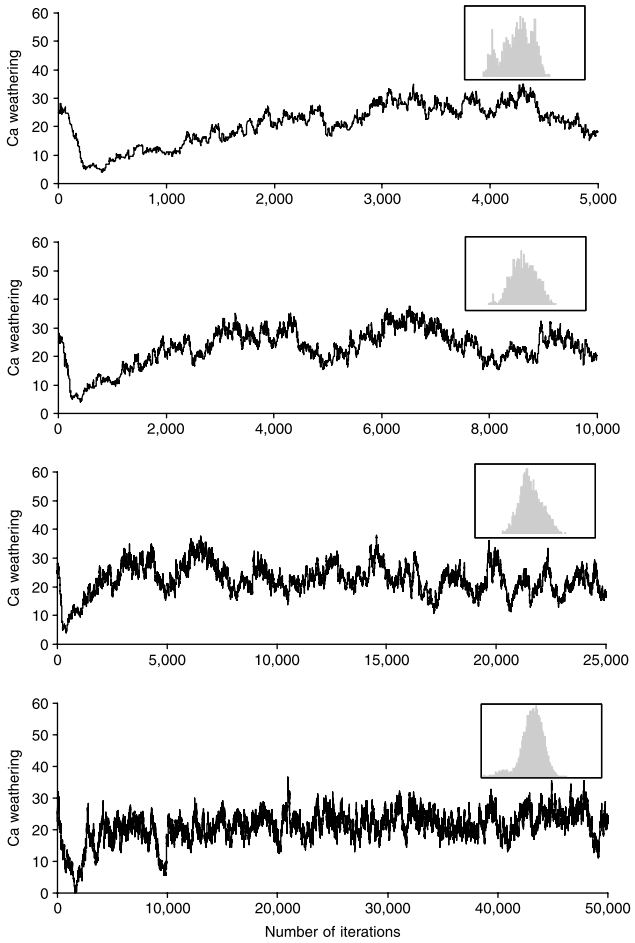

Figure 3 | Trace plots and probability distributions (insets) for calcium weathering rate (meq/m^2/yr) for 5,000, 10,000, 25,000 and 50,000 iterations.

possible to have the same size pool from an infinite number of combinations of these three correlated input parameters: as such, it is reasonable to fix two and allow the framework to vary only one. Assessment of uncertainty is still valid as it pertains to the total pool size but it will stem from the single parameter that is specified as a distribution.

In this study, MAGIC was insensitive to several input parameters. Traces (not shown) for stream temperature and stream carbon dioxide partial pressure did not contribute to the calibration process (they had very limited influence on model outputs) and accordingly the framework did not find any preferred values inside the distributions provided. In contrast, the stream aluminum disassociation coefficient converged quickly to an extremely narrow range, suggesting that there was a well-defined (preferred) value. Clearly optimum parameter values could be specified prior to the

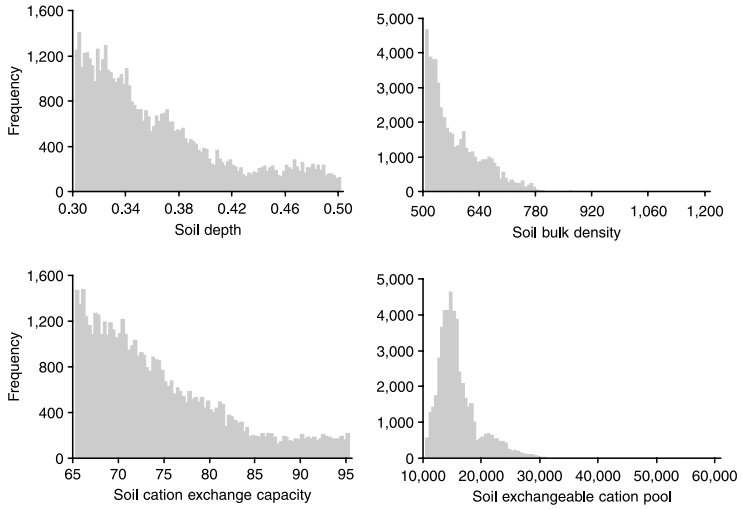

Figure 4 | Soil exchangeable cation pool (meq/m^2); estimated as soil depth × bulk density × cation exchange capacity (CEC). The *x* axis of each input parameter chart (soil depth, bulk density and CEC) is the minimum to maximum range specified (prior distributions: see Table 1). The framework converges (50,000 iterations) on the product of the correlated input parameters.

calibration process. A parameter is not generally a good candidate for inclusion in the automated calibration and uncertainty assessment process if the MCMC framework cannot converge or if it converges quickly to a single value. Either it has little or no impact for the current site configuration (varying its value does not impact model outputs) or it has no valid assessed uncertainty (one value is viable). In both cases, including such parameters would complicate the interpretation of the model results (and increase the computational burden).

Forecasts at Plastic Lake

Surface water concentrations of ANC are frequently used as an indicator of recovery from acid deposition. Numerous dynamic model applications in Europe and North America have predicted improvements in surface water ANC values in response to reductions in acid deposition levels (Larssen 2005; Posch *et al.* 2008). Regardless of the scenarios or processes being studied, e.g. changes in legislation or climate change, ANC forecasts are generally based on a few simulations per site (Aherne *et al.* 2003; Larssen *et al.* 2003) and do not include an estimate of uncertainty. Using the MCMC framework these results can be presented as distributions. The 5, 50 and 95 percentile forecasted ANC values at Plastic Lake when using 10,000 iterations and 11 years of water chemistry observations were –1.1, 11.9 and 29.2 μeq/L, respectively. In a risk assessment scenario, a realistic ANC target value of 40 μeq/L would be outside the 90% confidence bounds for 2060 under current legislation. This forecast agrees with previous studies at Plastic Lake (Aherne *et al.* 2007) which used similar model configurations and forecasted that ANC levels in Plastic Lake in 2060 will be 27.8 μeq/L, which is inside the 5 and 95 percentile values.

MCMC framework configuration parameters

The starting point had no perceptible impact on the performance of the framework or final results (when used with appropriate step size). This should be the case if the MCMC algorithm successfully searches the entire available parameter space. Starting values at the 5 and 95% values of input distributions were used (data not shown). Reinds *et al.* (2008) used MCMC sampling with random starting points and also observed little or no impact on the results for the Very Simple Dynamic model. In the current study, the initial value of many parameters was set to the mid-point of the input distribution (in practical terms, the distribution was set to bracket average field observations). In other cases, such as the soil aluminum dissolution constant, the distribution was chosen as a reasonable range based on field measurements in similar catchments.

In contrast to the starting point, the MCMC framework was sensitive to step size (proposal distribution). When a

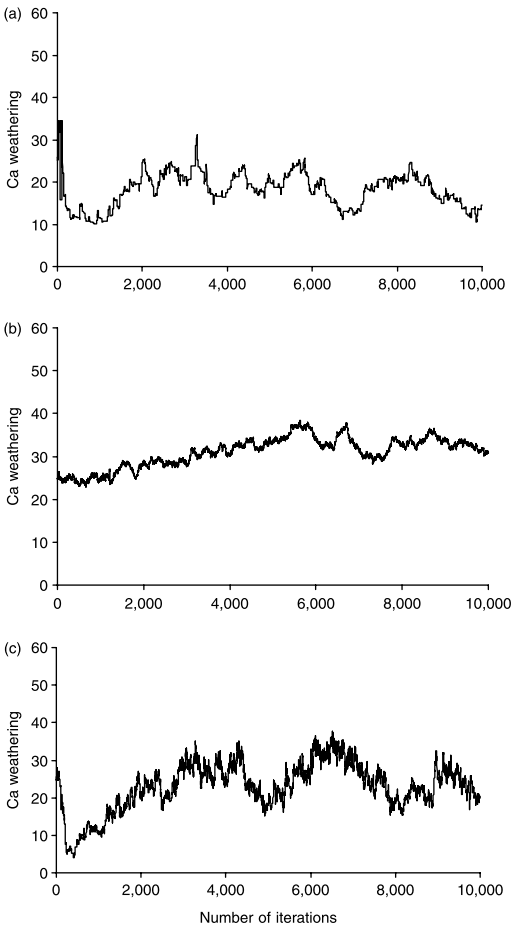

Figure 5 | Trace plots for calcium weathering rates depicting the influence of the framework parameter step size. In (a) the configured step size was 25–100% of the minimum to maximum range specified for each input parameter. In (b) the configured step size was 1–2% of the minimum to maximum range. In (c) the configured step size was 7–15% of the minimum to maximum range. The first 10,000 iterations are shown.

very small step size (1–2% of input distribution) or a very large step size (25–100% of input distribution) was specified many more iterations were required to find optimum parameter values (Figure 5). If enough iterations were run with extreme (small or large) step sizes the entire parameter space could be explored at the expense of framework performance. In contrast, a step size between 7 and 15% of the parameter distribution range required only 10,000 iterations to find optimum parameter values. When small step sizes are used, the acceptance rate of the random samples is high but, because each step is small, exploring the parameter space takes more iterations. When very large step sizes were used the acceptance rate of the random samples is lower, resulting in more candidate samples being rejected. The acceptance rates for calcium weathering in this study were 0.88, 0.06 and 0.30 for the small, large and selected (medium) step sizes, respectively (Figure 5). The importance of step size on the performance of the Metropolis–Hastings algorithms is well known (Gelman *et al.* 1995) and has led to the development of procedures to automatically select the optimal step size (Vrugt *et al.* 2005; 2008). In the current study, knowledge of the processes in MAGIC coupled with well-defined prior distributions ensured that a fixed step size (based on the parameter range) produced reasonable acceptance rates.

Data availability and regional considerations

One of the significant challenges to a regional application of the uncertainty framework is the amount (number of years) of observed stream chemistry data available at the study sites. When the number of observation years was decreased the uncertainty in the model output values increased (Figure 6). Likelihood functions determined against available observations of model outputs during the calibration period drive the selection of behaviorable input parameters; fewer years of observed data result in wider *a priori* distributions for input parameters (not shown) and uncertainty in model outputs. In general, the change in variability was more observable than a change in mean values, although on decreasing from 5 to two years of data the mean forecasts for ANC in 2060 moved from 15.9 to 3.5 μeq/L. Larssen *et al.* (2007) used variable subsets of data and saw variations in means and variability for model

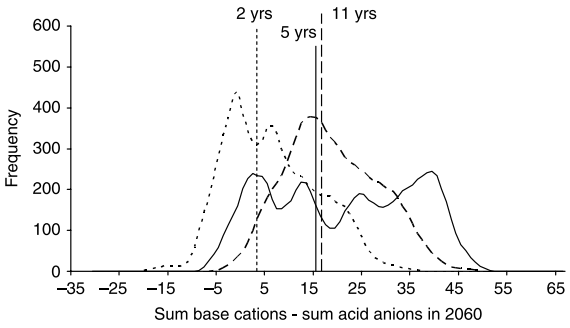

Figure 6 | Model forecasts for probability density for sum of base cations – sum of acid anions (μeq/L) in 2060 depicting the impact of truncated observed data on the MCMC framework (short dashed line: 2 years; continuous line: 5 years and long dashed line: 11 years). The vertical lines approximate the mean values for each data subset (11 yr ~ 16 μeq/L, 5 yr ~ 15.5 μeq/L, 2 yr ~ 3.4 μeq/L). The results are from 10,000 framework iterations.

outputs. The same first and last years were used, regardless of the subset, so the data spanned the same time period and therefore no comparison was made with years of decreasing acid deposition and years of increasing or stable acid deposition. Similar to Larssen *et al.* (2007), convergence improved when the widest possible temporal range was used, e.g. all data subsets included years 1984 and 1994 (for the period 1984–1994). This approach provided better results than selecting subsets of data from consecutive years (results not shown). As with all inverse theory methods, model parameters are conditioned upon observations of model outputs: as such, longer time series of observed surface water chemistry will result in greater certainty in input parameter space. Given that routine surface water sampling in Ontario has been common since the early 1980s the availability of observed data should not limit the regional application of such a framework.

A second challenge to a regional application of the uncertainty framework concerns the availability of site-relevant deposition data. The framework is dependent on a continuous time series of deposition data (missing values are interpolated). In this study deposition data were collected near the site. In a regional application this will not be the case and some adaptations will be required to accommodate deposition data from gridded maps. In contrast, the uncertainty framework uses a single year of soil chemistry (base saturation) data, not a time series. There may be concerns about the quantity of soil chemistry data due to the heterogeneity of soils in catchments and scarcity of soil pits but because the requirements of the framework are modest, it should not impact the viability of using the framework at a regional level.

CONCLUSION

It is paramount that model predictions in support of policy development include an assessment of uncertainty owing to model parameterisation. Moreover, probability distributions rather than single model forecasts will allow scientists and policy makers to use models in a risk analysis context where they have experience combining probabilities with consequences. In the current study an MCMC framework was used to investigate uncertainty in forecasts from MAGIC. The framework provided automatic calibration to observed data time series and updating of input parameter distributions (quantification of behavioral parameter space). Demonstrating that the framework's automated calibration process can produce reasonable parameter estimates strengthens the credibility of the model and therefore its forecasts. Despite the increased complexity of such frameworks, given knowledge of the model (MAGIC) and the framework, there are simplifications that can be employed to facilitate regional assessments.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Thorjorn Larssen for providing the MCMC framework and Jack Cosby for insightful discussions on MAGIC. This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program and an NSERC Discovery grant.

REFERENCES

Aherne, J., Dillon, P. J. & Cosby, B. J. 2003 Acidification and recovery of aquatic ecosystems in south-central Ontario, Canada: regional application of the MAGIC model. *Hydrol. Earth Syst. Sci.* **7**, 561–573.

Aherne, J., Futter, M. N. & Dillon, P. J. 2007 The impacts of future climate change and sulphur emission reductions on acidification recovery at Plastic Lake, Ontario. *Hydrol. Earth Syst. Sci.* **4**, 1–28.

Aherne, J., Larssen, T., Cosby, B. J. & Dillon, P. J. 2006 Climate variability and forecasting surface water recovery from acidification: modelling drought-induced sulphate release from wetlands. *Sci. Total Environ.* **365**, 186–199.

Aherne, J., Larssen, T., Dillon, P. J. & Cosby, B. J. 2004 Effects of climate events on elemental fluxes from forested catchments in Ontario, Canada: modelling drought-induced redox processes. *Water Air Soil Pollut. Focus* **4**, 37–48.

Beven, K. J. & Binley, A. M. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process.* **6**, 279–298.

Brooks, S. P. & Roberts, G. O. 1998 Convergence assessment techniques for Markov Chain Monte Carlo. *Stat. Comput.* **8**, 319–335.

Cosby, B. J., Ferrier, R. C., Jenkins, A. & Wright, R. F. 2001 Modelling the effects of acid deposition: refinements, adjustments and inclusion of nitrogen dynamics in the MAGIC model. *Hydrol. Earth Syst. Sci.* **5**, 499–517.

Cosby, B. J., Hornberger, G. M., Galloway, J. N. & Wright, R. F. 1985 Modeling the effects of acid deposition: assessment of a lumped parameter model of soil water and stream water chemistry. *Water Res. Res.* **21**, 51–63.

Dillon, P. J., Lusis, M., Reid, R. A. & Yap, D. 1988 Ten-year trends in sulphate, nitrate and hydrogen deposition in central Ontario. *Atmos. Environ.* **22**, 901–905.

Gallagher, M. & Doherty, J. 2007 Parameter estimation and analysis for a watershed model. *Environ. Modell. Software* **22**, 1000–1020.

Gelman, A., Roberts, G. O. & Gilks, W. R. 1995 Efficient metropolis jumping rules. *Bayesian Stat.* **5**, 599–607.

Hornberger, G. M., Cosby, B. J. & Galloway, J. N. 1986 Modeling the effects of acid deposition: uncertainty and spatial variability in estimation of long-term sulfate dynamics in a region. *Water Res. Res.* **22**, 1293–1302.

Jeffries, D. S., Clair, T. A., Couture, S., Dillon, P. J., Dupont, J., Keller, W., McNicol, D. K., Turner, M. A., Vet, R. & Weeber, R. 2003 Assessing the recovery of lakes in southeastern Canada from the effects of acid deposition. *Ambio* **32**, 176–182.

Kennedy, M. C. & O'Hagan, A. 2001 Bayesian calibration of computer models. *J. R. Statist. Soc. B* **63**, 425–464.

Kuczera, G. & Parent, E. 1998 Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *J. Hydrol.* **211**, 69–85.

Larssen, T. 2005 Model prognoses for future acidification recovery of surface waters in Norway using long-term monitoring data. *Environ. Sci. Technol.* **39**, 7970–7979.

Larssen, T., Brereton, C. & Gunn, J. M. 2003 Dynamic modeling of recovery from acidification of lakes in Killarney Park, Ontario, Canada. *Ambio* **32**, 244–248.

Larssen, T., Cosby, B. J. & Hogasen, T. 2004 Uncertainties in predictions of surface water acidity using the MAGIC model. *Water Air Soil Pollut. Focus* **4**, 125–137.

Larssen, T., Hogasen, T. & Cosby, B. J. 2007 Impact of time series data on calibration and prediction uncertainty for a deterministic hydrogeochemical model. *Ecol. Modell.* **207**, 22–33.

Larssen, T., Huseby, R., Cosby, B. J., Host, G., Hogasen, T. & Aldren, M. 2006 Forecasting acidification effects using a Bayesian calibration and uncertainty propagation approach. *Environ. Sci. Technol.* **40**, 7841–7847.

LaZerte, B. D. 1993 The impact of drought and acidification on the chemical exports from a minerotrophic conifer swamp. *Biogeochemistry* **18**, 153–175.

Locke, B. A. & Scott, L. D. 1986 *Studies of Lakes and Watersheds in Muskoka-Haliburton, Ontario: Methodology (1976–1985)*. Dorset Research Centre, Ontario Ministry of the Environment, Dorset, Ontario, Data report DR 86/4.

Lozano, F. 1987 *Physical and Chemical Properties of the Soils at the Southern Biogeochemistry Study Site*. Dorset Research Centre, Ontario Ministry of the Environment, Dorset, Ontario, Rep. BGC-018.

OME 1983 *Handbook of Analytical Methods for Environmental Samples*. Ontario Ministry of the Environment, Rexdale, Ontario, Technical Report.

Ouimet, R. & Duchesne, L. 2005 Base cation mineral weathering and total release rates from soils in three calibrated forest watersheds on the Canadian Boreal Shield. *Can. J. Soil Sci.* **85**, 245–260.

Page, T., Beven, K. J., Freer, J. & Jenkins, A. 2003 Investigating the uncertainty in predicting responses to atmospheric deposition using the model of acidification of groundwater in catchments (MAGIC) within a generalised likelihood uncertainty estimation (GLUE) framework. *Water Air Soil Pollut.* **142**, 71–94.

Posch, P., Aherne, J., Forsius, M., Fronzek, S. & Veijalainen, N. 2008 Modelling the impacts of European emission and climate change scenarios on acid-sensitive catchments in Finland. *Hydrol. Earth Syst. Sci.* **12**, 449–463.

Reinds, G. J., van Oijen, M., Heuvelink, G. B. M. & Kros, H. 2008 Bayesian calibration of the VSD soil acidification model using European forest monitoring data. *Geoderma* **146**, 475–488.

Rosenthal, J. S. 1994 Theoretical rates of convergence for Markov Chain Monte Carlo. In: *Conference Proceedings, Interface '94*.

Saloranta, T. M., Kamari, J., Rekolainen, S. & Malve, O. 2003 Benchmark criteria: a tool for selecting appropriate models in the field of water management. *Environ. Mngmnt.* **32**, 322–333.

Scheider, W. A., Reid, R. A., Locke, B. A. & Scott, L. D. 1983 *Studies of Lakes and Watersheds in Muskoka-Haliburton, Ontario: Methodology (1976–1982)*. Dorset Research Centre, Ontario Ministry of the Environment, Dorset, Ontario, Data Report DR 83/1.

Skeffington, R. A. 2006 Quantifying uncertainty in critical loads: (a) literature review. *Water Air Soil Pollut.* **169**, 3–24.

Sverdrup, H. & Warfvinge, P. 1993 Calculating field weathering rates using a mechanistic geochemical model–PROFILE. *J. Appl. Geochem.* **8**, 273–283.

van der Sluijs, J. P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J. & Risbey, J. 2005 Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: the NUSAP system. *Risk Anal.* **25**, 481–492.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W. & Verstraten, J. M. 2005 Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. *Water Res. Res.* **41**, W01017.

Vrugt, J. A., terBraak, C. J. F., Clark, M. P., Hyman, J. M. & Robinson, B. A. 2008 Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. *Water Res. Res.* **44**, W00B09.

Watmough, S. A., Aherne, J. A., Alewell, C. A., Arp, P., Bailey, S., Clair, T., Dillon, P., Duchesne, L., Eimers, C., Fernandez, I., Foster, N., Larssen, T., Miller, E., Mitchell, M. & Page, S. 2005 Sulphate, nitrogen and base cation budgets at 21 forested catchments in Canada, the United States, and Europe. *Environ. Monit. Assess.* **109**, 1–36.

Watmough, S. A. & Dillon, P. J. 2001 Base cation losses from a coniferous catchment in central Ontario, Canada. *Water Air Soil Pollut. Focus* **1**, 507–524.

Watmough, S. A. & Dillon, P. J. 2004 Major fluxes from a coniferous catchment in central Ontario, 1983-1999. *Biogeochemistry* **67**, 369–398.

Wright, R. F., Larssen, T., Camarero, L., Cosby, B. J., Ferrier, R. C., Helliwell, R. C., Forsius, M., Jenkins, A., Kopacek, J., Majer, V., Molden, F., Posch, M., Rogora, M. & Schopp, W. 2005 Recovery of acidified European surface waters. *Environ. Sci. Technol.* **39**, 64A–72A.

First received 2 October 2008; accepted in revised form 6 February 2009