Towards reduced uncertainty in catchment nitrogen modelling: quantifying the effect of field observation uncertainty on model calibration

Raat, K.J.; Vrugt, J.A.; Bouten, W.; Tietema, A.

DOI
10.5194/hess-8-751-2004

Publication date
2004

Published in
Hydrology and Earth System Sciences

Hydrology and Earth System Sciences, 8(4), *751–763* (2004) © EGU

Hydrology & Earth System Sciences

Towards reduced uncertainty in catchment nitrogen modelling: quantifying the effect of field observation uncertainty on model calibration

Klaasjan J. Raat, Jasper A. Vrugt, Willem Bouten and Albert Tietema

Centre for Geo-Ecological Research (ICG), Institute for Biodiversity and Ecosystem Dynamics (IBED) – Physical Geography, Universiteit van Amsterdam, Nieuwe Achtergracht 166, NL-1018WV Amsterdam, The Netherlands

E-mail for corresponding author: k.raat@science.uva.nl

Abstract

The value of nitrogen (N) field measurements for the calibration of parameters of the INCA nitrogen in catchment model is explored and quantified. A virtual catchment was designed by running INCA with a known set of parameters, and field 'measurements' were selected from the model run output. Then, using these measurements and the Shuffled Complex Evolution Metropolis algorithm (SCEM-UA), four of the INCA model parameters describing N transformations in the soil were optimised, while the measurement uncertainty was increased in subsequent steps. Considering measurement uncertainty typical for N field studies, none of the synthesised datasets contained sufficient information to identify the model parameters with a reasonable degree of confidence. Parameter equifinality occurred, leading to considerable uncertainty in model parameter values and in modelled N concentrations and fluxes. Fortunately, combining the datasets in a multi-objective calibration was found to be effective in dealing with these equifinality problems. With the right choice of calibration measurements, multi-objective calibrations resulted in lower parameter uncertainty. The methodology applied in this study, using a virtual catchment free of model errors, is proposed as a useful tool foregoing the application of a N model or the design of a N monitoring program. For an already gauged catchment, a virtual study can provide a point of reference for the minimum uncertainty associated with a model application. When setting up a monitoring program, it can help to decide what and when to measure. Numerical experiments indicate that for a forested, N-saturated catchment, a fortnightly sampling of NO_3 and NH_4 concentrations in stream water may be the most cost-effective monitoring strategy.

Keywords: INCA, nitrogen model, parameter uncertainty, multi-objective calibration, virtual catchment, experimental design

Introduction

Over the past 100 years, human activity has doubled the input of nitrogen (N) into terrestrial ecosystems (Vitousek *et al.*, 1997), causing environmental problems such as soil acidification, nitrate (NO_3) contamination of ground waters and eutrophication of lakes and streams. These problems have initiated intensive research in the field of N cycling, including the development of mathematical models describing N dynamics in soils and surface waters. These models provide a basis for integrating N transformation and transport processes and thus serve as an aid to understanding the fate of N in ecosystems. Moreover, models that simulate and predict N dynamics have become an indispensable tool for the abatement and prevention of N-related environmental problems (Neal *et al.*, 2002).

One such model is INCA (Wade *et al.*, 2002; Whitehead *et al.*, 1998a), a semi-distributed, physically-based model describing N dynamics in catchments. Recent investigations have demonstrated that INCA is able closely to predict N concentrations in rivers for a range of European catchments (e.g. Wade *et al.*, 2001). Unfortunately, these studies include little information on the uncertainty in the values of the model parameters used in the applications. However, as INCA is ultimately developed to explore the effects of changes in land use, N deposition and climate on N loads in catchments, there is a strong need for this kind of information.

Like almost any catchment model, many of the parameters of INCA cannot be measured directly but have to be inferred by a trial-and-error process that adjusts the parameter values to match the observed data. This process is called 'model calibration'. The aim of a model calibration is to reduce the uncertainty in the choice of parameter values (parameter uncertainty) while accounting for uncertainties in the measured input and output time series and uncertainties in the structural ability of the model to simulate the processes of interest (Thiemann *et al.*, 2001). Preferably, a calibration results in well-identified parameters with narrow uncertainty ranges around their optimum values. However, since catchment models are only an approximate description of reality and because the data used for calibration contain errors, estimates of parameters are error-prone (Vrugt *et al.*, 2002). As a consequence, well-identified parameters are often the exception rather than the rule.

A serious complication for the calibration of models describing N dynamics in catchments is the lack of reliable calibration data, especially when considering model parameters that describe the soil N transformations. Often, the only data used are concentrations of NO_3 in stream water, usually taken at weekly or fortnightly intervals and spanning a period of at most three years. Stream water NO_3 can be measured relatively easily, at low cost and with relatively high accuracy. However, these measurements may not contain useful information on the model parameters of interest and, as such, may be of limited value for model calibration. Measurements of N fluxes in soils, like nitrification or net mineralisation, often are informative to N model parameters, but these fluxes are difficult and costly to measure and are subject to large measurement errors. These large errors stem mainly from the heterogeneity of the soil, which complicates the estimation of N fluxes at a plot or catchment scale.

The Shuffled Complex Evolution Metropolis (SCEM-UA; Vrugt *et al.*, 2003a), is an effective and efficient search algorithm for the calibration of model parameters. Apart from finding the most suitable set of parameters, SCEM-UA aims at describing parameter uncertainty using a Bayesian inference framework. One of the desirable properties of this Bayesian framework is that the user can incorporate knowledge explicitly about the measurement errors (σ) of the calibration data into the estimation of the model parameters. The size of this measurement error determines the quality of the calibration data directly, and as such the final estimated uncertainty intervals of the parameters with the SCEM-UA algorithm.

The aim of this study was to explore the suitability of N field measurements for the calibration of parameters of the INCA model. The parameters of interest were four of the most relevant INCA parameters describing the N transformations in the soil-vegetation system of a well-drained, N-saturated forest. A virtual catchment was designed by running INCA with a known set of parameters, and field 'measurements' were selected from this model run output. These synthetically generated observations were subsequently used in combination with the SCEM-UA algorithm to retrieve the uncertainty intervals of the four INCA model parameters and to assess which measurement types contain the most information for the identification of the model parameters. To further explore the relationship between the quality of the calibration data and the uncertainty associated with the final parameter estimates, the measurement error σ was increased, stepwise, in subsequent optimisation runs.

Methods

INCA MODEL

A full description of INCA (version 1.6) appears in Wade *et al*. (2002) and Wade (2004) but a slightly modified version (version 1.7.1)was used to prevent numerical integration problems at low stream flows. Here, only those features of the INCA model are described which are necessary for a clear understanding of the results found in this study.

In short, INCA is a semi-distributed (lumped), physically-based model that simulates NO_3 and NH_4 concentrations in stream water by tracking water and N through the catchment soils and ground waters to the river. The soil-vegetation system in INCA is of primary importance, as N inputs and most N transformations take place there. As such, most of the parameters in INCA refer to processes in the soil-vegetation system. The groundwater zone only transports N; no N transformations are assumed to occur. Finally, the river system exports NO_3 and NH_4 out of the catchment, while taking into account in-river nitrification and denitrification.

The soil-vegetation system in INCA is represented by a single mixing model, which is an obvious simplification of reality. In addition to this simplification, denitrification and NO_3 plant uptake were assumed not to take place in the virtual catchment. This latter simplification is justified by the assumed low denitrification and NO_3 plant uptake fluxes in the well-drained, N saturated forest of Speuld, the Netherlands (Tietema *et al.*, 1993), which served as a model for the N cycling in the soils of the virtual catchment. As such, the only N fluxes in the soil-vegetation system taken into account in this study were atmospheric N deposition, gross NH_4 mineralisation, gross NH_4 immobilisation, nitrification, NH_4 plant uptake and NO_3 and NH_4 leaching. Whereas atmospheric deposition of NO_3 and NH_4 is input to the model, the other fluxes are calculated within the INCA model as follows (notation as in Wade *et al*., 2002; fluxes in kg-N km^{-2} day^{-1}):

$$\text{gross } NH_4 \text{ mineralisation} = C_{gmi} \cdot S_1 \cdot 100 \quad (1)$$

$$\text{gross } NH_4 \text{ immobilisation} = C_{gim} \cdot S_1 \cdot \frac{x_5}{V_{r,s} + x_{11}} \cdot 10^6 \quad (2)$$

$$\text{nitrification} = C_{nit} \cdot S_1 \cdot \frac{x_5}{V_{r,s} + x_{11}} \cdot 10^6 \quad (3)$$

$$NH_4 \text{ plant uptake} = C_{upt} \cdot S_1 \cdot S_2 \cdot \frac{x_5}{V_{r,s} + x_{11}} \cdot 10^6 \quad (4)$$

$$NH_4 \text{ leaching} = \frac{x_1 \cdot x_5 \cdot 86400}{V_{r,s} + x_{11}} \quad (5)$$

$$NO_3 \text{ leaching} = \frac{x_1 \cdot x_3 \cdot 86400}{V_{r,s} + x_{11}} \quad (6)$$

where C_{gmi} (kg-N ha^{-1} day^{-1}), C_{gim}, C_{nit} and C_{upt} (m day^{-1}) denote the gross NH_4 mineralisation, gross NH_4 immobilisation, nitrification and NH_4 plant uptake rate coefficients; x_5 and x_3 represent the NH_4 and NO_3 stores in the soil compartment (kg-N km^{-2}); $V_{r,s}$ is the soil water retention volume (m^3 km^{-2}); x_{11} is the soil water volume (m^3 km^{-2}); x_1 is the outflow of water from the soil (m^3 s^{-1} km^{-2}); S_1 signifies the soil moisture factor (-); S_2 is the seasonal plant growth index (-); and 100, 10^6 and 86400 are constants necessary for conversion to the correct units. Full definitions of $V_{r,s}$, x_{11}, x_1, S_1 and S_2 are given in Wade *et al.* (2002).

The NH_4 and NO_3 stores in the soil (x_5 and x_3) are calculated by integrating Eqns. (7) and (8):

$$\frac{dx_5}{dt} = NH_4 \text{ atmospheric deposition} + \text{gross } NH_4 \text{ mineralisation} - NH_4 \text{ leaching} - NH_4 \text{ plant uptake} - \text{nitrification} - \text{gross } NH_4 \text{ immobilisation} \quad (7)$$

$$\frac{dx_3}{dt} = NO_3 \text{ atmospheric deposition} + \text{nitrification} - NO_3 \text{ leaching} \quad (8)$$

None of the parameters in Eqns. (1 – 8) can be measured directly, instead they have to be inferred by model calibration. As there is some physical meaning to the hydrological parameters (x_1, x_{11}, $V_{r,s}$, S_1) and the seasonal plant growth index (S_2), appropriate values for these parameters can be assessed with relative confidence. In contrast, the rate coefficients in Eqns. (1 – 8) are highly conceptual, lack a clear physical meaning, and thus very little is known about suitable values for these parameters. As such, in the present study, focus lay on the calibration of the C_{gmi}, C_{gim}, C_{nit} and C_{upt} rate coefficients.

SCEM-UA

To estimate the values of the rate coefficients, the recently developed Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm was used. This algorithm is a modified version of the original SCE global optimisation algorithm developed by Duan *et al.* (1992) and uses a Bayesian inference scheme to estimate the best set of model parameters, along with its underlying posterior distribution. The SCEM-UA algorithm operates by selecting and modifying an initial population of parameter sets merging the strengths of a Markov Chain Monte Carlo (MCMC) algorithm developed by Metropolis *et al.* (1953), with the concepts of controlled random search (Price, 1987), competitive evolution (Holland, 1975) and complex shuffling (Duan *et al.*, 1992) to evolve the population of initial parameter sets to a stationary posterior target distribution.

Assuming that the error residuals between model and measurement are mutually independent, Gaussian distributed, with constant variance, the posterior density, or *likelihood*, of a parameter set θ_i for describing the observed data y is computed by SCEM-UA using the equation specified by Box and Tiao (1973):

$$L(\theta_i|y) = \exp\left[-\frac{1}{2}\sum_{j=1}^{N}\left|\frac{e(\theta_i)_j}{\sigma}\right|^2\right] \quad (9)$$

in which N signifies the number of measurements, σ denotes the measurement error deviation of the observations (‘measurement error’) and e represents the error residuals between model and measurement.

The size of the measurement error σ has important implications for SCEM-UA applications. Following Eqn. (9), an increment in the size of the measurement error will result in a wider range of parameter sets that will be considered acceptable in the fitting of the calibration data. In other words, large uncertainties in the measurements will result in a large uncertainty in the choice of parameter values and consequently in the model simulations. In line with this reasoning, the present study investigated how uncertainty in observations of N concentrations and fluxes affect the uncertainty in INCA parameters and simulations.

An important issue, when applying MCMC samplers like the Metropolis algorithm in SCEM-UA, is the convergence to a stationary posterior distribution. In theory, a MCMC sampler converges when the number of sampled parameter sets θ_i approaches infinity, that is $i \rightarrow \infty$. However, in practice one has to decide on how many draws to make with the sampler. To help decide, Gelman and Rubin (1992) developed a quantitative conversion diagnostic, the scale reduction factor $\sqrt{SR}$, based on within and between Markov chain variances. Following their recommendations, convergence to a stationary posterior distribution can be declared when $\sqrt{SR}$ drops below 1.2. When this criterion is not met, estimates of parameter distribution intervals derived from the final posterior distribution are only an approximation, and actual distribution intervals may be wider.

VIRTUAL CATCHMENT

The Doethie sub-catchment of the River Tywi system in South Wales (Whitehead *et al.*, 1998b) served as a model for the hydrology in the virtual catchment. The virtual catchment is a 2 km^2 watertight forested catchment that is drained by a single stream. Input time series (January 1991 – December 1998) of temperature, hydrologically effective rainfall (HER) and soil moisture deficit (SMD) were taken from the River Kennet system in southern England. The Speuld forest in the Netherlands (Raat *et al.*, 2002; Tietema *et al.*, 1993) served as a model for the N cycling in the soil-vegetation system. As such, the virtual catchment is considered N-saturated, receiving high levels of atmospheric N deposition, thereby resulting in high levels of NO_3 leaching.

Table 1 gives a complete list of the values of parameters used to characterise the virtual catchment. A model run with these 'true' parameters and the input data served as a reference run, or 'true' run, of the N cycling in the virtual catchment. The INCA output of this reference run is given in Fig. 1 and Table 2.

CALIBRATION DATA

Synthetic 'field measurements' were selected from the reference run output and included in the calibration datasets. The different calibration datasets included measurements of NO_3 and NH_4 concentrations in soil and stream water, and net mineralisation and net nitrification fluxes in the soil compartment. Measurements were selected on a fortnightly basis over a period of three years (July 1991 –June 1994). N fluxes were calculated as the 14-day sum of the daily fluxes calculated by INCA; net mineralisation was defined as the difference between the 14-day sum of gross NH_4 mineralisation and gross NH_4 immobilisation. Soil and stream water concentrations were selected from the model output every 14th day.

A summary of all calibration datasets and their short names as used in the text is given in Table 3. Note that no noise was added to the synthetic measurements as is sometimes done in studies on virtual systems (e.g. McIntyre and Wheater, 2004; Vrugt *et al.*, 2002). As such, the synthetic measurements are an exact representation of the catchment's state variables and processes.

PARAMETER OPTIMISATION AND UNCERTAINTY ASSESSMENT

The C_{gmi}, C_{gim}, C_{nit} and C_{upt} rate coefficients were optimised using the different calibration datasets. In addition, it was explored how the uncertainty ranges of the inversely estimated rate coefficients change with increasing measurement error.

In each application, the SCEM-UA algorithm was set to simultaneously optimise the four rate coefficients, using eight complexes and a population size of 240 (Vrugt *et al.*, 2003a). The error residual between model and measurement was calculated using a Simple Least Square objective function. If the scale reduction factor $\sqrt{SR}$ did not drop below 1.2 within the first 10 000 simulations, it was assumed that a stationary solution could not be found. For each rate coefficient, the feasible parameter space was a uniform distribution between 0 and 20 times the true value of the rate coefficient. This space can seen as relatively wide, given the fact that normally little information is available on the approximate values of these rate coefficients.

For each calibration dataset, in subsequent SCEM-UA optimisations, the measurement error σ (Eqn. 9) was increased from 0.1 to a maximum of 50% of the average value of the measurement of interest during the calibration period (July 1991–July 1994). The measurement error was defined as the uncertainty in field observations arising from the combined effect of analytical, sampling and support errors. Hence, a measurement error of 0.1% is a large underestimation of the uncertainties that are commonly present in actual field datasets. This very small error was used to verify whether the SCEM-UA algorithm is indeed able to infer the original rate coefficients used to generate

Table 2. Mean annual N fluxes in reference run (kg-N ha^{-1} yr^{-1})

NO_3 deposition	6.5
NH_4 deposition	20.0
NH_4 plant uptake	48.0
Gross NH_4 mineralisation	134.4
Gross NH_4 immobilisation	87.3
Net NH_4 mineralisation	47.1
Nitrification	15.6
Denitrification	0.0
NO_3 leaching	21.8
NH_4 leaching	3.3

Table 1. INCA parameter values used to characterise the virtual catchment. A model run with these 'true' parameters and the input data served as a reference run of the N cycling in the virtual catchment. See Wade *et al*. (2002) for parameter definitions.

Catchment characteristics		
Area	2.00	[km^2]
Land use type	Forest	
Maximum $V_{r,s}$ (depth · porosity)	0.16	[m]
Soilwater residence time	4.30	[day]
Groundwater residence time	56.0	[day]
Base flow index	0.43	[-]
River parameters		
River length	1000	[m]
Flow-a	0.04	[m^{-2}]
Flow-b	0.67	[-]
In-river denitrification	0.05	[day^{-1}]
In-river nitrification	0.20	[day^{-1}]
N deposition		
NO_3 dry deposition	6.0	[kg-N ha^{-1} yr^{-1}]
NO_3 wet deposition	0.5	[kg-N ha^{-1} yr^{-1}]
NH_4 dry deposition	12.2	[kg-N ha^{-1} yr^{-1}]
NH_4 wet deposition	7.8	[kg-N ha^{-1} yr^{-1}]
Soil hydrology		
Maximum temperature difference	4.5	[°C]
Soilwater deficit maximum	150	[mm]
Response to a 10°C temperature change	2.0	[-]
Base temperature response	30	[°C]
Initial snow pack depth	0	[mm]
Degree-day factor for snowmelt	3.0	[mm °C^{-1} day^{-1}]
Water equivalent factor	0.30	[-]
Snow depth / soil temperature factor	-0.025	[-]
Soil N transformation parameters		
Gross NH_4 mineralisation rate (C_{gmi})	**2.00**	[kg-N ha^{-1} day^{-1}]
Gross NH_4 immobilisation rate (C_{gim})	**0.14**	[day^{-1}]
Nitrification rate (C_{nit})	**0.025**	[day^{-1}]
Denitrification rate	0	[day^{-1}]
Vegetation parameters		
Plant growth start day	50	[julian day]
Plant growth period	253	[day]
Maximum yearly N plant uptake	70	[kg-N ha^{-1} yr^{-1}]
Plant NH_4 uptake rate (C_{upt})	**0.20**	[day^{-1}]
Plant NO_3 uptake rate	0	[day^{-1}]

the synthetic data. In the literature, little information is available on errors made in determining stream water chemistry. Yet, given good mixing of stream water, small measurement errors of 5–10% were assumed typical for NO_3 concentrations. Errors in NH_4 concentrations are probably somewhat higher (10–20%) as NH_4 concentrations in stream water are often low and close to detection limits (e.g. Langusch and Matzner, 2002; Whitehead *et al.*, 2002). Finally, mainly due to the heterogeneous nature of the soil, measurements of soil water N concentrations (e.g.

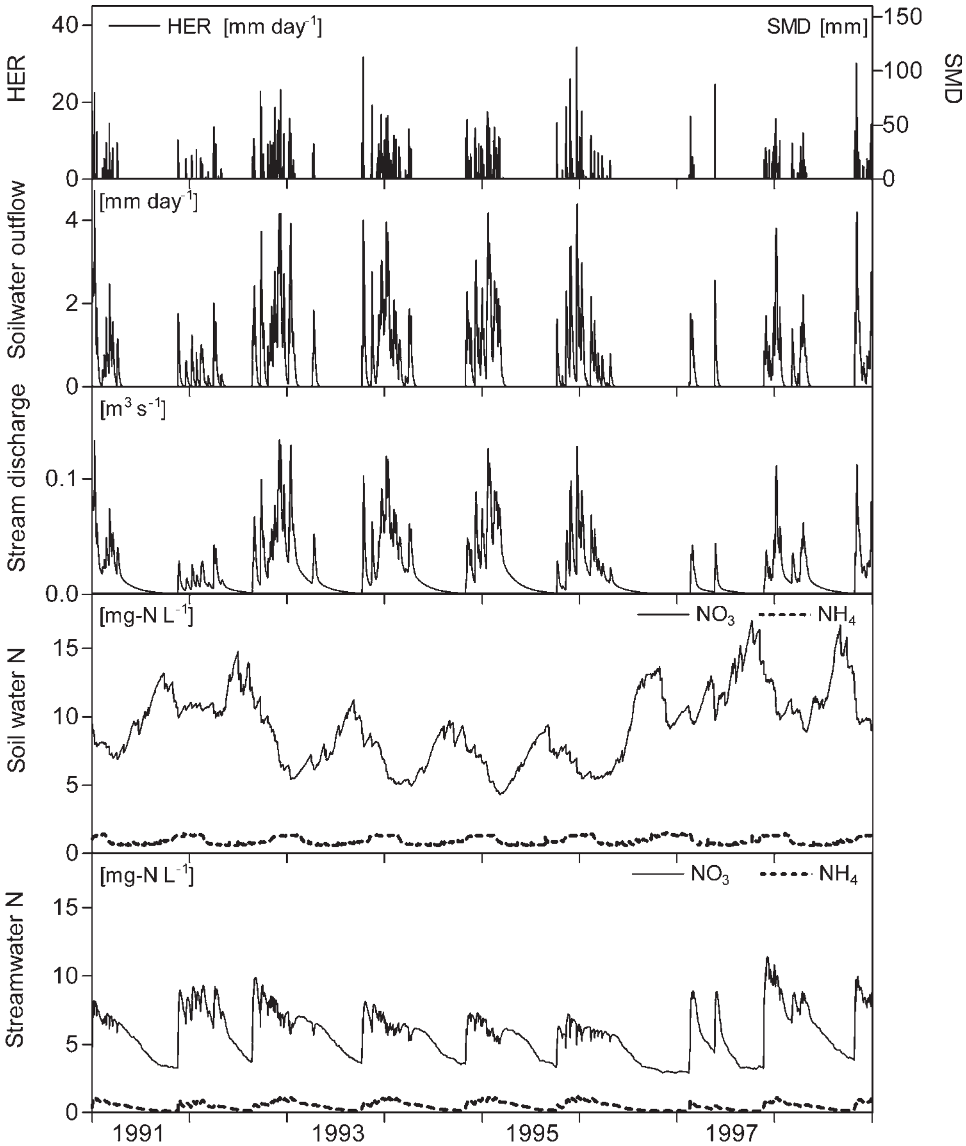

Fig. 1.*INCA-simulated stream discharge, soil water flow and N concentrations in soil water and streamwater corresponding to the parameters of the reference run.*

Table 3. Datasets used for calibration. Synthetic 'measurements' were selected fortnightly between July 1991 and June 1994 (3 years). The measurement error denotes the error that was assumed typical for real-world field measurements. See text for further details.

Short name	*Measurement type*	*Measurement error (%)*
streamNO_3	Streamwater NO_3 concentration	10
streamNH_4	Streamwater NH_4 concentration	20
soilNO_3	Soilwater NO_3 concentration	50
soilNH_4	Soilwater NH_4 concentration	50
NMI	Net NH_4 mineralisation	50
NIT	Net nitrification	50

Manderscheid and Matzner, 1995; Rothe *et al.*, 2002) and soil N fluxes (e.g. Laverman *et al.*, 2000; Tietema *et al.*, 1993) come with large errors of 20% or more. A summary of the measurement errors that were assumed typical for the measurements of the different calibration datasets is given in Table 3.

After each calibration, the distribution intervals of the rate coefficients (95% confidence level) were computed from the final SCEM-UA derived parameter sets in the posterior distribution. These parameter sets were subsequently used to compute the prediction uncertainty ranges associated with the INCA simulated N-concentrations and fluxes.

Results

NO_3 CONCENTRATIONS IN SOIL WATER (SOILNO$_3$) AND STREAM WATER (STREAMNO$_3$)

Both soilNO$_3$ and streamNO$_3$ were found to contain sufficient information to retrieve the original rate coefficients at a small measurement error of 0.1%. For both optimisations, convergence was met within 2000 simulations and the uncertainty in rate coefficient values was small (Table 4 for streamNO$_3$; results for soilNO$_3$ were similar to streamNO$_3$ and are not shown). Starting at a 1% measurement error, however, the SCEM-UA algorithm already experienced problems converging to a stationary posterior distribution. After 10 000 simulations, $\sqrt{SR}$ was still higher than 2.0 for soilNO$_3$. For streamNO$_3$, $\sqrt{SR}$ dropped below 1.2 after 2000 simulations but increased again to values between 1.2 and 2.0. At a 5% measurement error, $\sqrt{SR}$ was between 1.5 and 6 after 10 000 simulations with streamNO$_3$. Extending this optimisation run to 50,000 simulations did not improve the optimisation, as $\sqrt{SR}$ did not drop below 2.0.

For the 1% and 5% measurement error, estimates of the parameter distribution intervals (95% confidence level) for streamNO$_3$ are also listed in Table 4. Similar results were found for soilNO$_3$ and are not shown. Note in Table 4 that that as convergence criteria were not met for optimisations with the 1% and 5% measurement error, actual intervals may have been slightly wider. At a 1% measurement error, the proposed intervals are still narrow, but at a 5% measurement error they have become very wide. For example, C_{gmi} varied between 0 and a maximum of 23.7. This maximum value corresponds with a near 24 times over-estimation of the gross NH_4 mineralisation. INCA runs with the accepted rate coefficient sets showed that the sets indeed accurately simulate NO_3 concentrations in stream water, but that large uncertainties are associated with the simulations of NH_4 concentrations (soil and stream) and gross NH_4 mineralisation and gross NH_4 immobilisation fluxes (Fig. 2). Apparently, a wide variety of rate coefficient sets can adequately simulate NO_3 concentrations in stream water, while erroneously simulating N fluxes in the soil compartment.

NH_4 CONCENTRATIONS IN SOIL WATER (SOILNH$_4$) AND STREAM WATER (STREAMNH$_4$)

The NH_4 datasets showed approximate ($\sqrt{SR} \approx 1.4$ after 10 000 simulations; streamNH$_4$) or slow convergence ($\sqrt{SR} < 1.2$ after 8000 simulations; soilNH$_4$) at 0.1% measurement error. This problematic convergence may be due to the very strict parameter acceptance criteria associated with such a small measurement error. Under these strict conditions, the optimum region in the parameter space is likely to be very small, or 'narrow', making it difficult to locate.

Both streamNH$_4$ (Table 5) and soilNH$_4$ (not shown) were successful in retrieving C_{gmi} and C_{upt}, but C_{gim} and C_{nit} were not effectively confined. This was due to the near perfect correlation between C_{gim} and C_{nit} when optimising using NH_4 measurements. r equalled –1.00 for both streamNH$_4$ and soilNH$_4$, calculated from the last 2000 SCEM-UA simulations. This strong negative correlation indicated that an overestimation of C_{gim} (and a subsequent overestimation of gross NH_4 immobilisation) is compensated by an underestimation of the C_{nit} (and nitrification), thus rendering correct estimates of the amounts of NH_4 removed from the soil by these two processes. Hence, when using NH_4 concentrations (either in soil or stream), it is impossible to identify both C_{gim} and C_{nit}. Only information on the combined effect of both parameters can be retrieved.

Table 4. Number of simulations before convergence and rate coefficient distribution intervals (95% confidence level) for optimisation with fortnightly streamwater NO_3 concentrations (streamNO$_3$). The 95% confidence levels were calculated from the last 2000 simulations. Note that for these simulations $\sqrt{SR} > 1.2$ at 1 and 5% measurement errors, and that actual intervals may have been wider.

	Simulations before $\sqrt{SR} < 1.2$	C_{gmi}	C_{gim}	C_{nit}	C_{upt}
Reference		2.00	0.14	0.025	0.20
Feasible space		0 – 40	0 – 2.80	0 – 0.50	0 – 4.00
0.1%	2000	1.81 – 2.23	0.12 – 0.16	0.0246 – 0.0254	0.18 – 0.22
1.0%	*	1.40 – 2.65	0.10 – 0.20	0.022 – 0.030	0.15 – 0.29
5.0%	*	1.10 – 23.7	0.75 – 2.75	0.023 – 0.49	0.096 – 3.97

* Convergence not achieved within 10 000 simulations.

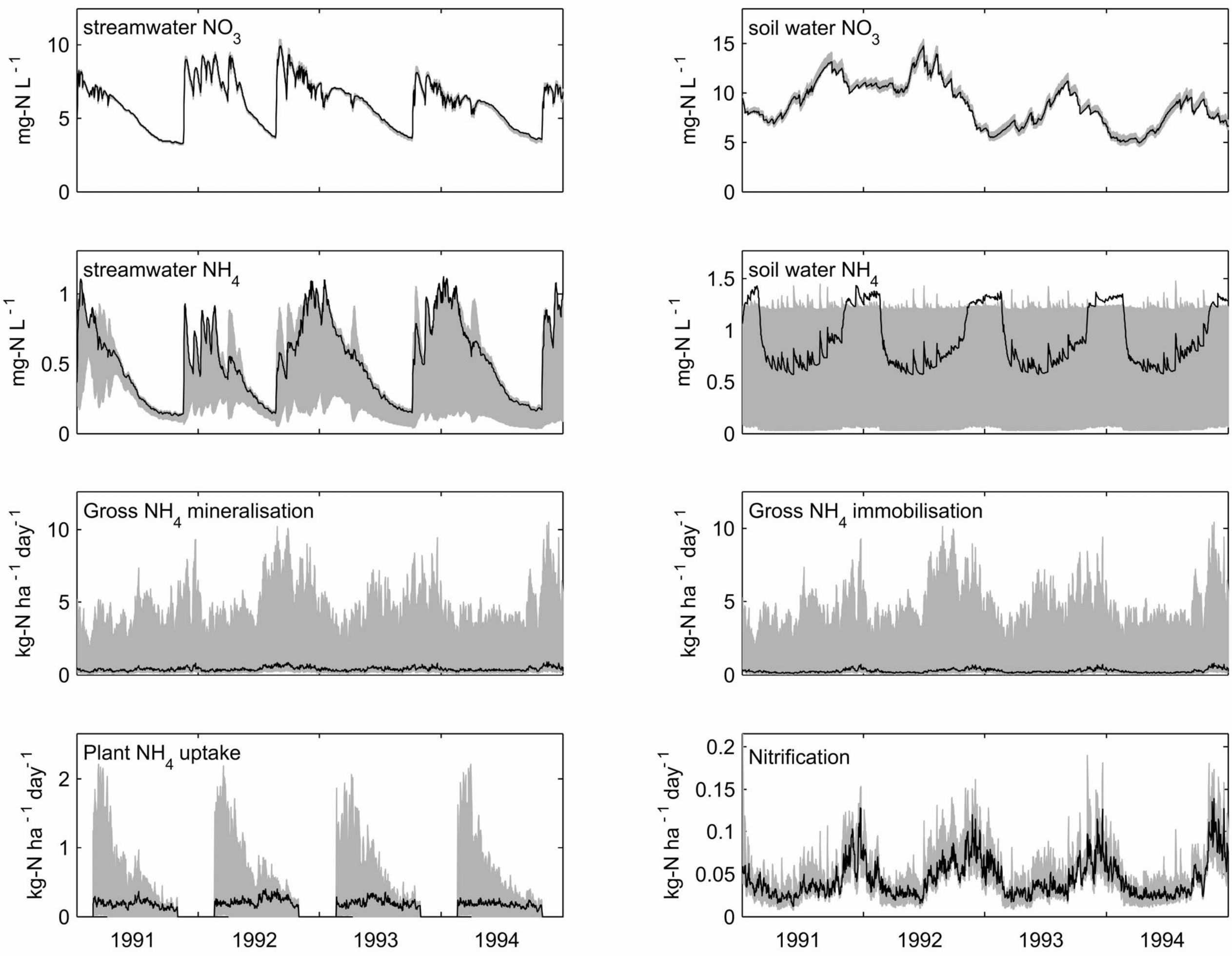

Fig. 2. *Uncertainty ranges (gray region, 95% confidence level) associated with calibration using 3-years of fortnightly streamwater NO_3 concentrations (streamNO$_3$) and a 5% measurement error. Notice that for this optimization $\sqrt{SR} > 1.2$, meaning that actual uncertainty ranges may have been wider. The solid line denotes the reference run.*

Table 5. Number of simulations before convergence and rate coefficient distribution intervals (95% confidence level) for optimisation with fortnightly streamwater NH_4 concentrations (streamNH$_4$). The 95% confidence levels were calculated from the last 2000 simulations. Note that for these simulations $\sqrt{SR} > 1.2$ at 0.1 and 20% measurement errors, and that actual intervals may have been wider.

	Simulations before $\sqrt{SR} < 1.2$	C_{gmi}	C_{gim}	C_{nit}	C_{upt}
Reference		2.00	0.14	0.025	0.20
Feasible space		0 – 40	0 – 2.80	0 – 0.50	0 – 4.00
0.1%	*	1.98 – 2.02	0 – 0.165	0 – 0.164	0.198 – 0.202
1.0%	3000	1.82 – 2.21	0 – 0.175	0 – 0.172	0.186 – 0.218
5.0%	4000	1.41 – 2.66	0 – 0.207	0 – 0.207	0.148 – 0.256
10%	3000	1.04 – 2.94	0 – 0.228	0 – 0.230	0.118 – 0.267
20%	*	1.35 – 3.23	0.02 – 0.215	0 – 0.214	0.120 – 0.296

* Convergence not achieved within 10,000 simulations.

Contrary to the NO_3 datasets, convergence diagnostics did not deteriorate with increasing measurement error for soilNH_4 and streamNH_4. At 10% measurement error, convergence was still met after 3500 simulations and C_{gmi} and C_{upt} were reasonably confined. Only when the measurement error was 20% or more, $\sqrt{SR}$ did not drop below 1.2 (minimum 3.0 and 1.5 for soilNH_4 and streamNH_4 respectively) and none of the parameter values could be identified with acceptable precision.

Table 5 lists the rate coefficient distribution intervals found for streamNH_4 (95% confidence level) for a 0.1 to 20% measurement error. Again, similar results were found for soilNH_4 and are not shown. Although C_{gim} and C_{nit} could not be confined effectively, streamNH_4 was found to be more effective in confining C_{gmi} and C_{upt} than streamNO_3. For example, at 5% measurement error C_{gmi} was confined between 1.41 and 2.66 by streamNH_4, whereas streamNO_3 led to C_{gmi} varying between 1.10 and 23.7. INCA runs with the accepted rate coefficient sets showed that at a measurement error of 20%, simulations were acceptable for NH_4 concentrations (soil and stream), gross NH_4 mineralisation and NH_4 plant uptake (Fig. 3). The uncertainties associated with the prediction of gross NH_4 immobilisation were considerable, and very large uncertainties accompanied the prediction of NO_3 concentrations (soil and stream) and nitrification.

NET MINERALISATION (NMI) AND NITRIFICATION(NIT) MEASUREMENTS

NMI and NIT were effective in constraining all four rate coefficients as long as the measurement error was not more than 5 (NIT) or 10% (NMI). At larger measurement errors, similar problems as for the NO_3 datasets were encountered. Table 6 shows the distribution intervals (95% confidence level) after calibration with NMI and NIT, respectively, using a 50% measurement error, a value typical for these types of measurements. Again, note that at this measurement error actual intervals may be wider as convergence criteria were not met. For both datasets, intervals are very wide for C_{gim},

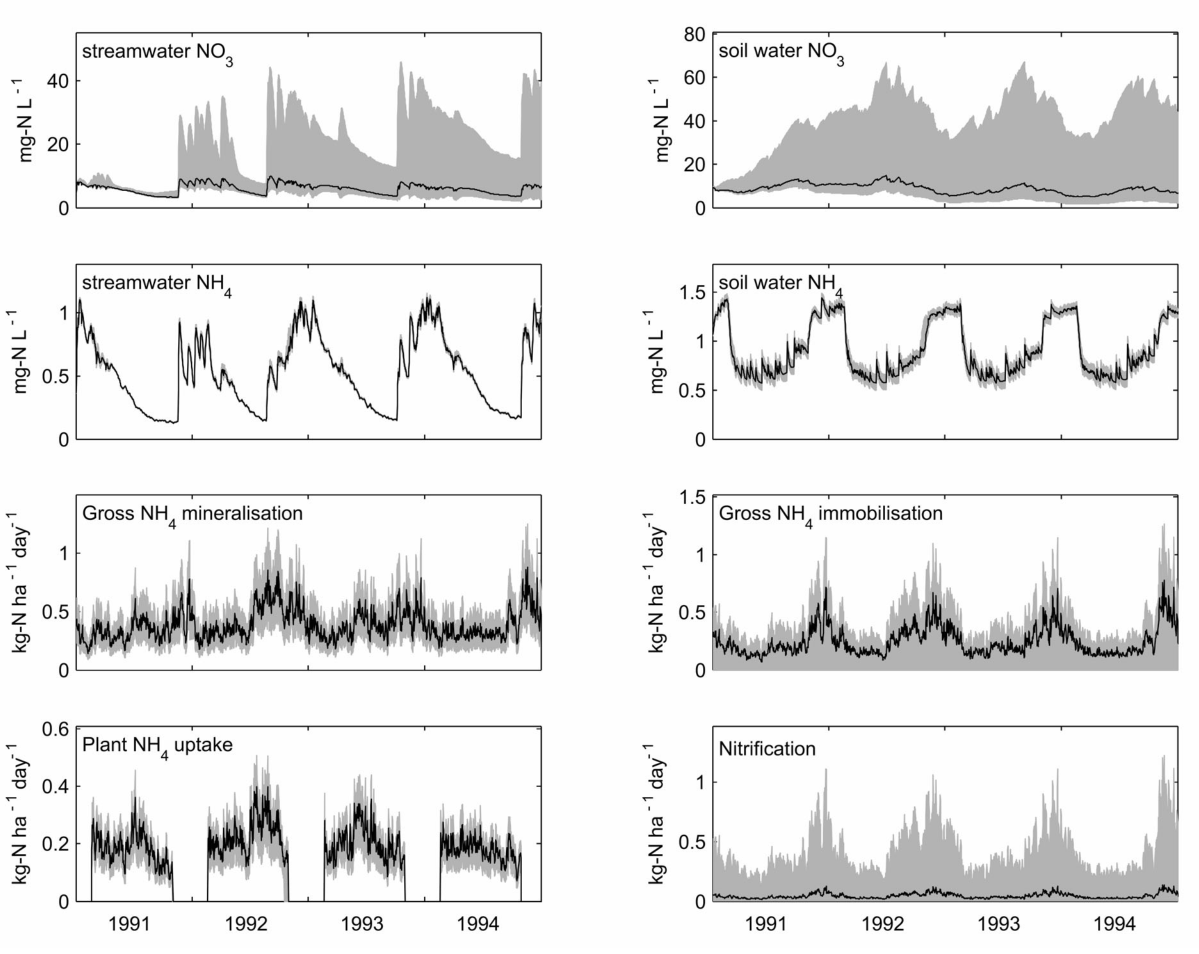

Fig.3. *Uncertainty ranges (gray region, 95% confidence level) associated with calibration using 3-years of fortnightly streamwater NH_4 concentrations (streamNH_4) and a 20% measurement error. The solid line denotes the reference run.*

Table 6. Rate coefficient distribution intervals (95% confidence level) for optimization with fortnightly net mineralisation (NMI) and nitrification (NIT) measurements, respectively, and a 50% measurement error. The 95% confidence levels were calculated from the last 2000 simulations. Note that for these simulations $\sqrt{SR} > 1.2$, and that actual intervals may have been wider.

	C_{gmi}	C_{gim}	C_{nit}	C_{upt}
Reference	2.00	0.14	0.025	0.20
Feasable space	0 – 40	0 – 2.80	0 – 0.50	0 – 4.00
NMI	0.99 – 4.95	0.09 – 2.80	0.01 – 0.50	0.24 – 3.99
NIT	0.64 – 4.60	0.47 – 2.80	0.10 – 0.50	0.24 – 3.99

C_{nit} and C_{upt}, but relatively small for C_{gmi}. For NMI, INCA runs accompanying the intervals showed adequate simulation of NH_4 plant uptake and nitrification, but poor agreement between modelled and measured soil water NO_3 concentrations and, especially, soil and stream water NH_4 concentrations (results not shown). For NIT, simulations were acceptable for NO_3 concentrations (soil and stream water) and all N fluxes in the soil compartment, but very poor for both soil and stream water NH_4 concentrations (results not shown).

EXTENDING DATASETS

The results presented in the previous sections illustrate the severity of the parameter estimation problem. When the measurement error, specified in the density criterion in Eqn. (9), is of the same order as that typically present in field observations, none of the datasets utilised for model calibration contain sufficient information to identify the four rate coefficients with a reasonable degree of confidence.

When confronted with these problems it seems reasonable to consider increasing the number of observations in the calibration dataset, either by extending the period of data collection or by increasing the measurement frequency. Additional calibrations using stream water NO_3 concentrations measured fortnightly for seven years (July 1991–June 1998) or measured daily for three years (July 1991–June 1994), showed that this strategy did not help to tackle the current problem. The seven-years record did not show any improvement compared to the original fortnightly three-years streamNO_3 dataset. A 1% measurement error still was the limit for correct inference of the rate coefficients. The daily record did show some improvements compared to streamNO_3, but the maximum acceptable measurement error still was not larger than 5%.

MULTI-OBJECTIVE OPTIMISATION

Combining different datasets may be more effective in reducing parameter uncertainty than extending the period of data collection or increasing the measurement frequency. To verify the validity of this hypothesis, the various types of datasets were combined to yield *multi-objective* datasets, which were subsequently used for parameter calibration.

Multi-objective datasets were constructed as follows. First, to enable equal weighing of different measurement types with different involved units, the measurements of the original *single-objective* datasets were scaled to a mean of 100 by dividing by the average value of the type of measurement (July 1991–June 1994) and multiplying by 100. Next, two or more of these scaled datasets were combined to form a multi-objective dataset. Please note that in optimisation runs with these multi-objective datasets, INCA output was scaled correspondingly.

SCEM-UA optimisations with these multi-objective datasets showed that measurements of NH_4 concentrations (soil or stream) play a key role in identifying the rate coefficients. For example, a multi-objective calibration using soil water NO_3 concentrations, net NH_4 mineralisation and nitrification measurements was successful only when the measurement error was 20% or less. Adding soil water NH_4 concentrations to the calibration dataset rendered successful calibrations, even when the measurement error was set as high as 50%. In this latter optimisation that used a measurement error typical for measurements in the soil compartment, all possible information on N in the soil was combined. As such, this run set the minimum uncertainty associated with the rate coefficient values when only measurements conducted in the soil compartment are available. These were 1.62–2.63, 0.10–0.21, 0.022–0.030 and 0.16–0.25 for C_{gmi}, C_{gim}, C_{nit} and C_{upt}, respectively (95% confidence intervals).

Finally, calibration using a combination of stream water NO_3 and NH_4 concentrations was found to be a reasonable alternative for using difficult, uncertain and costly soil N measurements. At the typical measurement error of 20%, uncertainty in rate coefficients was confined to 0.87–3.56,

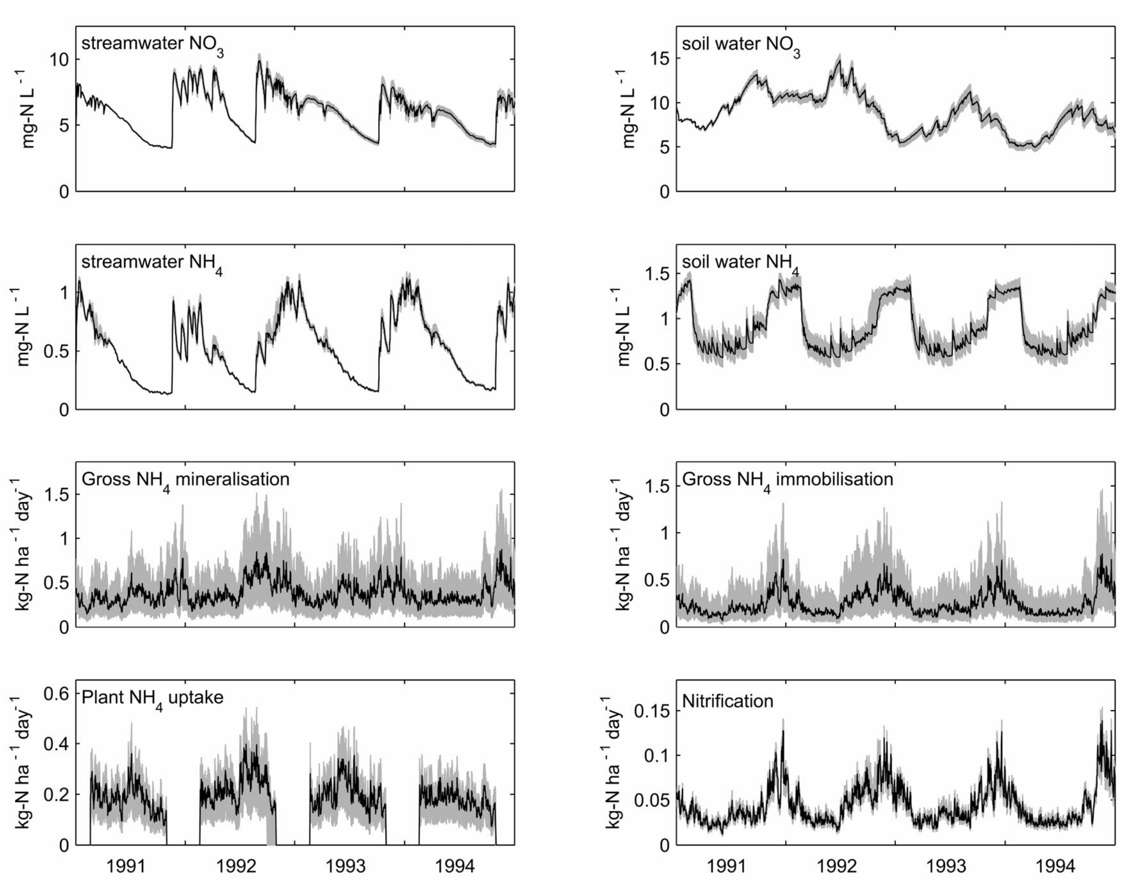

Fig. 4. *Uncertainty ranges (gray region, 95% confidence level) associated with calibration using a combination of streamwater NO_3 and NH_4 concentrations and a 20% measurement error. The solid line denotes the reference run.*

0.06–0.27, 0.021–0.027 and 0.08–0.30 for C_{gmi}, C_{gim}, C_{nit} and C_{upt} (95% confidence). INCA runs with the accepted rate coefficient sets (Fig. 4), showed small prediction uncertainty ranges in the simulation of NO_3 and NH_4 concentrations (soil and stream) and nitrification, and slightly larger uncertainty bounds for gross NH_4 mineralisation and gross NH_4 immobilisation.

Discussion

For the ideal situation where the INCA model structure is an exact representation of the system studied, four of the parameters describing N transformations in the soil were optimised using the SCEM-UA algorithm. The results demonstrated that, given typical measurement errors in N studies, datasets containing only one type of measurement (*single-objective calibration*) contain very limited information for the identification of the model parameters. Only calibrations using datasets of multiple measurements (*multi-objective calibration*) resulted in low parameter uncertainty and acceptable simulations of all N concentrations and fluxes in the soil and stream systems.

Regardless of the dataset used, parameter uncertainty for the single-objective calibrations was high: a wide variety of parameter sets could adequately predict the observed measurements. This phenomenon, named equifinality by Beven (1993), has been found in many hydrological studies (e.g. Beven and Freer, 2001; Duan *et al.*, 1992) and in some soil geochemical studies (Zak *et al.*, 1997; Zak and Beven, 1999). Recently, Schulz *et al.* (1999) showed that equifinality also exists for N budget models. Contrary to the Schulz *et al.* study, in which equifinality may have resulted from uncertainty in input data (rainfall and latent heat fluxes), measurement errors and the inability of the model to correctly describe the system of interest (model structural errors), the present results suggest that equifinality may also result from measurement errors alone. As such, if it is accepted that in nitrogen studies measurements will always come with errors, equifinality is endemic to the type of models used.

A closer look at the structure of the INCA model (Eqns. 1–8) provides insight into why equifinality may occur. It was mentioned already that the near perfect negative correlation between C_{gim} and C_{nit} when optimising using NH_4

concentrations (either in soil or stream) indicates that erroneous gross NH_4 immobilisation fluxes (due to erroneous C_{gim}) are compensated by (erroneous) nitrification fluxes (erroneous C_{nit}). Similar within-model compensation, or 'internal budgeting', may as well apply to other fluxes or when using other datasets for calibration. For example, in theory, when using NO_3 concentrations for calibration, a too low net NH_4 mineralisation may be compensated by a too high C_{nit} or too low C_{upt}, ensuring that the available NH_4 is transformed into NO_3 rather than taken up by plants. Of course, very dynamic, complex systems are being dealt with and, thus, internal budgeting is unclear. Yet, the many runs that provide good estimates of NO_3 or NH_4 concentrations while overestimating soil N fluxes, at least suggest that internal budgeting is an important mechanism causing equifinality.

Irrespective of the exact cause for equifinality, it is evident that none of the available measurements alone contain sufficient information to calibrate the parameters of interest. This does not make INCA a bad model, but does show that INCA and alike models have a high data requirement, making calibration difficult. The analysis showed that increasing the measurement frequency does not necessarily help to reduce parameter uncertainty. Seemingly, in the virtual catchment, dynamics in stream water NO_3 concentrations are almost equally well captured by fortnightly as by daily observations, resulting in only minor differences in information content of both datasets. As an alternative to intensifying measurements, a more productive way to reduce parameter uncertainty is to weigh different measurement sets in a multi-objective framework (e.g. Vrugt *et al.*, 2003b).

The use of a virtual catchment, free of model and input data errors, of course sets limits to the interpretation of the results in a real-world context. However, analyses like those presented in this study can serve as a useful tool preceding the application of an environmental model or the design of a monitoring programme (e.g. McIntyre and Wheater, 2004). Firstly, for an already gauged catchment, with a given amount, type and reliability of calibration data, the methodology applied here provides insight into the minimum uncertainty associated with a model application. Knowing this beforehand is important as it may prevent the modeller from an endless search for ever better parameter combinations. For example, the multi-objective calibration with stream water NO_3 and NH_4 concentrations set the minimum uncertainty associated with an INCA application to a N-saturated, forested catchment in which only stream water N concentrations were available. Note that these results are valid only when just the rate coefficient values are unknown, and that input and model structural errors are assumed absent. As such, it is indeed a very conservative, 'minimum', estimate of the model and parameter uncertainty, and actual uncertainties will be higher.

Second, when setting up a monitoring programme, the analyses presented in this paper can help to decide what and when to measure, especially if there is ample confidence in the model's capability to describe the system of interest. In the present catchment, stream water NO_3 and NH_4 concentrations were found almost as useful for model calibration as difficult and costly measurements in the soil compartment. Also, extending the period of data collection or increasing the measurement frequency hardly reduced parameter uncertainty. As such, three-years of fortnightly sampling of both NO_3 and NH_4 concentrations in the stream water may be the most cost-effective monitoring strategy for a N-saturated, forested catchment. Again, note that these results are only valid for the given conditions, that is when only the rate coefficients are unknown. When more or other parameters are uncertain, or when a catchment contains more than one land use type, (a combination of) other measurements or a different measurement frequency could be more appropriate.

Conclusions

Even for the ideal situation where the INCA model structure is an exact representation of the system studied, calibration of soil N parameters is difficult due to parameter equifinality. Single-objective calibrations, using only one type of measurement, render large uncertainty in both parameter values and modelled N concentrations and fluxes. Increasing the measurement frequency or extending the period of data collection does not necessarily help to reduce this uncertainty. Calibration using multiple sets of measurements, however, is an effective way to deal with the equifinality problems. With the right choice of calibration measurements, a multi-objective calibration results in low parameter uncertainty and proper modelling of the N cycle.

The methodology applied in this study, using a virtual catchment that is free of model errors, can serve as a useful tool to provide a point of reference for the minimum uncertainty associated with a model application. In addition, this methodology can aid the design of a N monitoring programme. The numerical experiments indicate that for a forested, N-saturated catchment, a fortnightly sampling of NO_3 and NH_4 concentrations in stream water may be the most cost-effective monitoring strategy.

Acknowledgements

The authors thank Dan Butterfield for writing a command line version of INCA and Andrew Wade for providing input time series. Koos Verstraten and two anonymous reviewers are thanked for their valuable comments on earlier versions of this paper. This research was supported by the European Commission (Project EVK1-1999-00011). In addition, Klaasjan Raat was supported financially by IBED, Universiteit van Amsterdam.

References

Beven, K.J. and Freer, J., 2001. A dynamic TOPMODEL. *Hydrol. Process.,* **15**, 1993–2011.

Beven, K. J., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Adv. Water Resour.,* **16**, 41–51.

Box, G. E.P. and Tiao, G.C., 1973. *Bayesian inference in statistical analyses.* Addison-Wesley Publishing Company, Reading, Massachusetts, USA.

Duan, Q. Y., Sorooshian, S. and Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.,* **28**, 1015–1031.

Gelman, A. and Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Stat. Sci.,* **7**, 457–472.

Holland, J., 1975. *Adaptation in natural and artificial systems.* Univerisity of Michigan Press, Ann Harbor, USA..

Langusch, J.J. and Matzner, E., 2002. N fluxes in two nitrogen saturated forested catchments in Germany: dynamics and modelling with INCA. *Hydrol. Earth Syst. Sci.,* **6**, 383–394.

Laverman, A.M., Zoomer, H.R., Van Verseveld, H.W. and Verhoef, H.A., 2000. Temporal and spatial variation of nitrogen transformations in a coniferous forest soil. *Soil Biol. Biochem.,* **32**, 1661–1670.

Manderscheid, B. and Matzner, E., 1995. Spatial and temporal variation of soil solution chemistry and ion fluxes through the soil in a mature Norway spruce (*Picea-Abies* (L) Karst) stand. *Biogeochemistry,* **30**, 99–114.

McIntyre, N.R. and Wheater, H.S., 2004. Calibration of an in-river phosphorus model: prior evaluation of data needs and model uncertainty. *J. Hydrol.,* **290**, 100–116.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E., 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.,* **21**, 1087–1091.

Neal, C., Jarvie, H.P., Wade, A.J. and Whitehead, P.G., 2002. Water quality functioning of lowland permeable catchments: inferences from an intensive study of the River Kennet and upper River Thames. *Sci. Total Envir.,* **282**, 471–490.

Price, W.L., 1987. Global optimization algorithms for a CAD workstation. *J. Optimiz. Theory App.,* **55**, 133–146.

Raat, K.J., Draaijers, G.P.J., Schaap, M.G., Tietema, A. and Verstraten, J.M., 2002. Spatial variability of throughfall water and chemistry and forest floor water content in a Douglas fir forest stand. *Hydrol. Earth Syst. Sci.,* **6**, 363–374.

Rothe, A., Huber, C., Kreutzer, K. and Weis, W., 2002. Deposition and soil leaching in stands of Norway spruce and European Beech: Results from the Hoglwald research in comparison with other European case studies. *Plant Soil,* **240**, 33–45.

Schulz, K., Beven, K.J and Huwe, B., 1999. Equifinality and the problem of robust calibration in nitrogen budget simulations. *Soil Sci. Soc. Amer. J.,* **63**, 1934–1941.

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

Tietema, A., Riemer, L., Verstraten, J.M., Van der Maas, M.P., Van Wijk, A.J. and Van Voorthuyzen, I., 1993. Nitrogen cycling in acid forest soils subject to increased atmospheric nitrogen input. *Forest Ecol. Manage.,* **57**, 29–44.

Vitousek, P.M., Aber, J.D., Howarth, R.W., Likens, G.E., Matson, P.A., Schindler, D.W., Schlesinger, W.H. and Tilman, D.G., 1997. Human alteration of the global nitrogen cycle: Sources and consequences. *Ecol. Appl.,* **7**, 737–750.

Vrugt, J.A., Bouten, W., Gupta, H.V. and Sorooshian, S., 2002. Toward improved identifiability of hydrologic model parameters: The information content of experimental data. *Water Resour. Res.,* **38**, 1312.

Vrugt, J.A., Bouten, W., Gupta, H.V. and Sorooshian, S., 2003a. A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrological model parameters. *Water Resour. Res.,* **39**, 1201.

Vrugt, J.A., Gupta, H.V., Bastidas, L.A., Bouten, W. and Sorooshian, S., 2003b. Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Wat. Resour. Res.,* **39**, 1214.

Wade, A.J., Durand, P., Beaujouan, V., Wessel, W.W., Raat, K.J., Whitehead, P.G., Butterfield, D., Rankinen, K. and Lepisto, A., 2002. A nitrogen model for European catchments: INCA, new model structure and equations. *Hydrol. Earth Syst. Sci.,* **6**, 559–582. (See also Errata. *Hydrol. Earth Syst. Sci.*, **8**, 858–859.)

Wade, A.J., Soulsby, C., Langan, S.J., Whitehead, P.G., Edwards, A.C., Butterfield, D., Smart, R.P., Cook, Y. and Owen, R.P., 2001. Modelling instream nitrogen variability in the Dee catchment, NE Scotland. *Sci. Total Envir.,* **265**, 229–252.

Whitehead, P.G., Lapworth, D.J., Skeffington, R.A. and Wade, A.J., 2002. Excess nitrogen leaching and C/N decline in the Tillingbourne catchment, southern England: INCA process modelling for current and historic time series. *Hydrol. Earth Syst. Sci.,* **6**, 455–466.

Whitehead, P.G., Wilson, E.J. and Butterfield, D., 1998a. A semi-distributed Integrated Nitrogen model for source assessment in Catchments (INCA): Part I - model structure and process equations. *Sci. Total Envir..,* **210/211**, 547–558.

Whitehead, P.G., Wilson, E.J., Butterfield, D. and Seed, K., 1998b. A semi-distributed Integrated Nitrogen model for source assessment in Catchments (INCA): Part II - application to large river basins in south Wales and eastern England. *Sci. Total Envir.,* **210/211**, 559–583.

Zak, S.K., Beven, K.J. and Reynolds, B., 1997. Uncertainty in the estimation of critical loads: A practical methodology. *Water Air Soil Poll.,* **98**, 297–316.

Zak, S.K. and Beven, K.J., 1999. Equifinality, sensitivity and predictive uncertainty in the estimation of critical loads. *Sci. Total Envir.* **236**, 191–214.