Application of a Markov Chain Monte Carlo calibration and uncertainty framework to a process-based integrated nitrogen model (INCA)

B. M. Jackson[a], *H. S. Wheater*[a], *N. McIntyre*[a] *and P. Whitehead*[b]

[a]Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, U.K.;
[b]Department of Geography, The University of Reading, Reading RG6 6AB, U.K.
E-mail: b.m.jackson@imperial.ac.uk

Abstract:

As part of the production of an integrated modelling system for lowland permeable catchments, a stochastic framework is being developed to allow quantification of uncertainty in the representation of catchment response and impacts of management scenarios, and to investigate how capable various conceptual models are of adequately characterising water flow, nitrate and phosphorous transport given a reasonably calibrated and, where applicable, physically realistic parameter set. To address these considerations, a collection of stochastic routines, including Markov chain Monte Carlo capabilities, have been integrated with a semi-distributed nitrogen model. This Integrated Nitrogen in Catchments model (INCA) simulates flow, nitrate and ammonium and tracks the temporal variations in flow and nitrogen mass operating in both land and river phases. This paper discusses some of the issues and initial results arising from the first application of Markov chain Monte Carlo (MCMC) to scenarios utilising the INCA model. Performance is illustrated with data from the Kennet catchment in southern England. The results demonstrate the power of Markov chain Monte Carlo methods to quantitatively examine the interrelationship between model structure, parameter identifiability and data support, but also the reliance of MCMC and other heuristic methodologies on objective function choices and model robustness.

Keywords: Markov chain Monte Carlo, calibration, catchment modelling, nitrate

1. INTRODUCTION

An integrated nitrogen model has been developed to investigate the fate and distribution of nitrogen in the aquatic and terrestrial environment. This Integrated Nitrogen in CAtchments model (INCA) simulates flow, nitrate and ammonium over the catchment scale, coupling land processes and in-river processes. Dilution, natural decay and biochemical transformation processes are included in the model as well as interactions with plant biomass. It is semi-distributed to account for spatial variations in land use, human impacts, effluent discharges and varying deposition levels, and produces daily estimates of the stream water flow and nitrate and ammonium concentrations, in addition to estimates of annual, land-use specific, N fluxes. The original model was described in Whitehead et al. (1998), while more recent additions to the model structure are contained in [1]. The

Sensitivity Analysis of Model Output
Kenneth M. Hanson and François M. Hemez, eds.
Los Alamos National Laboratory, 2005; http://library.lanl.gov/

model has been successfully applied to a range of catchments in the U.K. and Europe, and is still being refined to extend its applicability to a variety of catchment management needs. However, little work on parameter sensitivity and identifiability has been carried out on it to date.

A detailed study of these issues should aid the successful calibration of further catchment applications by highlighting the most significant parameters and allowing informed decisions as to the areas in which experimental resources and measurements should be allocated. For catchment management purposes, there is also a need to provide measures of the uncertainty present due to measurement errors in the inputs, parametric uncertainty, and issues related to model conceptualisation. To address these considerations, methods for propagating uncertainty, analysing parameter sensitivity and model structure, and optimisation are placed within a subjective probability framework, along with a collection of appropriate "objective functions" to specify criteria for successful calibration. The more efficient methodologies utilise heuristic guidance to explore parametric spaces and model output distributions through an automatic semi-random exploration of the parameter space. Included in these routines are Markov chain Monte Carlo methods (using Metropolis Hastings formulae), used to sample parametric and uncertain quantities. The framework permits both parametric and model structural uncertainty to be interrogated, and allows effective calibration and confidence predictions through optimisation of model inputs to fit observations or other criteria, with explicit consideration of effects of data uncertainty.

MCMC methods possess the general virtue of simulation methods, with information regarding parametric probability distributions easily collected along with optimal parameter sets. However, other sampling methods generally fail when the posterior involves many variables or is otherwise intractable. Markov chain methods are capable of sampling from posterior distributions of arbitrary complexity, through the Metropolis Hastings algorithm, which provides simple conditions under which the chain will equilibrate to the required distribution. Since such methods sit naturally within a subjective probability framework, they are also capable of quantifying distortions produced on the outputs by noise. Such a capability is indispensable for rigourous analysis of an environmental model such as INCA, as the input is subject to extreme uncertainty.

The performance of the modelling framework is illustrated with data from the Kennet catchment in southern England. To understand the characteristics of both overall uncertainty and particular parametric sensitivities in INCA, the effect of changes in the parameters and inputs are examined using the Markov chain sampling described above. Response surfaces, in this case distributions of input parameters against single-valued measures of performance (derived from the output parameters and optimality criteria), are examined, and the biases caused by differing optima considered. The influence of such biases on subsequent decisions regarding parameter sensitivities and "optimal" parameter sets is examined. The efficiencies of differing Metropolis proposal functions applied to sample both the "minima" and entirety of a given response surface are also being investigated.

The results demonstrate the power of Markov chain Monte Carlo methods to quantitatively examine the inter-relationship between model structure, parameter identifiability

and data support, and also provide an efficient means of addressing the problem of calibration given large parameter sets and the presence of measurement error and other uncertainties. In the context of model development, however, the need for alternate, non-heuristically guided methodologies to be included in such stochastic tools is also demonstrated.

2. THE STOCHASTIC FRAMEWORK

Any model describing nutrient transport within a catchment, the result of complicated environmental processes with dependencies on both space and time, is necessarily a simplified representation of the phenomena being studied. This imposes a limit upon one's confidence in its responses or outputs, regardless of the accuracy of any input information. The input itself is subject to many sources of uncertainty, including measurement errors, absence of information, temporal and spatial variability, and incomplete understanding of underlying driving forces and mechanisms. Adequate spatial representation is particularly difficult, due to the intrinsic variability present within the environment, such as the continuous variation in soil properties and nitrogen inputs over space, and the difficulty of characterising properties in the subsurface.

To give a measure of confidence in scenario predictions, a reliable catchment modelling tool should provide measures of the uncertainty present due to measurement errors in the inputs, parametric uncertainty, and issues related to model conceptualisation, and be able to translate these measures into prediction confidence limits for management purposes [2]. Where models are still in development, stochastic analysis can also aid in identifying the components of model structures that are most significant in the simulation of nutrient dynamics in river systems, aspects that appear redundant, and the inter-relationship between model structure, parameter identifiability and data support. This allows informed decisions as to the areas in which experimental resources and measurements should be allocated. The relative importance of differing measurands over space is also important.

To address all these concerns within one framework, methods suitable for prediction uncertainty, model sensitivity to parameters and data error and calibration must be associated with appropriate catchment scale models. For successful calibration, collections of appropriate "objective functions" to specify optimal criteria, data processing capabilities for handling and analysing errors and guidelines for choosing calibration criteria and parameter distributions given specific modelling tasks must also be included.

3. UNCERTAINTY AND SENSITIVITY METHODOLOGIES

To address prediction uncertainty and model sensitivity, three possible Monte Carlo methodologies have been implemented: basic Monte-Carlo simulation, Latin hypercube sampling and Markov-chain Monte-Carlo techniques (Metropolis and Metropolis Hastings). The classic Monte Carlo method samples the input parameter space using the exact probability distribution assigned to it, which, given an exact mathematical description of the model, must converge eventually to the precise output distribution. The simulation's main impediment is its high computational cost, with the run numbers necessary for a successful analysis of a model's outputs typically running into the thousands [2]. A degree

of computational efficiency can be accomplished through the use of efficient input sampling methods, which may include heuristic search procedures (purposeful or partially informed searches using heuristic functions for guidance), or less informed approaches where segments of the probability distributions are split or stratified, and systematically explored. The latter approach is contained within the framework through inclusion of the Latin Hypercube method [3].

To formally include subjective probability to be used within the framework, Markov chain Monte Carlo methods (using Metropolis Hastings formulae) are used to sample parametric and uncertain quantities. These methods sample from the input, or posterior, distribution, and sit naturally within a subjective probability (Bayesian) framework. A Markov chain is a series of random variables $\{X(0), X(1), X(2), \ldots, X(N)\}$ for which the conditional probability of a transition from any state $X(i)$ to any other state $X(j)$ depends only on the current state, and not on any previous states. The construction of a Markov chain requires two basic ingredients, namely an initial distribution (a first approximation to the probability of being in each the states X(i)) and a transition matrix [4]. This transition matrix is a matrix of probabilities, defining all the associated probabilities (transition probabilities) of the chain moving from state $X(i)$ to state $X(j)$, $i, j \in (1, N)$.

Markov chain Monte-Carlo methods draw samples from a Markov chain rather than from the probability distribution $f(x)$. When constructed carefully, these can be very efficient approximators. Most Markov chain schemes in use today, such as the popular Gills sampler, are a variant of the Metropolis Hastings approach. For details, see Gilks et al. (1996). The Markov chain is constructed such that its equilibrium distribution is that of the posterior distribution of interest. In this context, such a distribution might be that of the INCA parameters conditional on measured "output" observations and optimality constraints, or the uncertainty present in a prediction given uncertainty in measurands and model structure.

Each state is visited the required number of times to satisfy the conditional distribution of the parameters given the data. This is achieved through satisfying appropriate conditions of reversibility (detailed balance) and ergodicity (Hastings 1970). By giving the microscopic dynamics of the Markov chain (that is, an algorithm that determines $X(i + 1)$ given $X(i)$), the transition matrix and consequently the (unnormalised) input distribution is implicitly fixed. Markov chain methods are capable of sampling from posterior distributions of arbitrary complexity, through the Metropolis Hastings algorithm, which provides simple conditions under which the chain will equilibrate to the required distribution [5]. They have been successfully applied in hydrological modelling by Kuczera and Parent [6].

The draws from the Markov chain are accomplished through variants of the pleasingly simple Metropolis Hastings formulae, involving proposals of candidate values through a proposal function and rejection/acceptance steps. This proposal function is constructed such that it implicitly defines the required conditional distributions, along with satisfying the necessary Markov chain conditions. At any time, it describes the current knowledge regarding parameter distributions, given initial knowledge and information from prior runs. By equating the posterior with Bayes' rule, it also allows potential for converging upon the "true" input distributions through incorporation of learnt information [7].

Bayes' rule is derived from basic axioms of probability. In the context of this work, it is best viewed in terms of updating belief in a hypothesis H given new evidence D. A posterior belief $P(H|D)$, giving the probability of hypothesis H after considering the effect of new data is calculated by multiplying the prior belief $P(H)$ by the likelihood $P(D|H)$ that D will occur if H is true. There is no fundamental distinction between observable quantities and parametric inputs to a model; both can be considered to be random quantities. The theorem can be written as follows,

$$p(H|D) = \frac{p(H)p(D|H)}{p(D)}. \tag{1}$$

A first quantification of $P(H)$ is provided before any data is gathered; this is the prior probability of H. In the context of calibrating a physically based model, the "hypotheses" are the parameter value probability distributions adopted before a simulation commences. These are generally determined subjectively in terms of prior beliefs or knowledge, such as what are realistic ranges of the parameters from previous knowledge of their properties and of the specific scenario situation. Physical constraints, such as non-negativity of concentrations, are also generally included.

In the case of continuous problems, the hypotheses become one continuous parametric distribution. To avoid confusion with the discrete case, this will be denoted by θ, and the outcomes (data) by y. These could be scalars or vectors. Prior beliefs are specified as a probability density function $p(\theta)$, while the outcomes conditional on the hypotheses are the conditional density $p(y|\theta)$, often referred to as the likelihood function [2]. This prior and conditional density fully specify the joint density $p(\theta, y)$ over all hypotheses and data,

$$p(\theta, y) = p(\theta)p(y|\theta) \tag{2}$$

The marginal distribution $p(y)$ of y can be calculated from this joint distribution by integrating over θ,

$$p(y) = \int_\theta p(\theta, y)d\theta = \int_\theta p(\theta)p(y|\theta)d\theta. \tag{3}$$

The posterior probability distribution is therefore given by

$$p(\theta, y) = \frac{p(y|\theta)p(\theta)}{\int_\theta p(y|\theta)p(\theta)d\theta}. \tag{4}$$

This is the continuous form of Bayes' Theorem. The denominator, or marginal probability, is easily calculated by recognising that it can, given all other quantities, be considered as a normalising constant.

4. CALIBRATION METHODOLOGIES

The Monte Carlo routines above are suitable for both sensitivity and uncertainty analysis, as they preserve distributional information, and yield information on the total parameter and output spaces. However, in calibration, one is usually interested in locating only a

limited part of the above distributions: generally the global optimum (given appropriate parameter constraints), or, where data error, model structure error, or multiple objectives prevent a single optimum being achievable or meaningful, sets of acceptable local or Pareto-optimal parameters. While full explorations combined with additional information can produce information on "optimal" regions of this space, they are rarely an efficient means of doing so. Therefore, extra routines suitable for calibration have been included in the framework. These are: the Levenburg Marquadt method [8] and the SCEA (Shuffled Complex Evolution Algorithm) [9], neither of which will be considered further in this paper, and finally the addition of simulated annealing to the Markov chain Monte Carlo scheme.

Simulated annealing is a heuristic search procedure based on the metaphor of how annealing works [7]. It aims to reach a global minimum through a procedure that incorporates a decreasing random component to avoid trapping at a local minimum, by allowing a non-improving move to a neighbour with a probability that decreases over time. Since the random component is decreasing, the magnitude of any non-improving change also becomes smaller with time. The rate of this decrease is determined by the cooling schedule, often an exponential decay (in keeping with the thermodynamic metaphor).

To calibrate a model effectively, the "objective" of the calibration must be specified. An objective may be singular, or include several independent criteria that may need to be traded off against each other. They are very application, as well as model, specific, but generally include measures of fit applied to the model output against observed data, and often criteria aimed at minimising risks or costs (economic, environmental, etc). If a model is manually calibrated, the objective may be stated qualitatively: fits may be obtained by eye and intuition then play a part in choosing appropriate calibrated parameter sets. For automated calibration, an 'objective function" or functions giving a mathematical definition of how good a solution is must be formally specified.

Multiple objectives can arise from multiple types of output, emphasising different aspects of model performance, and also from time or space series of one quantity. The latter is generally compacted into one measure through application of a norm (such as least squares). However, different measures of fit will favour different aspects of a series [10]. For example, fits to the variation of stream flow over time may favour the overall water balance, overall shape, or weight the calibration to good agreement of low flow or peak flows. The chosen measure of fit may vary according to the modelling task; flood management may require good estimates of peak flows and overall shape, whereas low flows may be of more important for agricultural management.

One approach is to aggregate the multiple objectives into one single objective function, and optimise to the single-valued best fit. The result is then strongly dependent on the aggregation, or weighting of the objectives. An increasingly common alternative is to employ the concept of Pareto optimality [11]. A set of parameters is said to be Pareto optimal if an improvement in any one criteria will lead to another criteria being degraded; no criteria dominates. This concept does not give a single solution, but rather a set of solutions called the Pareto optimal set. Parameter sets corresponding to these solutions are called non-dominated, and give a visual trade-off between competing objectives. The

user can then choose a solution according to his or her preference. However, in a guided calibration, or to explore posterior distributions conditional upon data, optimality must still be reduced to one measure, perhaps through a normalised weighting of the objectives. One popular means of achieving this is through calculating the coefficient of determination R^2 (widely known in the hydrological literature as the Nash-Sutcliffe efficiency criterion), given by

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}, \tag{5}$$

where y_i is the simulated value, x_i is measured value, $\overline{x}$ is the mean of the measured values and n is the number of samples. This is the measure used in the forthcoming application, although a variety of approaches are being explored and added to the toolbox capabilities.

5. APPLICATION

To test performance of the integrated model stochastic framework, Markov chain Monte Carlo methods were applied to a model application utilising data from the Kennet catchment in southern England. This is a groundwater-dominated catchment draining an area of 1164 km^2, with a chalk aquifer supplying approximately 95% of its water. As it has been a focus of a variety of water quality and ecological concerns, there is a relatively large amount of data available to compare model response against.

An initial goal was to provide the INCA model with an automated calibration routine in place of the manual calibration procedure used previously, with the aim of enhancing the reliability of calibrated parameters due to a more exhaustive exploration of the parameter space and shortening implementation time. A second purpose was to highlight the most significant parameters for such a calibration, and to identify areas in which model structure could be strengthened. Thirdly, the reliance of results on differing calibration criteria is being investigated.

A one-year simulation period was chosen, from 1^{st} January 1998 to 31^{st} December 1998. Daily precipitation and air temperature were provided, and hydrologically effective rainfall along with soil moisture deficits obtained from MORECS (the U. K. Met Office Rainfall and Evaporation Calculation System). This extracts the relevant quantities from atmospheric data using a Penman-Monteith type routine. Geospatial and other relevant information (e.g. farming practices, proportional land use, base flow indexes, dry deposition data) were obtained from national databases and other sources where available. A hand-calibrated parameter set provided other parameters needed by the model, and was also used as the starting state for the Markov chain Monte Carlo simulations.

To address the first two purposes of the performance test, twenty-six parameters were varied within ranges constrained by existing calibration guidelines and literature values. These parameters are shown in Table 1.

Two scenarios were considered: the first treated the 26 parameters as spatially homogeneous in both land and river phases, and the second allowed for heterogeneity. This heterogeneous scenario had 6 unique land types, and divided the river Kennet into 25 “reaches”, or contiguous lengths. Each reach is then associated with a subcatchment, and

Table 1. Calibration/Sensitivity Parameters examined in Markov Chain Monte Carlo Kennet simulations, January-December 1998.

Name	Units	Distribution	Minimum	Maximum
initial soil flow	$m^3 s^{-1}$	land use	0.0	1.0
initial groundwater flow	$m^3 s^{-1}$	land use	0.0	0.1
initial soilwater nitrate	$mg.l^{-1}$	land use	0.0	10.0
initial groundwater nitrate	$mg.l^{-1}$	land use	0.0	8.0
initial soilwater ammonium	$mg.l^{-1}$	land use	0.0	2.0
initial groundwater ammonium	$mg.l^{-1}$	land use	0.0	1.0
initial soil drainage volume	m^3	land use	10^5	2×10^7
initial ground drainage volume	m^3	land use	10^6	10^8
initial in-stream flow rate	$m^3 s^{-1}$	top reach	0.0	2.0
initial in-stream nitrate	$mg.l^{-1}$	top reach	0.0	10.0
initial in-stream ammonium	$mg.l^{-1}$	top reach	0.0	2.0
denitrification rate	$m.day^{-1}$	land use	0.01	19.0
nitrogen fixation	$kg.ha^{-1}day^{-1}$	land use	0.0	0.0001
plant nitrate uptake	$m.day^{-1}$	land use	0.0	162.0
nitrification rate	$m.day^{-1}$	land use	1.0	54.0
mineralisation	$kg.ha^{-1}day^{-1}$	land use	1.0	292.0
immobilisation rate	$m.day^{-1}$	land use	0.0	1.0
ammonium addition rate	$kg.ha^{-1}day^{-1}$	land use	0.0	100.0
plant ammonium uptake	$m.day^{-1}$	land use	0.0	162.0
Reactive zone residence time	$days$	land use	0.5	5.0
Groundwater residence time	$days$	land use	10.0	200.0
Maximum soil water retention†	m	land use	0.0	1.0
velocity flow a parameter (Qa)	-	by reach	0.001	0.2
velocity flow b parameter Qb	-	by reach	0.3	0.99
Denitrification rate	day^{-1}	by reach	0.04	0.09
Nitrification rate	day^{-1}	by reach	0.1	5.0

this association allows for the coupling of in-river and land processes (for more details, see [12]). This discretisation in space resulted in 215 unique parameters. The objectives for calibration were taken to be least squares fits to in-river flow, nitrate, and ammonium concentrations, with the Nash-Sutcliffe criterion providing a means to weight these appropriately for heuristic guidance.

The first implementation of the tool was only partially successful, as the model was not robust over the entire parametric space. Such problems are common in a first application of a full sensitivity analysis to a model, as an automated routine is likely to discover (by brute force) subtle instabilities in a model formulation, and also extract any parameter constraints that have not been explicitly specified. This interfered with the ability of the heuristically guided Markov chain Monte Carlo and calibration methodologies to explore model input and output distributions.

When tested on sub-sets of parameters where the model was robust, Markov chain

Monte Carlo analysis results were encouraging, and the annealing approach to calibration substantially reduced run-time needed to locate optima. The hand calibrated and automated river flow optima were almost identical; for this particular only two parameters were significant. However, in most reaches the automated calibration improved upon the nitrate concentration optima by a factor of 3 to 6 (using a least squares objective function). An example, showing data, hand and automated calibrations is shown in Figure 1. Ammonium is disregarded for comparison purposes, as the hand-calibration did not seek to optimise this. Figure 2 shows plots of the most sensitive parameter against the

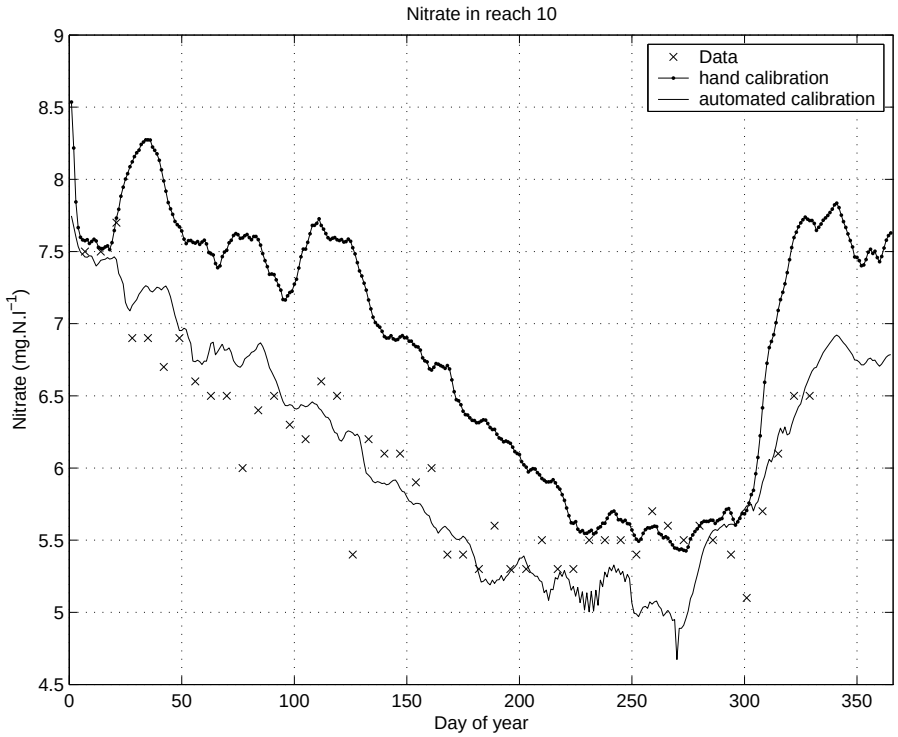

Figure 1. "Hand" versus automated calibration for nitrate in reach 10.

least squares nitrate objective function, for several reaches along the river. Figure 3 shows the combined posterior distribution of two parameters; groundwater drainage volume and initial groundwater nitrate concentration, conditioned on a data fit measure derived from least squares fits to measured in-river nitrate concentrations and flows. Equal importance was given to each reach, and the Nash-Sutcliffe efficiency measure used for normalisation purposes. A flat prior was used with a simple random walk Metropolis proposal function, and the chain was visited two million times. Quantitative convergence diagnostics have not yet been included, partly due to the debates surrounding the issue [**?**]. However, a preliminary diagnostic, dividing the chain into four sub-intervals, showed almost identical distributions. This, along with the long length of the chain, provides a strong argument for acceptance.

The influence of differing data fit measures on the above posterior distribution has also been examined. However, the presence of small instabilities in model responses for all the

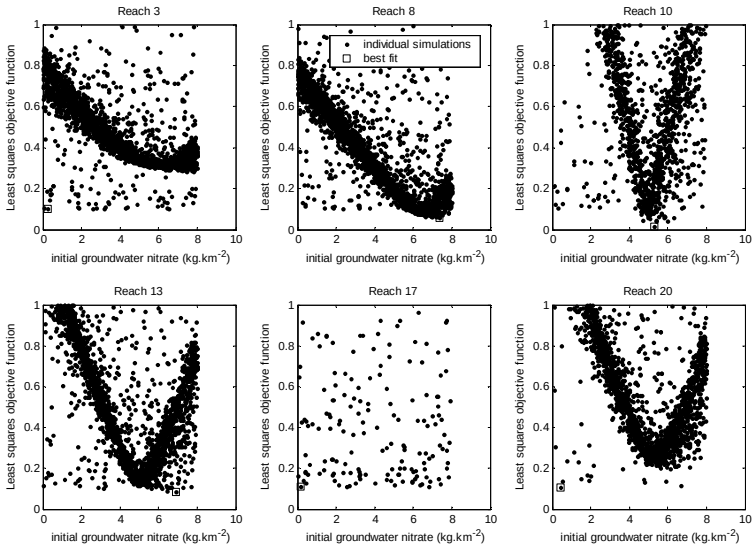

Figure 2. Response of in-river nitrate objective function against initial groundwater nitrate values over space.

posteriors somewhat obscured the biases caused by differing optima. This is continuing to be addressed as the model is refined.

6. CONCLUSIONS

A framework for stochastic analysis of catchment scale modelling scenarios, utilising Markov chain Monte Carlo along with other methodologies, has been developed and integrated with an existing nitrogen in catchments model (INCA).

It has become evident that, while MCMC provides an efficient means of investigating various conditional distributions and model responses, its relevance as an aid to developing models is limited unless it is complemented with cruder, set search methods and tools to identify structural problems and problematic parameter sets. This is a consequence of its reliance, in this context, on heuristic searches, which demand a certain degree of smoothness within the explored response surfaces.

The framework presented here is being extended to include a range of multi-variate analysis tools to investigate and isolate non-viable parameter combinations and structural issues. It is our view that this will be a useful, and arguably necessary, aid to any environmental model that has not already undergone substantial automated testing in its current form.

Implementation of the Markov chains produced successful results in parameter regions with stable model response, and has demonstrated the ability of the Metropolis

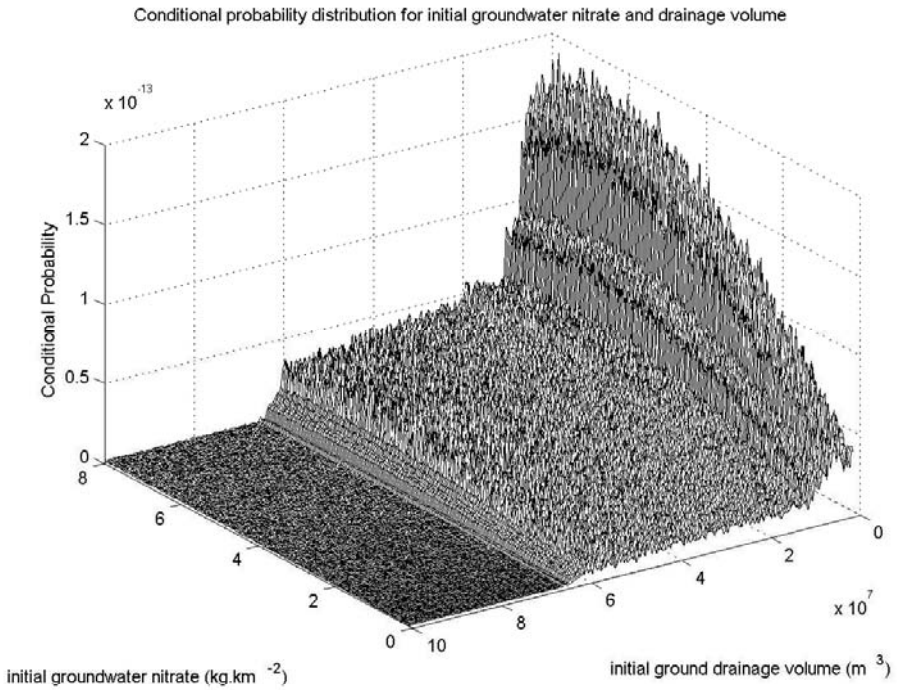

Figure 3. Posterior distribution of initial groundwater nitrate and drainage volume, conditional upon normalised least squares in-river flow and nitrate levels.

Hastings scheme to efficiently recover conditional distributions given appropriate data. To further test its efficiency against other methodologies, and parameterise its heuristics appropriately, model robustness is being tested and improved, and further information on parameter interactions investigated.

Further work is seeking to extend the robustness of the INCA model under automated calibration, through model component changes and further constraints upon parameters, and investigations into how conditional distributions are affected by aspects of model response and the optimality criteria imposed upon an analysis are ongoing. Proposal functions constructed with the aim of allowing sampling to be efficiently weighted towards subsets of the distribution are also being examined.

ACKNOWLEDGMENTS

This work was supported by two United Kingdom bodies: the National Environment Research Council, as part of its safety lowland catchment research programme (LOCAR), and the Environment Agency. Their support is gratefully acknowledged.

REFERENCES

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

2. N. McIntyre, B. M. Jackson, H. S. Wheater, and S. Chapra. Numerical efficiency in Monte Carlo simulations- a case study of a river thermodynamic model. *Journal of Environmental Engineering, ASCE*, (in press).
3. G. S. Fishman. *Monte Carlo: concepts, algorithms, and applications*. Springer Verlag, New York, 1996.
4. R. M. Neal. Probabilistic inference using Markov Chain Monte Carlo methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, 1993.
5. W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.
6. G. Kuczera and E. Parent. Monte-Carlo assessment of parameter uncertainty in conceptual catchment models. *Journal of Hydrology*, 211 (1-4):69–85, 1998.
7. C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 1999.
8. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.
9. Q. Duan, S. Soorooshian, and V. Gupta. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28(4):1015–1031, 1992.
10. K. J. Beven. Calibration, validation and equifinality in hydrological modelling. In M. G. Anderson and P. D. Bates, editors, *Validation in Hydrological Modelling*. John Wiley and Sons, Chichester, U.K., 2001.
11. K. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, 1999.
12. P. G. Whitehead, E. J. Wilson, and D. Butterfield. A semi-distributed Integrated Nitrogen model for multiple source assessment in CAtchments (INCA): Part I-model structure and process equations. *Sci. Tot. Environ.*, 210/211:547–558, 1998.