



Technical note: The CREDIBLE Uncertainty Estimation (CURE) toolbox: facilitating the communication of epistemic uncertainty

Trevor Page¹, Paul Smith^{1,2}, Keith Beven¹, Francesca Pianosi³, Fanny Sarrazin⁴, Susana Almeida⁵, Liz Holcombe³, Jim Freer^{6,7}, Nick Chappell¹, and Thorsten Wagener⁸

¹Lancaster Environment Centre, Lancaster University, Lancaster, UK

²waternumbers, Lancaster, UK

³Department of Civil Engineering, Bristol University, Bristol, UK

⁴Department of Computational Hydrosystems, Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany

⁵Atkins Global, Warrington, UK

⁶School of Geographical Sciences, Bristol University, Bristol, UK

⁷Global Institute for Water Security, University of Saskatchewan, Saskatchewan, Canada

⁸Institute for Environmental Science and Geography, University of Potsdam, Potsdam, Germany

Correspondence: Keith Beven (k.beven@lancaster.ac.uk)

Received: 7 October 2022 – Discussion started: 17 November 2022

Revised: 23 March 2023 – Accepted: 29 May 2023 – Published: 11 July 2023

Abstract. There is a general trend toward the increasing inclusion of uncertainty estimation in the environmental modelling domain. We present the Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE) Uncertainty Estimation (CURE) toolbox, an open-source MATLABTM toolbox for uncertainty estimation aimed at scientists and practitioners who are not necessarily experts in uncertainty estimation. The toolbox focusses on environmental simulation models and, hence, employs a range of different Monte Carlo methods for forward and conditioned uncertainty estimation. The methods included span both formal statistical and informal approaches, which are demonstrated using a range of modelling applications set up as workflow scripts. The workflow scripts provide examples of how to utilize toolbox functions for a variety of modelling applications and, hence, aid the user in defining their own workflow; additional help is provided by extensively commented code. The toolbox implementation aims to increase the uptake of uncertainty estimation methods within a framework designed to be open and explicit in a way that tries to represent best practice with respect to applying the methods included. Best practice with respect to the evaluation of modelling assumptions and choices, specifically including epistemic uncertainties, is also included by the incorporation of a condition tree that

allows users to record assumptions and choices made as an audit trail log.

1 Introduction

Environmental simulation models are used extensively for research and environmental management. There is a general trend toward the increasing inclusion of uncertainty estimation (UE) in the environmental modelling domain, including applications used in decision-making (Alexandrov et al., 2011; Ascough et al., 2008). Effective use of model estimates in decision-making requires a level of confidence to be established (Bennett et al., 2013), and UE is one element of determining this. Another required element is an assessment of the conditionality of any UE, i.e. the conditionality associated with the implicit and explicit choices and assumptions made during the modelling and UE process, given the information available (e.g. Rougier and Beven, 2013).

Here, we present the Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE) Uncertainty Estimation (CURE) toolbox, an open-source MATLABTM toolbox for UE associated with environmental simulation models. It is aimed at scientists and practitioners with some modelling experience

who are not necessarily experts in UE. The toolbox structure is similar to that of the SAFE (Sensitivity Analysis For Everybody) toolbox (Pianosi et al., 2016) in that it allows more experienced users to modify and enhance the code and to add new UE methods. The implementation of the toolbox also aims to increase the uptake of UE methods within a framework designed to be open and explicit in a way that tries to represent best practice; more specifically, we refer to best practice with respect to applying the various UE methods included as well as best practice with respect to being explicit about modelling choices and assumptions.

As the focus of the toolbox is UE for simulation models, often with relatively complex structures and many model parameters, the toolbox employs a range of different Monte Carlo methods. These are used for the forward propagation of uncertainties by sampling from input and parameter distributions defined a priori, for forward UE, or for the estimation of refined model structures and/or associated posterior parameter distributions when conditioned on observations (conditioned UE). The methods included span both formal statistical and informal approaches to UE, which are demonstrated using a range of modelling applications set up as workflow scripts that provide examples of how to utilize toolbox functions. As noted in the comments in the code, many of the workflows can be linked to the description of methods in Beven (2009).

Formal statistical and informal methods are included because there are no commonly agreed upon techniques for UE in environmental modelling applications, as evidenced by continuing debates and disputes in the literature (e.g. Clark et al., 2011; Beven et al., 2012; Beven, 2016; Nearing et al., 2016). The lack of a consensus on the most appropriate UE method is to be expected given that the sources of uncertainty associated with environmental modelling applications are dominated by a lack of knowledge (epistemic uncertainties; e.g. Refsgaard et al., 2007; Beven, 2009, 2016; Beven and Lane, 2022) rather than solely by random variability (aleatory uncertainties). Rigorous statistical inference applies to the latter, but it might lead to unwarranted confidence if applied to the former, especially where some data might be disinformative in model evaluation (e.g. Beven and Westenberg, 2011; Beven and Smith, 2015; Beven, 2019; Beven and Lane, 2022).

Assessing the impact of epistemic uncertainties for environmental modelling requires assumptions about their nature (which are difficult to define); thus, the output from any UE will be conditional upon these assumptions. This poses the question of what is good practice with respect to evaluating assumptions and choices made during the modelling process and what is good practice with respect to communicating the meaning of any subsequent analyses (Walker et al., 2003; Sutherland et al., 2013; Beven et al., 2018b; see also the TRACE (TRANSPARENT and Comprehensive Ecological model documentation) framework of Grimm et al., 2014, for documentation on the modelling process). Beven and Al-

cock (2012) suggest a condition tree approach that records the modelling choices and assumptions made during analyses and, thus, provides a clear audit trail (e.g. Beven et al., 2014; Beven and Lane, 2022). The audit trail consequently provides a vehicle that promotes transparency, best practice and communication with stakeholders (Refsgaard et al., 2007; Beven and Alcock, 2012). To encourage best practice, the process of defining a condition tree and recording an audit trail has been made an integral part of the CURE toolbox via a condition tree graphical user interface (GUI).

Other freely available toolboxes for forward UE include the Data Uncertainty Engine (DUE; http://harmonirib.geus.info/due_download/index.html, last access: 2 July 2023; Brown and Heuvelink, 2007) and the SIMLAB toolbox (<https://ec.europa.eu/jrc/en/samo/simlab>, last access: 2 July 2023; Saltelli et al., 2004). For conditioned UE, the following freely available toolboxes are accessible: GLUEWIN (https://joint-research-centre.ec.europa.eu/macro-econometric-and-statistical-software/econometric-software/gluewin_en, last access: 2 July 2023; Ratto and Saltelli, 2001), UCODE 2014 (<http://igwmc.mines.edu/freeware/ucode>, last access: 2 July 2023; Poeter et al., 2014), the MATLABTM Framework for Uncertainty Quantification (UQLAB; <http://www.uqlab.com>, last access: 2 July 2023; Wagener and Kollat, 2007), the Interactive Probabilistic Predictions software (<http://www.probablisticpredictions.org>, last access: 2 July 2023; McInerney et al., 2018) and the DREAM toolbox (http://faculty.sites.uci.edu/jasper/files/2015/03/manual_DREAM.pdf, last access: 2 July 2023; Vrugt et al., 2008, 2009; Vrugt, 2016). The reader is also referred to the broader review of uncertainty tools undertaken by the Uncertainty Enabled Model Web (UNCERTWEB) European research project (Bastin et al., 2013), which includes tools supporting elicitation, visualization, uncertainty and sensitivity analysis. While links exist for these toolboxes, it is not clear if all continue to be maintained and supported. The CURE toolbox presented here is open source and brings together formal and informal modelling methodologies, underpinned by different philosophies, that users are encouraged to explore via the example workflows (Table 1). It also offers a method not included in previous toolboxes (i.e. the coupled generalized likelihood uncertainty estimation and limits of acceptability, GLUE-LoA, method; Beven, 2006; Blazkova and Beven, 2009; Hollaway et al., 2018; Beven et al., 2022) and explicitly sets out to encourage best practice regarding the conditionality of modelling results using the condition tree approach.

2 Choosing a workflow

Table 1 lists the example workflows included in the first release of the CURE toolbox and the methods employed, with references to published papers where the methods have been

Table 1. Toolbox workflow examples and the uncertainty estimation methods employed.

Workflow	Uncertainty estimation method	Example model	Short description
1	Forward (independent sampling)	CHASM ^{1,b}	Application of uniform sampling of statistical distributions
2	Forward (independent sampling)	HYMOD ²	Fuzzy parameter distributions
3	Forward (dependent sampling – copula)	HYMOD	Marginal gamma distributions and rank-correlation-derived copula
4	Forward ^a (dependent sampling – copula)	LISFLOOD ^{3,b}	Application of covariant model ⁴ fitted as copula for inflows to LISFLOOD
5	Conditioned; adaptive metropolis MCMC ^{4,5,6}	HYMOD	Single-chain MCMC; formal likelihood
6	Conditioned; DREAM ⁷	HYMOD	Multi-chain MCMC; formal likelihood
7	Conditioned; DREAM ⁸	HYMOD	Multi-chain MCMC with thresholding of informal likelihood measure
8	Conditioned; DREAM	PROTECH ^{4,b}	Multi-chain MCMC using thresholding of informal likelihood measure
9	Conditioned; GLUE ⁹	HYMOD	GLUE using threshold of informal likelihood measure
10	Conditioned; GLUE	PROTECH ^b	GLUE using threshold of informal likelihood measure.
11	Conditioned; GLUE-LoA	HYMOD	GLUE using a single variable limits of acceptability
12	Conditioned; GLUE-LoA ¹⁰	PROTECH ^b	GLUE using multi-variable limits of acceptability
13	Analysis of rainfall–runoff observations	Event analysis ¹¹	Derivation of limits of acceptability based on event runoff coefficients
14	Conditioned; GLUE-LoA	Dynamic TOPMODEL ¹¹	GLUE using limits of acceptability based on runoff coefficients

^a In this example, the input was sampled in a forward uncertainty analysis, but the LISFLOOD model was conditioned in a prior analysis. ¹ Almeida et al. (2017). ² Wagener et al. (2001). ³ Neal et al. (2013). ⁴ Haario et al. (2001). ⁵ Roberts and Rosenthal (2001). ⁶ Roberts and Rosenthal (2009). ⁷ Differential evolution adaptive metropolis (Vrugt, 2016). ⁸ Sadegh and Vrugt (2014). ⁹ Generalized likelihood uncertainty estimation (GLUE; Beven and Binley, 1992). ¹⁰ Blazkova and Beven (2009).

^b Owing to long model run times, this example uses pre-run simulation output. ¹¹ Beven et al. (2022).

applied. A variety of workflows covering forward UE and both formal statistical and informal methods of conditioned UE are given. Figure 1 provides an illustration of the choices that might be made in deciding on a workflow within the CURE toolbox (see also the earlier decision trees of this type

in Pappenberger et al., 2006, and Beven, 2009). Forward UE methods (workflows 1 and 2) must be used when there are no observational data with which to condition the model output. The outcomes will then be directly dependent on the assumptions about the prior distributions and covariation of

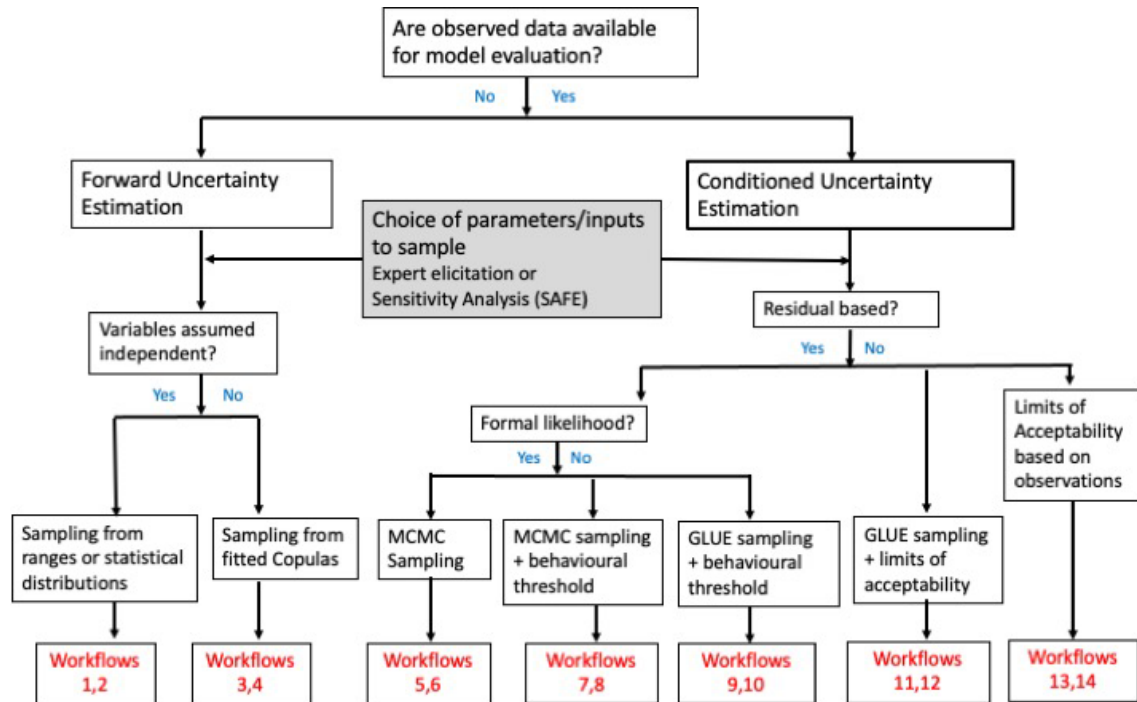


Figure 1. The decision tree guiding users towards different methodologies and workflows.

parameters and input variables. Copula methods are used to sample covariates (workflows 3 and 4). In the case of both forward and conditioned UE workflows, input uncertainties are parameterized to be applied as ranges or distributions, for example, as multipliers or an additive bias applied when the model is run.

When observational data are available, formal statistical likelihood methods (workflows 5 and 6) will be most appropriate in cases where any model residuals can be assumed to be aleatory and represented by a simple stochastic model. Where such assumptions are difficult to justify because of epistemic sources of uncertainty, there is a choice between approximate Bayesian estimation (ABC) using Markov chain Monte Carlo (MCMC) sampling and GLUE methods. Within ABC, a threshold of acceptability for some informal summary measure of performance is chosen. MCMC sampling is implemented using the DREAM code described in Vrugt (2015); the reader is also referred to Vrugt (2016) for a more recent description. This aims to produce an ensemble of model parameter sets comprising the samples from the final iterations of the DREAM algorithm (defined by the user) that are considered to be equally probable (workflows 7 and 8). Convergence of the sampling can be tested using the Gelman and Rubin (1992) diagnostic statistic.

Within GLUE each model is associated with a likelihood measure that initially reflects sampling of the assumed prior distributions and is then modified during the conditioning process. GLUE allows for different ways of updating the

likelihood measure, including both Bayesian multiplication and fuzzy operators (Beven and Binley, 1992, 2014). Uniform independent priors across specified ranges are often assumed when there is a lack of robust knowledge about the parameters but, as in the options for the forward UE workflows, other prior distributions can be used. Deciding on whether a model is acceptable or behavioural can again be based on some informal summary measure of performance (workflows 9 and 10) or some predefined limits of acceptability (workflows 11 and 12). A particular case of defining limits of acceptability for rainfall–runoff models based on historical event runoff coefficients as a way of reflecting epistemic uncertainties in observed input and output is included (workflows 13 and 14). Vrugt and Beven (2018) demonstrated an adaptive sampling methodology for applying the limits of acceptability ($DREAM_{(LoA)}$) that aims to find feasible samples that satisfy all of the limits applied. The DREAM algorithm used in workflows 7 and 8 can be adapted to be used in this way.

It should be noted that the examples associated with each workflow are intended to be illustrative. They cannot all be described in detail in this publication, which is intended to introduce the toolbox. However, the MATLABTM code is freely available and can be easily adapted by users for their own application. Extensive comments are included in each workflow to aid this process.

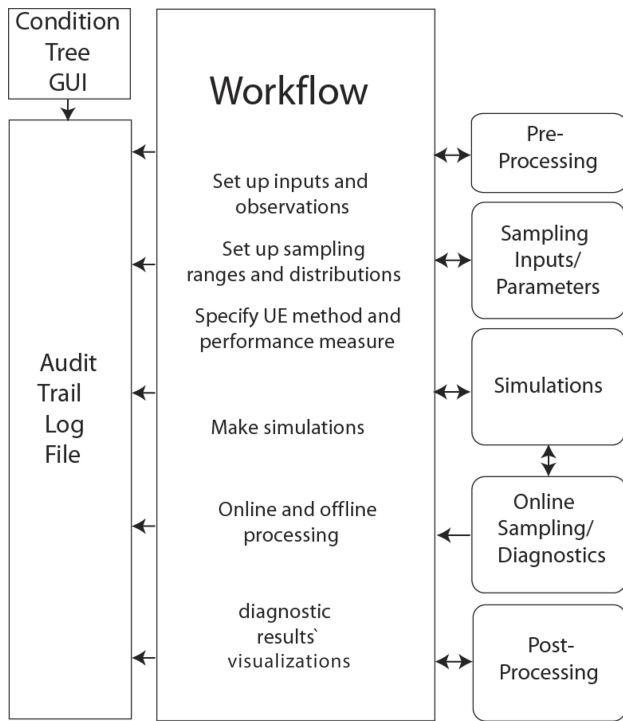


Figure 2. Overall structure of the CURE toolbox.

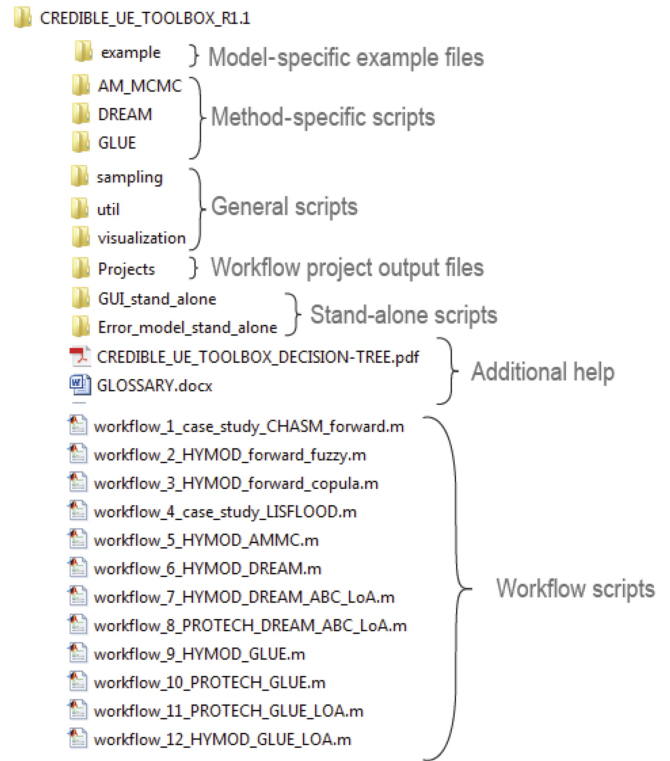


Figure 3. Outline of the folder structure of the CURE toolbox.

3 The CURE toolbox, version 1.0, structure

The CURE toolbox essentially has two linked structures. There is an overall structure with which the user interacts throughout the analysis (Fig. 2) and an underlying folder structure (Fig. 3) containing the toolbox functions and example model-specific files. The toolbox folder structure has specific folders for the UE methods where method-specific functions are collated (e.g. method-specific sampling and diagnostics and visualization) and for the individual example modelling applications (i.e. the model functions and input files as well as any links to any “external models”, such as models not coded as a MATLAB™ function but which can be executed from the command line). Folders also exist for general (i.e. not method-specific) sampling methods, visualizations and utility functions. Additionally, there are project folders for each example workflow where audit trail logs, diagnostics and results are written.

The functions for general sampling of parameter distributions (e.g. uniform, low-discrepancy or Latin hypercube sampling of the large number of supported distributions) are common with the SAFE toolbox of Pianosi et al. (2016). In addition, and of particular importance for forward uncertainty analysis, the sampling functions have been extended to represent parameter and forcing-input dependencies using copulas (e.g. Workflow 3 in Table 1 uses copula sampling based on results from previous analyses to describe parameter dependencies for forward uncertainty propaga-

tion). Other specific sampling functions are associated with the adaptive sampling (“online” sampling) for Markov chain Monte Carlo (MCMC) approaches, implemented using the DREAM algorithm of Vrugt (2016), where distributions and correlation structures are modified as the chain(s) evolve. Modelling diagnostics, both numerical and graphical, are provided for both online adaptive sampling and “offline” methods (i.e. those that are not adaptively sampled within a given method). In the case of online MCMC methods, visualization of the evolution of the states of the chain(s) and tests for convergence to stationary distributions are included (e.g. Fig. 4a, b).

In the case of formal statistical likelihood methods (e.g. Evin et al., 2013, 2014, and the recent “universal likelihood” of Vrugt et al., 2022), residual model fitting can be carried out interactively, using command line prompts, and can form part of a workflow (or can be used in a stand-alone manner). The approach uses Box–Cox transformations, which provide flexibility in transforming the data to remove heteroscedasticity and non-normality (Box and Cox, 1964), and also provides for fitting an autoregressive model of suitable order in an iterative way, as proposed by Beven et al. (2008). Figure 4c and d, for example, show the use of the residual model-fitting visualizations in Workflow 5. The visualizations also serve as an approximate check of the residual model assumptions when analysing posterior simulations.

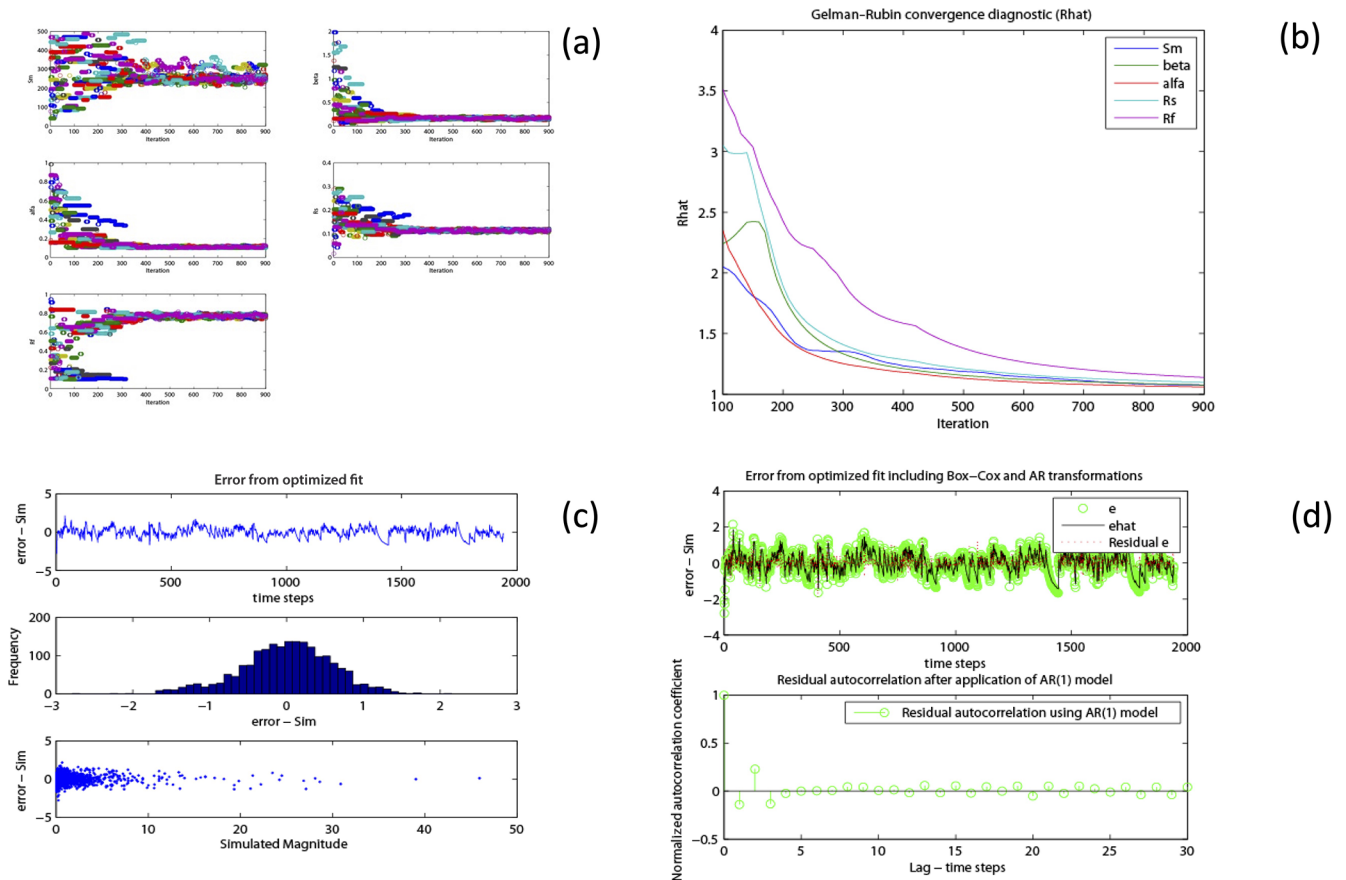


Figure 4. Visualization of simulation diagnostics in conditioning of HYMOD parameters in Workflow 5 using DREAM with a formal likelihood: (a) the evolution of 12 chains using DREAM, (b) the evolution of the Gelman–Rubin convergence statistic for five parameters and (c, d) the visualization of structural parameters during residual model fitting.

For the GLUE methods (see Beven and Binley, 1992, 2014; Beven and Freer, 2001; Beven et al., 2008; Beven and Lane, 2022), diagnostics are included for the exploration of the acceptable parameter space and which criterion (or criteria) and at which time steps (or locations) simulations were rejected. There are also method-specific and generic toolbox functions for the visualization and presentation of simulation results and associated uncertainties (e.g. see Fig. 5 for the application in Workflow 1). Results are both alphanumeric and graphical; alphanumeric results (including those from diagnostic statistics and summary variables where appropriate) can be automatically written to the audit trail log, and plots are saved to the project folder.

4 Condition tree implementation within CURE

An important part of any CURE toolbox application is the way that users can explore and document modelling choices, assumptions and uncertainties using the condition tree GUI (e.g. Fig. 6). The GUI aids in the elicitation of primary modelling uncertainties, their likely sources and how they are to

be treated during the analysis. It is also designed to elicit other important choices and assumptions, including those regarding elements of the analysis assumed to be associated with insignificant uncertainties and perhaps treated deterministically; for example, where only one model structure is considered or where uncertainties are assumed negligible for certain elements or are perhaps subsumed into other uncertain elements. Similar to the incorporation of UE, the condition tree would be completed, ideally, as an integral part of any modelling application and can help in the definition of an appropriate workflow structure. This is particularly important in considering epistemic sources of uncertainty. We fully understand that non-probabilistic approaches to UE remain controversial (e.g. Nearing et al., 2016) but have, in the past, demonstrated that the assumptions required to use formal statistical methods (e.g. the recent paper of Vrugt et al., 2022) may lead to overconfidence in the resulting inference when epistemic uncertainties are important (Beven and Smith, 2015; Beven, 2016). Because epistemic uncertainties are the result of a lack of knowledge, their nature and impacts cannot be defined easily. That means that, effectively, there

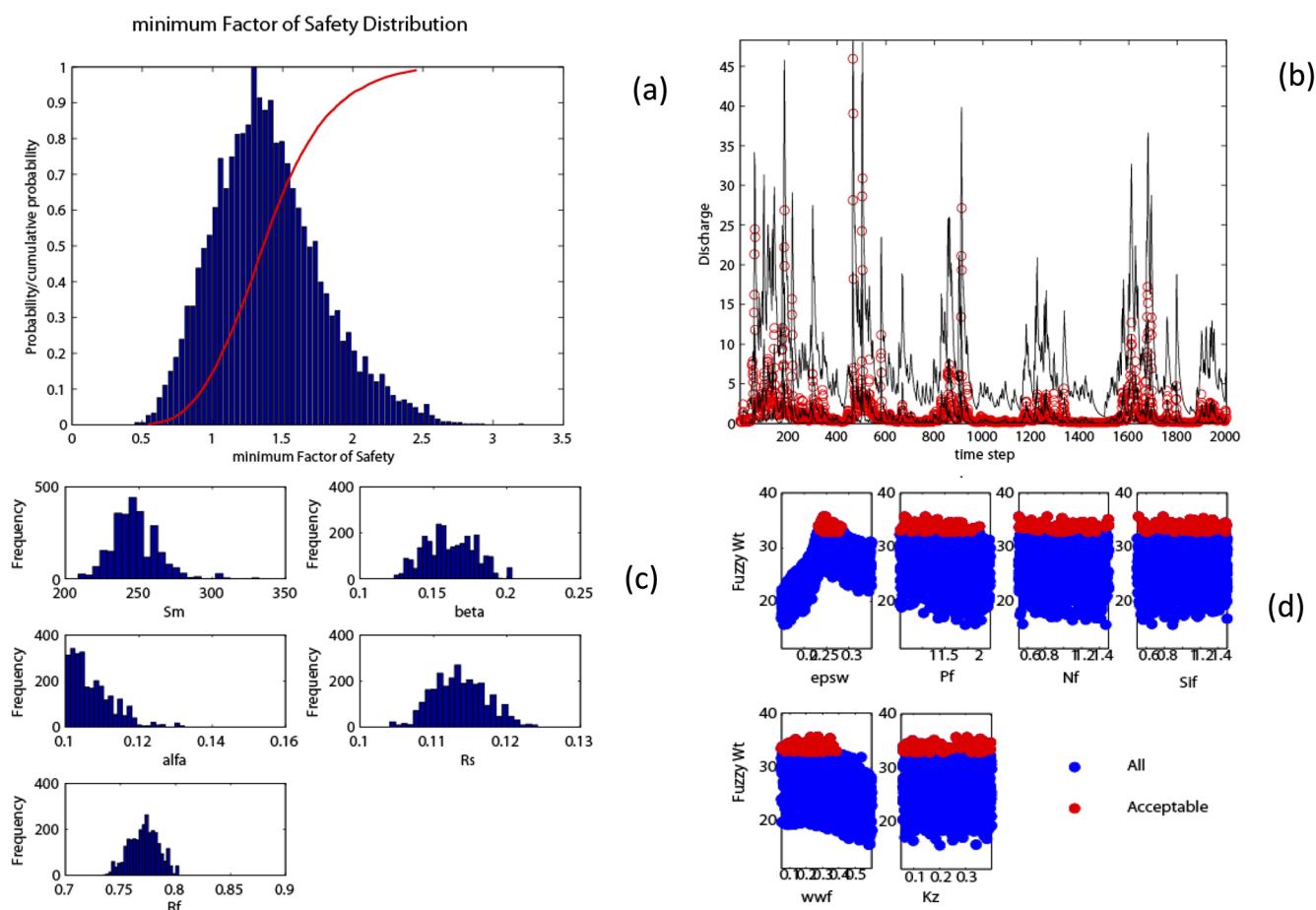


Figure 5. Visualization of results: (a) the distribution of the simulated minimum factor of safety from a forward UE using the CHASM landslide model in Workflow 1; (b) the 5th, 50th and 95th percentiles of simulated discharge (black lines) and observed discharge (red circles) from an MCMC-conditioned UE method using HYMOD; and (c) the posterior parameter distributions for the same CHASM example and (d), both showing all and acceptable parameter sets from a GLUE analysis using the PROTECH model (Workflow 10).

can be no right answer (e.g. Beven et al., 2018a, b; Beven and Lane, 2022); therefore, the recording of assumptions in the audit trail for analysis should be a requisite of any analysis to allow for later evaluation by others.

The GUI takes the form of a number of simple, sequential dialogue boxes in which the user is asked to enter text. In the initial release of the toolbox there are five primary dialogue boxes covering the following:

1. project aims and model(s)/model structures considered;
2. modelling uncertainties – overview, covering the model structure, parameters, input and observations for model conditioning;
3. uncertainties – observations for model conditioning – specific, covering the associated uncertainties and basis for assessing simulation performance;
4. uncertainties – input – specific, covering the sampling strategy, distributions and dependencies;

5. uncertainties – parameters – specific, covering the choice of parameters, sampling strategy, distributions and dependencies.

The information, elicited using the dialogue boxes, can be automatically written to the project audit trail log during the initial phase of entry; the audit trail log remains editable as the user defines their own workflow and during any subsequent modifications to the analysis contained within a workflow.

5 Defining a workflow

An a priori consideration of modelling uncertainties via the condition tree is an optional first step to help choose and structure an appropriate workflow. The decision tree in Fig. 1 can also be used as a guide in this respect. These are complemented by the toolbox documentation and help text, which are available via the workflows and functions. Documentation and help are in the form of targeted comments within

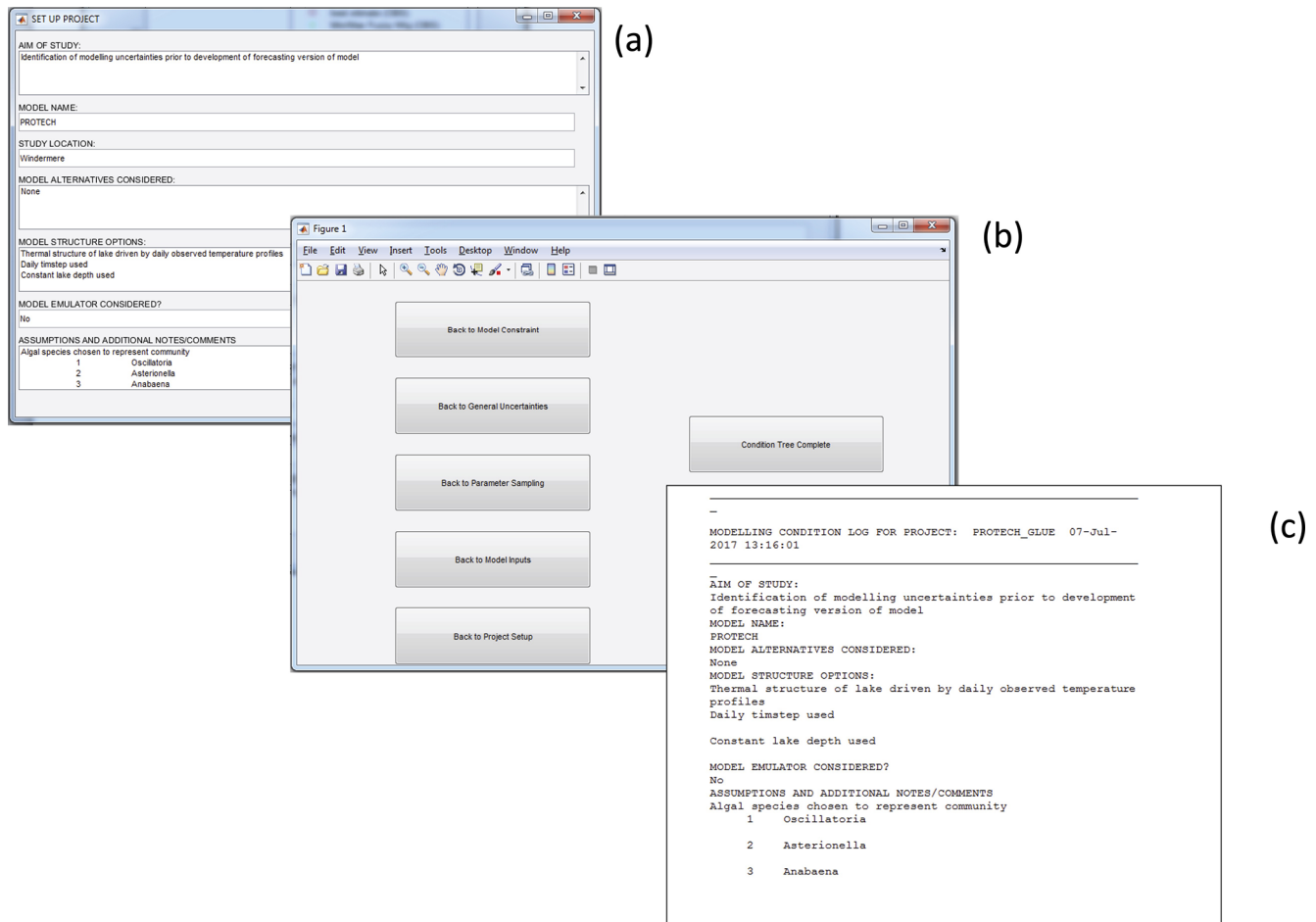


Figure 6. Condition tree example: GUI dialogue box for (a) the project set-up, (b) the condition tree navigation pane and (c) part of an example audit trail log.

the code, and function header text is available by typing the help “function name” in the command line (e.g. headers may include a definition of function variables and references for a specific UE method). Each workflow is also linked, where possible, to the relevant chapters of Beven (2009); these are specified in the header text of each workflow script. Clarification of the terminology used in the help and documentation is provided by a glossary of terms included as part of the toolbox.

It is assumed that the user has completed any necessary pre-processing analyses such as forcing-input uncertainty assessment and disinformation screening (e.g. Beven and Smith, 2015) as well as an assessment of uncertainties associated with conditioning observations, where used. An exception is the interactive toolbox facility for fitting residual models mentioned earlier, where formal statistical likelihoods are to be used.

The example workflows have been chosen to span the UE methods included in the toolbox and, in some cases, provide comparison of different UE methods for similar modelling

applications. The structure of the workflows themselves includes the primary steps to be “populated” that are outlined in the following.

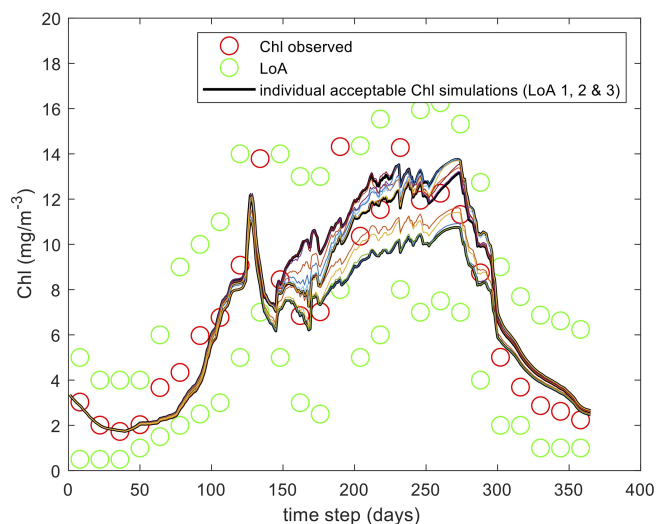
The condition tree GUI project set-up and interactive dialogue boxes are as follows:

1. set up input and observations;
2. set up parameter ranges, distributions and sampling strategy;
3. define performance measure (if conditioned UE);
4. simulations (online or offline; MATLABTM function or external model);
5. post-processing, including diagnostics, results, propagation and visualization of uncertainty.

Associated with these main steps, example workflows include automatic “text writes” that are appended to the audit trail log for each analysis. These include specific choices that are made when implementing steps 1–5 above, such as the

Table 2. Parameters and uniform distribution sampling ranges for the application of the PROTECH model to Lake Windermere (Workflow 12 application).

Parameter	Meaning	Min	Max
EPSW	Background light extinction coefficient	0.15	0.35
P_f	Growth rate factor for phosphorus	0.5	2.5
N_f	Growth rate factor for nitrate	0.5	1.5
Si_f	Growth rate factor for silica	0.5	1.5
K_z	Vertical effective eddy diffusion coefficient	0.05	0.4
WW_f	Waste water treatment works adjustment factor for phosphorus	0.05	0.6

**Figure 7.** PROTECH application to the Lake Windermere example in Workflow 12: observed chlorophyll data (red circles), limits of acceptability (green circles), and predictions of models that satisfy all of the chlorophyll, R-type and CS-type algae limits.

ranges of parameter values used and their distributions, the sampling strategy employed, and diagnostic and simulation results.

In general, users will not need to modify any toolbox functions; they will only need to build a workflow. However, given the requirement for online simulation performance to be assessed for MCMC methods as well as the many permutations of performance measures and ways of combining them where multiple criteria are used, users are also required to specify the function that returns an overall measure of individual simulation performance. In addition, where external models are to be used for online approaches, additional modifications may be required for modification of input/parameter files using some form of wrapper code.

6 An example workflow

The CURE workflows can be applied to a wide range of geoscience applications, including the water science examples

set out in Table 1. In particular, CURE is well suited to the specification of assumptions about epistemic uncertainties, to conditioning using uncertain observational data and to rejectionist approaches to model evaluation (see also Beven et al., 2018a, b, 2022; Beven and Lane, 2022). Here, we provide some more detail on the application of the PROTECH model within such a multi-variable rejectionist conditioning framework (Workflow 12 in Table 1). The full workflow and output are given in the Supplement.

PROTECH is a lake algal community model that has been applied to predict concentrations for functional classes of algae in Lake Windermere in Cumbria, UK (Page et al., 2017). It is a 1D model with water volumes related to the lake bathymetry and runs with a daily time step. In this case, the model is provided in an executable form and was run off-line for randomly sampled parameter sets; thus, the workflow takes the simulated output files as input. The model requires flow, weather and nutrient information as input. A reduced set of six parameters was sampled, as in Table 2 (see Page et al., 2017, for a more complete analysis). Model evaluation is based on limits of acceptability for three variables: chlorophyll and the concentrations of R-type and CS-type algae. Figure 7 shows the resulting chlorophyll output for the surviving models from the analysis after evaluation against all three sets of limits of acceptability. The full workflow and resulting audit trail and output figures are presented in the Supplement.

7 Toolbox evolution

The toolbox structure is such that new methods can be easily added, and it will be subject to ongoing development and augmentation with additional workflow examples. It is hoped that the CURE toolbox will contribute to the ongoing development and testing of UE methods and good practice with respect to their application. In particular, the condition tree approach could be further developed via feedback from toolbox users and end users of the conditional uncertainty estimates. The toolbox is freely available for non-commercial research and education from <https://www.lancaster.ac.uk/lec/sites/qnfm/credible> (Page et al., 2021).

Code availability. The CURE MATLAB™ toolbox, version 1.0, is an open-source MATLAB™ code hosted at Lancaster University. It can be downloaded from <https://www.lancaster.ac.uk/lec/sites/qnfm/credible> (Page et al., 2021) (contact n.chappell@lancaster.ac.uk) and was first made available in 2021.

Data availability. All of the data needed to run the examples contained in this paper are supplied with the CURE toolbox, version 1.0 (Page et al., 2021).

Author contributions. TP and KB were involved in the conceptualization of the CURE toolbox and the development of the example applications. TP, PS, FP and FS were responsible for the software development. TP and KB wrote the original draft of the paper. All of the other authors were involved in the development of applications within the Natural Environment Research Council CREDIBLE project, which was the original motivation for the development of the SAFE and CURE toolboxes, and in reviewing and editing the paper. TW was the principal investigator of CREDIBLE. The use of the CURE in the NERC Q-NFM project was supported by Nick Chappell, who also established the CURE website.

Competing interests. At least one of the (co-)authors is a member of the editorial board of *Hydrology and Earth System Sciences*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. This work was supported by the Natural Environment Research Council (NERC) "Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation" project (CREDIBLE; grant no. NE/J017450/1) and the NERC "Quantifying the likely magnitude of nature-based flood mitigation effects across large catchments" project (Q-NFM; grant no. NE/R004722/1).

Financial support. This research has been supported by the Natural Environment Research Council (grant nos. NE/J017299/1 and NE/J017450/1).

Review statement. This paper was edited by Fabrizio Fenicia and reviewed by Tobias Krueger and one anonymous referee.

References

- Alexandrov, G. A., Ames, D., Bellocchi, G., Bruen, M., Crout, N., Erechchoukova, M., Hildebrandt, A., Hoffman, F., Jackisch, C., Khaiteer, P., Mannina, G., Matsunaga, T., Purucker, S. T., Rivington, M., and Samaniego, L.: Technical assessment and evaluation of environmental models and software: letter to the Editor, *Environ. Model. Softw.*, 26, 328–336, 2011.
- Almeida, S., Holcombe, E. A., Pianosi, F., and Wagener, T.: Dealing with deep uncertainties in landslide modelling for disaster risk reduction under climate change, *Nat. Hazards Earth Syst. Sci.*, 17, 225–241, <https://doi.org/10.5194/nhess-17-225-2017>, 2017.
- Ascough II, J. C., Maier, H. R., Ravalico, J. K., and Strudley, M. W.: Future research challenges for incorporation of uncertainty in environmental and ecological decision-making, *Ecol. Model.*, 219, 383–399, <https://doi.org/10.1016/j.ecolmodel.2008.07.015>, 2008.
- Bastin, L., Cornford, D., Jones, R., Heuvelink, G. B. M., Pebesma, E., Stasch, C., Nativi, S., Mazzetti, P., and Williams, M.: Managing uncertainty in integrated environmental modelling: The UncertWeb framework, *Environ. Model. Softw.*, 39, 116–134, <https://doi.org/10.1016/j.envsoft.2012.02.008>, 2013.
- Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Sempelt, R., Voinov, A. A., Fath, B. D., and Andreassian, V.: Characterising performance of environmental models, *Environ. Model. Softw.*, 40, 1–20, <https://doi.org/10.1016/j.envsoft.2012.09.011>, 2013.
- Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, 6, 279–298, 1992.
- Beven, K., Smith, P., Westerberg, I., and Freer, J.: Comment on "Pursuing the method of multiple working hypotheses for hydrological modeling" by P. Clark et al., *Water Resour. Res.*, 48, 1–5, 2012.
- Beven, K. J.: A manifesto for the equifinality thesis, *J. Hydrol.*, 320, 18–36, 2006.
- Beven, K. J., *Environment Modelling: An Uncertain Future?* Routledge: London, 2009.
- Beven, K. J.: EGU Leonardo Lecture: Facets of Hydrology – epistemic error, non-stationarity, likelihood, hypothesis testing, and communication, *Hydrolog. Sci. J.*, 61, 1652–1665, <https://doi.org/10.1080/02626667.2015.1031761>, 2015.
- Beven, K. J. and Alcock, R.: Modelling everything everywhere: a new approach to decision making for water management under uncertainty, *Freshwater Biol.*, 56, 124–132, <https://doi.org/10.1111/j.1365-2427.2011.02592.x>, 2012.
- Beven, K. J. and Binley, A. M.: GLUE, 20 years on, *Hydrol. Process.*, 28, 5897–5918, <https://doi.org/10.1002/hyp.10082>, 2014.
- Beven, K. J. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems, *J. Hydrol.*, 249, 11–29, 2001.
- Beven, K. J. and Lane, S.: On (in)validating environmental models. 1. Principles for formulating a Turing-like Test for determining when a model is fit-for purpose, *Hydrol. Process.*, 36, e14704, <https://doi.org/10.1002/hyp.14704>, 2022.
- Beven, K. J. and Smith, P. J.: Concepts of Information Content and Likelihood in Parameter Calibration for Hydrologi-

- cal Simulation Models, *ASCE J. Hydrol. Eng.*, 20, A4014010, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000991](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000991), 2015.
- Beven, K. J. and Westerberg, I.: On red herrings and real herrings: disinformation and information in hydrological inference, *Hydrol. Process.*, 25, 1676–1680, <https://doi.org/10.1002/hyp.7963>, 2011.
- Beven, K. J., Smith, P. J., and Freer, J. E.: So just why would a modeller choose to be incoherent?, *J. Hydrol.*, 354, 15–32, <https://doi.org/10.1016/j.jhydrol.2008.02.007>, 2008.
- Beven, K. J., Lamb, R., Leedal, D. T., and Hunter, N.: Communicating uncertainty in flood risk mapping: a case study, *Int. J. River Basin Manage.*, 13, 285–296, <https://doi.org/10.1080/15715124.2014.917318>, 2014.
- Beven, K. J., Almeida, S., Aspinall, W. P., Bates, P. D., Blazkova, S., Borgomeo, E., Freer, J., Goda, K., Hall, J. W., Phillips, J. C., Simpson, M., Smith, P. J., Stephenson, D. B., Wagener, T., Watson, M., and Wilkins, K. L.: Epistemic uncertainties and natural hazard risk assessment. 1. A review of different natural hazard areas, *Nat. Hazards Earth Syst. Sci.*, 18, 2741–2768, <https://doi.org/10.5194/nhess-18-2741-2018>, 2018a.
- Beven, K. J., Aspinall, W. P., Bates, P. D., Borgomeo, E., Goda, K., Hall, J. W., Page, T., Phillips, J. C., Simpson, M., Smith, P. J., Wagener, T., and Watson, M.: Epistemic uncertainties and natural hazard risk assessment – Part 2: What should constitute good practice?, *Nat. Hazards Earth Syst. Sci.*, 18, 2769–2783, <https://doi.org/10.5194/nhess-18-2769-2018>, 2018b.
- Beven, K. J., Lane, S., Page, T., Hankin, B., Kretzschmar, A., Smith, P. J., and Chappell, N.: On (in)validating environmental models. 2. Implementation of the Turing-like Test to modelling hydrological processes, *Hydrol. Process.*, 36, e14703, <https://doi.org/10.1002/hyp.14703>, 2022.
- Blazkova, S. and Beven, K. J.: A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, *Water Resour. Res.*, 45, W00B16, <https://doi.org/10.1029/2007WR006726>, 2009.
- Box, G. E. P. and Cox, D. R.: An analysis of transformations, *J. Royal Stat. Soc. Ser. B*, 26, 211–252, 1964.
- Brown, J. D. and Heuvelink, G. B. M.: The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables, *Comput. Geosci.*, 33, 172–190, 2007.
- Clark, M. P., Kavetski, D., and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, 47, W09301, <https://doi.org/10.1029/2010WR009827>, 2011.
- Evin, G., Kavetski, D., Thyer, M., and Kuczera, G.: Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration, *Water Resour. Res.*, 49, 4518–4524, 2013.
- Evin, G., Thyer, M., Kavetski, D., McInerney, D., and Kuczera, G.: Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity, *Water Resour. Res.*, 50, 2350–2375, 2014.
- Gelman, A. and Rubin, D. B.: Inference from iterative simulation using multiple sequences (with discussion), *Statist. Sci.*, 7, 457–472, 1992.
- Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S. A., Liu, C., Martin, B. T., Meli, M., Radchuk, V., Thorbek, P., and Railsback, S. F.: Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE, *Ecol. Model.*, 280, 129–139, <https://doi.org/10.1016/j.ecolmodel.2014.01.018>, 2014.
- Haario, H., Saksman, E., and Tamminen, J.: An adaptive Metropolis algorithm, *Bernoulli*, 7, 223–242, 2001.
- Hollaway, M. J., Beven, K. J., Benskin, C. M. W. H., Collins, A. L., Evans, R., Falloon, P. D., Forber, K. J., Hiscock, K. M., Kahana, R., Macleod, C. J. A., Ockenden, M. C., Villamizar, M. L., Wearing, C., Withers, P. J. A., Zhou, J. G., and Haygarth, P. M.: Evaluating a processed based water quality model on a UK headwater catchment: what can we learn from a ‘limits of acceptability’ uncertainty framework?, *J. Hydrol.*, 558, 607–624, <https://doi.org/10.1016/j.jhydrol.2018.01.063>, 2018.
- McInerney, D., Thyer, M., Kavetski, D., Bennett, B., Lerat, J., Gibbs, M., and Kuczera, G.: A simplified approach to produce probabilistic hydrological model predictions, *Environ. Model. Softw.*, 109, 306–314, 2018.
- Neal, J., Keef, C., Bates, P., Beven, K. J., and Leedal, D.: Probabilistic flood risk mapping including spatial dependence, *Hydrol. Process.*, 27, 1349–1363, <https://doi.org/10.1002/hyp.9572>, 2013.
- Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., and Weijs, S. V.: A Philosophical Basis for Hydrologic Uncertainty, *Hydrolog. Sci. J.*, 61, 1666–1678, <https://doi.org/10.1080/02626667.2016.1183009>, 2016.
- Page, T., Smith, P. J., Beven, K. J., Jones, I. D., Elliott, J. A., Maberly, S. C., Mackay, E. B., De Ville, M., and Feuchtmayr, H.: Constraining uncertainty and process-representation in an algal community lake model using high frequency in-lake observations, *Ecol. Modell.*, 357, 1–13, <https://doi.org/10.1016/j.ecolmodel.2017.04.011>, 2017.
- Page, T., Smith, P. J., Beven, K. J., Pianosi, F., Sarrazin, F., Almeida, S., Holcombe, E., Freer, J., Chappell, N., and Wagener, T.: The CURE Uncertainty Estimation Matlab Toolbox, Version 1.0, <https://www.lancaster.ac.uk/lec/sites/qnfm/credible> (last access: 3 July 2023), 2021.
- Pappenberger, F., Harvey, H., Beven, K. J., Hall, J., and Meadowcroft, I.: Decision tree for choosing an uncertainty analysis methodology: a wiki experiment <http://www.floodrisknet.org.uk/methods> <http://www.floodrisk.net>, *Hydrol. Process.*, 20, 1099–1085, <https://doi.org/10.1002/hyp.6541>, 2006.
- Pianosi, F., Rougier, J., Freer, J., Hall, J., Stephenson, D. B., Beven, K. J., and Wagener, T.: Sensitivity Analysis of environmental models: a systematic review with practical workflows, *Environ. Model. Softw.*, 79, 214–232, 2016.
- Poeter, E. P., Hill, M. C., Lu, D., Tiedeman, C. R., and Mehl, S.: UCODE_2014, with new capabilities to define parameters unique to predictions, calculate weights using simulated values, estimate parameters with SVD, evaluate uncertainty with MCMC, and More: Integrated Groundwater Modeling Center Report Number GWMI 2014-02, <https://pubs.er.usgs.gov/publication/70159674> (last access: 3 July 2023), Colorado, USA, 2014.
- Ratto, M. and Saltelli, A.: Model assessment in integrated procedures for environmental impact evaluation: software prototypes, *GLUEWIN User’s Manual, Estimation of human impact in the presence of natural fluctuations (IMPACT)*, Deliverable 18. Joint Research Centre of European Commission (JRC), Institute for

- the Protection and Security of the Citizen (ISIS), Ispra, Italy, 2001.
- Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., and Vanrolleghem, P. A.: Uncertainty in the environmental modelling process – A framework and guidance, *Environ. Model. Softw.*, 22, 1543–1556, 2007.
- Roberts, G. O. and Rosenthal, J. S.: Optimal scaling for various Metropolis-Hastings algorithms, *Statist. Sci.*, 16, 351–367, <https://doi.org/10.1214/ss/1015346320>, 2001.
- Roberts, G. O. and Rosenthal, J. S.: Examples of Adaptive MCMC, *J. Comput. Graph. Stat.*, 18, 349–367, <https://doi.org/10.1198/jcgs.2009.06134>, 2009.
- Rougier, J. and Beven, K. J.: Model limitations: the sources and implications of epistemic uncertainty, in: Risk and uncertainty assessment for natural hazards, edited by: Rougier, J., Sparks, S., and Hill, L., Cambridge University Press, Cambridge, UK, 40–63, <https://doi.org/10.1017/CBO9781139047562.004>, 2013.
- Sadegh, M. and Vrugt, J. A.: Approximate Bayesian computation using Markov Chain Monte Carlo simulation: DREAM(ABC), *Water Resour. Res.*, 50, 6767–6787, <https://doi.org/10.1002/2014WR015386>, 2014.
- Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M.: How to Use SIMLAB in Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, Wiley, ISBN 0-470-87093-1, 2004.
- Sutherland, W. J., Spiegelhalter, D., and Burgman, M. A.: Twenty tips for interpreting scientific claims, *Nature*, 503, 335–337, 2013.
- Vrugt, J.: Markov chain Monte Carlo Simulation Using the DREAM Software Package: Theory, Concepts, and MATLAB Implementation, https://bpb-us-e2.wpmucdn/faculty.sites.uci.edu/dist/f/94/files/2015/03/manual_DREAM.pdf (last access: 3 July 2023), 2015.
- Vrugt, J. A.: Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation, *Environ. Model. Softw.*, 75, 273–316, <https://doi.org/10.1016/j.envsoft.2015.08.013>, 2016.
- Vrugt, J. A. and Beven, K. J.: Embracing Equifinality with Efficiency: Limits of Acceptability Sampling Using the DREAM_(LOA) algorithm, *J. Hydrol.*, 559, 954–971, 2018.
- Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson B. A.: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, 44, W00B09, <https://doi.org/10.1029/2007WR006720>, 2008.
- Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Higdon, D., Robinson, B. A., and Hyman, J. M.: Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *Int. J. Nonlin. Sci. Numer. Simul.*, 10, 273–290, 2009.
- Vrugt, J. A., de Oliveira, D. Y., Schoups, G., and Diks, C. G.: On the use of distribution-adaptive likelihood functions: Generalized and universal likelihood functions, scoring rules and multi-criteria ranking, *J. Hydrol.*, 615, 128542, <https://doi.org/10.1016/j.jhydrol.2022.128542>, 2022.
- Wagener, T. and Kollat, J.: Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox, *Environ. Model. Softw.*, 22, 1021–1033, <https://doi.org/10.1016/j.envsoft.2006.06.017>, 2007.
- Wagener, T., Boyle, D. P., Lees, M. J., Wheeler, H. S., Gupta, H. V., and Sorooshian, S.: A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, 5, 13–26, <https://doi.org/10.5194/hess-5-13-2001>, 2001.
- Walker, W. E., Harremoës, P., Rotmans, J., Van der Sluijs, J. P., Van Asselt M. B. A., Janssen, P., and Krayer von Krauss, M. P.: Defining Uncertainty A Conceptual Basis for Uncertainty Management in Model-Based Decision, Support, *Integrat. Assess.*, 4, 5–17, 2003.