https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

Hydrology and Earth System Sciences
Discussions

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

Trevor Page[1], Paul Smith[1,2], Keith Beven[1], Francesca Pianosi[3], Fanny Sarrazin[4], Susana Almeida[5], Liz Holcombe[3], Jim Freer[6] and Thorsten Wagener[7]

[1] Lancaster Environment Centre, Lancaster University, Lancaster, UK

[2] Waternumbers, Lancaster, UK

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

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

[5] Atkins Global, Bristol, UK

[6] School of Geographical Sciences, Bristol University, UK and Global Institute for Water Security, University of Saskatchewan, Canada

[7] Institute for Environmental Science and Geography, University of Potsdam, Germany

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

Abstract. There is a general trend for increasing inclusion of uncertainty estimation in the environmental modelling domain. We present the CREDIBLE Uncertainty Estimation (CURE) Toolbox, an open source MATLAB™ toolbox for uncertainty estimation aimed at scientists and practitioners that 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 utilise 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 in applying the methods included. Best practice in 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 for increasing inclusion of uncertainty estimation (UE) in the environmental modelling domain,

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

Hydrology and Earth System Sciences
Discussions
Open Access

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. see Rougier and Beven, 2013).

Here we present the CREDIBLE Uncertainty Estimation (CURE) Toolbox; an open source MATLAB™ toolbox for UE associated with environmental *simulation* models. It is aimed at scientists and practitioners with some modelling experience, that are not necessarily experts in UE. The toolbox structure is similar to that of the SAFE toolbox (Sensitivity Analysis For Everyone; Pianosi *et al.*, 2016) such 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. That is, best practice in applying the various UE methods included as well as best practice in 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 forward propagation of uncertainties by sampling from *a priori* defined input and parameter distributions, for *forward UE,* or in 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 utilise toolbox functions.

Formal statistical and informal methods are included as there are no commonly agreed 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, 2015; Nearing *et al.*, 2016). The lack of 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. see Refsgaard *et al.*, 2007, Beven, 2009; Beven *et al.* 2016; Beven and Lane, 2022) rather than solely natural variability (*aleatory* uncertainties). Rigorous statistical inference applies to the latter, but might lead to unwarranted confidence if applied to the former,

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

Hydrology and Earth System Sciences Discussions
Open Access EGU

especially where some data might be disinformative in model evaluation (e.g. see Beven and Westerberg, 2011; Beven and Smith, 2015).

Assessing the impact of epistemic uncertainties for environmental modelling requires assumptions about their nature (which are difficult to define) such that the outputs from any UE will be conditional upon these assumptions. This poses the question of what is good practice in evaluating assumptions and choices made during the modelling process and what is good practice in communicating the meaning of any subsequent analyses (Walker *et al*., 2003; Sutherland *et al*., 2013; see also the TRACE framework of Grimm *et al*., (2014) for documenting the modelling process). Beven and Alcock (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. see Beven *et al*., 2014). 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 include for *forward UE:* the Data Uncertainty Engine[1] (DUE; Brown *et al*., 2007) and the SIMLAB[2] toolbox (Saltelli *et al*., 2004); and for *conditioned UE:* GLUEWIN [3], UCODE 2014[4] (Poeter *et al*., 2014), the Monte Carlo Analysis Toolbox[5] (MCAT; Wagener and Kollat, 2007), the MATLAB UQLAB[6] and the DREAM[7]Toolbox (Vrugt *et al*., 2008; 2009; Vrugt, 2016). See also the broader review of uncertainty tools undertaken by UNCERTWEB (Bastin *et al*., 2013) which includes tools supporting elicitation, visualisation, uncertainty and sensitivity analysis. 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 includes a method not included in previous toolboxes (i.e. the GLUE-Limits of Acceptability (LoA) method; Beven, 2006; Blazkova and Beven, 2009; Hollaway *et al.,* 2018) and explicitly sets out to encourage best practice regarding the conditionality of modelling results using the condition tree approach.

[1]*http://harmonirib.geus.info/due_download/index.html;* [2]*https://ec.europa.eu/jrc/en/samo/simlab;*
[3]*https://ec.europa.eu/jrc/en/econometric-statistical-software;* [4]*http://igwmc.mines.edu/freeware/ucode;*
[5]*http://www.imperial.ac.uk/environmental-and-water-resource-engineering/research/software/;* [6]*http://www.uqlab.com;*
[7]*http://faculty.sites.uci.edu/jasper/files/2015/03/manual_DREAM.pdf*

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

2 The CURE Toolbox

The CURE Toolbox essentially has two linked structures. There is an overall structure with which the user interacts throughout the analysis (Figure 1) and an underlying folder structure (Figure 2) 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, diagnostics and visualisation) and for the individual example modelling applications (i.e. the model functions and input files as well as any links to any '*external models*': i.e. models not coded as a MATLAB^{TM} function, but which can be executed from the command line). Folders also exist for general (i.e. not method-specific) sampling methods, visualisations 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. 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 propagation). Other specific sampling functions are associated with the adaptive sampling ("*on-line*" sampling) for Markov Chain Monte Carlo (MCMC) approaches where distributions and correlation structures are modified as the chain(s) evolve. Modelling diagnostics, both numeric and graphical are provided for both on-line adaptive sampling and "*off-line*" methods (i.e. those which are not adaptively sampled within a given method). In the case of on-line MCMC methods, visualisation of the evolution of the states of the chain(s) and tests for convergence to stationary distributions are included (e.g. Figure 3 a and b).

Functions used in fitting error models, as outlined below, also serve as an approximate check of error model validity when analysing posterior simulations (e.g. Figure 3 c and d). For the GLUE methods (see Beven and Binley, 1992, 2014), diagnostics are included for exploration of the acceptable parameter space and which criterion (or criteria) and at which timesteps (or locations) simulations were rejected. There are also method-specific and generic toolbox functions for visualisation and presentation of simulation results and associated uncertainties (e.g. see Figure 4). 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.

Table 1 lists the example workflows included in the first release of the CURE Toolbox and the methods employed.

Table 1 – Toolbox workflow examples and uncertainty estimation methods employed

Workflow	Uncertainty estimation method	Example Model	Shot Description
1	Forward (independent sampling)	CHASM[1,§]	Case study - uniform sampling of statistical distributions

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

2	Forward (independent)	HYMOD[2]	Fuzzy parameter distributions
3	Forward (dependent sampling - copula)	HYMOD	Marginal gamma distributions and rank correlation derived copula
4	Forward[€] (dependent sampling - copula)	LISFLOOD[3, §]	Case study - statistical 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-ABC[8] – LoA	HYMOD	Multi-chain MCMC; informal treatment with ABC[ψ] & LoA[¥]
8	Conditioned; DREAM-ABC - LoA	PROTECH[4, §]	Multi-chain MCMC; informal treatment with ABC & LoA
9	Conditioned; GLUE[9]	HYMOD	GLUE using threshold of informal likelihood for acceptance
10	Conditioned; GLUE	PROTECH[§]	GLUE using threshold of informal likelihood for acceptance
11	Conditioned; GLUE-LoA[10]	PROTECH[§]	GLUE using multi-variable LoA
12	Conditioned; GLUE-LoA	HYMOD	GLUE using single-variable LoA

[€] In this example the inputs were 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] Generalised Likelihood Uncertainty Estimation, Beven and Binley (1992); [10] Blazkova and Beven (2009); [§] Owing to long model run times this example uses pre-run simulation output; [ψ] Approximate Bayesian Computation; [¥]Limits of Acceptability.

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

Hydrology and Earth System Sciences Discussions

3 Condition Tree Implementation

As part of any CURE toolbox application, users can explore and document modelling choices, assumptions and uncertainties using the condition tree GUI (e.g. Figure 5). 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. Similarly to the incorporation of UE, the condition tree would, ideally, be completed as an integral part of any modelling application and can help in the definition of an appropriate workflow structure.

The GUI takes the form of a number of simple, sequential dialogue boxes where the user is asked to enter text. In the initial release of the toolbox there are 5 primary dialogue boxes covering:

1. Project aims and model(s)/model structures considered
2. Modelling uncertainties - *overview:* model structure, parameters, inputs, observations for model conditioning
3. Uncertainties - observations for model conditioning - *specific*: associated uncertainties and basis for assessing simulation performance
4. Uncertainties - inputs - *specific*: sampling strategy, distributions, dependencies
5. Uncertainties - parameters - *specific*: choice of parameters, sampling strategy, distributions, 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.

4 Defining a workflow

The *a priori* consideration of modelling uncertainties via the condition tree is an *optional* first step to help structure a workflow; an additional aid is provided in the form of a *decision-tree* (e.g. Figure 6) to guide the user to example toolbox workflows which most closely match their application and to help in the choice of UE method(s) (Pappenberger *et al*., 2006). These complement 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 the code and function *header text* available by typing *help "function name"* at 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): 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.

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

Hydrology and Earth System Sciences
Open Access
Discussions

It is assumed that the user has completed pre-processing analyses such as forcing input error assessment and disinformation screening (e.g. Beven and Smith, 2015) as well as an assessment of errors associated with conditioning observations where used. An exception is the toolbox facility for fitting error models when formal statistical likelihoods are to be used. This is done interactively, using command line prompts, and can form part of a workflow (or used stand-alone). Error model fitting utilises functions which transform the data to remove heteroscedasticity and non-normality using Box-Cox transformations, which cover a broad continuum of transformations (Box and Cox, 1964), and autocorrelation (by fitting an autoregressive model of suitable order) in an iterative way as proposed by (Beven *et al*., 2008; see Figure 3 c and d for example error model fitting visualisations).

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' as follows:

1. Condition tree GUI: project setup and interactive dialogue boxes
2. Set up inputs and observations
3. Set up parameter ranges, distributions and sampling strategy
4. Define performance measure (if conditioned UE)
5. Simulations (*on-line* or *off-line*; MATLAB™ function or '*external model'*)
6. Post-processing: diagnostics, results, propagation and visualisation of uncertainty

Associated with these main steps, example workflows include *automatic* 'text writes' which are appended to the audit trail log. These include specific choices which are made when implementing steps 1-5 above: such as the ranges of parameter values used and their distributions, the sampling strategy employed as well as 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 *on-line* simulation performance to be assessed for MCMC methods, and the many permutations of performance measures, and ways of combining them where multiple criteria are used, users are also required to modify an additional function that returns an overall measure of individual simulation performance. In addition, where 'external' models are to be used for on-line approaches, additional modifications may be required for modification of input/parameter files, using some form of *wrapper code*.

5 An Example Application

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, it is well suited to the specification of assumptions about espitemic uncertainties, conditioning using uncertain observational data, and rejectionist approaches to model evaluation (see also Beven et al., 2016; 2022a,b). Here we provide some more detail on the application of the PROTECH model within such a rejectionist conditioning framework (Workflow 11 in Table 1). The full workflow and outputs are given in the Electronic Supplement to this article.

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

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, so that the workflow takes the simulated output files as inputs. The model requires flow, weather and nutrient information as inputs. A reduced set of six parameters were 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 outputs 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 Electronic Supplement.

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

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

6 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 in 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/default.htm.

Acknowledgements

This work was supported by the Natural Environment Research Council project: Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE); grant number NE/J017450/1.

Computer Code Availability

The CURE Matlab Toolbox an Open Source Matlab Code hosted at Lancaster University and can be downloaded at https://www.lancaster.ac.uk/lec/sites/qnfm/credible/default.htm (contact n.chappell@lancaster.ac.uk). It was first made available in 2021.

Author Contributions

Trevor Page and Keith Beven were involved in the conceptualisation of the CURE Toolbox and the development of the example applications. Trevor Page, Paul Smith, Francesca Pianosi and Fanny Sarrazin provided the software development. Trevor Page and Keith Beven wrote the original draft of the paper. All the other authors were involved in the development of applications within the NERC CREDIBLE project that was the original motivation for the development of the SAFE and CURE Toolboxes, and in reviewing and editing the paper. Thorsten Wagener was the principal investigator of CREDIBLE.

References

Alexandrov, G. A., Ames, D., Bellocchi, G., Bruen, M, Crout, N., Erechtchoukova, M., Hildebrandt, A., Hoffman, F., Jackisch, C., Khaiter, P., Mannina, G., Matsunaga, T., Purucker, S. T., Rivington and M., Samaniego, L., Technical assessment and evaluation of environmental models and software: letter to the Editor. *Environmental Modelling and Software*, 26 (3), pp.328-336, 2011.

Almeida, S., Holcombe, E. A., Pianosi, F., & Wagener, T., Dealing with deep uncertainties in landslide modelling for disaster risk reduction under climate change. *Natural Hazards and Earth System Sciences*, 17(2), 225-241. doi:http://dx.doi.org/10.5194/nhess-17-225-2017, 2017.

Ascough, J.C. II, Maier, H.R., Ravalico, J.K. and Strudley, M.W., Future research challenges for incorporation of uncertainty in environmental and ecological decision-making, *Ecological Modelling*, Volume 219, Issues 3–4, 10 Pages 383-399, ISSN 0304-3800, http://dx.doi.org/10.1016/j.ecolmodel.2008.07.015, 2008

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., Seppelt, R., Voinov, A. A., Fath, B. D., Andreassian, V.. Characterising performance of environmental models, *Environmental Modelling and Software*, 40, 1-20, ISSN 1364-8152, http://dx.doi.org/10.1016/j.envsoft.2012.09.011, 2013.

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, *Environmental Modelling & Software*, 39, 116-134, ISSN 1364-8152, http://dx.doi.org/10.1016/j.envsoft.2012.02.008, 2013.

Beven, K J, A manifesto for the equifinality thesis, *J. Hydrology*, 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. *Hydrol. Sci. J.* 10.1080/02626667.2015.1031761, 2015.

Beven, K. and Binley, A., The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Process*.,6,279-298, 1992.

Beven K. J., Smith P. J., and Freer, J.E., So just why would a modeller choose to be incoherent?, *Journal of Hydrology*, 354(1–4), 15-32. ISSN 0022-1694, http://dx.doi.org/10.1016/j.jhydrol.2008.02.007, 2008.

Beven, K J and Westerberg, I, On red herrings and real herrings: disinformation and information in hydrological inference, *Hydrological Processes (HPToday)*, 25, 1676–1680, DOI: 10.1002/hyp.7963, 2011.

Beven, K. J. and Alcock, R., Modelling everything everywhere: a new approach to decision making for water management under uncertainty*, Freshwater Biology,* 56, 124-132, doi:10.1111/j.1365-2427.2011.02592.x, 2012.

Beven, K., Smith, P., Westerberg, I., Freer, J., Comment on "Pursuing the method of multiple working hypotheses for hydrological modeling" by P. Clark et al. *Water Resources Research*. 48, 11, 1-5, 2012.

Beven, K. J., and Binley, A. M., GLUE, 20 years on. *Hydrol. Process.* 28(24):5897-5918, DOI: 10.1002/hyp.10082.

Beven, K J, Lamb, R, Leedal, D T, and Hunter, N, 2014, Communicating uncertainty in flood risk mapping: a case study, *Int. J. River Basin Manag.*, 13(3):285-296, DOI:10.1080/15715124.2014.917318, 2014.

Beven, K. J., and Smith, P. J., Concepts of Information Content and Likelihood in Parameter Calibration for Hydrological Simulation Models, *ASCE J. Hydrol. Eng*., DOI: 10.1061/(ASCE)HE.1943-5584.0000991, 2015.

Beven, K. J., Aspinall, W. P., Bates, P. D., Borgomeo, E., Goda, K., Hall, J. W., Page, T., Phillips, J. C., Rougier, J. T., Simpson, M., Stephenson, D. B., Smith, P. J., Wagener, T., and Watson, M. Epistemic uncertainties and natural hazard risk assessment – Part 1: A review of the issues, Nat. Hazards Earth Syst. Sci. Discuss., 3, 7333-7377, doi:10.5194/nhessd-3-7333-2015, 2016.

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. *Hydrological Processes*, in press, 2022a.

Beven, K. J., Lane, S., Page, T., Hankin, B, Kretzschmar, A., Smith, P. J., Chappell, N., On (in)validating environmental models. 2. Implementation of the Turing-like Test to modelling hydrological processes, *Hydrological Processes*, in press, 2022b.

Blazkova S., 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 Resources Research. 45, p. W00B16, 2009.

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

Box, George E. P., Cox, D. R., An analysis of transformations. *Journal of the Royal Statistical Society*, Series B 26 (2): 211–252, 1964.

Brown J.D., Heuvelink G. B.M., The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables; *Computers and Geosciences*. 33(2), 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(9), doi:10.1029/2010WR009827, 2011.

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. *Ecological Modelling*, 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(2), 223-242.

Hollaway, M.J., Beven, K.J., Benskin, C.McW.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., Haygarth, P.M., 2018, Evaluating a processed based water quality model on a UK headwater catchment: what can we learn from a 'limits of acceptability' uncertainty framework?, *J. Hydrology.* Doi: 10.1016/j.jhydrol.2018.01.063, 2001.

Neal J., Keef C., Bates P., Beven K.J. and Leedal, D., Probabilistic flood risk mapping including spatial dependence. *Hydrol. Process.* 27, 1349–1363. 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, *Hydrological Sciences Journal*, DOI:10.1080/02626667.2016.1183009, 2016.

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.floodrisknetorg.uk/methodshttp://www.floodrisk.net. *Hydrological Processes*, 20 (17), 1099-1085. doi.org/10.1002/hyp.6541, 2006..

Pianosi, F., Sarrazin, F., Wagener, T., A Matlab toolbox for Global Sensitivity Analysis, *Environmental Modelling & Software*, 70, 80-85, 2015.

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, Colorado, USA, 2014.

Refsgaard JC, van der Sluijs JP, Højberg AL, Vanrolleghem PA., Uncertainty in the environmental modelling process – A framework and guidance. *Environmental Modelling and Software*, 22, 1543-1556, 2007.

Roberts G. O. and Rosenthal J. S., Optimal scaling for various Metropolis-Hastings algorithms. *Statist. Sci*. 16 (4), 351-367. doi:10.1214/ss/1015346320, 2001.

Roberts G. O. and Rosenthal J. S., Examples of Adaptive MCMC, *Journal of Computational and Graphical Statistics*, 18:2, 349-367, DOI: 10.1198/jcgs.2009.06134, 2009.

Rougier, J and Beven, K J, Model limitations: the sources and implications of epistemic uncertainty, in Rougier J, Sparks, S and Hill, L, *Risk and uncertainty assessment for natural hazards*, Cambridge University Press: Cambridge, UK, 40-63, 2013.

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

Sadegh, M. and Vrugt J.A., Approximate Bayesian computation using Markov Chain Monte Carlo simulation: DREAM(ABC), *Water Resources Research*, 50(8), 6767-6787. doi:10.1002/2014WR015386, 2014.

Saltelli A., Tarantola S., Campolongo F., Ratto M., How to Use SIMLAB in *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley, ISBN0-470-87093-1, 2004.

Sutherland, W. J., Spiegelhalter, D., & Burgman, M. A., Twenty tips for interpreting scientific claims. *Nature*, *503*(7476): 335-337, 2013.

Vrugt, J.A., Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, 75, 273-316. https://doi.org/10.1016/j.envsoft.2015.08.013, 2016.

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 Resources Research*, 44, W00B09. doi: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. *International Journal of Nonlinear Sciences and Numerical Simulation,* 10(3), 273-290, 2009.

Wagener, T. and Kollat, J., Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox. *Environmental Modelling and Software*, 22(7), 1021-1033. 10.1016/j.envsoft.2006.06.017, 2007.

Wagener, T., Boyle, D., Lees, M., Wheater, H., Gupta, H., and Sorooshian,S., A framework for development and application of hydrological models. *Hydrol. Earth Syst. Sci*., 5, 13-26, 2001.

Walker WE, Harremoës P, Rotmans J, Van der Sluijs JP, Van Asselt MBA, Janssen P and Krayer von Krauss, MP., Defining Uncertainty A Conceptual Basis for Uncertainty Management in Model-Based Decision, Support, *Integrated Assessment*, 4(1), 5-17, 2003.

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

Hydrology and Earth System Sciences
Open Access
Discussions

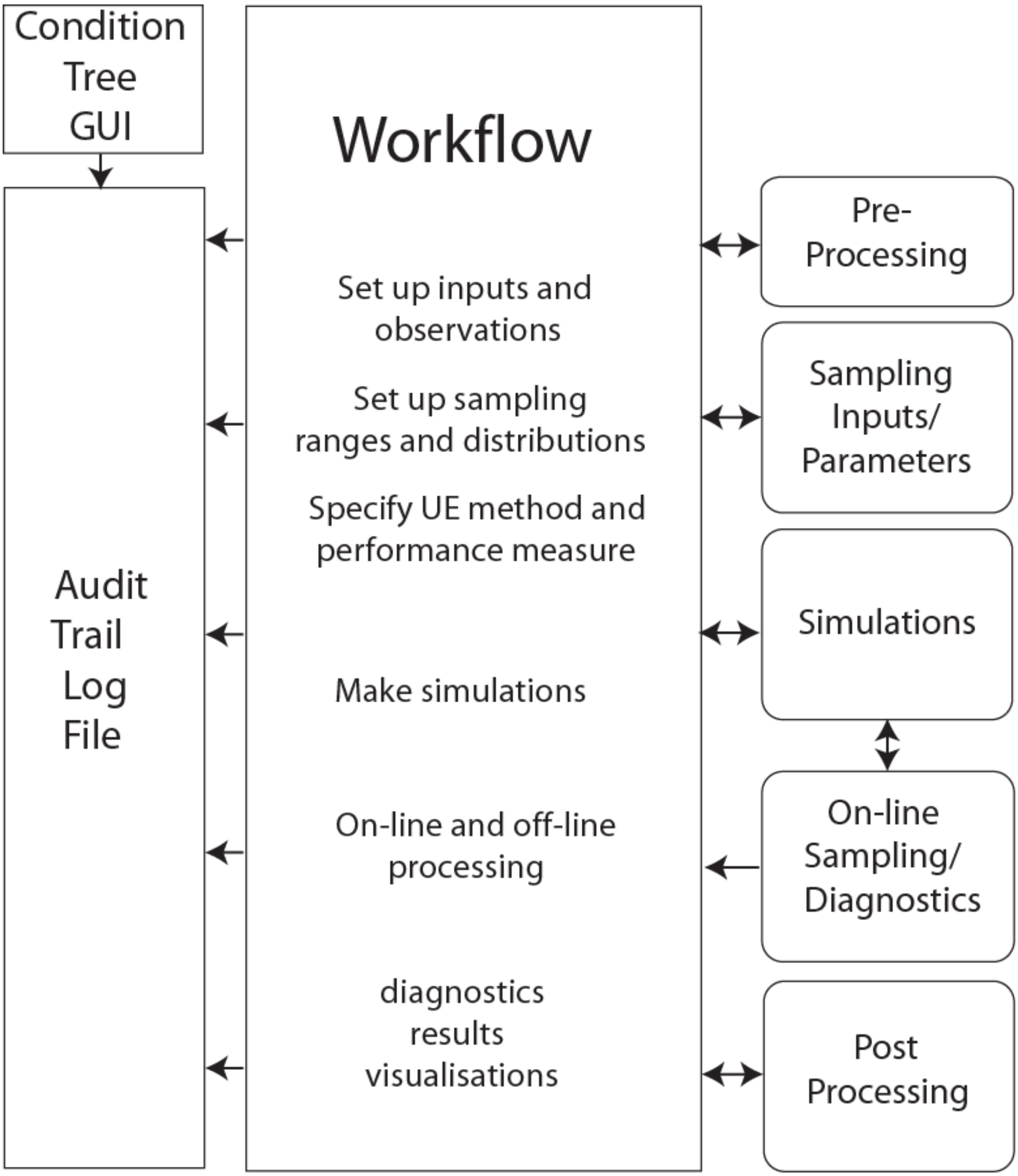


Figure 1 CREDIBLE Uncertainty Estimation Toolbox: overall structure

320

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

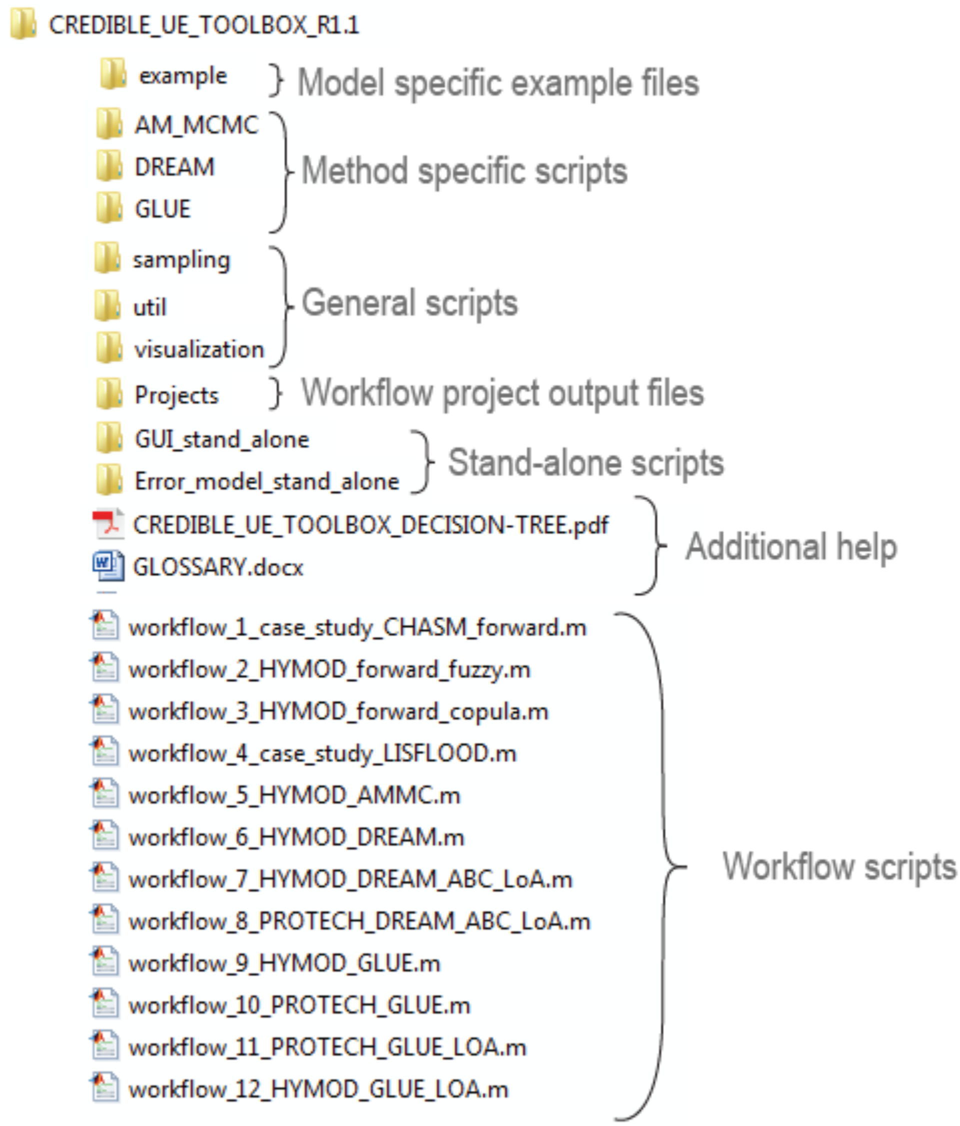


Figure 2 Outline folder structure of the CURE Toolbox

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

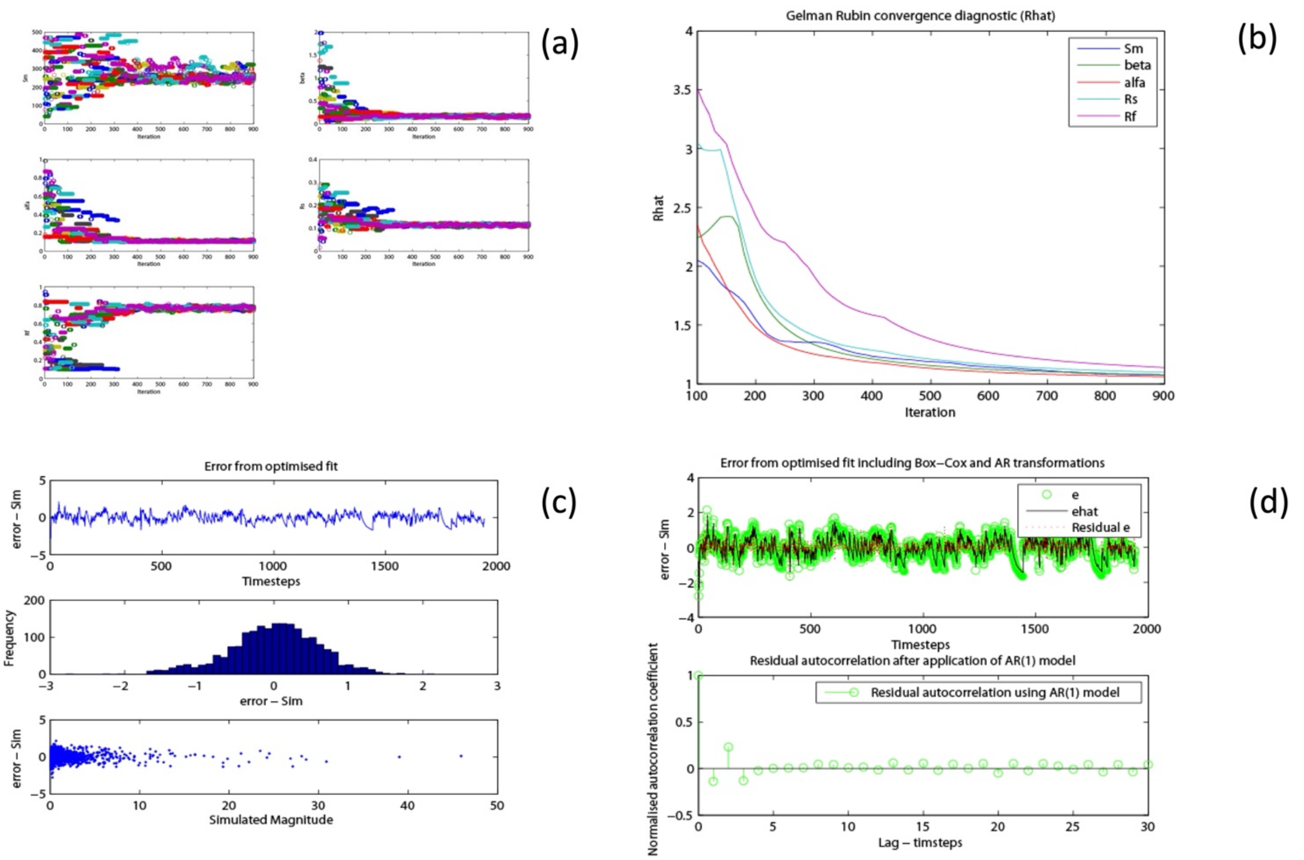


Figure 3 Visualisation of simulation diagnostics: (a) the evolution of 12 chains using DREAM; (b) evolution of the Gelman Rubin convergence statistic for 5 parameters; (c) & (d) visualisation of error structure during error model fitting;

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

Hydrology and Earth System Sciences
Open Access
Discussions

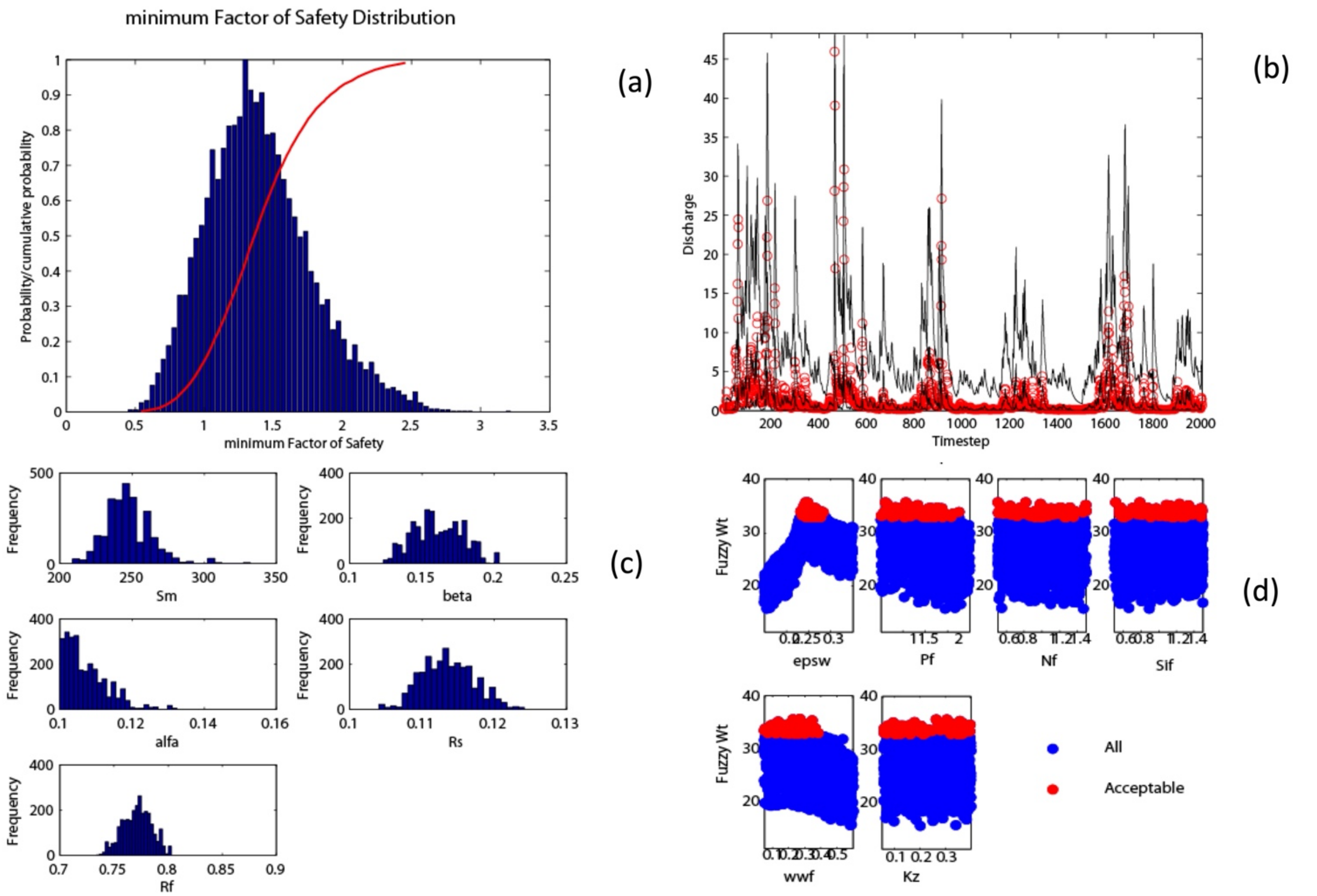


Figure 4 Visualisation of results: (a) The distribution of simulated minimum factor of safety from a forward UE using the CHASM landslide model; (b) 5th; 50th and 95th percentiles of simulated discharge (black lines) and observed discharge from an MCMC conditioned UE method using HYMOD; (c) Posterior parameter distributions for the same example as (b) and (d) dotty plots showing both all and acceptable parameter sets from a GLUE analysis using the PROTECH model.

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

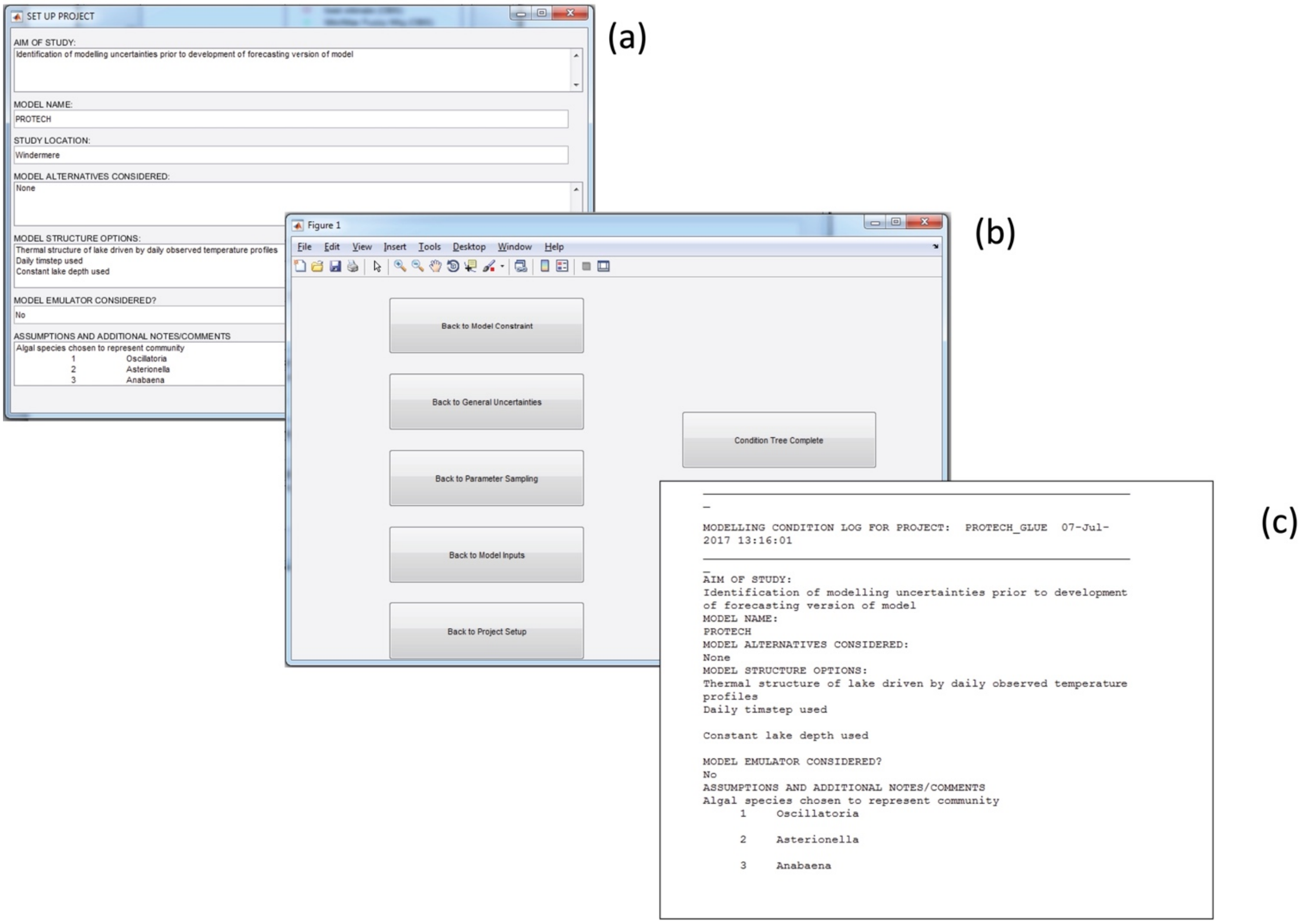
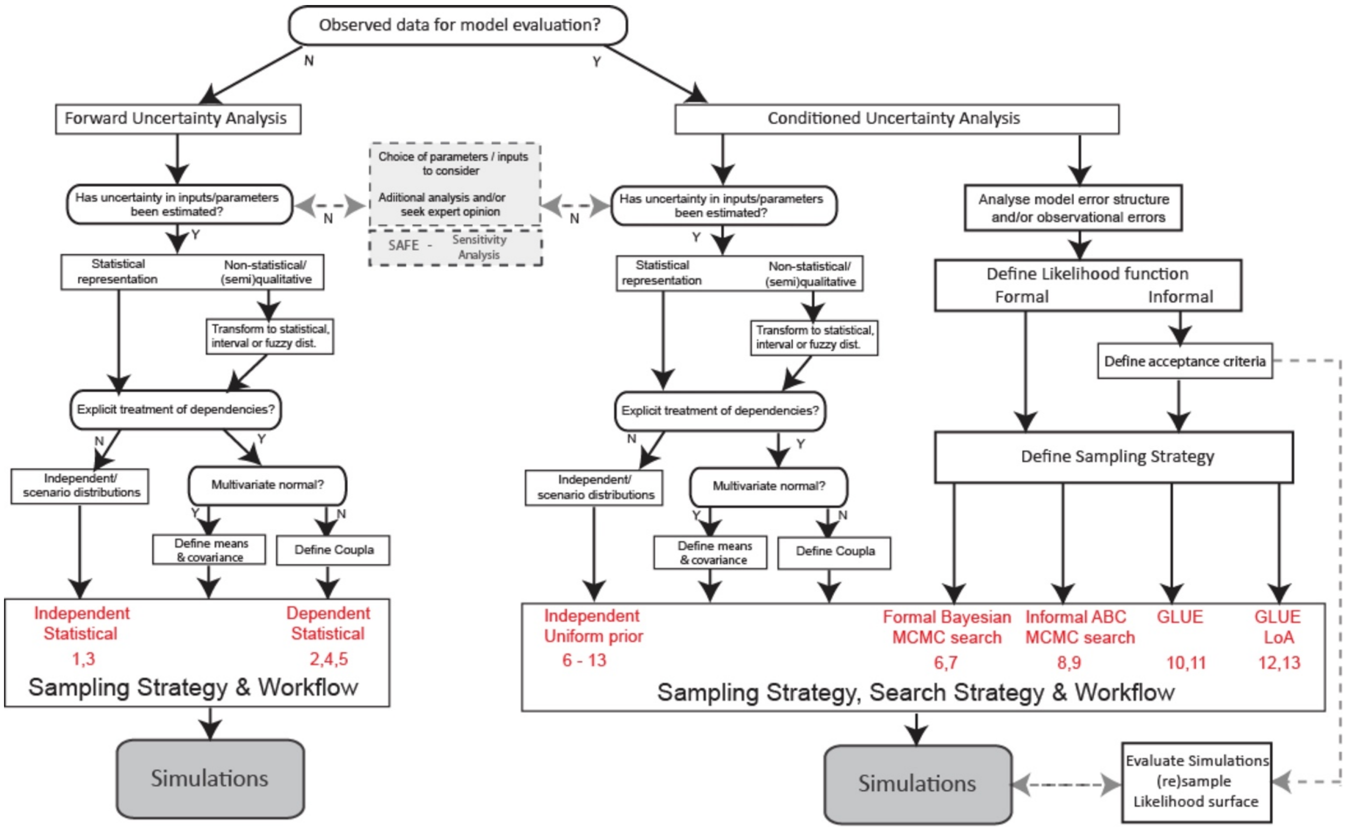


Figure 5. Condition Tree example: GUI dialogue box for (a) Project Setup, (b) the Condition Tree tree navigation pane and (c) part of an example audit trail log.

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

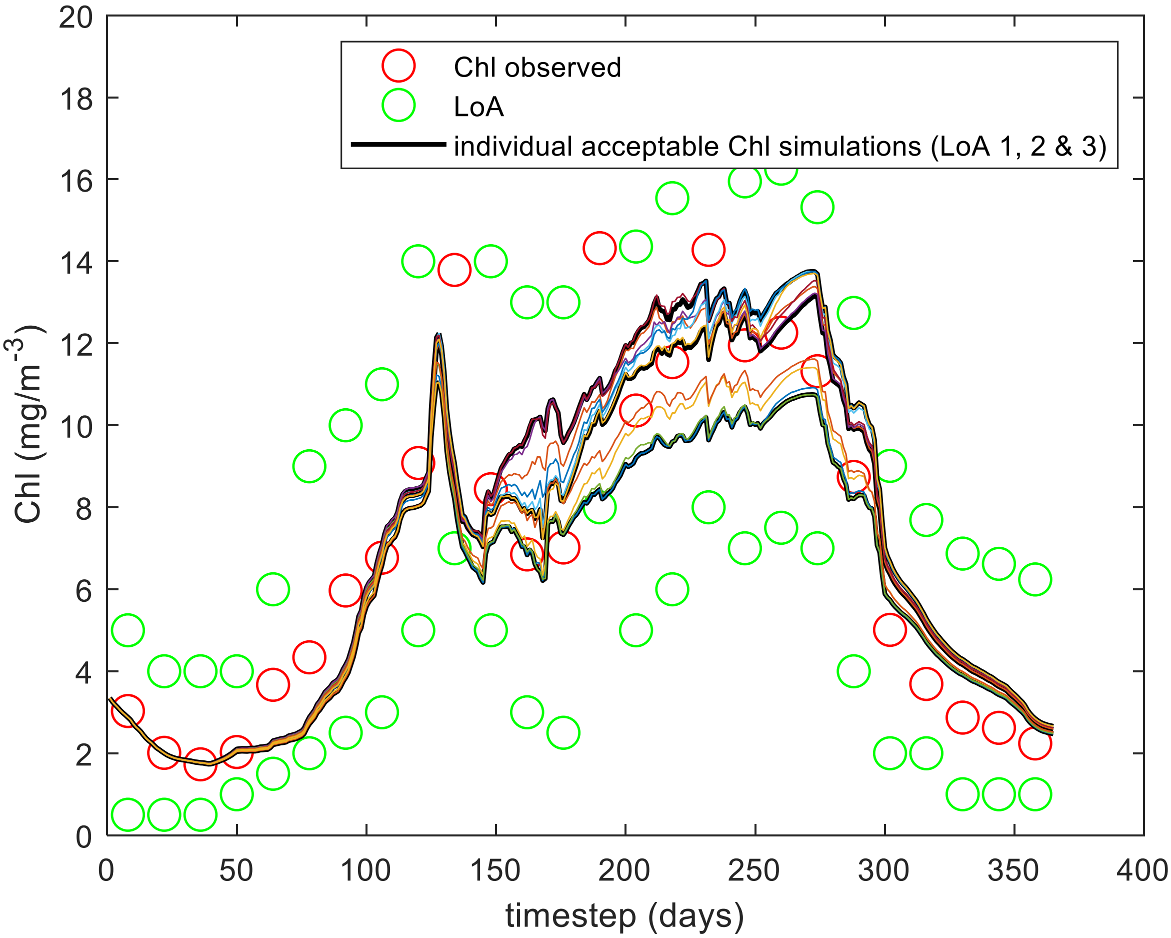
Hydrology and Earth System Sciences
Open Access
Discussions



345 **Figure 6. Decision tree guiding users towards different methodologies and workflows**

https://doi.org/10.5194/hess-2022-349
Preprint. Discussion started: 17 November 2022

Hydrology and Earth System Sciences
Discussions



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