Teaching hydrological modelling: Illustrating model structure uncertainty with a ready-to-use computational exercise

Wouter J. M. Knoben[1] and Diana Spieler[2]

[1]University of Saskatchewan Coldwater Laboratory, Canmore, Alberta, Canada
[2]Institute of Hydrology and Meteorology, Technische Universität Dresden, Dresden, Germany

Correspondence: diana.spieler@tu-dresden.de

Abstract. Estimating the impact of different sources of uncertainty along the modelling chain is an important skill graduates are expected to have. Broadly speaking, educators can cover uncertainty in hydrological modelling by differentiating uncertainty in data, model parameters and model structure. This provides students with insights on the impact of uncertainties on modelling results and thus on the usability of the acquired model simulations for decision making. A survey among teachers in the Earth and environmental sciences showed that model structural uncertainty is the least represented uncertainty group in teaching. This paper introduces a computational exercise that introduces students to the basics of model structure uncertainty through two ready-to-use modeling experiments. These experiments require either Matlab or Octave, and use the open-source Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT) and the open-source Catchment Attributes and Meteorology for Large-sample Studies (CAMELS) dataset. The exercise is short and can easily be integrated into an existing hydrologic curriculum, with only a limited time investment needed to introduce the topic of model structure uncertainty and run the exercise. A trial application at the Technische Universität Dresden (Germany) showed that the exercise can be completed in two afternoons and that the provided setup effectively transfers the intended insights about model structure uncertainty.

1 Introduction

The ability to use computer models to provide hydrologic predictions is a critical skill for young hydrologists (Seibert et al., 2013; Wagener and McIntyre, 2007). Model use is so widespread that students will have to generate, use or present modelling results at some point in their professional career (Seibert et al., 2013). A very wide range of different models currently exists and it is arguably less important for students to learn how to use any specific model than to be taught general modelling concepts. Students should have some understanding of different modelling philosophies, learn to use different model types and be aware of the strengths and limitations of hydrologic modelling (Wagener et al., 2012). Given the societal need to provide hydrologic predictions far into the future and the unknown (Kirchner, 2006), a core competence for young professionals is knowing how to provide such predictions in a scientifically sound manner.

Understanding of uncertainty in the modelling process is key to interpreting model results (e.g. Pechlivanidis et al., 2011; Blöschl and Montanari, 2010; Beven et al., 2011; Mendoza et al., 2015, among many others). Modelling uncertainties can be roughly classified as relating to the input and evaluation data, the estimation or calibration of model parameters, and the choice of equations that make up the model structure. These concepts should be an integral part of the hydrologic curriculum (Wagener et al., 2012; AghaKouchak et al., 2013; Thompson et al., 2012) in a teaching structure that includes student-driven, hands-on exercises that reinforce the taught concepts (Thompson et al., 2012). A survey among 101 teachers in the earth and environmental sciences (see Supplementary Materials) shows large differences in how much time is spent on teaching hydrologic modelling in general, whether model-related uncertainty is part of the course and, if so, which aspects of uncertainty are covered. Based on the survey, model structural uncertainty is the least represented uncertainty aspect in teaching. The main reason named for not covering model-related uncertainty is a lack of time, while the lack of good teaching materials is the second-most common explanation. Just 6% of respondents that did not cover uncertainty in their classes stated that the topic would be covered in another course.

Selecting a model that faithfully represents current and future hydrologic conditions in a given catchment is critical for accurate long-term projections of water availability. In other words, one requires "the right answers for the right reasons" (Kirchner, 2006). The difficult task of finding an appropriate model structure, i.e. the combination of which hydrologic processes are included in a model, which equations are used to describe these processes and how model states and fluxes are connected, can be referred to as model structure uncertainty and is a significant source of overall modeling uncertainty (Di Baldassarre and Montanari, 2009). Model structure uncertainty is being investigated with increasing numbers of models in increasingly varied selections of catchments (e.g. Perrin et al., 2001; Butts et al., 2004; Duan et al., 2006; van Esse et al., 2013; Knoben et al., 2020; Spieler et al., 2020) and results are consistent: model choice matters and selecting an inappropriate model for a given catchment can lead to simulations of questionable quality. For a variety of reasons, the suitability of a given model for the task at hand is not always the main driver in model selection (Addor and Melsen, 2019) and it is not unlikely that students will encounter such cases in both academia and practice. Hands-on experience with model structure uncertainty in a classroom setting will prepare students for when they will need to interpret modeling results in their future careers.

Thoughtful interpretation of model results is among many other skills that are expected of young hydrologists (see for example Table 1 in Seibert et al., 2013). However, finding or creating course materials that cover all these expected skills and incorporating these materials into an existing curriculum is time-consuming, as is updating existing materials with new knowledge. This time is consequently not spent on preparing delivery of the material (Wagener et al., 2012). Wagener et al. (2012) therefore introduces the Modular Curriculum for Hydrologic Advancement (MOCHA), in which educators from many different countries freely share hydrologic course materials in a modular manner. Each module addresses a specific topic and can theoretically be inserted into an existing curriculum with very little effort. Although the MOCHA project has been inactive for some time, the principle of freely shared, self-contained teaching modules can be of great use to the teaching community and is experiencing a revival in platforms such as HydroLearn (https://www.hydrolearn.org).

Seibert and Vis (2012) provide a stand-alone version of the *Hydrologiska Byråns Vattenavdelning* (HBV) model that is a good example of the MOCHA philosophy in practice. The software is specifically modified for teaching and comes with documentation and descriptions of various teaching goals. This is a so-called lumped conceptual hydrologic model that relies on empirical equations to describe catchment processes and on calibration to find its parameter values. Although there is debate about the usefulness of such models for predictions under change (see e.g. Archfield et al., 2015), there are good reasons to use them as teaching tools provided that the limitations of these tools are clearly communicated to the students. Conceptual models tend to be much easier to set up and run than their more physics-based, spatially distributed counterparts; they generally have fewer lines of code and internal dynamics that are easier to grasp than those of physics-based models; and they continue to be widely used for practical applications. These characteristics mean that limited teaching time is spent on using and analyzing models rather than setting up the models; that students have more opportunity to explore internal model dynamics instead of focusing on model outputs only; and that students obtain a firm understanding of the type of tools they are likely to encounter in positions outside of academic research (Seibert and Vis, 2012).

This paper introduces a set of computational exercises designed to give students hands-on experience with model structure uncertainty and to encourage critical thinking about how the results of a modelling study can be interpreted. Our goal is to increase the frequency with which model structure uncertainty is taught to (under-)graduates and to reduce the time investment required for educators to do so. The exercises use two conceptual model structures applied to two carefully selected catchments to illustrate various important lessons about hydrologic model selection. Briefly, these lessons focus on the need to carefully interpret aggregated performance metrics, the dangers of applying models in new places based on performance elsewhere and the need to consider if a model's internal structure is an appropriate representation of the catchment at hand. The goals, learning objectives and materials of this module are described in more detail in Section 2. The exercises are described in Section 3 (ready-to-use student handouts can be found in Supplementary Materials S5). Section 4 briefly summarizes the benefits of using these exercises and describes a trial application of this module at the Technische Universität Dresden, Germany. The exercises require access to and knowledge of either Matlab or Octave. Course materials can be downloaded through GitHub: https://github.com/wknoben/Dresden-Structure-Uncertainty.

2 Description

This section describes the main teaching objectives (Section 2.1), the catchment data and models used (Sections 2.2.1 and 2.2.2 respectively), an overview of provided materials, requirements and install instructions (Section 2.3), and suggestions on how to integrate these exercises into an existing curriculum (Section 2.4).

2.1 Objectives and outline

The main goal of this teaching module is to facilitate teaching of model structure uncertainty in hydrology. Learning objectives are conveyed through comparative analysis between model results generated by students, using two conceptual model structures and two catchments. Both models and catchments have been specifically selected out of a sample of 40+ models and 500+ catchments for the lessons that can be conveyed by each comparative exercise. Note that detailed understanding of the selected models and catchments is not a goal in itself; they are only intended to convey the learning objectives specified in this section. The catchments and models are described in sections 2.2.1 and 2.2.2 respectively.

Common ways of evaluating a hydrologic model's performance involves calculating some aggregated score that expresses the similarity between observations and simulations of a given state or flux (typically streamflow). Examples are root-mean-squared error (RMSE), the Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970) or the Kling-Gupta efficiency (KGE; Gupta et al., 2009) scores. Such approaches are common in many hydrologic disciplines but, as this teaching module is intended to show, are not guaranteed to identify whether a model is an appropriate representation of the catchment under consideration. In other words, the value of the efficiency score does not indicate whether a model produces "the right answers for the right reasons" (Kirchner, 2006). Note that this teaching module does not cover the difficult issue of defining when a given efficiency score can be called adequate, i.e. setting a minimum score the model must achieve before its simulations are considered of sufficient quality for further consideration. This requires careful use of benchmarks (e.g. Garrick et al., 1978; Seibert, 2001; Schaefli and Gupta, 2007; Seibert et al., 2018) that dictate expectations for model performance, which is outside the scope of this module. Instead, this module uses the common interpretation that higher efficiency scores indicate more accurate models, in the sense that simulations with higher efficiency scores more closely resemble observations than the simulations from models with lower efficiency scores. The models and catchments in this module are selected so that in one of the catchments both models achieve very similar KGE scores despite the models having very different structures, while in the other catchment the models achieve very different KGE scores. This is intended to convey the following lessons to students (KGE scores and summary of these take home messages in Figure 1):

1. Model choice matters. Because all models are "hydrologic models" it is an easy assumption to make that the choice of model is largely one of taste or convenience, rather than one of suitability for the task at hand. Comparing the performance of both models in catchment 8109700 shows that this is not the case: the choice of model strongly affects the accuracy of obtained simulations. In this particular case, the catchment experiences periods of no flow which model m03 can simulate but model m02 cannot.

2. Models with very different structures can achieve virtually identical efficiency scores in a given catchment. Comparing the performance of both models in catchment 12145500 shows that both achieve similar KGE scores. Logically only one (or neither) of the models can be an appropriate representation of the hydrologic conditions in this catchment. This comparison shows that achieving high efficiency scores in a given catchment is no guarantee that the model accurately represents the dominant processes in the catchment.

3. Reinforcing the previous point, comparing the performance of model m03 across both catchments shows that the model achieves higher efficiency scores than model m02 in both places, while the catchments themselves are structurally very different (catchment descriptions are shown as part of the suggested exercises). Logically, model m03 may be an accurate representation of the hydrologic conditions in one of the catchments, but not both. This again shows that high efficiency scores are no guarantee of having used the "right" model.

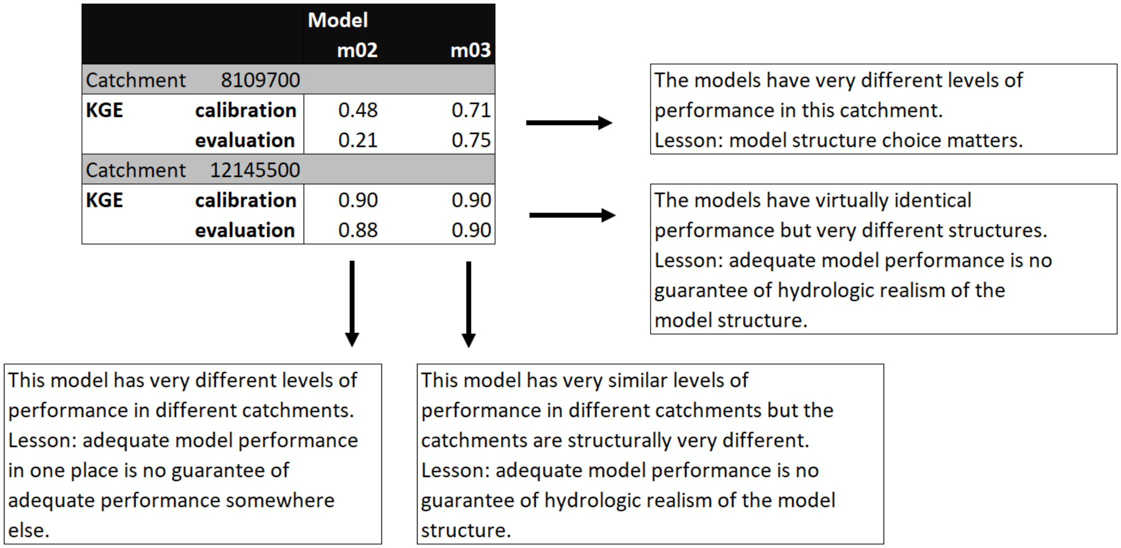


Figure 1. Using comparative assessments to transfer lessons about hydrologic model structure uncertainty.

4. Choosing a model based on past performance should be done with care. Comparing the performance of model m02 across both catchments shows that the model performance is very different in both places and that having a "successful" model for one catchment is no guarantee that this model will perform equally well somewhere else.

2.2 Catchments and models

Wagener et al. (2012) outlines a need for multi-media tools that support teaching in hydrology, specifically mentioning *"a model base with algorithms that the students can download and use to support their homework assignments or in terms projects (Wagener et al, 2004). Such algorithms need to be accompanied by sufficient documentation and data examples."* This module uses open-source data to allow straightforward application in assignments and projects. Catchment data is selected from the Catchment Attributes and Meteorology for Large-Sample studies (CAMELS, Addor et al., 2017). Models are selected from the Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT, Knoben et al., 2019). Data and models are described in more detail in the following sections.

2.2.1 CAMELS catchment data

The CAMELS dataset (Addor et al., 2017) provides meteorological forcing data and a variety of catchment attributes for 671 river basin in the contiguous United States. The catchments upstream of Middle Yegua Creek near Dime Box, Texas (USGS gauge ID: 08109700), and Raging River near Fall City, Washington (USGS gauge ID: 12145500), are used in this module. Our example hand-outs (Supplementary Materials S5) direct students to data that forms the basis of, and expands on, the catchment descriptions given below.

Middle Yegua Creek is a water-limited catchment (aridity fraction = 1.3, with the aridity fraction calculated as mean annual precipitation divided by mean annual potential evapotranspiration) with a corresponding low runoff ratio (0.11; mean annual runoff divided by mean annual precipitation), low mean runoff (0.3 mm/d) and on average 30 days with no observed streamflow. Precipitation is sporadic (on average 294 days have < 1 mm precipitation) and mostly concentrated in autumn with little to no snowfall. The catchment is relatively large (615 km^2) with little variation in elevation (mean slope = 6 m/km). Vegetation cover consists mostly of cropland, shrubs and low trees.

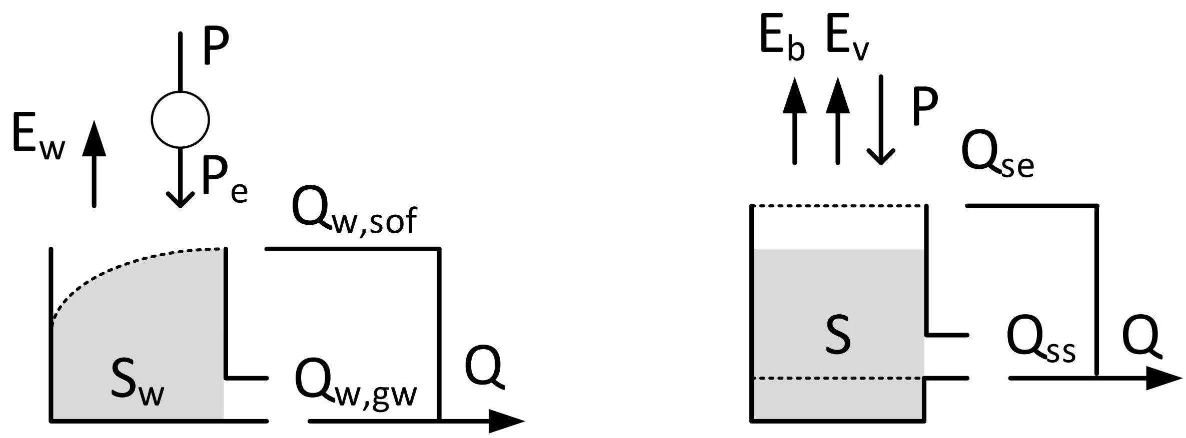
Raging River is an energy-limited catchment (aridity fraction = 0.37) with a high runoff ratio (0.68), high mean runoff (3.9 mm/d) and observed streamflow on all days in the record. Precipitation occurs regularly (180 days with < 1 mm precipitation) and is winter-dominated although snowfall is rare (precipitation as snow fraction = 0.04). The catchment is comparatively small (80 km^2) and steep (mean slope = 86 m/km). Vegetation cover consists nearly exclusively of mixed forests.

The exercises use Daymet meteorological forcing data that is provided as part of the CAMELS data set (Newman et al., 2015; Addor et al., 2017). Precipitation is part of the source data and time series of potential evapotranspiration are estimated using the Priestley-Taylor method (Priestley and Taylor, 1972).

2.2.2 MARRMoT models

The Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT; Knoben et al., 2019) contains Matlab/Octave code for 46 conceptual models implemented in a single framework. Each model requires standardized inputs and provides standardized outputs. This means that data preparation and experiment analysis scripts have to be prepared only once, after which running and comparing different model structures becomes trivial. The toolbox is supported by extensive documentation, divided into the main paper describing the toolbox setup; the Supplementary Materials to that paper describing each model, flux equation and default parameter ranges; a User Manual that provides guidance for practical issues such as installation, use and modification or creation of models and fluxes; and comments included as part of the actual computer code.

This course uses MARRMoT models m02 and m03 (names refer to consistent identifiers used in all MARRMoT documentation). Both have a single state variable and 4 calibration parameters but very different internal mechanics (Figure 2). Both models require time series of precipitation P and potential evapotranspiration E_P as input. Briefly, model m02 is part of the Flex-Topo approach (Savenije, 2010) and intended to represent the dominant hydrologic fluxes in a West-European wetland/riparian zone. This model has a single state representing catchment storage and uses 4 parameters to conceptualize interception by vegetation (turning precipitation P into flux P_e), surface overland flow $Q_{w,sof}$ from a variable contributing source area, and groundwater flow $Q_{w,gw}$. Evaporation E_w occurs at the potential rate. Model m03 is part of a study that develops a model for a semi-arid catchment in Western Australia (Jothityangkoon et al., 2001). This model also has a single state representing catchment storage but uses its 4 parameters to conceptualize a difference between bare soil evaporation E_b and transpiration by vegetation E_v, saturation excess overland flow Q_{se} if maximum storage is exceeded, and threshold-based subsurface flow Q_{ss}. Full details of both models, including state equations and flux parametrizations, can be found in the Supplementary Materials of Knoben et al. (2019).



(a) MARRMoT model m02

(b) MARRMoT model m03

Figure 2. Wiring schematics of MARRMoT models selected for this teaching module. Schematics are reproduced from Figures S2 and S3 in the Supplement of Knoben et al. (2019) under CC BY 4.0. More in-depth model descriptions can also be found in this Supplement. Students are directed to these descriptions as part of Exercise 1.

2.3 Materials, requirements and installation

2.3.1 Provided course materials

All materials can be downloaded from https://github.com/wknoben/Dresden-Structure-Uncertainty. Provided are:

- Example exercise hand-outs, including instructions to obtain and install MARRMoT (hand-outs are also shown in Supplementary Materials S5 for convenience);
- Prepared data for the second exercise (data for the first exercisee are part of the MARRMoT install);
- An example script for model calibration for the second part of the exercise for use by educators;
- Calibrated parameter sets for both models that result in the KGE scores shown in Figure 1 for use by educators.

Note that these materials are sufficient to run the exercises with minimal effort. They do not include lecture materials to introduce the topic of model structure uncertainty. Such materials should provide a level of background and detail appropriate to the curriculum the exercises are inserted into, which will vary between curricula.

2.3.2 Software requirements

Requirements for running MARRMoT are either Matlab with the optimization toolbox installed, or Octave. MARRMoT was developed on Matlab version 9.2.0.538062 (R2017a), with the Optimization Toolbox Version 7.6 (R2017a) and tested on

Octave version 4.4.1 with the "optim" package (Knoben et al., 2019). Note that the calibration workflow example (workflow example 4) differs slightly between Matlab and Octave 4.4.1 (see Section 7 in the MARRMoT User Guide on Github for more details about running MARRMoT in Octave). There are no differences in workflow example 4 between Matlab and Octave 5.2.0, thanks to a recent update to MARRMoT (M. K. Türkeri, personal communication, 2020).

2.3.3 Install instructions

Detailed step-by-step install instructions for MARRMoT are included in our provided hand-out for Exercise 1. Briefly, download or fork and clone the MARRMoT source code on https://github.com/wknoben/MARRMoT. Next, remove the folder "Octave" if Matlab will be used. Open Matlab or Octave and ensure that all MARRMoT folders are added to the Matlab/Octave path. MARRMoT is now ready to be used. Data necessary for Exercise 1 are part of the MARRMoT install. Data necessary for Exercise 2 are part of the GitHub repository that accompanies this paper and should be distributed to the students by the teacher.

2.4 Integration in existing curriculum

Assuming the existing curriculum provides access to and instruction in either Matlab or Octave, integrating these exercises into the curriculum could happen along the following lines. The exercises would be preceded by a lecture that introduces the concept of model structure uncertainty. We direct the reader to e.g. Perrin et al. (2001), Clark et al. (2011b) and Knoben et al. (2020) and the references therein for potentially useful sources to populate lecture materials with.

Next, our two proposed exercises can be run. Broad descriptions are provided in Section 3 while ready-to-use students handouts are included as part of the GitHub repository and in the Supplementary Materials (Section S5). These exercises can be used as provided, or adapted to include more or different learning objectives. Distributing the data that underpins these exercises can either be done by referring the students to the GitHub repository that accompanies this manuscript, or by downloading the data and sharing these with the students in an alternative manner. Our example exercises include all instructions needed to obtain and install the MARRMoT source code. Students are then able to work through the exercises and will use MARRMoT to calibrate both models for both catchments, obtaining the Kling-Gupta Efficiency scores shown in Figure 1. Our proposed exercises (see Supplementary Materials S5) contain guiding questions that will help the students draw the correct lessons from a four-way comparison of these scores, so that they arrive at the learning objectives outlined in Section 2.1.

Finally, a concluding lecture can focus on how to effectively deal with model structure uncertainty. Such approaches could, for example, be (1) designing a model from the ground up for a specific combination of catchment and study purpose rather than relying on an off-the-shelf model structure (e.g. Atkinson et al., 2002; Farmer et al., 2003; Fenicia et al., 2016), (2) quantifying model structure uncertainty through the use of model inter-comparison (e.g. Perrin et al., 2001; van Esse et al., 2013; Spieler et al., 2020), (3) setting more objective limits for when efficiency scores are considered acceptable by defining benchmarks that provide a context of minimum and maximum expected model performance (e.g. Schaefli and Gupta, 2007; Seibert, 2001; Seibert et al., 2018; Knoben et al., 2020), (4) defining which model is most appropriate through evaluation metrics that go beyond the use of aggregated efficiency scores and rely on, for example, multiple metrics or data sources (e.g.

Gupta et al., 2008; Kirchner, 2006; Clark et al., 2011a, b), or (5) applying model-selection or model-averaging techniques to effectively select or combine models with the appropriate strengths for a given study purpose (e.g. Neuman, 2003; Rojas et al., 2010; Schöniger et al., 2014; Höge et al., 2019).

3 Exercises

Sections 3.1 and 3.2 contain summaries of student handouts that can be used to run the two computational exercises. The student handouts themselves contain step-by-step instructions and guiding questions that take students from the comparison of KGE scores to the intended learning objectives. The full handouts can be found in section S.5 of the Supplementary Materials to this paper. To facilitate modification by educators, PDFs and LaTeX source files of the student handouts that describe these exercises are also available as part of the GitHub repository.

3.1 Exercise 1: MARRMoT basics

It is recommended to first run an exercise on an individual basis that introduces students to the MARRMoT framework. In the example exercise that is provided as part of the module's materials, students are asked to go through MARRMoT's four workflow examples and think critically about each example and possible ways to improve it. Download and installation of the toolbox are part of the exercise. The learning objectives for this exercise are for students to:

– Gain basic understanding of MARRMoT functionality;

– Be able to calibrate a hydrologic model and create diagnostic graphics that show the simulation results.

To achieve the learning objectives, students are asked to work through MARRMoT's four provided workflows. Workflow example 1 shows an example of running a MARRMoT model from scratch, using a single catchment and a single parameter set. The example includes loading and preparing climatic forcing, selecting one of the MARRMoT models to use, defining the models parameters and initial states, choosing settings for the numerical solver, running the model with the specified forcing and settings and analysis of the model simulations with the KGE objective function and qualitative plots. The predefined parameter values in this example are not well chosen and students are asked to vary the values and see how the simulations change. Manual sampling of parameter values naturally leads into workflow example 2.

Workflow example 2 replaces the arbitrarily chosen single parameter set with a random sampling procedure, using the provided parameter ranges that are part of MARRMoT. Results are visualized through qualitative plots. Students are asked to consider if a different model structure might not be more suitable than the pre-selected model and are directed to the MARRMoT documentation to investigate which other options are available in the toolbox. Students are asked to select a different model and re-run this workflow example, leading into workflow example 3.

Workflow example 3 shows how the code can be adapted to easily run multiple different MARRMoT models, with different numbers of parameters and state variables, from a single script. The example also includes code for visualization of the ensemble simulations. The example uses a randomly selected parameter set which is unlikely to give very good simulations.

Students are asked to think of how to improve simulations and are asked to investigate the evaporation simulations as well as streamflow simulations.

Workflow example 4 shows an example of model calibration and evaluation and forms the basis for the model structure uncertainty exercises. Students are asked to adapt this script based on the code provided in workflow example 1 and 3 and are asked to consider better ways to initialize model storage values.

3.2 Exercise 2: Model structure uncertainty

This second exercise can be completed individually or in groups and gives the students hands-on experience with model structure uncertainty. The exercise asks students to calibrate two models (as introduced in Section 2.2.2) for two catchments (as introduced in Section 2.2.1), evaluate the resulting model simulations and think critically about the implications of their findings for model structure uncertainty. If working with groups, a possible approach would be to have each group work with a certain combination of model first, and bring the groups together for a discussion of their findings after. Groups will reach different conclusions based on which model and catchment they were assigned and a class-wide discussion is critical to impart the take home messages of this module because these can only be obtained by comparing the calibration and evaluation results across catchments and models (see Figure 1). The learning objectives for this exercise are for students to:

- Be able to navigate model documentation and the inner workings of hydrologic model code;
- Critically think about the relationship between model structure, catchment structure and model calibration and evaluation procedures, and in doing so arrive at the understanding outlined in Section 2.1.

As the first part of exercise 2, students are asked to familiarize themselves with the catchments and models. Catchment data are provided in the file "Part 2 - catchment data.mat" that is part of the provided materials. Students are asked to create some exploratory figures of the meteorological data and streamflow observations, and to take a look at the catchment descriptors that are provided as part of the CAMELS data set (Addor et al., 2017). Familiarity with the models is obtained by referencing the MARRMoT documentation and an initial sensitivity analysis.

Next, students are asked to calibrate both models for both catchments, using Workflow example 4 as a basis for their code. This part of the exercise can take some time, partly due to the need to setup calibration and evaluation scripts and partly due to the time needed for the optimization algorithm to converge. This makes the second exercise well-suited for a homework assignment or for a brief introduction to the balance between accuracy of the optimization algorithm and its convergence speed.

Finally, students are asked to compare their calibration results in four different ways (see Figure 1). By comparing the KGE scores of both models for catchment 08109700, students are expected to find that an inappropriate model choice negatively impacts the accuracy of streamflow simulations. By comparing the KGE scores of both models for catchment 12145500, students are expected to understand that models with very different internal mechanics can generate equally accurate streamflow simulations (in KGE terms). By comparing the performance of model m02 across both catchments, students will see that adequate model performance in one place is no guarantee that the model will perform well in a different catchment. By comparing the performance of model m03 across both catchments, students are expected to realize that good efficiency scores do not necessarily mean that the model structure faithfully represents the dominant process in a catchment, because the model can logically do so for one catchment, but not both. Students are asked to formalize these insights into three take home messages.

4 Discussion

4.1 Benefits

The main goal of the proposed exercises is to provide students with hands-on experience of the concept of model structure uncertainty. This section outlines various other benefits of using the course for both educators and students.

As mentioned, both models and catchments have been specifically selected out of much larger samples for the lessons they can convey. The exploratory work needed to do so (calibrating 40+ models for 500+ catchments) would typically be well outside of what is feasible for teaching preparation. With this selection already made, educators may spend their limited time on preparation of delivery of the course materials, without having to also spend time creating the exercises. The suggested exercises expose students to a variety of different concepts that easily transfer to other disciplines and topics, such as navigating peer-reviewed literature and model documentation, and working with open-source data, open-source software and version control through Github. Understanding of MARRMoT and model structure uncertainty can be leveraged into term projects or theses, providing students with a certain amount of modelling experience before their projects start. The recent publication of multiple CAMELS data sets covering the United States (Addor et al., 2017), Chile (Alvarez-Garreton et al., 2018), Brazil (Chagas et al., 2020) and Great Britain (Coxon et al., 2020) can, for example, provide the necessary data for such projects.

4.2 Trial application at TU Dresden

This course was run over the span of two afternoons at the Technische Universität Dresden (Germany) during June 2019. During the first afternoon, attendees attended a 1-hour seminar about model structure uncertainty and spent approximately 2.5 hours working on Exercise 1. During the second afternoon, attendees spent approximately 4 hours on Exercise 2. Students were divided into groups and each group initially worked with a single combination of one model and one catchment. The second afternoon included a final classroom discussion to tie the insights from individual groups together. The course was attended by both students (2 PhD & 4 MSc students) and faculty members (5) outside of the regular curriculum. The intent of the course was to trial prototype exercises which could potentially be included in the curriculum of the "Hydrological Modelling" module of the Hydrology Master Program. The curriculum of the Hydrology Master at the Technische Universität Dresden covers the application of numerical tools for the planning and management of hydrological and water management systems, planning and implementing measurement networks and campaigns, data analysis, working with geographic information systems and various modelling techniques. At the point when this course was/would be held, students are expected to have some basic coding experience and have seen and used a simple hydrologic model before.

MARRMoT proved to be an easy-to-use tool for this particular exercise. All participants were able to download, install and use MARRMoT within minutes, using the instructions provided in our example handouts. MARRMoT's four workflow examples proved sufficiently documented for the students to quickly grasp the basic modelling chain (data preparation, model set up, model run, analysis of simulations) and satisfactorily complete Exercise 1. Exercise 2 required students to set up and run their own model calibration scripts. Again all students were able to adjust MARRMoT's workflow examples with only minimal guidance and produce the expected results. A script showing a possible way to complete Exercise 2 is part of the provided materials for educators.

The attendees were asked to fill in a short anonymous feedback form after the course was completed. The form had several questions that had to be answered on a 1 to 5 scale and three open questions. A summary of responses (4 MSc students, 1 PhD student, 3 Postdoctoral Fellows; senior faculty members provided verbal feedback) is shown in Figure 3. Attendees unanimously reported that the course was easy to follow and complete, and that the main messages were clear. Various attendees specifically noted in their open questions that the exercises were helpful for better understanding the material covered during the seminar, showing the importance of hands-on exercises to reinforce learning objectives (Thompson et al., 2012). The number of models and catchments used in the exercises was sufficient and attendees were able to improve their understanding of the implications of model structure and parameter choice. Various attendees also noted that the initial setup for sharing modelling results of Exercise 2 between the different groups was somewhat unwieldy. Consequently, the provided example handout for Exercise 2 is set up to work for an individual student and avoids the need to define groups and share results.

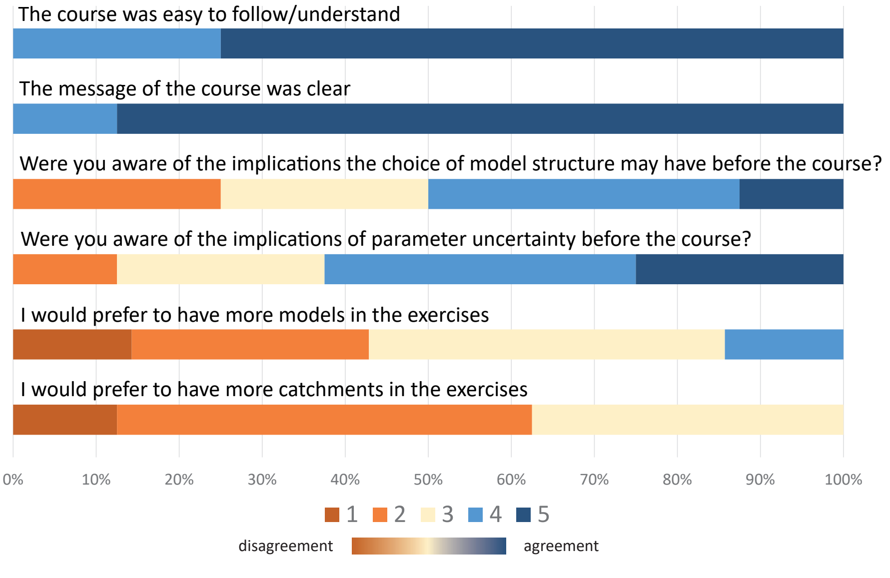


Figure 3. Responses to the feedback form distributed after the course. Only responses to the questions that had to be answered on a one to five scale are shown; responses to open questions are summarized in the main text. One indicates disagreement and five indicates agreement.

5 Conclusions

Understanding uncertainties in the modelling process is an important skill for graduates, and necessary to interpret the results from any modelling exercise. An informal survey circulated amongst educators in the earth sciences suggests that model structure uncertainty is less often part of the curriculum than data and parameter uncertainty are. This paper introduces a set of ready-to-use computational exercises that can be used to introduce the concept of model structure uncertainty to students. Running the module requires either Matlab or Octave. The module uses open-source hydrometeorological data for 2 catchments and open-source model code for 2 models, specifically selected out of a much larger sample of catchments and models for the lessons these pairings can convey. Students are tasked to calibrate both models for both catchments and to evaluate the calibrated models using data that was not used for calibration. Students are then asked to do a four-way comparison that will show that: (1) model choice matters, as in one of the catchments both models achieve very different levels of performance; (2) adequate model performance expressed as efficiency scores is no guarantee of hydrologic realism of the models, as in one of the catchments both models achieve very similar levels of performance, despite having very different internal mechanics; (3) the same applies when a single model achieves adequate performance scores in two different catchments, as logically the model may be realistically representing one of the catchments but not both; and (4) that adequate model performance in one catchment does not guarantee that this model will work well everywhere, as the performance of one of the models is very different in both catchments. A trial application of this module at the Technische Universität Dresden suggests that the module can effectively transfer these insights in the span of two afternoons. Data, model code, example exercise sheets and example code to complete the exercise are provided in a GitHub repository so that educators wanting to teach model structure uncertainty can focus on the delivery of these materials, rather than on creating them.

Code and data availability.

Catchments used in this work are part of the CAMELS dataset (Newman et al., 2015; Addor et al., 2017) which can be downloaded from https://ral.ucar.edu/solutions/products/camels. Course materials can be downloaded from Github: https://github.com/wknoben/Dresden-Structure-Uncertainty. The most recent version of MARRMoT can be downloaded from the "master" branch on Github: https://github.com/wknoben/MARRMoT

Author contributions. DS conceived the idea for the workshop that led to this publication, secured the funding and organized the workshop. WJMK selected the models and catchments. WJMK and DS created the course materials and wrote the paper. DS created the survey with help from WJMK. DS analyzed the survey results.

Competing interests. The authors declare they have no competing interests.

Acknowledgements. This work was partly funded by the European Social Fund (ESF) (grant 100270097). We are thankful for the Open Access Funding by the Publication Fund of the TU Dresden.

References

Addor, N. and Melsen, L. A.: Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models, Water Resources Research, 55, 378–390, https://doi.org/https://doi.org/10.1029/2018WR022958, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR022958, 2019.

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrology and Earth System Sciences, 21, 5293–5313, https://doi.org/10.5194/hess-2017-169, 2017.

AghaKouchak, A., Nakhjiri, N., and Habib, E.: An educational model for ensemble streamflow simulation and uncertainty analysis, Hydrology and Earth System Sciences, 17, 445–452, https://doi.org/10.5194/hess-17-445-2013, 2013.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies - Chile dataset, Hydrology and Earth System Sciences, 22, 5817–5846, https://doi.org/10.5194/hess-22-5817-2018, https://www.hydrol-earth-syst-sci-discuss.net/hess-2018-23/https://www.hydrol-earth-syst-sci.net/22/5817/2018/, 2018.

Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W. H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S., and Over, T.: Accelerating advances in continental domain hydrologic modeling, Water Resources Research, 51, 10 078–10 091, https://doi.org/10.1002/2015WR017498, 2015.

Atkinson, S. E., Woods, R. A., and Sivapalan, M.: Climate and landscape controls on water balance model complexity over changing timescales, Water Resources Research, 38, 50–1–50–17, https://doi.org/https://doi.org/10.1029/2002WR001487, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002WR001487, 2002.

Beven, K., Smith, P. J., and Wood, A.: On the colour and spin of epistemic error (and what we might do about it), Hydrology and Earth System Sciences, 15, 3123–3133, https://doi.org/10.5194/hess-15-3123-2011, 2011.

Blöschl, G. and Montanari, A.: Climate change impacts - Throwing the dice?, Hydrological Processes, 24, 374–381, https://doi.org/10.1002/hyp.7574, 2010.

Butts, M. B., Payne, J. T., Kristensen, M., and Madsen, H.: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, Journal of Hydrology, 298, 242–266, https://doi.org/https://doi.org/10.1016/j.jhydrol.2004.03.042, https://www.sciencedirect.com/science/article/pii/S0022169404002471, the Distributed Model Intercomparison Project (DMIP), 2004.

Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D., and Siqueira, V. A.: CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil, Earth System Science Data, 12, 2075–2096, https://doi.org/10.5194/essd-12-2075-2020, https://essd.copernicus.org/articles/12/2075/2020/, 2020.

Clark, M. P., Kavetski, D., and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resources Research, 47, https://doi.org/https://doi.org/10.1029/2010WR009827, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010WR009827, 2011a.

Clark, M. P., McMillan, H. K., Collins, D. B. G., Kavetski, D., and Woods, R. a.: Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses, Hydrological Processes, 25, 523–543, https://doi.org/10.1002/hyp.7902, http://doi.wiley.com/10.1002/hyp.7902, 2011b.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Earth System Science Data, 12, 2459–2483, https://doi.org/10.5194/essd-12-2459-2020, https://essd.copernicus.org/articles/12/2459/2020/, 2020.

Di Baldassarre, G. and Montanari, A.: Uncertainty in river discharge observations: a quantitative analysis, Hydrology and Earth System Sciences, 13, 913–921, https://doi.org/10.5194/hess-13-913-2009, https://hess.copernicus.org/articles/13/913/2009/, 2009.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H., Gusev, Y., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, Journal of Hydrology, 320, 3–17, https://doi.org/https://doi.org/10.1016/j.jhydrol.2005.07.031, https://www.sciencedirect.com/science/article/pii/S002216940500329X, the model parameter estimation experiment, 2006.

Farmer, D., Sivapalan, M., and Jothityangkoon, C.: Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: Downward approach to water balance analysis, Water Resources Research, 39, https://doi.org/https://doi.org/10.1029/2001WR000328, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001WR000328, 2003.

Fenicia, F., Kavetski, D., Savenije, H. H. G., and Pfister, L.: From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, Water Resources Research, 52, 954–989, https://doi.org/10.1002/2015WR017398, 2016.

Garrick, M., Cunnane, C., and Nash, J. E.: A criterion of efficiency for rainfall-runoff models, Journal of Hydrology, 36, 375–381, https://doi.org/10.1016/0022-1694(78)90155-5, 1978.

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations : elements of a diagnostic approach to model evaluation, Hydrological Processes, 3813, 3802–3813, https://doi.org/https://doi.org/10.1002/hyp.6989, 2008.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, http://dx.doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Höge, M., Guthke, A., and Nowak, W.: The hydrologist's guide to Bayesian model selection, averaging and combination, Journal of Hydrology, 572, 96–107, https://doi.org/10.1016/j.jhydrol.2019.01.072, https://linkinghub.elsevier.com/retrieve/pii/S0022169419301532, 2019.

Jothityangkoon, C., Sivapalan, M., and Farmer, D.: Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development, Journal of Hydrology, 254, 174–198, https://doi.org/10.1016/S0022-1694(01)00496-6, http://linkinghub.elsevier.com/retrieve/pii/S0022169401004966, 2001.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resources Research, 42, https://doi.org/10.1029/2005WR004362, http://doi.wiley.com/10.1029/2005WR004362, 2006.

Knoben, W. J., Freer, J. E., Fowler, K. J., Peel, M. C., and Woods, R. A.: Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT) v1.2: An open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations, Geoscientific Model Development, 12, 2463–2480, https://doi.org/10.5194/gmd-12-2463-2019, 2019.

Knoben, W. J., Freer, J. E., Peel, M. C., Fowler, K. J., and Woods, R. A.: A Brief Analysis of Conceptual Model Structure Uncertainty Using 36 Models and 559 Catchments, Water Resources Research, 56, 1–23, https://doi.org/10.1029/2019WR025975, 2020.

Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D., and Rajagopalan, B.: How do hydrologic modeling decisions affect the portrayal of climate change impacts?, Hydrological Processes, 1095, 1071–1095, https://doi.org/10.1002/hyp.10684, 2015.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, https://linkinghub.elsevier.com/retrieve/pii/0022169470902556, 1970.

Neuman, S. P.: Maximum likelihood Bayesian averaging of uncertain model predictions, Stochastic Environmental Research and Risk Assessment (SERRA), 17, 291–305, https://doi.org/10.1007/s00477-003-0151-7, http://link.springer.com/10.1007/s00477-003-0151-7, 2003.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, 2015.

Pechlivanidis, I., Jackson, B., McIntyre, N., and Wheater, H.: Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications, Global NEST, 13, 193–214, 2011.

Perrin, C., Michel, C., and Andréassian, V.: Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, Journal of Hydrology, 242, 275–301, https://doi.org/10.1016/S0022-1694(00)00393-0, 2001.

Priestley, C. H. B. and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Monthly Weather Review, 100, 81–92, https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, http://journals.ametsoc.org/doi/abs/10.1175/1520-0493{%}281972{%}29100{%}3C0081{%}3AOTAOSH{%}3E2.3.CO{%}3B2, 1972.

Rojas, R., Kahunde, S., Peeters, L., Batelaan, O., Feyen, L., and Dassargues, A.: Application of a multimodel approach to account for conceptual model and scenario uncertainties in groundwater modelling, Journal of Hydrology, 394, 416–435, https://doi.org/10.1016/j.jhydrol.2010.09.016, 2010.

Savenije, H. H. G.: "Topography driven conceptual modelling (FLEX-Topo)", Hydrology and Earth System Sciences, 14, 2681–2692, https://doi.org/10.5194/hess-14-2681-2010, http://www.hydrol-earth-syst-sci.net/14/2681/2010/, 2010.

Schaefli, B. and Gupta, H. V.: Do Nash values have value?, Hydrological Processes, 21, 2075–2080, https://doi.org/10.1002/hyp.6825, http://jamsb.austms.org.au/courses/CSC2408/semester3/resources/ldp/abs-guide.pdf, 2007.

Schöniger, A., Wöhling, T., Samaniego, L., and Nowak, W.: Model selection on solid ground: Rigorous comparison of nine ways to evaluate Bayesian model evidence, Water Resources Research, 50, 9484–9513, https://doi.org/10.1002/2014WR016062, http://doi.wiley.com/10.1002/2014WR016062, 2014.

Seibert, J.: On the need for benchmarks in hydrological modelling, Hydrological Processes, 15, 1063–1064, https://doi.org/10.1002/hyp.446, http://doi.wiley.com/10.1002/hyp.446, 2001.

Seibert, J. and Vis, M. J.: Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, Hydrology and Earth System Sciences, 16, 3315–3325, https://doi.org/10.5194/hess-16-3315-2012, https://www.hydrol-earth-syst-sci.net/16/3315/2012/, 2012.

Seibert, J., Uhlenbrook, S., and Wagener, T.: Preface Hydrology education in a changing world, Hydrology and Earth System Sciences, 17, 1393–1399, https://doi.org/10.5194/hess-17-1393-2013, http://www.hydrol-earth-syst-sci.net/17/1393/2013/, 2013.

Seibert, J., Vis, M. J. P., Lewis, E., and van Meerveld, H.: Upper and lower benchmarks in hydrological modelling, Hydrological Processes, 32, 1120–1125, https://doi.org/10.1002/hyp.11476, http://doi.wiley.com/10.1002/hyp.11476, 2018.

Spieler, D., Mai, J., Craig, J. R., Tolson, B. A., and Schütze, N.: Automatic Model Structure Identification for Conceptual Hydrologic Models, Water Resources Research, 56, e2019WR027 009, https://doi.org/https://doi.org/10.1029/2019WR027009, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019WR027009, e2019WR027009 10.1029/2019WR027009, 2020.

Thompson, S. E., Ngambeki, I., Troch, P. A., Sivapalan, M., and Evangelou, D.: Incorporating student-centered approaches into catchment hydrology teaching: A review and synthesis, Hydrology and Earth System Sciences, 16, 3263–3278, https://doi.org/10.5194/hess-16-3263-2012, 2012.

van Esse, W. R., Perrin, C., Booij, M. J., Augustijn, D. C. M., Fenicia, F., Kavetski, D., and Lobligeois, F.: The influence of conceptual model structure on model performance: a comparative study for 237 French catchments, Hydrology and Earth System Sciences, 17, 4227–4239, https://doi.org/10.5194/hess-17-4227-2013, https://hess.copernicus.org/articles/17/4227/2013/, 2013.

Wagener, T. and McIntyre, N.: Tools for teaching hydrological and environmental modeling, Computers in Education Journal, 17, 16–26, 2007.

Wagener, T., Kelleher, C., Weiler, M., McGlynn, B., Gooseff, M., Marshall, L., Meixner, T., McGuire, K., Gregg, S., Sharma, P., and Zappe, S.: It takes a community to raise a hydrologist: The Modular Curriculum for Hydrologic Advancement (MOCHA), Hydrology and Earth System Sciences, 16, 3405–3418, https://doi.org/10.5194/hess-16-3405-2012, 2012.