Teaching hydrological modelling: Illustrating model structure uncertainty with a ready-to-use teaching module

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 insight 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 teaching module that introduces students to the basics of model structure uncertainty through two ready-to-use exercises. The module is short and can easily be integrated into an existing hydrologic curriculum, limiting the time investment needed to teach this aspect of modelling uncertainty. A trial application at the Technische Universität Dresden (Germany) showed that the exercises can be completed in less than two afternoons and that the provided setup effectively transfers the intended insights about model structure uncertainty. The module requires either Matlab or Octave, and uses 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.

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

5 (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

10 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.

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

15 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

20 for some time, the principle of freely shared, self-contained teaching modules can be of great use to the teaching community.

Seibert and Vis (2012) provide a stand-alone version of the *Hydrologiska Byråns Vattenavdelning* (HBV) model that is a good example of the MOCHA philisophy 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

25 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

30 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 teaching module 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 with making this module

35 available 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 course uses two conceptual model structures applied to two carefully selected catchments to illustrate various important lessons about hydrologic modelling. The teaching module is described in more detail in Section 2 and relies on a modeling framework that can be used with Matlab and Octave. Ready-to-use exercises are presented in Section 3. Section 4.3 describes a trial application of this module at the Technische Universität Dresden,

5   Germany. Course materials can be downloaded through GitHub: https://github.com/wknoben/Dresden-Structure-Uncertainty.

2   Teaching module description

This section describes the main teaching objectives, the catchment data and models used, and an overview of provided materials, requirements and install instructions.

2.1   Objectives and outline

10   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. The catchments and models are described in sections 2.2.1 and 2.2.2 respectively.

15   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 a correct representation of the catchment under consideration. In

20   other words, the value of the efficiency score does not indicate whether a model produces "the right results 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

25   module. Instead, this module uses the common interpretation that higher efficiency scores indicate more (mathematically) 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

30   (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
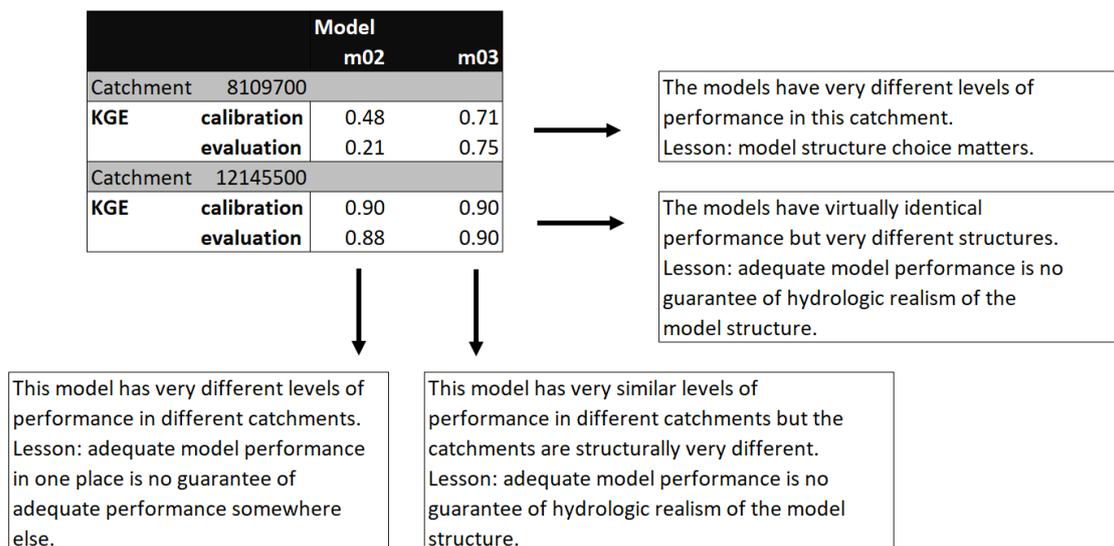
Figure 1. Using comparative assessments to transfer lessons about hydrologic model structure uncertainty.

performance of both models in catchment c8109700 shows that this is not the case: the choice of model strongly affects the accuracy of obtained simulations.

2. Models with very different structures can achieve virtually identical efficiency scores in a given catchment. Comparing the performance of both models in catchment c12145500 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). This again shows that high efficiency scores are no guarantee of having used the "right" model.

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

5 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.

10 Middle Yegua Creek is a water-limited catchment (aridity fraction = 1.3) with a corresponding low runoff ratio (0.11), 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.

15 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.

This module uses Daymet meteorological forcing data that is provided as part of the CAMELS data set (Newman et al.,

20 2015; Addor et al., 2017). Precipitation is part of the source data. Time series of potential evapotranspiration are estimated using the Priestley-Taylor method (Priestley and Taylor, 1972). Forcing data and streamflow observations for both catchments are shown in Figure 2.

2.2.2 MARRMoT models

The Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT; Knoben et al., 2019) contains Matlab/Octave

25 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

30 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 3). Both models require time series of precipitation P and potential evapotranspiration E_P as input. Briefly, model m02 is part
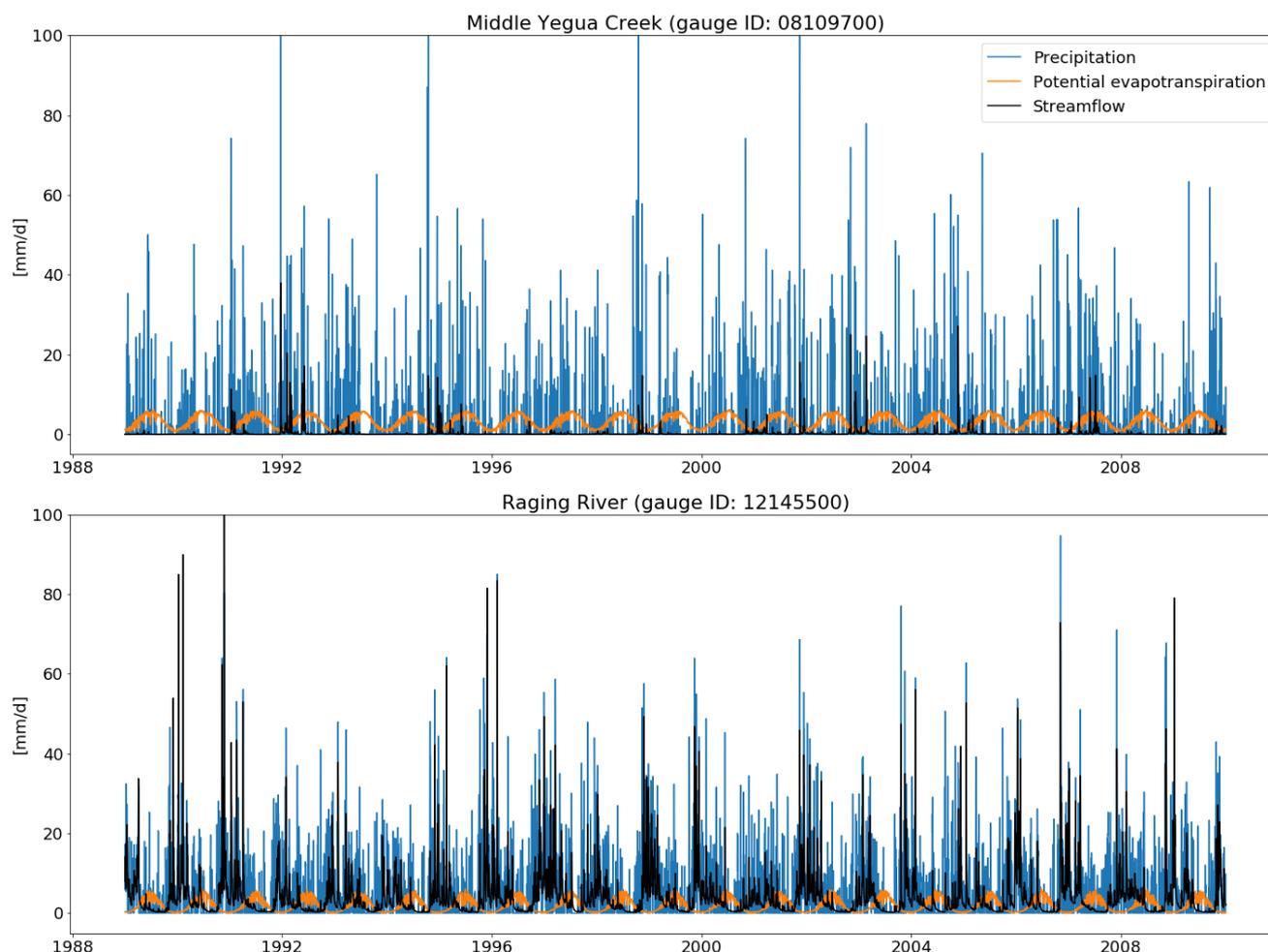
Figure 2. Time series of precipitation, potential evapotranspiration and streamflow for both catchments.

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
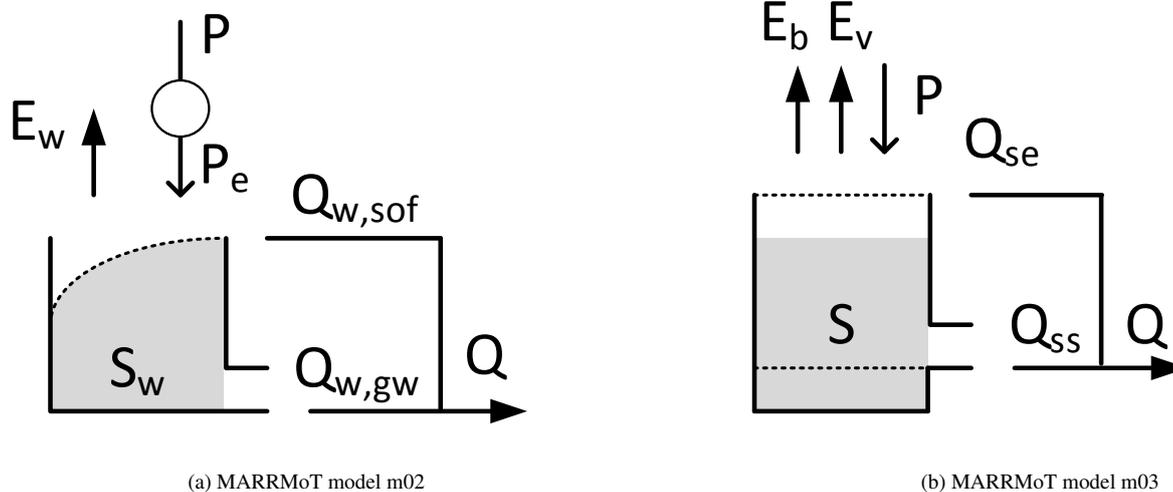
Figure 3. Wiring schematics of MARRMoT models selected for this teaching module. Schematics are reproduced from Figures S2 and S3 in
the Supplement of Knoben et al. (2019) under CC BY 4.0. More in-depth model descriptions can also be found in this Supplement. Students
are directed to these descriptions as part of Exercise 1.

2.3 Materials, requirements and installation

2.3.1 Provided course materials

The teaching module can be downloaded from https://github.com/wknoben/Dresden-Structure-Uncertainty. Provided are:

– Example exercise sheets for the first and second part of the course, including instructions to obtain and install MAR-
5   RMoT;

– Prepared data for the second part of the course (data for the first part of the course are part of the MARRMoT install);

– An example script for model calibration for the second part of the exercise;

– Calibrated parameter sets for both models that result in the KGE scores shown in Figure 1.

2.3.2 Software requirements

10   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
15   5.2.0, thanks to a recent update to MARRMoT (M. K. Türkeri, personal communication, 2020).

2.3.3 MARRMoT install instructions

Install instructions are straightforward. First, 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.

5 3 Exercise examples

The exercises described in the following sections are based on how this module was run during a 2-day workshop at the Technische Universität Dresden. These exercises are intended as examples of how this module may be used. To facilitate modification by educators, the GitHub repository that contains the module's files also contains LaTeX source files of student handouts that describe these exercises.

10 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:

15   – 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 20 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 25 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 differ-30 ent 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
5 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 both models for each catchment, evaluate the resulting model
10 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
15 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.

As the first part of exercise 2, students are asked to familiarize themselves with their assigned catchment and model. Catch-
20 ment 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 model 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.
25 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 calibration 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
30 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
5   messages.

4   Discussion

4.1   Benefits

The main goal of this teaching module is to introduce students to the concept of model structure uncertainty. This section outlines various other benefits of using the course for both educators and students.

10   This teaching module has been designed as a stand alone package and can be inserted into any hydrology course with minimal effort. Educators can thus 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
15   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   Possible follow-up teaching topics

20   This section briefly outlines several possible follow-up topics that relate closely to the material presented in this course. This module avoids the question of benchmarking model performance, that is, defining a level of efficiency scores above which model results are considered acceptable (see e.g. Schaefli and Gupta, 2007; Seibert, 2001; Seibert et al., 2018; Knoben et al., 2020). Reasons for using benchmarks and ways to implement such benchmarks can be a follow-up teaching topic. Further teaching topics could also include model development based on understanding of the catchment at hand (see e.g. Fenicia et al.,
25   2016; Jothityangkoon et al., 2001), diagnostic model evaluation (e.g. explicit focus on discovering where the model might be improved; Gupta et al., 2008), and process-based model evaluation (e.g. moving away from aggregated efficiency scores and assessing whether the model's process representations are correct; Clark et al., 2011).

4.3   Trial application at TU Dresden

This course was run as a two-day workshop at the Technische Universität Dresden (Germany) during June 2019. On both
30   days, only the afternoon was used for this course. On the first day, attendees attended a 1-hour seminar about model structure

uncertainty and spent approximately 2.5 hours working on Exercise 1. On the second day, 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 day included a final classroom discussion to tie the insights from individual groups together. The course was attended by both students and faculty members.

5   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. 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.

10   The attendees were asked to fill in a short anonymous feedback form after the course was completed. Attendees unanimously reported that the course was easy to follow and complete, and that the main messages were clear. Various attendees specifically noted 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). Various attendees also noted that the initial setup for sharing modelling results of Exercise 2 between the different groups was somewhat unwieldy. Consequently, the
15 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. The MARRMoT toolbox has been used by one of the attendees as part of their final research project, suggesting that this course is an easy way to introduce students to uncertainty concepts and hydrologic modelling in general.

5   Conclusions

Understanding uncertainties in the modelling process is an important skill for graduates, and necessary to interpret the results
20 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 ready-to-use teaching module that can be used to introduce the concept of model structure uncertainty to students. 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
25 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
30 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
5   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
10 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., 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.

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.

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., 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, 2011.

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.

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.

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 statespace 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.

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.

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.

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.

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.

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.

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
5 System Sciences, 16, 3405–3418, https://doi.org/10.5194/hess-16-3405-2012, 2012.