Author:
Knoben, Wouter

Title:
Investigating conceptual model structure uncertainty

Progress in large-sample comparative hydrology

This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Knoben, Wouter

Title:
Investigating conceptual model structure uncertainty

Investigating conceptual model structure uncertainty

Progress in large-sample comparative hydrology

By

WOUTER JOHANNES MARIA KNOBEN

Department of Civil Engineering
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Engineering.

MARCH 2019

Word count: 41,951

ABSTRACT

Hydrological modelling is complicated by several sources of uncertainty: observations of climatic forcing and streamflow will have errors, model parameters are not always properly identifiable, and the choice of model structure itself can be difficult. A model structure, i.e. the collection of equations that represent certain hydrologic states and fluxes, can be seen as a hypothesis about how a given catchment behaves. Different models are different hypotheses and choosing the appropriate model is thus important for many applications, e.g. runoff prediction in ungauged basins, climate change impact assessment, and prediction of extremes such as floods and droughts. However, many different models have been created and differences and similarities between models are not well understood. In an ideal situation, we would (1) select a representative sample of testing conditions and (2) a representative sample of models (that are (3) implemented in a coding framework that ensures an objective comparison between models can be made), (4) use an objective and extensive testing scheme to find between-model differences and similarities and thus improve our understanding of model structure uncertainty. None of these four steps are straightforward. Focussing on a sub-type of hydrologic models known as conceptual models, this thesis contributes to each of the steps in turn, although in every case many challenges remain.

The findings from any experiment are conditional on the study set-up and conclusions can only be generalised if it is clear how the tested sample relates to all possible cases. In hydrology, selecting a representative sample of catchments is difficult because it is not well known how hydrologic conditions vary across the world. Hydrologic conditions can be separated into climatic conditions and catchment attributes such as soil type and geology. Global data are available for several climatic variables that are hydrologically relevant (e.g. precipitation, temperature), but global data for hydrologically meaningful catchment attributes are more difficult to find. This thesis investigates the possibility of a hydrologically-informed climate classification and shows that a system based on three dimensionless numbers contains sufficient information to group hydrologically similar regimes on a global scale. These numbers do not account for a region's annual number of rainfall seasons, even though this strongly influences within-year hydrologic behaviour. Further analysis shows that on a global scale, the number of rainfall seasons is a relevant indicator on approximately 7% of the Earth's land surface. These results give us some idea of how representative a sample of catchments is of the global range of hydro-climatic conditions. An open challenge on this topic is expanding our approach to a global catchment classification that includes catchment attributes as well as climatic conditions, but, for now, data limitations place that out of reach.

In a model-comparison study, the next challenge is selecting a representative sample of models. There is however a wide variety of models available and no clear basis for defining both the total model space (i.e. all possible model configurations) and a representative selection within that space (i.e. several model structures that are in some way considered to cover different areas of the total model space). This thesis assumes that an iterative approach is appropriate, where a large number of models is used in a first model-comparison study. In a next pass, results from the first round can inform model selection to reduce the total number of structures and identify whether the first model sample leaves any obvious gaps in the total model space uncovered. To facilitate objective model comparison, a new open-source modelling framework has been created that currently includes 46 different conceptual hydrological models. The framework follows several best practices in model development and through its modular nature provides a stable ground for model-comparison studies. Each modelling decision can be isolated in turn and its impact assessed. Expansion of the framework is straightforward, so that it can easily be used for iterative model-comparison studies.

Any large-sample study must necessarily sacrifice study depth to some extent, in return for larger sample sizes. This thesis investigates model structure uncertainty using 36 conceptual models, three different objective functions and 559 catchments (for which time series of climatic forcing and a variety of catchment attributes are known). To keep the analysis manageable, for each combination of model, catchment and objective function, only a single parameter set is calibrated. For the majority of cases at least two but up to 25 models achieve very similar efficiency values, although these are not the same models for each catchment. Contrary to expectations, there is no obvious relation between the number of model parameters and either calibration performance, evaluation performance and performance change between both periods. Instead, the model structure seems to dictate whether a model will do well for a given objective function and whether it will perform better or worse (relative to the other models) under certain flow regimes. The relation between model performance and catchment attributes is inconclusive, but models can be relatively neatly grouped based on their (weak) correlation with catchment attributes. This suggests that certain shared model structure elements lead to similar model performance in specific types of catchments, although this hypothesis remains untested.

Taken together, several important steps towards a comprehensive and generalizable model-comparison study have been made in this thesis. A variety of challenges remain, from the need for a catchment classification scheme to a combined assessment of model structure, data and parameter uncertainty. However, these are grand challenges which the hydrologic community has been working on for several decades. The results presented here contribute to progress towards these goals and indicate several promising, practical next steps that can be taken in future work.

People look down on stuff like geography and meteorology, and not only because they're standing on one and being soaked by the other. They don't look quite like real science†. But geography is only physics slowed down with a few trees stuck on it, and meteorology is full of excitingly fashionable chaos and complexity. And summer isn't a time. It's a place as well. Summer is a moving creature and likes to go south for the winter.

† That is to say, the sort you can use to give something three extra legs and then blow it up.

TERRY PRATCHETT, FEET OF CLAY

With the formalities out of the way, let me say that none of this work would have happened as it did without the generous support of my supervisors, Ross Woods and Jim Freer. You've always given me the right balance between support and freedom but, most importantly, you've also made this whole thing fun. I don't think there has been a single meeting where we didn't have a laugh about something and I appreciate that enormously. Special thanks go to Ross, for your mentorship during those moments where I did struggle and for the thorough care with which you've gone through what I expect is well over 100,000 words during the past 4 years. Jim, thanks for keeping me aware of the bigger picture when I tend to get lost in interesting details and for your genuine care about my project and well-being.

I greatly appreciate Murray Peel for hosting me for three months in Melbourne and for making time to answer my incessant questions. Keirnan Fowler, thanks for your great input into my work and for how you and Kelly made me feel welcome in the boiling heat of Australian summer. Barney Dobson, thanks for being a solid friend from the very first day we met, for helping me navigate British society and for the countless questions you've answered over the years. Wouter Berghuijs, thanks for your interest in my work and progress, and for introducing me to the wider hydrologic community. To my friends and colleagues in Bristol, Melbourne and other places, thanks the fun times and the good chats.

To my family, thank you for your never-ending support and belief that I could do this. I still don't understand why you all seem to think this was a good idea, but I'm glad that you do. Last, and most, I need to mention Marijke Hesemans. You encouraged me to start this thing even though it meant being apart for a long time, which turned out to be the right thing to do. You've overcome so much in moving here so we could be together again, which turned out to be the right decision. And you're willing to join me for the next step in this adventure, which I hope will turn out to be even better than anything that has come before. I'm more thankful for your love and support than I know how to say.

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE: ....

LIST OF FIGURES

INTRODUCTION

Sections 1.1.1 and 1.1.2 are based on a summary of discussions with Rosie Lane, Ida Westerberg, Jan Seibert, Stan Schymanski, Charlie Luce and Martyn Clark.

1.1 Background

Hydrology is a science that revolves around a simple question: what happens to rain [252]? Simple as this questions sounds, answers to it are often vague. Experiments in the field and laboratories can improve our understanding of the real-world processes that define hydrologic behaviour, and ever increasingly, computer models are used to complement this understanding. These models serve a variety of roles that can be broadly categorized into virtual laboratories [345] and decision support systems [40]. As the former, models are used to improve our understanding of the real world by testing assumptions or hypotheses based on field or laboratory observations. As the latter, models are used to provide simulations that inform a variety of decisions, ranging from mitigation of hydrological extremes (i.e. floods and droughts), to long-term planning of water resources, to enforcement of water quality standards. Both approaches involve extrapolation into the unknown. The measurements that underlie our models are limited in time and space [e.g. 34] but humanity's need for accurate hydrological predictions extends into situations where measurements are unavailable (ungauged catchments) and where measurements are impossible (the future) [40]. It is therefore critical that the tools which we use to make predictions, i.e. our hydrological models, are well understood and appropriate for the question they are applied to.

1.1.1 The model development process

Different types of hydrologic computer models exist and they can be broadly categorized in various ways. We can for example distinguish between lumped and distributed models (depending on their spatial discretisation), between conceptual and physics-based models (depending on whether the equations are more empirical or describe the physics of the system), between deterministic and stochastic models (depending on whether there is one outcome or several with associated probabilities) [40], and various other distinctions are possible as well [248]. Distinguishing between models is generally not quite as clear cut as these definitions might imply and there are for example semi-distributed models that divide a catchment into several different elements (e.g. through topographic indices [49]), without resorting to a fully distributed approach (e.g. THALES, [124]). Similarly, many models blur the line between conceptual and physics-based models [e.g. 194], and to a certain extent all models are arguably empirical because even the physics-based models are simplified representations of reality. Regardless, in all cases model development follows (or at least, should follow) the same basic steps.

Many authors have summarized the model development steps [e.g. 37, 40, 72, 128]. There are differences in terminology and in the number of steps they define, but they agree on the general structure of the modelling process. First, the modeller develops a *perceptual model* of a catchment, which is a personal and potentially very complex understanding of how the catchment responds to climatic forcing. The perceptual model is not fixed and changes with increased understanding of the system (see for example the change in understanding of the experimental Maimai catchment in New Zealand as more studies are conducted [210]). Next, the modeller develops a *conceptual model*, which is a simplification of the perceptual model that describes the most relevant processes and their interactions. Based on this, the modeller creates a *mathematical model*. This is a definition of the state variables and units, and of the equations that describe the changes in state variables (i.e. the mathematical functions that represent the hydrological fluxes described in the conceptual model). The modeller translates the mathematical model into computer code and specifies a way in which the model equations are solved, creating a *numerical model*. After this, any model parameters are assigned a value either through parameter estimation (if the parameters are assumed to relate to measurable real-world phenomena) or calibration (i.e. adjusting parameter values in order to optimize a score on a given (set of) objective function(s)). For an application of this process, see for example the steps taken to create models for several catchments in Luxembourg [104–106] or those that inform the landscape-specific FLEX-Topo models [274].

1.1.2 Models as hypotheses

Regardless of the final shape of the model, any hydrological model is in essence a hypothesis about how a given catchment works. However, it is not straightforward to map our understanding of a catchment (i.e. the perceptual and conceptual models) onto a model structure (i.e. the mathematical model) [35, 36, 73, 280]. There is no agreement on which hydrologic processes are important (in general, and for a given study purpose such as floods or droughts, because which processes are important is often dependent on the catchment), nor on which type of model (e.g. empirical or physics-based, lumped or distributed) should be used, nor on which specific equations should be used to represent a given hydrologic process (see e.g. the enormous variety of evaporation [211] and snow equations [97]). Many models have been created [298, 337] and for any given place several different model structures tend to emulate observations of streamflow equally well [47, 112, 254], a problem termed *equifinality*. However, simply reproducing observations sufficiently accurately is not enough. A model must produce the right results because it is an appropriate representation of the catchment under consideration. Otherwise, the model is nothing more than a "mathematical marionette", reproducing streamflow observations it has seen during model calibration but not to be trusted when streamflow simulations are required that extrapolate beyond calibration conditions [173, 175].

Unfortunately, models of open systems (such as a river basin) can never be validated, i.e. proven to be correct [239]. However, models can be tested under a variety of conditions and the results tell us something about the appropriateness of the assumptions that underlie the model. Klemeš' four-step model evaluation framework [176] assesses how generally applicable a calibrated model is. The tests show how the model performs under climatic conditions that are different from those during calibration, how the model performs in a similar climate but different catchment, and how it performs in a different catchment with different climatic conditions (for an example application, see [265]). Similarly, in the case of model structure equifinality, different models can be treated as alternative working hypotheses [38, 73] that can be subjected to a variety of tests. If a model fails an evaluation test, we can use this as an opportunity for learning [13, 38]: through diagnosing the reason why the model failed, improvements to the model are possible [83, 90, 131]. If a model passes our evaluation tests, we have increased our confidence in the model's ability and we can provide better justification for using the model during our research and work.

1.1.3 Modelling uncertainty

Model performance is influenced by various sources of uncertainty and these complicate our search for finding the most appropriate model for a given place. Three main sources of
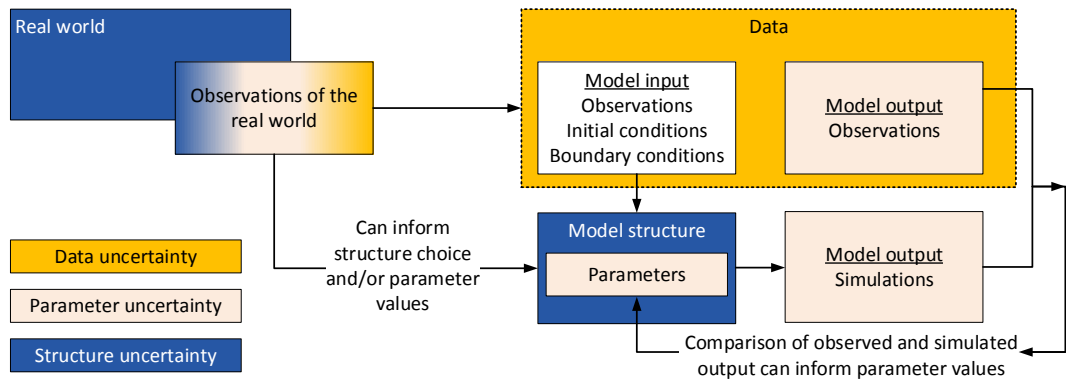
FIGURE 1.1. Overview of the main uncertainties influencing hydrologic modelling. Observations of the natural world inform all our modelling efforts, but do not necessarily fully overlap with reality. Data uncertainty, parameter uncertainty and model structure uncertainty are consequences of this mismatch between reality and observations of reality.

uncertainty can be distinguished [e.g. 40, 43, 214, 248, 335]: data uncertainty (including uncertainty in initial and boundary conditions), parameter uncertainty and model structure uncertainty (Figure 1.1). Further uncertainties are natural variability [248] and "unknown unknowns" [91] which we ignore for the moment, because these are respectively not directly related to the modelling process and impossible to quantify. There is considerable debate within the hydrological community about the nature of uncertainty and appropriate terminology [e.g. 41, 227]. This section contains a practical, rather than philosophical, discussion about the uncertainties that complicate hydrologic modelling.

Data uncertainty originates from inaccurate or incomplete measurements of the real world. Data uncertainty can impact both model use through input data, initial conditions and boundary conditions, and model set-up when it influences model calibration [e.g. 43, 248]. Models generally require observations of the real world (e.g. time series of climatic forcing, digital elevation maps showing topography) in order to be used. However, these measurements are fragmented in space and time (not everything is measured everywhere and always), have their own unique properties (e.g. relating to the relevant time scale for observations and time-variable error properties of different observations) and are also susceptible to a variety of errors [214]. For example, the natural world is continuous but measurements might be taken at fixed locations (e.g. rain gauges) and spatial heterogeneity is thus not fully represented in the measurements. This can lead to a mismatch in scale between available data and model domain. The catchment is an

often used scale of interest for models, but data might only be (accurately) available in a certain radius around each measurement station, measurement stations themselves tend to be scattered with unmonitored space between them, and measurement stations are completely absent in the vast majority of locations [55]. Measurement conditions can be non-stationary and introduce bias (e.g. a stream that overflows its banks and thus partly bypasses the discharge measurement station during high flows leads to underestimation of high flows). There may also be malfunctions of the equipment and human-induced errors [214].

Parameter uncertainty originates from difficulties with estimating the most appropriate parameter set for a given model and catchment. Ideally, parameter values would be estimated directly from field measurements [e.g. 309] but effectively translating field observations to model parameters has not yet been realised (partly due to the scale mismatch of data availability and model domain mentioned before). Model parameters are therefore often calibrated: i.e. parameter values are iteratively changed until a sufficient similarity between observations and model simulations is obtained (or some other convergence criterion is reached). This weakens the connection between model behaviour and reality, because there is no guarantee that the obtained parameter values relate to the actual conditions in the catchment. In the best case scenario, parameter sets are identifiable [94, 332], unique [34, 94], and stable [94]. Identifiability indicates that each parameter has a clearly defined optimal value. If this is not the case, the model might be insensitive to this parameter's value or the parameter's optimal value might be difficult to determine due to interactions with a another parameter. In either case, it is difficult to determine a parameter set that fulfils the uniqueness criterion. Uniqueness requires that the data constrains parameter values into a single time-invariant parameter set for a catchment (i.e. parameter values that are not conditional on the calibration conditions). Stability indicates that (small) errors in the input data (sometimes referred to as dis-informative data [51]) should not have disproportionate (large) impacts on the optimal parameter values. Further complications arise through over-parametrisation [10, 254], in cases where the degrees of freedom of the model (i.e. the number of parameters) are higher than what can be robustly estimated from the data. Parameter estimation is also affected by the geometry of the model's parameter and objective function space, which can be unnecessarily complex due to non-robust numerical implementation of the model [72, 150, 160, 167], and the choice of optimization method used to search this space [16, 248].

Model structure uncertainty originates from lacking or incomplete knowledge about how the real world should be represented by a model, and from the fact that even the most complex model will still be an abstract representation of reality [239]. Even though ideally the conceptual model of a catchment is based on a perceptual model, the perceptual model

5

is often based on local observations that cannot directly explain catchment-wide behaviour. This makes it difficult to relate local observations, catchment-averaged descriptors and model behaviour [104]. It has been suggested to instead focus on emergent processes at the catchment scale rather than scaling up local observations [6, 53]. However, incomplete knowledge about hydrologic processes, whether they are local and detailed or catchment-wide emergent processes, makes it difficult to represent them in a model. Examples are the functioning of preferential flow paths and travel time distributions [39], the often high non-linearity between precipitation and runoff [53] and the apparent non-stationarity of hydro-meteorological systems [53, 220, 333]. Non-stationarity in particular leads to modification of the physical structure of the system over longer time scales and the model structure that best represents the catchment can thus change over time [113, 128]. There is also little agreement on how different hydrologic processes can be mathematically represented [104, 148].

Summarizing, the availability and accuracy of observations in a catchment influence our view of how a catchment works and thus which processes we think a model of the catchment should include. However, we are usually unable to directly measure any parameters used in mathematical descriptions of these processes and must thus calibrate the model parameters instead. Unfortunately, obtaining appropriate parameter values through calibration is often difficult (in part due to data uncertainty) and we are thus faced with a trade-off between a model that uses as few parameters as is justified by the data, and a model that uses enough parameters to include all the hydrological processes we think are relevant. Data availability is limited for many places and the number of required model parameters (for all important processes) regularly exceeds the number of parameters that can be robustly estimated. This leads to a situation where models can be calibrated to optimize a score on a given objective function but this mathematical optimum is not necessarily the hydrological optimum [10, 109, 148]. That is, model calibration gives parameter values that maximize some efficiency metric, but are not necessarily a representation of catchments processes (conditional on the - doubtful - assumption that model parameters relate to catchment processes in the first place). However, because models are intended to be a representation of reality, a good model should produce the right results because it accurately represents the catchment processes (i.e. catchment-appropriate parameter values), rather than being a chance result of a mathematical curve-fitting exercise (i.e. parameter values that maximize some efficiency score) [37, 73, 125, 148, 173, 176].

The impacts of multiple uncertainties on the modelling process are generally assessed through sensitivity analysis (SA) and uncertainty analysis (UA) methods. The main benefits of SA and UA include include being able to learn about the data being used [51, 159], learn about the model as a whole [106] and to produce more reliable and robust simulations by

Data   Tools   Methods   Results

| **1.** Representative sample of testing conditions |
| **2.** Representative sample of models | → | **3.** Model-comparison framework | → | **4.** Methodology (research questions, experimental protocol, account for data & parameter uncertainty) | → | **5.** Understanding of model structure uncertainty |
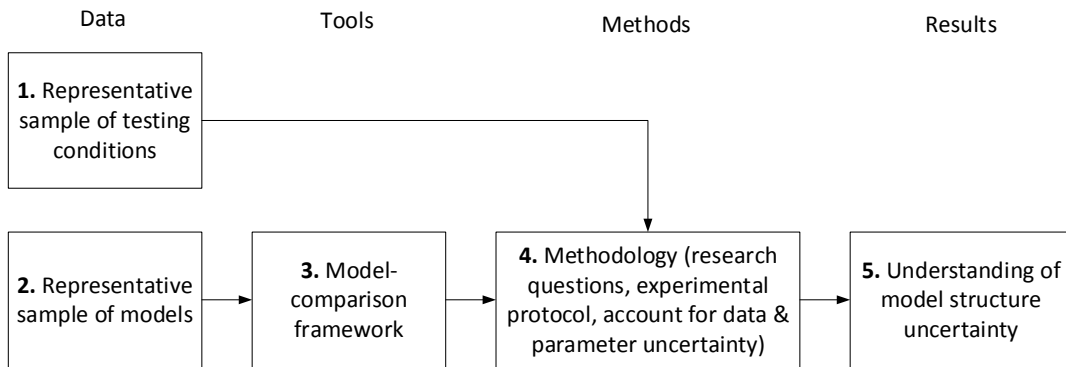
FIGURE 1.2. Set-up of the ideal model inter-comparison study. This framework results in generalizable understanding of model structure uncertainty.

providing ranges of possible simulations (through ensembles of parameter sets, or models, or otherwise) rather than deterministic values [159]. These methods can for example be used to quantify the uncertainty in input data [e.g. 165, 166, 214, 267], streamflow observations [e.g. 213, 214], parameters [e.g. 43, 47] and model structure [e.g. 128, 267, 325, 330]. For the specific purpose of understanding model structure uncertainty and the differences between various model structures, multi-model comparison frameworks have been developed. Examples include the Modular Modeling System [MMS, 191], the Rainfall-Runoff Modelling Toolbox [RRMT, 331], the Framework for Understanding Structural Errors [FUSE, 78], SUPERFLEX [103, 162] and the Structure for Unifying Multiple Modeling Alternatives [SUMMA, 75, 76].

1.2   Problem analysis

As evident from Section 1.1.3, the three uncertainty aspects that complicate hydrologic modelling have received considerable attention. However, it is not yet common practice to address model structure uncertainty in hydrologic studies, there is no clear guidance that helps choosing an appropriate model, and in many studies the choice for a given model is motivated by concerns other than the appropriateness of the model for a given catchment and study purpose [2]. It is known *that* different model structures behave differently, but *how* differences in model structures lead to differences in simulation capabilities is currently not well understood [73, 75, 78, 128] and research on this topic so far has been fragmented and lacks an overarching methodology [128].

Figure 1.2 shows a general set-up of a model inter-comparison study, which would increase our understanding of model structure uncertainty. Because it is logically impos-

sible to test all possible combinations of models and catchments [227], we must perform studies using a reduced sample of models and catchments. For generalizable results, this sample must be representative of the full range of catchments and models we are interested in. Additionally, the methodology must ensure that our research questions can be unambiguously answered. This ideal framework therefore specifies the use of a representative sample of catchments and models and accounts for other sources of uncertainty that might interfere with the assessment of model structure uncertainty. Many models exist [e.g. 298] and various authors have investigated between-model differences [e.g. 88, 98, 111, 143, 185, 222, 254, 293, 325]. These studies have taken place over a wide range of catchments, but little is known about how well these catchments cover the global range of hydro-climatic conditions. No study as of yet has been able to follow a framework as shown in Figure 1.2 and model structure uncertainty assessments have been quite ad hoc and poorly reproducible [65]. This thesis focuses on structuring this process and clarifying model structure uncertainty, specifically that of lumped conceptual hydrological models (also known as bucket models).

1.2.1   Representative sample of testing conditions

Box 1 (Figure 1.2) states that a representative sample of testing conditions is required. It is prohibitively expensive to test model performance using the entire world as a sample. In many cases it is also impossible to evaluate the model simulations against observations, because the observations simply do not exist: climate data is generally available at the global scale but not necessarily at the spatial or temporal resolution of interest, streamflow observations are unavailable for many catchments worldwide, and hydrological states and fluxes inside the catchment (e.g. extent of saturated catchment area, actual evaporation) are rare. We thus need to select a subset of the world where the necessary observations are available as our testing sample, such that we can extend findings from this subset to the wider world. However, hydrology struggles with organizing the catchments we work in in a meaningful way. Studies generally report some characteristics of the catchment(s) under consideration (e.g. mean precipitation, average slope) but we have no framework that tells us the full possible range of these characteristics, nor is it clear which of these characteristics are meaningful to report. We need a catchment classification system that lets us structure and organize results from different studies, but so far this does not exist [209, 334].

There are two distinct challenges in creating a catchment classification. The first is the lack of high-quality global data on potentially relevant characteristics. Hydrologic similarity can be expressed as a 3-dimensional space with axes for climatic forcing, catchment

structure and catchment response [334]. Climatic data are commonly available on the global scale, although the spatial and temporal resolution varies [e.g. 226]. Measurements of streamflow response are increasingly available on larger scales, but are by no means available everywhere [e.g. 93, 127, 317], and measurements of water quality and internal catchment states and fluxes are rare outside of densely monitored experimental catchments [e.g. 64]. Data on the catchment structure (e.g. local topography, geology, vegetation, soils,) tend to suffer from large uncertainties and are fragmented across different institutes [e.g. 4].

The second challenge is a lack of understanding of what makes catchments different or similar. At a certain (high) level of detail, all catchments are structurally different from each other (i.e. "uniqueness of place" [34]). However, this is in part an issue of scale and relevance, because not all differences are meaningful at the scale of interest. Instead of looking at only the catchment structure, emergent processes have been suggested as an important feature of defining catchment similarity [145].

There is currently little theory underlying the choice of metrics used to summarize a catchment's climatic conditions, catchment attributes and streamflow response, and how these three aspects influence each other. Recent studies have relied on ad hoc selection of data and used statistical clustering methods to generate groups of similar catchments [e.g. 79, 187, 275, 276]. These results show that catchments can exhibit similar streamflow behaviour but the relation with catchment attributes is not conclusive [3]. Given the lack of clarity on this subject, we limit the scope of this thesis to catchments with limited human influences. The presence of dams, reservoirs, abstractions etc. interferes with the natural tendency of hydrological processes and obscures the relationship between climatic forcing, catchment structure and catchment response. Although predictions of future water resources are most critical for regions with human populations, understanding hydrologic similarity in natural catchments is an easier place to start.

1.2.2 Representative sample of models

Box 2 (Figure 1.2) states that a representative sample of models is required. We already constrain the scope of this thesis to lumped conceptual models only, but this still leaves an enormous variety of models too choose from. Several authors have attempted to create overviews of hydrological models in books [e.g. 296–298] and wiki-pages [e.g. 341] but the simple conclusion is that maintaining such a list is impossible. There are too many models [337] and even though we can in theory perfectly know and define the "model space" [37], in practice this is impossible.

Therefore, selecting a representative sample of models for a model inter-comparison

study is not straightforward. Ideally, we would create a model classification scheme akin to the catchment classification scheme mentioned earlier. However, model classification faces the same challenges as catchment classification: we don't have the data to quantify the full "model space" and we don't understand well enough what makes models similar or different in a practical sense [73, 75, 78, 128]. It is also highly plausible that whether models are functionally different depends on the testing conditions used. In a climate with no rain and runoff, every model that produces no runoff is effectively the same. In a similar vein, in a snow-dominated catchment a model with a snow component will be more appropriate than a model without one, but in a catchment without an annual snow pack these two models might be functionally equivalent. Of course these examples are obvious, but they do highlight that what consists of a representative sample of models is partly conditional on the climatic conditions under consideration. It is plausible that this reasoning also applies to catchment attributes and the objective functions used to quantify model performance. It is thus critical to frame model selection both in terms of model space (i.e. which models are available) and in terms of the conditions the models are intended to represent.

No such framing currently exists and lacking an objective way to select representative model samples, model comparison studies to date have generally used a large selection of existing model structures [e.g. 254, 293], ensemble approaches that decompose a few different model structures into several constitutive components and combine these into various new models [e.g. 78, 335] or relied on generic model elements (i.e. fluxes, stores, delays) to generate new model structures [e.g. 103, 184]. Other approaches include step-wise model building (expanding a simple model) [e.g. 309] or stepwise model breakdown (simplifying a complex model) [e.g. 154]. These approaches have generated various new insights, but these focus mainly on the fact *that* models are different and that model choice can strongly influence study conclusions. We are not yet near true understanding of *why* models respond differently to climate inputs [73, 75, 78, 128] and whether models can be grouped according to their respective response to inputs. The main challenge remains unchanged: it is still unknown what the most appropriate model structure for a given problem is [78, 128].

1.2.3 Model comparison framework

Box 3 (Figure 1.2) indicates a need for model comparison frameworks. These frameworks are used to ensure objective testing of model structures. They provide the flexibility needed to isolate individual modelling decisions and to test different hypotheses about model behaviour [78, 103]. Most importantly, they ensure that the comparison is objective, by standardizing those elements of model set-up (e.g. coding of individual model elements,

numerical implementation, calibration algorithm) that can create artificial differences between models if they are not strictly controlled for. Using a model comparison framework ensures that any differences in model outcome can be attributed to only differences in the model structures [65, 103].

Various model comparison frameworks have been created. Obvious candidates for use in this work are RRMT [331], FUSE [78] and SUPERFLEX [103, 162] because each incorporates a suite of coded conceptual models, which are the focus of this thesis. At the same time, each framework has its own limitations which might reduce their usefulness for the work here. RRMT does not separate model equations (i.e. the Ordinary Differential Equations that describe the change in model stores) and the numerical approximation used to solve these equations. The chosen numerical scheme is Explicit Euler, which is a common choice in hydrologic modelling but has several drawbacks [e.g. 72, 160, 285]. However, the combined implementation of model equations and numerical scheme requires a thorough overhaul of RRMT if a different numerical approximation method were to be used. FUSE and SUPERFLEX avoid these numerical complications. FUSE uses four "parent models" to provide equally plausible modelling options for the store configurations in the upper and lower soil layer, and for a variety of fluxes. The user chooses which equations to combine into a new model. However, the overall model structure is fixed (into an upper soil layer connected to a lower soil lower layer) and expanding this to allow for a wider variety of model structures requires significant changes to the code. Similarly, SUPERFLEX uses a single "master model" in which the user turns components on or off. This master model is based on a given hypothesis of how hydrologic behaviour can be expressed as a model and changing the code to allow different hypotheses (e.g. introduce a new connection between two stores that is not in the master model) would again require significant work. Summarizing, the choice of model comparison framework is as much dependent on the research questions one is interested in as the choice of any model is. Different frameworks are suitable for different questions and the final choice requires careful consideration.

1.2.4 Methodology that accounts for other uncertainty sources

Box 4 (Figure 1.2) specifies some concerns regarding the study's methodology. As mentioned, three main sources of uncertainty influence hydrological modelling: data uncertainty, parameter uncertainty and model structure uncertainty. To properly investigate model structure uncertainty, we need a way to mitigate the impacts of data and parameter uncertainty. Various methods have been developed that do this [e.g. 185, 214, 248]. However, a major concern in the ideal study set-up is the practical impact of large sample sizes. Uncertainty analysis generally involves increased sampling of e.g. different parameter

values or permutations of the input data. Combined with our aim to use a representative sample of catchments and a representative sample of models, available time and our (limited) ability to summarize, comprehend and visualize the study results are real concerns. This is a known limitation of large-sample studies, where one has to sacrifice study depth (e.g. properly account for parameter and data uncertainty) for study breadth (e.g. using a large sample of places and models) [130]. The objectives in this thesis are thus necessarily shaped by what is reasonably possible in a limited amount of time.

1.3   Aims and objectives

Better understanding of the models we use to predict future water resources is critical but we have identified that the ideal model comparison study that could facilitate this understanding is currently out of reach. The aim of this thesis is to address some of the current shortcomings in large-sample model comparison studies and improve our understanding of model structure differences. We limit the scope of this thesis to lumped conceptual hydrological models and to catchments with limited human influences. Ultimately, we hope to contribute to answering a critical question in hydrologic modelling: what is the most appropriate model structure for a given objective? This aim divides the research in this thesis into different parts, centred around the following broad questions:

1. What is a representative sample of testing conditions?

2. What is a representative sample of models?

3. How can we ensure a flexible framework for testing conditions?

4. What can we learn about model structure uncertainty?

It is unlikely that we can resolve all challenges involved and we can therefore only address these questions in part. The following research objectives contribute to answering these questions:

- Given that data is lacking for a full catchment classification but that climate data is widely available, create a way to quantify the global hydro-climate. This will give us a meaningful way of assessing how representative any given sample of catchments is compared to the global variety of hydro-climatic conditions.

- Review the currently available conceptual hydrologic models and create a testing framework that allows objective comparison of different lumped model structures.

Given that (too) many models currently exist, it is unlikely that our review can cover all relevant model structures. Therefore our testing framework must be available to, and extendible by, other researchers.

- Perform an exploratory large-sample model comparison study, (i) quantifying model structure uncertainty across a wide range of catchments, (ii) quantifying differences between the models in our sample in simulating streamflow under a range of different conditions and objective functions, and (iii) connecting both aspects to investigate whether certain models are inherently better suited for certain places or for certain objectives and why that is the case.

1.4 Dissertation outline

Figure 1.3 summarizes the contributions of this work in the context of the ideal model structure comparison study in Figure 1.2. Contributions of individual chapters are briefly described in this section.
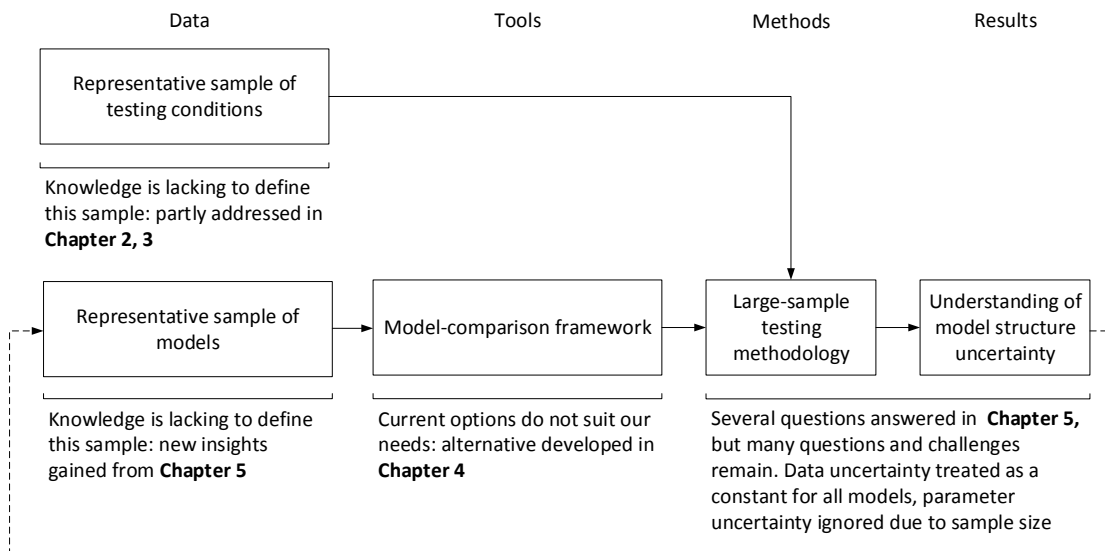
FIGURE 1.3. Overview of how the work in thesis contributes to improving our ability to perform large-sample model structure comparison studies. Because defining a representative sample of models is currently not possible, an iterative procedure might be adopted where results based on a large sample of models inform the selection of a smaller representative model sample in future work.

Chapter 2: Hydrologic climate classification This chapter introduces a new hydrologically-informed climate classification scheme. This climate classification scheme shows how climatic forcing can be summarized in an informative but parsimonious way, shows how global streamflow regimes relate to the global climate patterns, and can be used to quantify how much of the global hydro-climate any sample of catchments covers.

Chapter 3: Global precipitation modality Our hydrologic climate classification does not specifically account for the number of rainfall seasons regions experience. This chapter investigates how common the occurrence of more than one rainfall season is on a global scale, to assess whether this is a major limitation of the scheme developed in Chapter 2.

Chapter 4: Multi-model comparison framework The modelling process is complex and models can only objectively be compared in a model-comparison framework that ensures that all models are coded in the same way. This chapter introduces a new open-source modelling framework for this purpose. It is based on a review of conceptual hydrologic models and currently includes 46 different model structures.

Chapter 5: Model structure uncertainty The new modelling framework is used as part of a large-sample study. This chapter shows the results of calibrating 36 models for streamflow simulation in 559 catchments that cover a wide variety of hydro-climates, using three different objective functions that focus on high, low and combined flow simulation.

Chapter 6: Conclusions and outlook This chapter summarizes the work performed and outlines several promising avenues for further research.

Appendix A Supporting analyses for Chapter 2, showing the result of manual quality control of each of the 1103 catchments used, development of an empirical variant of the standard Wilcoxon statistical test, extended significance testing results and an overview of the geographical spread of the 1103 catchments.

Appendix B Supporting documentation for the modelling framework presented in Chapter 4, giving detailed descriptions of the equations that make up each of the 46 models, an overview of how these model equations are implemented as re-usable modular computer code, a similar overview for the 7 different routing schemes implemented in the modelling framework, and tables with generalized parameter ranges that can be used to ensure fair model comparison studies.

Appendix C Supporting information for Chapter 5, showing the changes in climatic conditions between model calibration and evaluation periods, additional figures that repeat

analysis done in the main chapter for different objective functions, and correlation figures that summarize the relationships shown in the individual scatter plots.

Appendix D Additional publication that appeared as a blog post on the website of the Young Hydrologic Society.

Appendix E Author's CV.

1.5 Note on definitions

Confusingly, within the hydrologic literature two different meanings are attached to the phrase "conceptual model". The first uses the term to mean a distinct step in the building process of any model: it is used to either mean the theoretical understanding of how a given catchment functions [e.g. 128] or the collection of equations that quantify this understanding [e.g. 40, 72]. The second refers to a specific type of hydrological model [e.g. 167, 309], also know as bucket-style models [e.g. 301], with at their core a soil moisture accounting routine (they are in fact also occasionally referred to as Explicit Soil Moisture Accounting models [p.16 in 40]). In this thesis, "conceptual model" is used to mean bucket-style hydrologic models, unless explicitly stated otherwise.

HYDROLOGIC CLIMATE CLASSIFICATION

Abstract

Classification is essential in the study of natural systems, yet hydrology has no formal way to structure the climatic forcing that underlies hydrologic response. Various climate classification systems can be borrowed from other disciplines but these are based on different organizing principles than a hydrological classification might need. This work presents a hydrologically informed way to quantify global climates, explicitly addressing the shortcomings in earlier climate classifications. In this work, causal factors (climate) and hydrologic response (streamflow) are separated, meaning that our classification scheme is based only on climatic information and can be evaluated with independent streamflow data. Using gridded global climate data, we calculate three dimensionless indices per grid cell, describing annual aridity, aridity seasonality, and precipitation-as-snow. We use these indices to create several climate groups and define the membership degree of 1,103 catchments to each of the climate groups, based on each catchment's climate.

Streamflow patterns within each group tend to be similar, and tend to be different between groups. Visual comparison of flow regimes and Wilcoxon two-sample statistical tests on 16 streamflow signatures show that this index-based approach is more effective than the often-used Köppen-Geiger classification for grouping hydrologically similar catchments. Climate forcing exerts a strong control on typical hydrologic response and we show that at the global scale both change gradually in space. We argue that hydrologists should consider the hydro-climate as a continuous spectrum defined by the three climate indices, on which all catchments are positioned and show examples of this in a regionalisation context.

2.1 Introduction

Classification is an essential step in understanding natural phenomena, as evidenced by globally agreed-upon classification schemes in many different disciplines and a strong expressed need for a catchment classification scheme in hydrology [e.g. 209, 334]. Well-known classification examples are the periodic table that chemistry uses to group elements with similar properties [e.g. 278] and Linnaean taxonomy as used in biology to group organisms based on similarity of their characteristics [e.g. 89]. Classifying phenomena into groups with similar characteristics allows transfer of knowledge from well-observed members of the group to members about which less is known. In hydrology, defining similarity between catchments plays a crucial role in enabling predictions in ungauged basins [334].

In complex systems as are common in earth sciences, classification is not straightforward [209]. Many different classification schemes are available, each with a different focus or underlying principles, and the choice for one is often motivated by a study's particular needs. All classification schemes however aim to group those elements of a system that are similar, and separate them from groups of other elements that are in some significant way different from the others. For example, soils can be classified with an international system based on their diagnostic horizons, properties and materials [151], but various national systems are used as well [e.g. 21, 140, 307]. Lakes can be classified on a variety of characteristics; e.g. thermal properties [108, 149], mixing properties [193], trophic status [61] or a combination of hydrological, chemical and biological properties [156]. Similarly, different schemes are available to classify vegetation, e.g. by using plants' survival strategy [126], or as a hierarchical scheme based on leaf cover area akin to Linnaean taxonomy [327], or as a function of dominant prevailing climate known as life zones [142].

Catchments are a common object of study within hydrology. The need for a catchment classification scheme [e.g. 209, 334] is usually interpreted as defining catchment similarity based on hydrological response, presumed drivers of the streamflow response, or a

combination of both. Wagener et al. [334] lists possible options for classification based on hydro-climatic region, catchment structure or functional catchment response. Hydrologic similarity (i.e. grouping similar catchments) follows from mapping the relationship between these aspects. An early example of a global classification of river regimes [133] defines 15 different typical annual streamflow patterns across the globe. Increases in data availability have allowed more detailed regional studies covering e.g. Australia [170] and the US [15]. Looking just at causal factors underlying streamflow, examples of regional classifications exist based on soil characteristics [196] and climate [29]. Many studies combine both approaches, using causal factors, such as average aridity, average catchment slope and land use, together with streamflow characteristics, often in the form of streamflow signatures such as mean flow and slope of the flow duration curve, to group similar catchments [e.g. 79, 187, 275, 276, 348]. Whereas a wide variety of metrics and models are used to describe catchment structure and functional response, there seems to be at least some consensus on how hydro-climatic aspects can be conceptualised: available water (precipitation) and energy (temperature, evaporation) interact within the catchment to control the water balance. Understanding of this principle has led to the Budyko-curve on an annual scale [58] and shown the importance of within-year variation of climate [e.g. 219].

In catchment classification studies, climate is often considered in a basic form (e.g. annual average aridity) or in direct relation to streamflow response (e.g. runoff ratio, streamflow elasticity) but recent work shows that a more nuanced approach that describes the influence of climatic input on typical flow regimes might be appropriate [3, 4, 29]. With three dimensionless numbers that summarize the climate's aridity, precipitation timing and snowiness, typical flow regimes in the US can be classified into 10 distinct groups [29]. Addor et al. [4] present an extended set of US catchments, including information about each catchment's climate (using three very similar indices), topography, soils and vegetation. In later work [3], they correlate this information with streamflow signatures for each catchment and find that climate, as expressed by the three indices, has the strongest correlation with streamflow signature values for this set of USA data. Information about climate, even expressed as three simple numbers, can thus be used to explain broad streamflow patterns.

Several global climate classifications exist, but these are mostly bio-climatic in origin, and thus do not explicitly include those aspects of climate regimes that are important influences on hydrology. The original Köppen scheme (work by Köppen in the late 19th and early 20th century) is for an important part based on observations of vegetation, which could be used as a proxy for prevailing climate in times when large-scale climate data was unavailable [250]. Köppen's classification inspired several other classification

19

schemes that tried to improve the correspondence between climate zones and observed global vegetation patterns [117, 319, 320]. These schemes use hierarchical rules, mainly based on temperature and to a lesser extent precipitation thresholds, to define climatic zones. They are still regularly updated with new data [e.g. 27, 183, 250]. Vahl's climatic divisions [268] attempt to address the arbitrary nature of Köppen's thresholds and certain mismatches between the classification scheme and observations, by using fewer hierarchical divisions and introducing precipitation probabilities. Holdridge Life Zones [142] and the Thornthwaite classification [319] move away from using mainly temperature and precipitation for classification, although they are still bioclimatic in origin. Holdridge uses a combination of precipitation, potential evapotranspiration, humidity, altitude and latitude to define biomes. Thornthwaite attempts to address the perceived arbitrary nature of the Köppen-Geiger thresholds and to create a more rational classification scheme. Thornthwaite uses climate observations from the USA to create a classification approach that relies on a precipitation-effectiveness index, a moisture index, thermal efficiency index and the absolute value of potential evapotranspiration. However, despite these improved alternatives, the original Köppen-Geiger scheme remains widely used today.

Currently, the main available climate classifications suffer from significant shortcomings when applied to hydrology. Haines et al. [133] tested the ability of the Köppen-Geiger classification to predict typical global runoff regimes and found some relationship between climate zones and flow regimes, but also considerable spread in the data: a flow regime might occur in many climate zones, and a single climate zone might contain many flow regimes. Based on recent work [3, 29] in the USA, we can hypothesise that this is likely because Köppen-like climate classification schemes lack hydrologically relevant detail, in the form of the interaction between water and energy availability, climate seasonality and snow pack formation. Thornthwaite's classification comes close to addressing this, but is only based on USA data and untested in its accuracy for predicting global hydrologic regimes. Additionally, Thornthwaite already noted that "variations in the heat factor of climate do not generally result in the development of sharply defined boundaries between vegetation formations" and that "the boundaries separating tropical, mesothermal, microthermal and subpolar climates are vague and ill-defined" [318], as a point of potential improvement for classification schemes. Traditionally, classification maps include sharp, unrealistic, boundaries between different classes. More recently, advances in data sciences [e.g. 286] have led to more nuanced classification schemes in hydrology where catchments can belong to several classes at the same time, but with differing degrees of membership to each class [e.g. 276].

This study addresses an identified need for a global hydrologically-informed climate classification scheme, that (i) corresponds to observed similarities and differences in

observed hydrological response, (ii) avoids introducing artificial boundaries between classes. We choose to address climate alone, without consideration of catchment characteristics, as a first step to developing a more general catchment classification.

2.2 Data

This study first uses gridded climate data to summarize the world's climate with several climate indices and uses these to define different climate clusters. Then, 1103 catchments are associated with the appropriate climate clusters using the catchments' locations and boundaries, after which streamflow data from the catchments is used to evaluate the hydrological usefulness of the climate clusters.

2.2.1 Climate data

This study uses monthly average climate values from the CRU TS v3.23 data set [137], for the climatic variables precipitation (P), number of rain days per month (N, defined as days with P > 0.1mm), temperature (T) and potential evapotranspiration (Ep). These data are available at a 0.5° x 0.5° resolution for the Earth's land areas, excluding Antarctica. The data set offers so-called primary variables, which include P, N and T, that are a re-analysis of station observations and existing climatology. The secondary variables, such as Ep, are estimated from the primary variables. Ep is estimated with a variant of the Penman-Monteith formula [7, 137].

Ep values are missing for approximately 7.3% of global land cells due to incomplete coverage of the wind speed data needed for Ep estimation. Ep values are highly spatially correlated (average autocorrelation coefficient at 1 grid cell distance = 0.99 in latitude direction, average 0.72 in longitude direction) and most of the missing values are bordered by cells for which Ep values are available. Nearly all missing values can be filled with a weighted nearest-neighbour approach, apart from several small islands that are too isolated for correlations to be a useful approach.

For this study, P/T/N/Ep data for 1984-2014 are averaged per month to find a typical year (e.g. the typical January P is the average of all 30 January P values from 1984 to 2014), to approximate the typical annual variation in all four climate variables.

2.2.2 Streamflow data

The Global Runoff Data Centre [316] manages a large database of river discharge data. This study uses a subset of data known as Pristine River Basins that contains daily streamflow data for 1182 gauging stations world-wide for the study period 1984-2014. The
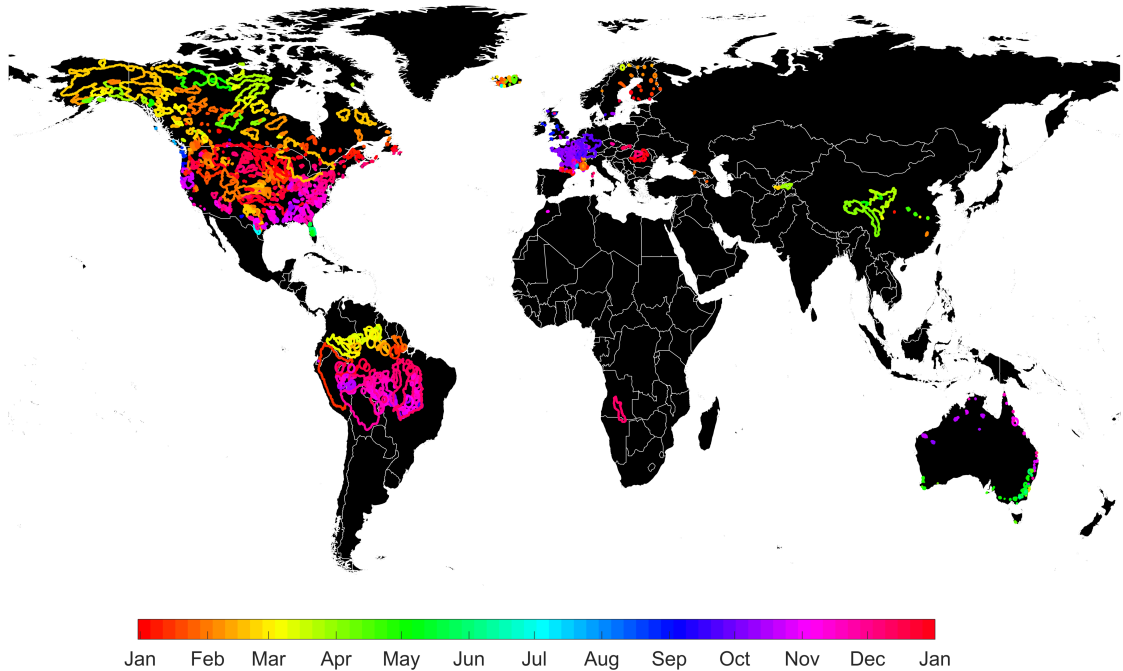
FIGURE 2.1. Location and boundaries (if available, circles with size relative to catchment area are used when not) of GRDC Pristine catchments. Catchments for which no boundary data is available are only used if the aproximate catchment length is smaller than or equal to a climate correlation threshold length and removed from the analysis if larger. Colouring indicates approximate start of the hydrological year, here defined as the 120 days before the time of the 61-day average maximum flow.

catchments in this dataset are asserted to have minimal development and river regulations and diversions. In addition, records for each catchment cover at least 20 years (overview of record lengths in Appendix A.2.3) and exceed a certain accuracy threshold [315]. We applied quality assurance procedures to the data and as a result 79 catchments were excluded (details in Appendix A.2.2), leaving 1103 stations for use in this study.

Catchment boundary information is available for 718 of the GRDC Pristine Basins [314]. The remaining 449 catchments in the Pristine Basins set vary in area from 0.69 km2 to 4,680,000 km^2, with median 596 km^2. Larger catchments can cover many grid cells and without information on the catchments' boundaries it is impossible to tell how varied the climate within each catchment is. Therefore we include only those catchments with an area smaller than or equivalent to the approximate area of 9 grid cells, for which the climate at each gauge's location might reasonably be considered representative of the climate in the

whole catchment. This limits both the uncertainty about the prevailing climate in these catchments, and the number of catchments that must be excluded from further analysis (details in Appendix A.2.1).

We create a typical streamflow year for each catchment from daily streamflow data, by taking the median flow for each Julian day (e.g. the typical Jan-1 flow refers to the median of all available Jan-1 for a particular place). Using the median flow decreases the influence of extreme events in the data and is preferable to the mean because of the skewed nature of flow variability. We also align all flow records in time so that t=1 coincides with the start of the hydrological/water year for each location, because it is easier to visually assess similarities between flow patterns when distinctive features such as the seasonal flow peaks are aligned. Catchments are spread across both hemispheres, so hydrological years are preferable to calendar years for comparison purposes. By convention the water year in the Northern Hemisphere runs from October to September [e.g. 24, 322] however for the Southern Hemisphere both April to March [24] and July to June [324] are in use. While conventions such as these can be useful on a small scale, on a global scale these are too general. Therefore, we use a 61-day moving window to find the period of maximum flow in a typical year for each catchment, and assume that the water year has started 120 days before this point (Figure 2.1). Both numbers are determined through trial-and-error and were found to give the best results (in terms of ease of visual comparison of flow similarity) for the data used in this study, but should be revisited if a more comprehensive data set is available.

2.3 Method

This study creates a climate classification scheme that summarises global climate patterns as a causal factor of global streamflow response. Causal factors (climate) and response (streamflow) are separated, meaning that our classification scheme is based on only climatic information and can be independently evaluated with streamflow data. First, we summarize the global climate with several gridded dimensionless indices (section 2.3.1). These climatic indices are clustered into fuzzy groups with a fuzzy c-means clustering algorithm (section 2.3.2) to define several climates that are representative of the land surface. We then evaluate the correspondence of the climatic clusters with global streamflow response, testing the hypothesis that locations within a cluster experience similar flow patterns while locations in different clusters show different streamflow regimes. We evaluate this both qualitatively through comparing typical seasonal flow patterns within and between climate clusters, and quantitatively through streamflow signature values and statistical tests. We compare the effectiveness of our climatic clustering with that

of the Köppen-Geiger classification, testing the hypothesis that our scheme improves on another often-used method (section 2.3.3). Last, we investigate the potential of abandoning climate classes or clusters and show the benefits of viewing the global hydro-climate as a continuum rather than a patchwork of different classes (section 2.3.4).

2.3.1 Dimensionless climate indices

The climate at any given location influences the processes near the land surface and those concerning precipitation and evaporation. The balance between available water and energy determines whether water will remain on land or be returned to the atmosphere. Periods with lower temperatures can lead to snow pack formation, and precipitation intensity can influence whether water will infiltrate into the soil or become surface runoff. However, precipitation and temperature (and by extension potential evapotranspiration) patterns are variable throughout the year and precipitation and temperature peaks are not necessarily in phase. It is thus plausible that our indices need to cover not only annual averages, but also provide a measure of the seasonal variability of climate variables. This leads to the hypothesis that, in addition to the total annual precipitation, five different climate aspects might be hydrologically relevant [e.g. 3, 4, 29, 219, 346, 347]: (i) the annual average aridity, specifying the ratio of available energy and water; (ii) the seasonality in aridity, indicating if seasonal water and energy distributions are in or out of phase; (iii) the fraction of precipitation that falls as snow, indicating whether precipitation will be (temporarily) stored on the land surface; (iv) the average rainfall intensity, showing whether rainfall will exceed infiltration rates and thus produce surface runoff; and (v) the seasonality of rainfall intensity, indicating whether infiltration excess runoff is more likely to occur in certain parts of the year.

We limit this work to aridity and snow indices for several reasons. First, although precipitation intensity can vary significantly across the world, its impact on local hydrology (i.e. whether rain infiltrates or becomes surface flow) depends on local catchment characteristics. Accounting for global differences in soil types and other catchment characteristics is considered beyond the scope of this work. Second, the CRU TS climate data set lacks information on the sub-monthly time scale, and precipitation intensity can thus only be quantified by dividing the monthly precipitation totals by the number of rain days per month (days with P ≥ 0.1mm). Both the annual average and seasonality of this approximate intensity are strongly inversely correlated with the annual average aridity (Spearman rank correlation coefficient R < -0.8 across all land cells) and thus are unlikely to add any significant new information at the global scale. Similarly, we considered using the absolute annual average precipitation [mm/y] as a metric, but this is strongly correlated with annual

average aridity (R = 0.74). Several tests during clustering (not shown for brevity) confirm that these metrics indeed add very little independent information at the global scale. Third, earlier work [3, 4, 29] shows that average and seasonal aridity indices and one snow index are strongly related to seasonal streamflow patterns, without considering rainfall intensity or absolute precipitation totals.

Using CRU TS climate data averaged into a typical year (section 2.2.1), we calculate three climate indices for each 0.5° land cell. We use a version of Thornthwaite's moisture index MI (eq. 2.1) [343] to express average aridity (I_m, eq. 2.2) and its seasonality ($I_{m,r}$, eq. 2.3), and a numerical implementation of the fraction of annual precipitation that occurs as snowfall (f_s, eq. 2.4) [347]. These indices have been used for climate classification before but not in this particular combination (e.g. Willmott & Feddema [343] for MI; Berghuijs et al. [29] for f_s). These indices describe the processes of interest using bounded intervals, which is useful for interpretation and clustering analysis.

$$MI(t) = \begin{cases} 1 - \frac{E_p(t)}{P(t)}, & if\ P(t) > E_p(t) \\ 0, & if\ P(t) = E_p(t) \\ \frac{P(t)}{E_p(t)} - 1, & if\ P(t) < E_p(t) \end{cases} \tag{2.1}$$

$$I_m = \frac{1}{12} \sum_{t=1}^{t=12} MI(t) \tag{2.2}$$

$$I_{m,r} = max(MI(t)) - min(MI(t)) \tag{2.3}$$

$$f_s = \frac{\sum P(T(t) \leq T_0)}{\sum_{t=1}^{t=12} P(t)} \tag{2.4}$$

P(t), Ep(t) and T(t) are mean monthly observations of precipitation, potential evapotranspiration and temperature in the CRU TS data set. T_0 is a threshold temperature below which precipitation is assumed to occur as snow, here set at 0°C. The annual average moisture index I_m has range [-1, 1] where -1 indicates the most arid (water-limited) conditions and 1 indicates the most humid (energy-limited) conditions. The moisture index seasonality $I_{m,r}$ has range [0, 2] where 0 indicates that there are no intra-annual changes in the water/energy budget and 2 indicates that the climate switches between fully arid ($I_m = -1$) and fully saturated ($I_m = 1$) within a single year. f_s has range [0,1] where 0 indicates no snowfall in a year and 1 that all precipitation falls as snow. Note that $f_s = 0$ does not imply that the temperature does not go below the threshold temperature T_0, but merely that during this period no precipitation occurs. The indices rely on similar information and express phenomena with similar underlying causes (e.g. seasonality of aridity might be caused by a strong summer-winter contrast, which may also increase

the likelihood of snowfall) so some correlation between the indices is unavoidable. The Spearman rank correlation between I_m and $I_{m,r}$ is 0.27, between Im and f_s 0.27, and between $I_{m,r}$ and f_s 0.37. These are considered to be sufficiently independent for use in this study, because each index has a different physical interpretation.

2.3.2 Selecting representative climates for comparison with the Köppen-Geiger classification

Traditional climate classification schemes use distinct boundaries between climate classes [e.g. 117, 183, 250, 320], but Thornthwaite already pointed out that climates change gradually in space and distinct boundaries do not do this justice [318]. However, sharp boundaries are a logical and inescapable result of the classification method that underlies Köppen-like classifications. In this work, we argue that the global hydro-climate should be seen as a continuous spectrum and that imposing boundaries on this spectrum should generally be avoided. However, for illustration purposes we use an automated fuzzy c-means clustering algorithm [52] to select several representative points in the climate space described by our three indices. Each location (grid cell in the global data) belongs with a certain degree of membership to each representative climate, based on the similarity of each location's climate index values to the climate in each representative point. Memberships can vary from 0 (the location does not belong to this representative climate at all) to 1 (the location's climate is the same as the representative climate), with the possibility for a location to belong simultaneously to several representative climates. Using these representative climates, it is straightforward to compare how similar the hydrologic regimes are for locations with the same Köppen-Geiger class compared to locations with the same representative climate.

While the fuzzy c-means algorithm can objectively create clusters from data, it does require human input in finding the appropriate settings and determining the appropriate number of clusters. We use Matlab's c-means implementation (function *fcm*) in a multi-start framework to account for the inherent randomness resulting from its use of random initial cluster centroids. Before clustering, we standardize the values of our climate indices so that each has a range [0,1], to avoid biasing the clustering procedure towards the index with the largest range. The fuzzy c-means procedure uses a so-called fuzzifier parameter to allow data points to belong to different clusters through fuzzy membership. This parameter can be used to decrease the influence of data points that are near the boundaries between two clusters when determining the cluster centroid positions [286]. This value is kept at its default value of 2. The number of representative climates was determined through trial-and-error, by performing the clustering procedure with 2 to 30 clusters and analysing

the resulting climate clusters. We did not use any river flow data to either create or help choose the number of climate clusters. We chose 18 clusters for communication purposes in this study, because this provides an adequate amount of detail but does not create overly specific geographically-focussed clusters. However, we emphasise that our key goal is the identification of climate indices for global hydrology, rather than the set of 18 clusters.

2.3.3 Effectiveness of hydrologic grouping based on representative climates versus Köppen-Geiger classes

We use GRDC river flow data for 1103 catchments to compare how well hydrologic regimes can be grouped based on our representative climates. We also group the same catchments based on their Köppen-Geiger climate class, to assess whether our approach improves upon this alternative. The success of this grouping exercise is determined with a qualitative approach to investigate typical streamflow patterns per group and a quantitative approach to investigate differences between streamflow signatures in each cluster. First, we define the membership degree of 1103 catchments to all 18 representative climates, using the catchment-averaged values of our three climate indices. For catchments without boundary information we assume that the climate at the outlet location is representative of the whole catchment. We can then show the typical flows per representative climate, using every catchment's membership degree to determine how closely the climate in each catchment resembles that of each representative climate. We assess the typical flows in a qualitative way.

We also assess the differences between flows per representative climate quantitatively through streamflow signatures and statistical tests. Olden and Poff [237] categorize 171 streamflow signatures into five main types relating to flow event magnitude, frequency, duration, timing and rate of change, distinguishing between high and low flow conditions within the first three categories. This study uses 16 signatures that cover these 5 categories (Table 2.1), mainly following recommendations from Kuentz et al. [187] and Addor et al. [4]. For each catchment, we calculate a signature's value per hydrological year and then take the average of these yearly signature values. We repeat this for all 16 signatures. Correlation analysis (not shown here for brevity) indicates that each signature contains some independent information although there is duplication of information as well. We consider this acceptable for our purposes because the signatures are only used to evaluate the two classification schemes and are not part of the classification methods themselves. The classification thus remains unbiased by potential duplicate information in the streamflow signatures.

Our null hypothesis is that there are no significant differences between signature

27

values calculated for flows in different representative climates. The alternative is that there are differences between signature values of flows per representative climate, which indicates that our climate classification scheme can tell us something informative about the hydrologic response. The Wilcoxon two-sample test [336, 342] is a suitable statistical test to compare a signature's values between two climate clusters, because the test assumes no knowledge of the distribution and parameters of the total population, and allows comparing samples with very different sizes. It allows testing of distributions (e.g. the values of a signature calculated for 70 catchments in climate I and 115 catchments in climate II) with $H_0 : \mu_1 = \mu_2$. We have modified this test to allow weighted samples as well (details in Appendix A.3.1). We apply this test to all climate cluster pairs and for all signatures. The sheer number of tests makes it likely that we will find significant differences through chance alone. We therefore investigate the number of signatures for which a climate pair is statistically different: if a pair is different for 16 out of 16 signatures, we can assume that typical streamflow for these pairs is different. If a statistical difference is only found for 1 out 16 signatures, it is more likely that we have found this result through chance. We repeat this analysis on the catchment grouping created based on the Köppen-Geiger class of each catchment and comment on the differences.

2.3.4 Beyond catchment grouping and towards climatic assessment on a continuous spectrum

In addition to being a quantified way to communicate the climate of hydrological systems, these indices can be used as a rational way to transfer hydrological information from gauged to ungauged basins. This can also be a starting point to define more powerful hydrological similarity metrics, eventually resulting in a hydrological catchment classification scheme. In the second part of this paper we briefly explore the predictive power of the three climate indices. Each catchment is treated as ungauged in turn, and we use climatic similarity as a very preliminary flow prediction method. Climatic similarity is expressed as those catchments that (1) belong to the same Köppen-Geiger class, (2) belong to the same climate cluster, or (3) are nearby based on standardized Euclidean distance in climate index space (so that every index has range [0,1]) expressed by the I_m, $I_{m,r}$ and f_s indices. In the latter case, we investigate both (3a) distance-based weighting of all catchments and (3b) distance-based weighting of the five catchments that are climatically the most similar to the "ungauged" catchment. We estimate both the flow regime of each "ungauged" catchment and values for the 16 signatures. The accuracy metric used to compare estimated and observed flow regimes is the Kling-Gupta Efficiency (KGE) [129]. The metric used to compare estimated and observed signature values is the absolute error.

Table 2.1: Overview of the hydrological signatures used in this study. Signatures are calculated for every hydrological year available for each catchment, after which we take the mean for each signature across all hydrological years available for a catchment. Numbering in the leftmost column refers to Figure 2.7.

	Signature	Unit	Description	Source
Magnitude				
1	Mean flow	[mm/d]	Mean of daily flow	-
11	Q5	[mm/d]	5th percentile of daily flow	[187]
12	Q95	[mm/d]	95th percentile of daily flow	[187]
14	Skewness	[-]	Mean divided by median of daily flow	[187]
2	Baseflow index	[-]	Baseflow fraction of total flow	[132]
4	High flow discharge	[-]	90th percentile divided by median flow	[187]
Frequency				
10	No flow frequency	[-]	Normalized average frequency of no flow (number of days with 0 flow)	-
8	Low flow frequency	[-]	Normalized average frequency of low flow (number of days with flow < 0.2*mean)	[237, 339]
6	High flow frequency	[-]	Normalized average frequency of high flow (number of days with flow > 9*median)	[237, 339]
Duration				
9	No flow duration	[-]	Normalized average duration of no flow (number of consecutive days with 0 flow)	-
7	Low flow duration	[-]	Normalized average duration of low flow (number of consecutive days < 0.2*mean)	[237, 339]
5	High flow duration	[-]	Normalized average duration of high flow (number of consecutive days > 9*median)	[237, 339]
Timing				
16	Half flow date	[-]	Fraction of year when 50% flow occurs	[81]
15	Half flow interval	[-]	Fraction of year in which 25th to 75th percentile flow occurs	[81]
Rate of change				
3	Flow duration curve slope	[-]	FDC slope between 33rd and 66th percentile in log space	[348]
13	Rising limb density	$[d^{-1}]$	Number of rising limbs divided by time that hydrograph is rising	[275]

2.4 Results

2.4.1 Approximating climatic gradients with representative climates

Figure 2.2 shows that values for the three climate indices (annual average aridity, I_m; the seasonal change in aridity, $I_{m,r}$; and the fraction of precipitation as snowfall, f_s) generally change gradually in space (Figure 2.2c-e for individual indices, 2.2b for a map combining all three indices into a single global overview). The presence of mountain ranges leads to relatively sharp transitions in climate (e.g. Canadian Rockies, Andes, European Alps, Himalayas). Large areas of deserts are visible in red. These are arid locations with a high potential evapotranspiration compared to available precipitation, only small seasonal changes in this ratio and no snowfall. Very wet regions (dark green) are centred mostly around the equator. These are areas with a continual water surplus and low snowfall. Traditionally this climate is associated with tropical rain forests but other areas (e.g. Scotland, Japan, northern New Zealand) show similar index values, even if the underlying climatic drivers are different in absolute terms. Regions in bright green and yellow show transitional zones between constantly arid and constantly wet regions. The transitional zones experience strong seasonality in their water-energy balance, either through clearly defined wet and dry seasons (seasonal rain), through summer and winter patterns (seasonal changes in potential evapotranspiration) or a combination of both. Blue and pink regions indicate places where nearly all precipitation occurs as snowfall. Figure 2.2a further shows that climates with low seasonality concentrate near both ends of the aridity (I_m) axis (bright red, dark green) and that annual average aridity is not necessarily an accurate representation of month-to-month aridity, especially in cases where the annual water and energy budgets are approximately balanced ($I_m = 0$).

In section 2.4.2, we will investigate whether our index-based classification is better suited for grouping hydrologically similar regimes than the Köppen-Geiger classification is. For a straightforward comparison with the Köppen-Geiger classes, we here define 18 representative climates in our continuous climate-index space. These give a representative sample of the climate on the land surface. Figure 2.3a shows that 18 clusters approximate the climatic gradients in Figure 2.2b well, but the continuous variation of climate in space makes it impossible to create completely homogeneous classes where every location has a climate that strongly resembles that of the representative point it belongs to. Each grid cell is coloured based on the climate cluster that the cell belongs to with the highest degree of membership, here called the "main cluster" for each cell. Figure 2.3c shows how high this main membership degree is. A membership threshold of 0.5 is commonly seen as the cell belonging exclusively to its main cluster [286]. Large areas of high membership degree
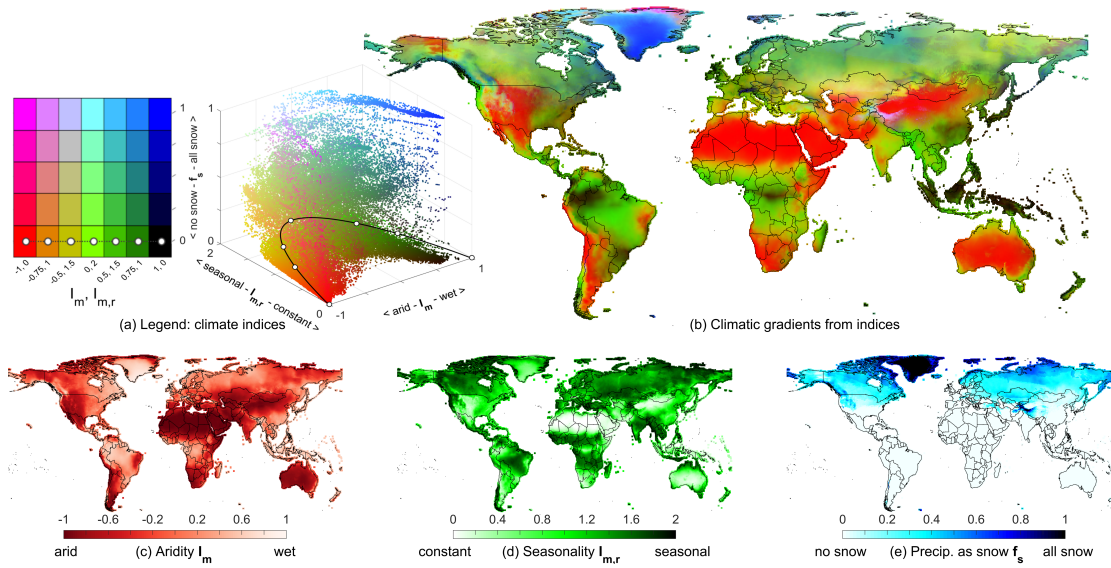
FIGURE 2.2. Overview of average climate index values calculated for 1984-2014. (a) Climate index legend to help interpret figure 2.2b, showing how values on the three climate index axes determine the final RGB colour. The 3D-plot includes all land cells shown in Figure 2.2b. The coloured square shows the colour scheme at 7 pre-determined points in 5 different I_m, $I_{m,r}$ planes along the f_s axis. (b) World map with each 0.5° resolution grid cell with local average aridity (red), aridity seasonality (green) and fraction precipitation as snowfall (blue) determining the RGB colour scale. (c-e) Plots of average aridity I_m (red), aridity seasonality $I_{m,r}$ (green) and fraction of precipitation as snow f_s (blue) respectively, showing how each index varies globally.

values are visible (blue) and mainly occur away from cluster boundaries. However, the gradual nature of the changes in climate indices makes it difficult to classify all cells in homogeneous clusters, as evidenced by the large number of cells that have membership degrees <0.5 for their main cluster. These cells can be thought of as belonging to multiple clusters simultaneously. With 18 clusters, slightly over half (50.4%) of all land cells have membership degrees >0.5 for their main cluster.

Climate cluster centroids (Figure 2.3b) are not distributed uniformly in the climate index space and the centroid marker size (larger size indicates that a higher number of land cells have that cluster as their main cluster) shows that certain climates are more prevalent than others. The centroids approximate the pattern of all individual cells in climate index space (Figure 2.2b), showing where this pattern is dense and comparatively sparse. This is a result of the clustering procedure trying to maximise within-cluster

(a) Cluster with highest membership degree

(b) Climate cluster centroids in climate index space

(c) Degree of membership to main cluster

(d) Cluster size and membership distribution per cluster
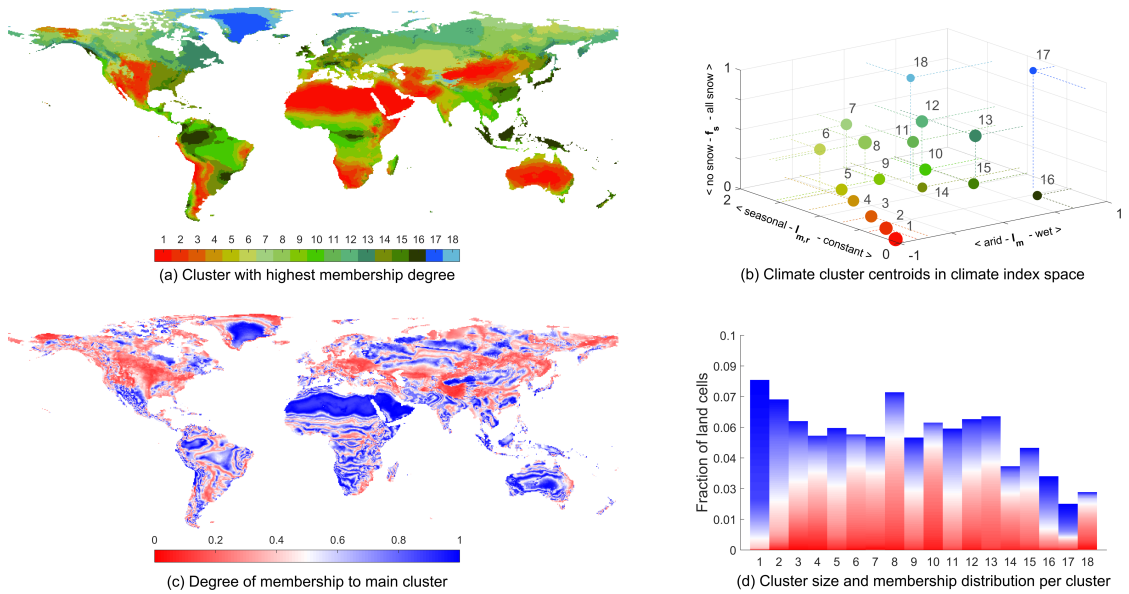
FIGURE 2.3. Results of fuzzy c-means clustering performed on climatic indices. (a) The cluster to which a cell belongs with the highest degree of membership (here called main cluster). (b) Location of climate cluster centroids in climate index space, with marker size corresponding to the number of cells for which a cluster is the main cluster. (c) The degree of membership with which each cell belongs to its main cluster; membership of each cell to the remaining 17 non-main clusters is by definition lower than its membership to the main cluster. (d) Number of cells for which a cluster is the main cluster (bar height) and degree of membership distribution per cluster (bar shading, legend in Figure 2.3c).

similarity and between-cluster differences. In the absence of clearly defined clusters/groups in the data, as is the case with the gradual changes in climate, the algorithm will struggle to draw appropriate boundaries between clusters and reverts to positioning the cluster centroids in response to point density. Figure 2.3d quantifies the number of cells for which each climate cluster is the main cluster and the degree of membership to the main cluster. Hot and very arid deserts (clusters 1 and 2) are both common and well-defined. Clusters 16 and 17 are on the other extremes (being very wet and snow-dominated respectively) and are also well-defined but contain fewer cells. In most clusters however, membership degrees are generally lower (< 0.5, red shading), because locations tend to lie between several representative climate points. Clusters 1, 2, 16 and 17 are relatively well-defined because their climates can be roughly approximated with terms as "always" and "no" (e.g. climate 1: always arid, no seasonality and no snow). The other clusters are all positioned at

some non-extreme point on each climate index axis, and this makes it impossible to draw distinct appropriate boundaries between different climatic zones in these cases.

2.4.2 Effectiveness of hydrologic grouping

2.4.2.1 Comparison of climatic gradients and Köppen-Geiger classes

The proposed new climate indices do not map directly onto Köppen-Geiger classes. The subclasses of the tropical (A) and arid (B) Köppen-Geiger main classes are relatively distinct from one another in the climate space defined by indices I_m, $I_{m,r}$ and f_s, whereas the subclasses of the colder temperate (C), continental (D) and polar (E) classes cover relatively similar regions in climate index space (Figure 2.4). This can be seen around the equator, in North-Africa, the Middle-East and most of Australia, where the Köppen-Geiger map (Figure 2.4b) is similar to the climate index map (Figure 2.4a). These regions are either very dry (through a combination of high temperatures and low precipitation) or very wet (resulting from very high precipitation) and see virtually no snowfall. These characteristics are captured well through the threshold approach in the Köppen-Geiger classification scheme. The hydrologically relevant nuances of precipitation differences in colder climates are not well captured in the Köppen-Geiger scheme. This can be seen in e.g. the Eastern USA, Alaska, Greenland, most of Northern Europe and Russia, where the Köppen-Geiger boundaries are nearly exclusively determined by temperature thresholds. Different degrees of relative water availability and snow pack formation are lost in this classification. While the thresholds are an appropriate choice to define vegetation zones, as is the original goal of the Köppen-Geiger scheme, this approach is less relevant from a hydrological point of view. The climate indices contain more hydrologically relevant information, as the following sections will show.

2.4.2.2 Qualitative comparison of grouped streamflow regimes

Grouping the typical flow regime of all catchments according to the catchments' climate indices (Figure 2.5) shows that seasonal flow patterns gradually evolve along climate gradients. Clusters 4, 14 and 15 are similar with respect to the aridity seasonality $I_{m,r}$ and snow f_s metrics but are progressively less arid (I_m metric). As a result of this increased water availability, the clusters' typical flow patterns look similar but average flows become progressively higher. Clusters 1, 2, 3, 4, 5 and 6 are similarly arid (I_m) and low on snow (f_s) but their aridity is progressively more seasonal ($I_{m,r}$). The latter clusters thus occasionally experience a water-surplus, even if on average these places are severely water-limited. As a result, the average flow is low for all clusters, but a progressively higher seasonal flow
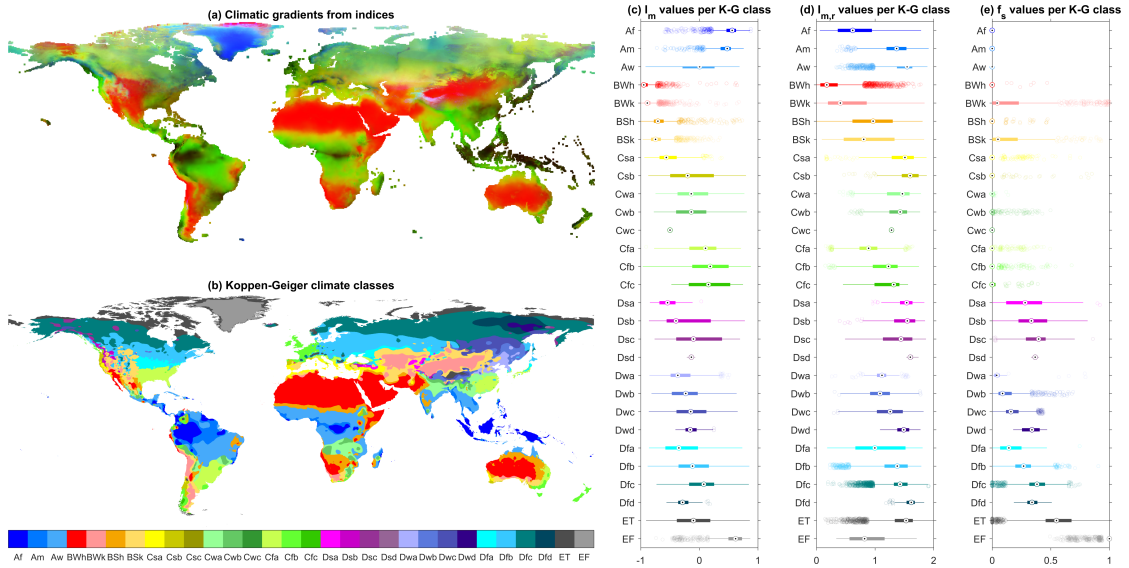
FIGURE 2.4. Comparison of the Köppen-Geiger climate classification and the global distribution of climate index values. (a) Global distribution of climate index values, as shown in Figure 2.2b. (b) Köppen-Geiger climate classification [250]. (c-e) Box plots of average aridity (I_m), seasonality of aridity ($I_{m,r}$) and fraction of precipitation as snow (f_s) values per Köppen-Geiger class.

peak can be seen. Clusters 6, 8, 11, 12 and 13 have similar values for the snow (f_s) and seasonality ($I_{m,r}$) metrics but are progressively less arid. As a result of this increased water availability, average flows become progressively higher and the main flow peak (likely resulting from snow melt since $f_s > 0$ at the cluster centroids) becomes progressively more pronounced.

Typical flow features such as average flow magnitude and flow peak height and shape are distinctly different between clusters, but climate can only inform us about average seasonal patterns. For example, the flow peak shape in snow dominated climates (e.g. clusters 13, 12, 11) shows a much sharper rise and decline than elsewhere, presumably due to snow storage and melt processes. In warmer but not water-limited climates (e.g. 16, 15, 10) the flow peak rises and declines gradually, presumably as a result of seasonal changes in water surplus. However, within each cluster a wide variety of flows are included and what is true on average for the cluster, is not necessarily true for a single specific catchment. In a catchment classification context, climate is an important driver of hydrologic processes but the influence of the catchment itself (e.g. topography, vegetation, anthropogenic influence) cannot be ignored. This is however considered beyond the scope of this work.
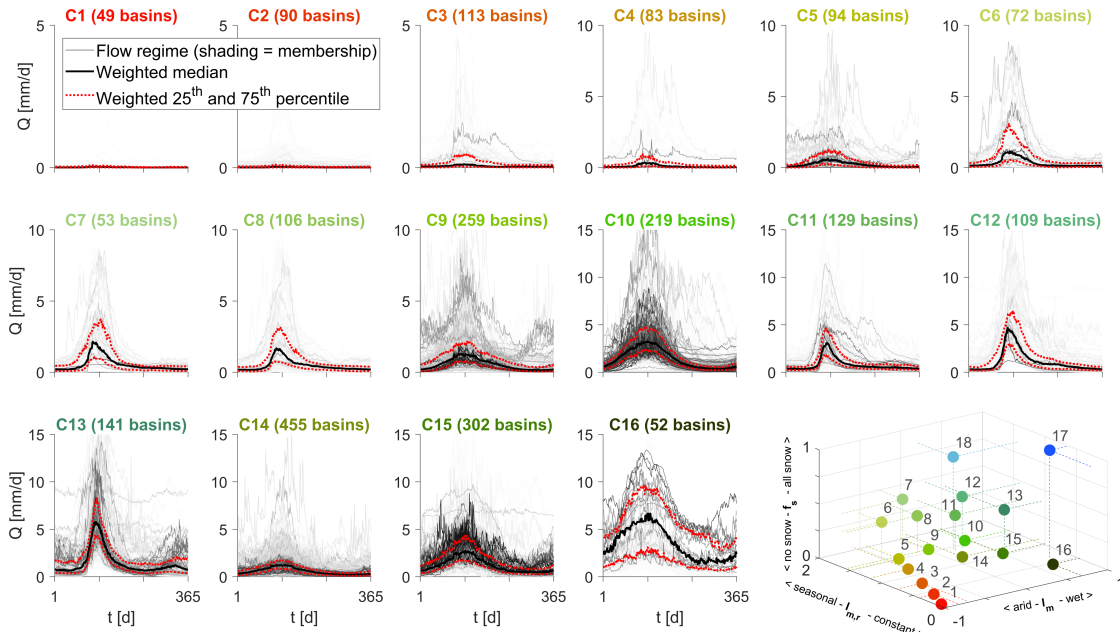
FIGURE 2.5. Typical flow regime for catchments grouped by climate cluster with the membership-weighted weighted median in black and the weighted 25th and 75th percentiles in red. Only catchments with a minimum membership of 0.10 or higher are shown, with darker lines corresponding to higher membership degrees. Includes all 1103 unique catchments, although catchments may appear in multiple climate plots. Title colouring corresponds to climate cluster centroids (figure 2.3a, 2.3b). Clusters 17 and 18 are not shown because the data lacks climate-specific flow records for these clusters.

Figure 2.4 showed that Köppen-Geiger main classes A and B show strong correspondence with our more arid and wet representative climates (e.g. climates 1-4, and 10, 15, 16 respectively). This pattern repeats with respect to grouping flow regimes by Köppen-Geiger classes (Figure 2.6): grouped flows for subclasses in the tropical (A) zone are very similar to the flows in representative climates 16, 15 and 10 (compare Figure 2.5), which have low aridity and no snowfall. The flows in the arid (B) subclasses are similar to our arid clusters 1, 2, 3 and 4. However, subclasses of C, D and E climates do not seem to group flow patterns in any meaningful way. To aid in this comparison, each flow record is coloured according to our catchment-averaged climatic index values and within main classes A and B the colouring seems relatively consistent. In climates C, D and E however, catchments with very different hydro-climates are lumped in each subclass and don't reveal any obvious typical flow pattern. E.g. subclass ET (polar tundra) contains flow patterns ranging from

being nearly zero all-year round (orange), to very high, snow melt-dominated regimes (blue-green). The snow melt regimes are not as obviously grouped in the Köppen-Geiger classes as they are in the climate-index clusters (compare Figure 2.5 and Figure 2.6).
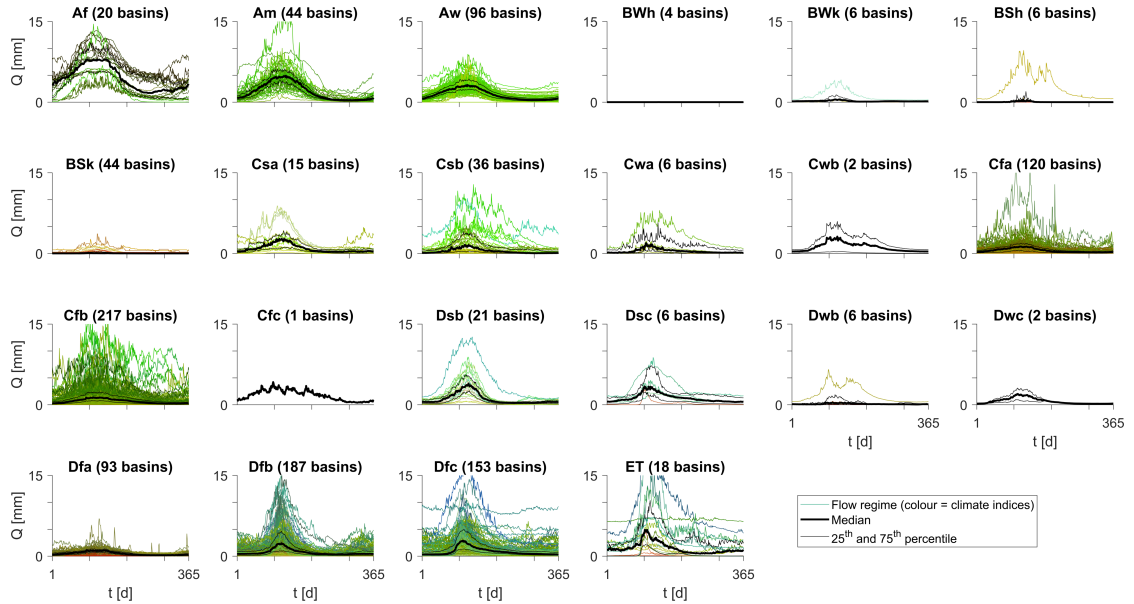
FIGURE 2.6. Typical annual flow for all 1103 catchments per Köppen-Geiger climate class with the median (black) and 25th and 75th percentiles (red). Typical flows from individual catchments are coloured by the catchment's climate index values as used with the climate index approach.

2.4.2.3   Quantitative comparison of grouped streamflow signatures

We use catchment membership degrees to create a weighted average streamflow signature value for 16 different streamflow signatures for each of the 18 representative climates. Statistical tests show that 145 out of the 153 possible combinations of two representative climates are statistically different at a 0.01 significance level. Another 7 out of 153 pairs are different at a 0.1 significance level and only a single pair shows no significant difference (p-value of 0.28; Figure 2.7d). Figure 2.7a shows an example of weighted average signature values per representative climate, here showing results for the average_flow signature (overview of all signatures is given in Appendix A.3.2). A clear gradient is visible in the climate space, with the signature value increasing primarily as aridity decreases and secondarily as seasonality increases. Figure 2.7b shows the results of an empirical Wilcoxon test to determine the statistical significance of the differences in average_flow signature

values between all representative climates. This procedure uses the average_flow signature value for each catchment, coupled with the catchment's membership degree (Figure 2.7c) to each representative climate, to estimate an empirical p-value (details in Appendix A.3.1). Most representative climates have statistically different average_flow signature values at a 0.01 level (dark blue shading), but not all climate pairs are significantly different based on this single signature (white and red shades). Figure 2.7d shows the lowest p-value per climate pair across all 16 signatures and shows that 148 out of 153 climate pairs are different at the 0.05 significance level (bottom-left section of the figure). The top-right part of the figure shows the number of signatures for which the empirical p-value is below 0.05. The prevalence of darker shades indicates that climate pairs are statistically different for multiple signatures, indicating that our clustering approach can indeed group catchments with similar flow characteristics.

Climate pairs 6-8, 7-8, 6-7, 4-5 and 17-18 are not statistically different on any of the signatures. If there are statistical differences to be found, either these differences manifest in flow characteristics not captured in the chosen signatures, or we lack climate-specific (high membership) flow records to construct an image of how a typical flow pattern for each representative climate looks. It is unlikely that the signatures are poorly chosen because they are adequate to distinguish between all other climate pairs (Figure 2.7d). Lack of climate-specific flow records is a likely explanation in the case of representative climate 18 (only 4 catchments with membership > 0.1; Figure 2.7c) and 17 (only 1 catchment has membership > 0.1; Figure 2.7c). Similarly, climates 6, 7 and 8 are close together in climate space and membership degrees of all catchments to each of these three representative climates are quite similar (Figure 2.7c). It is likely that the 1103 catchments lack the diversity that would allow the signatures to distinguish better between these three representative climates. This same explanation might be applied to climates 4 and 5. The alternative to these explanations is that there are no statistical differences between the typical flows of these representative climates; i.e. our assumption that the typical flow regime should be different between these representative climates, because the catchments associated with each climate have different hydro-climates, is false. However, given the success of the method with other climate pairs that are close together in climate_space (e.g. 10-14, 11-12, 1-2), lack of climate-specific flow records seems the more likely explanation.

Even though the Köppen-Geiger classification has more climate classes than our climate-index method, analysis of signature values shows that grouping catchments by their dominant Köppen-Geiger climate class leads to fewer distinguishable differences in typical flow patterns (Figure 2.7). Using the average_flow signature as an example, catchments in classes A and B seem to be sorted well according to their signature values (Figure 2.7e.) This is not the case in classes C, D and E, where the boxplots for subclasses
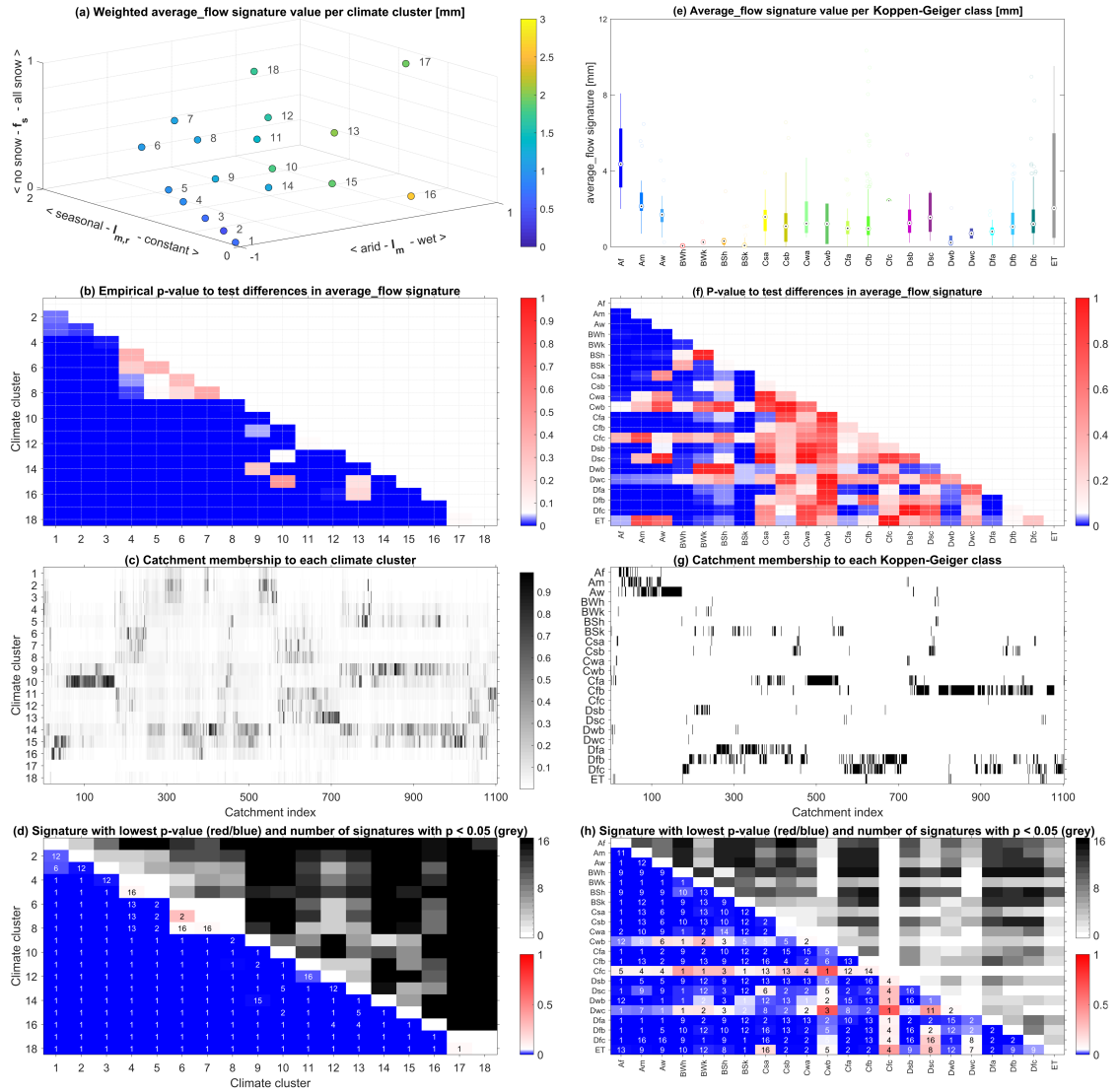
FIGURE 2.7. (a-d) Quantitative differences between grouped flows regimes using climate clusters, (e-h) quantitative differences between grouped flows regimes using Köppen-Geiger classes. (a, e) Value of the average_flow signature per climate cluster/Köppen-Geiger class. Similar plots for all signatures in Appendix A.3.2. [a] Values are calculated as a weighted average from all 1103 catchments, with weights being each catchment's membership to a cluster. Numbers refer to climate clusters. [e] Box plot colour refers to the legend in Figure 2.4b. (Caption continues on next page.)

tend to overlap (e.g. for Cfa, Cfb, Dfb and Dfc, all of which include 90 catchments or more).

FIGURE 2.7. (Continued from previous page.) (b, f) Statistical tests to determine whether values of the average_flow signature per cluster/Köppen-Geiger class are statistically different. Blue shades show p < 0.05, white shades show 0.05 < p < 0.10 and red shades show p > 0.10. [b] Results of an Empirical Wilcoxon test (Appendix A.3.1) used with cluster grouping. [f] Regular Wilcoxon test used with Köppen-Geiger grouping. (c, g) Membership degree of catchments (x-axis) to climate cluster/Köppen-Geiger classes (y-axis). [c] Darker shades show that a catchment belongs more strongly to a given cluster, and thus contributes more to the average cluster signature value. [g] All memberships to Köppen-Geiger classes are 1. (d, h) Bottom-left shows the lowest p-value from all 16 signatures, i.e. the largest significant difference between two groups. The number in each cell shows for which signature this lowest p-value is found (see Table 2.1 for numbering). Top-right grey shading shows for how many out of 16 signatures we find a significant difference (p < 0.05). See Appendix A.3.3 for p-values of the individual signatures that were used to create the top-right grey shading. [d] Results of cluster grouping. [h] Results of Köppen-Geiger grouping.

A Wilcoxon test confirms that statistically significant differences occur less frequently for the classes C, D and E than for classes A and B (Figure 2.7f). Using all signatures, we can find statistically significant differences in signature values between most Köppen-Geiger classes (Figure 2.7h). Where we do not, we likely lack enough catchments in our dataset for that subclass to make proper statistical inferences (i.e. Cwb, Cfc, Dwc; Figure 2.7g). However, in many cases in classes C, D and E we only find statistically significant differences in a few out of all 16 signatures (compare grey shades in Figure 2.7d and Figure 2.7h). This supports the idea that the temperate (C), continental (D) and polar (E) Köppen-Geiger classes are not well suited to grouping hydrological flow regimes.

2.4.3 Beyond climate grouping and towards a continuous representation of climates

Figure 2.8a shows the results of treating each catchment as ungauged in turn and using a climatic similarity approach to estimate the flow regime of this "ungauged" catchment. Comparing the effectiveness of Köppen-Geiger classes and our climate clusters, the clusters are somewhat more effective for estimating typical flow regimes. However, it is not our intent to advocate replacing one set of climate groups with another. Avoiding groups/clusters and using hydro-climatic similarity only, can have strong benefits compared to the classes/clusters approach. Using just 5 climatically similar basins to estimate the "ungauged" flows shows a significant increase in the number of basins where KGE values of
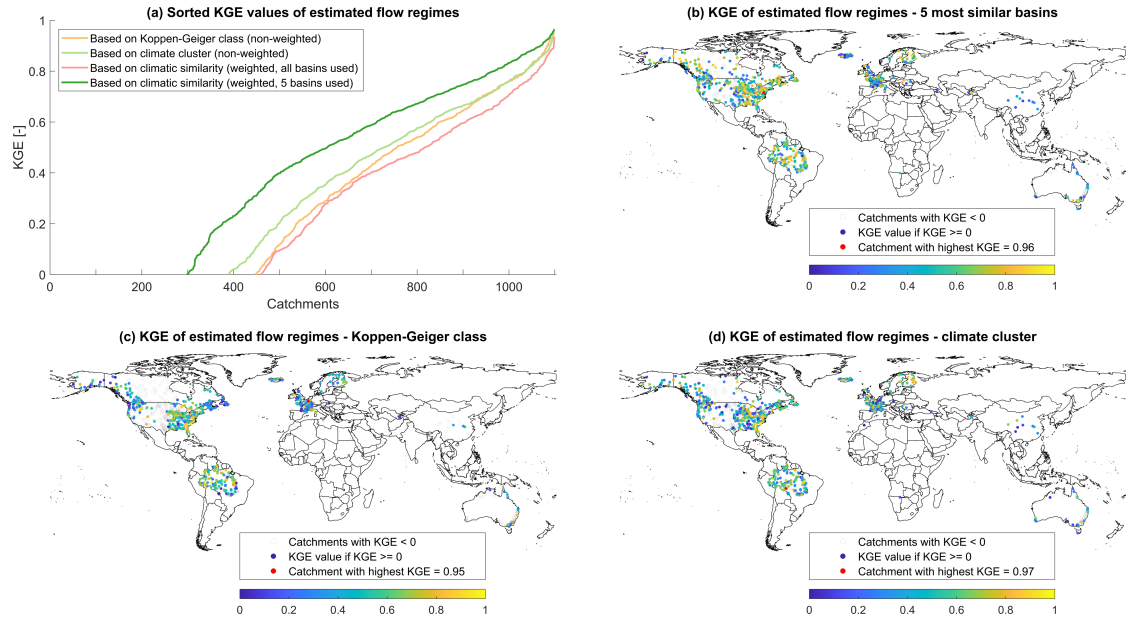
FIGURE 2.8. Estimating flow regimes through various definitions of climatic similarity, treating all 1103 catchments as if they were ungauged in turn. (a) Overview of sorted KGE values for estimated typical flow regimes, based on various similarity metrics. KGE values below 0 are not shown for clarity. (b-d) Geographical location of catchments and KGE value for estimated flow regime for each. Catchments with KGE < 0 shown in grey. [b] Estimation of "ungauged" catchment as the weighted daily mean of the regimes in the 5 climatically most similar catchments. [c] Estimation of "ungauged" catchment as the daily mean of the regimes in the same Köppen-Geiger class. [d] Estimation of "ungauged" catchment as the daily mean of the regimes in the same climatic cluster.

the estimated regime exceeds 0. Quality of the flow estimates does depend on the number of samples used: using climate-weighted records from 1102 catchments to estimate each "ungauged" catchment leads to worse results than using either Köppen-Geiger classes or climate clusters. In these cases, most catchments are dissimilar from the "ungauged" one and these flow records dilute the estimate through sheer numbers (even if any individual catchment has a low weight). Using a small number of climatically similar basins overcomes this issue (within this data set, climatic similarity consistently outperforms climate clusters when fewer than 150 catchments are used to estimate any "ungauged" regime – the best results are obtained when 3-10 climatically similar catchments are used).

A similar pattern is revealed when climatic similarity is used to estimate signature values for "ungauged" catchments (Figure 2.9). Climatic similarity of a few catchments to
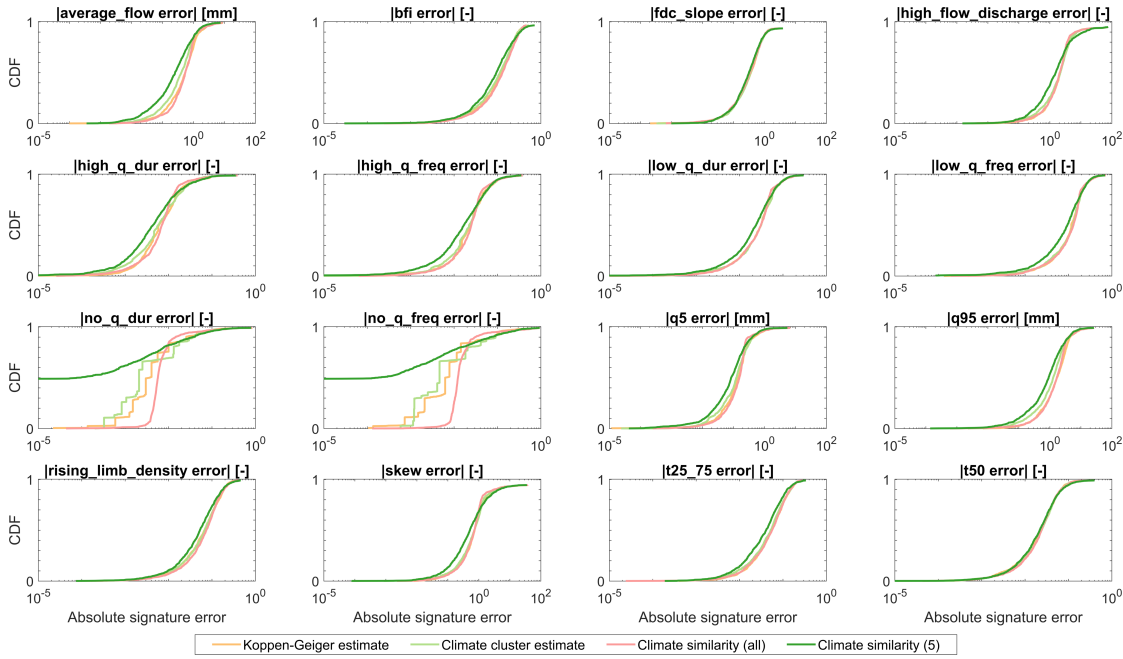
FIGURE 2.9. Cumulative Distribution Functions of absolute errors in signature value estimation. Signatures are estimated by treating each catchment as ungauged in turn and using climatic similarity to find donor catchments. Signature values from the donor catchments are averaged to estimate the signature's value in the "ungauged" catchment. Climatic similarity is expressed as belonging to the same Köppen-Geiger class (orange), the same climate cluster (light green), or as Euclidean distance in climate index space expressed by the I_m, $I_{m,r}$ and f_s indices. In the latter two cases, the estimated signature value for the "ungauged" catchments are based on distance-based weighting of all catchments (red) or the five catchments that are climatically the most similar to the "ungauged" catchment (dark green).

the target catchment generally gives the lowest errors across the ensemble of catchments. However, there is considerable spread in performance between the various signatures. Estimates of signatures associated with the magnitude of various parts of the water balance (e.g. average flow, q5, q95) and duration and frequency of high/low/no flow events seem to benefit the most from using index-based similarity over other options. Improvements are smaller for signatures related to timing (t50, t25_75) and rate of change (rising limb density, slope of the flow duration curve). Using climatic similarity of a few basins to estimate signatures also occasionally results in a higher occurrence of larger errors than other methods (e.g. for the skew signature). There is a delicate balance between using all available catchments (weighted by climatic similarity) and using just a few climatically

41

very similar catchments to create estimates. Using more catchments decreases the risk of selecting a small number of climatically similar but structurally (in terms of vegetation, geology, etc.) different catchments as donors, but also has the potential to dilute the quality of estimate through the sheer number of dissimilar basins included. Using fewer catchments can be very accurate due to the absence of dissimilar basins in the estimate, but leaves one vulnerable for differences in the catchment structure which this approach does not account for. In general, results seem to indicate that using climatic similarity expressed through indices is a promising avenue for catchment classification. With refinement and introduction of catchment characteristics into the procedure, this approach to transfer knowledge between gauged and ungauged catchments can potentially be a powerful tool for prediction in ungauged basins.

2.5 Discussion

This work presents a hydrologically-motivated alternative to traditional climate classification schemes, accounting for gradual changes in climate and the influence that has on flow regimes and streamflow signatures. This addresses two criticisms of traditional classification schemes used for hydrology, namely that their underlying motivation is not hydrological and the subjective nature of the number of classes and their distinct boundaries. Although we define 18 representative climates, these are intended as a communication device only, to enable straightforward comparison between our method and the Köppen-Geiger classification. Section 2.4.3 shows that clear benefits can be gained by using a continuous hydro-climatic spectrum instead.

We find that three simple climate indices, that quantify a location's average aridity (I_m), the seasonal range of water-versus-energy availability ($I_{m,r}$) and the fraction of precipitation that occurs as snowfall (f_s), are good indicators for finding similar hydrological regimes on a continuous scale. To further illustrate this, Figure 2.10 shows the degree to which all catchments belong to representative climates 16, 15, 9 and 8 respectively, and how the typical flows that are strongly associated with each cluster look. From climate 16 to 15 to 9, the index values indicate progressively more arid climates, with increasing aridity seasonality and constant (nearly zero) snowfall. The corresponding streamflow regimes become lower on average as a result of increasing average aridity, with lower low flows resulting from the increase in aridity seasonality. From cluster 9 to 8, aridity and seasonality remain constant, but snowfall increases. The corresponding streamflow regimes in climate 8 are similar to those in climate 9 (both on average and during low flows) but have a much sharper high flow peak as a result of snow accumulation and melt processes. This reinforces the hypothesis that gradual changes in climatic conditions

lead to gradual changes in seasonal streamflow patterns and can be of importance during catchment classification and catchment similarity studies. Most catchment characteristics can be described on a continuous scale (e.g. area, elevation, slope, porosity, conductivity, degree of vegetation cover, leaf area index) and these results suggest that climate should be treated in the same way, rather than using discrete classes.
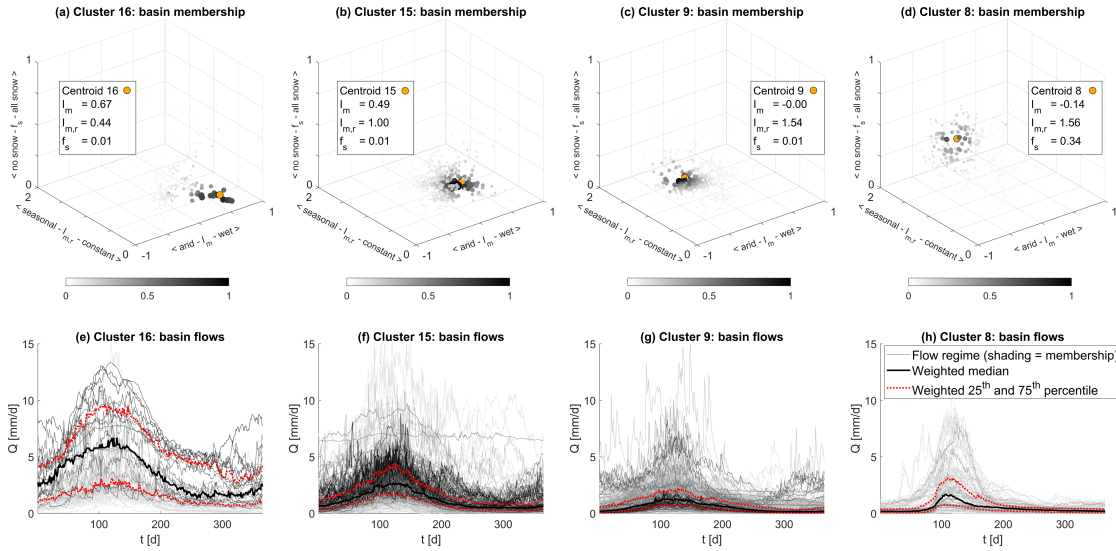
FIGURE 2.10. (a-d) Membership degree [0,1] of catchments to clusters 16, 15, 9 and 8 respectively (shading) and the catchment-averaged values for the three climate indices that describe aridity (I_m), seasonality of aridity ($I_{m,r}$) and fraction of precipitation as snowfall (f_s). (e-h) Typical streamflow in catchments with climates similar to representative points 16, 15, 9 and 8 respectively (flows shaded by their degree of membership to the cluster) with the weighted median (black) and 25th and 75th percentile (red).

Our findings are in line with earlier work on the relation between seasonal streamflow patterns and climate [3, 4, 29] and with work on the suitability of the Köppen-Geiger classification for mapping global flow regimes [133]. An important difference is that both Berghuijs et al. [29] and Haines et al. [133] create climatic classes by grouping flow regimes, whereas this work uses streamflow data only to evaluate the appropriateness of our climate indices in relation to hydrologic regimes. The specific climate indices we have chosen are slightly different from those used in Berghuijs et al. [29] and Addor et al. [4], but they are intended to capture the same climatic aspects (aridity, seasonality and snow). Both those studies are regional, focussing on the contiguous USA and our results indicates that their general findings (i.e. that 3 climate indices can be used in defining hydrologic

similarity) might be applicable on the global scale as well. Haines et al. [133] present a global classification of river flows based on monthly streamflow data and find considerable spread in how their proposed regimes relate to Köppen-Geiger classes, similar to our results. Haines et al. [133] also compare their result to an earlier study by Beckinsale [26], who adapted the Köppen-Geiger classification to apply to river regimes, and find that "many of the 'different' regimes proposed by Beckinsale are in practice found not to be significantly different at the world scale" [133]. This is a consequence of their choice to cap the number of possible regime classes, such that all classes contained a significant (but unspecified) number of observed flows and were consistent with known geographic features. Analysing river flows on a continuous spectrum, such as proposed in this work, rather than using discrete groups would avoid this problem and allow rarely occurring regimes to be somewhere on this spectrum as well. However, we emphasize that our work is not intended as a river regime classification scheme (which would necessarily involve accounting for a catchment's characteristics as well), but rather presents a hydrologically-relevant way of accounting for the influence of global climates in such a catchment classification.

2.5.1 On geographical proximity of the catchments

Geographical proximity of catchments could explain similarity between typical flows as well as climatic similarity, but the GRDC catchments are spread out enough in climatic and geographical space that this plays only a small role. Typical correlation lengths for hydrologic similarity are in the order of 100 to 200 km [e.g. 63, 122, 305]. Within the GRDC data set, approximately 1.3% of catchment pairs are within this distance from one another. This can explain certain similarities in flow patterns per climate cluster (Figure 2.5), because geographically close catchments are likely to have high membership degrees to the same representative climate(s). However, nearly all representative climates include catchments with high degrees of membership from at least two continents and all climates contain catchments that are far enough apart to ignore spatial correlation (see Appendix A.4). In Figure 2.11, the typical flows in representative climate 15 are separated by continent (columns) and degree of membership (rows). Within a column, flows are relatively similar in pattern and size, which could be explained by relative geographical proximity (although the catchments still span several 100s of kilometres). Across columns however, especially above 0.50 membership degree, the flows on each of the four continents are remarkably similar. This reinforces the idea that similar climatic conditions lead to relatively similar flow patterns.
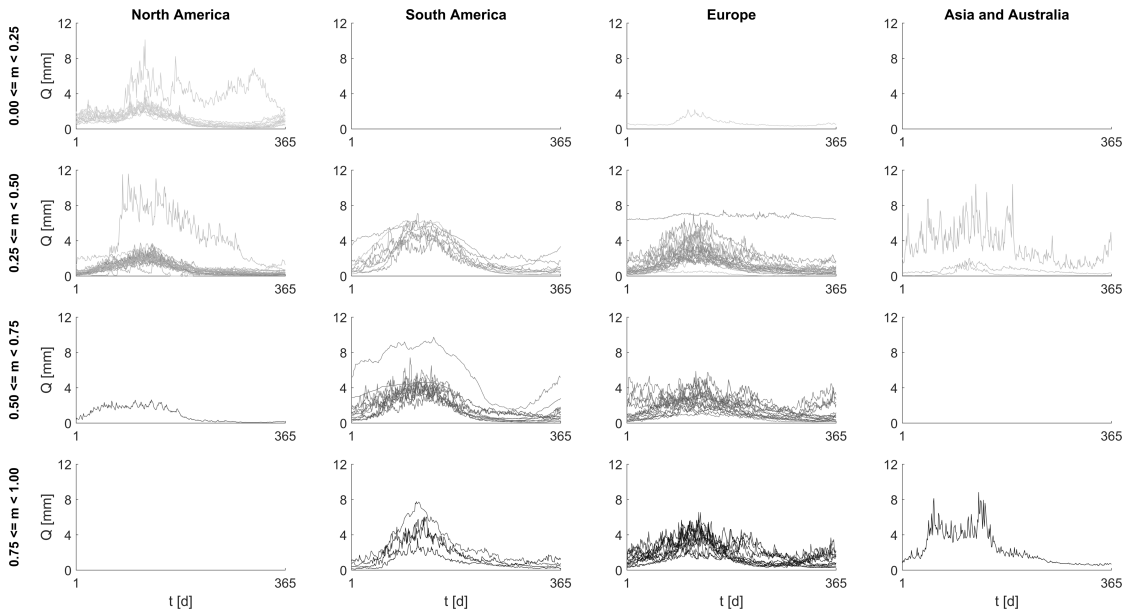
FIGURE 2.11. Flows from GRDC catchments that have climate cluster 15 as their main cluster (i.e. their highest degree of membership to any cluster is to cluster 15) separated by continent (columns) and degree of membership (rows). Flow shading corresponds to degree of membership of each individual catchment to cluster 15.

2.5.2 On the choice of climate indices

Our climate indices express the annual average water and energy budget, the seasonality of water and energy availability and the fraction of precipitation that occurs as snowfall. These indices relate to similar climatic attributes as earlier regional studies in the US [3, 4, 29] have used but we use different equations for aridity and the seasonality of aridity. Our choices are motivated by both practical concerns and a need to find indices that confer relevant information on a global scale. Indices on a bounded interval are easy to visualise (in the case of our 3D climate index space, Figure 2.2a) and straightforward to normalize to a [0,1] interval. The latter is useful for clustering analysis and regionalisation, both within this study and for potential later work. Traditionally, aridity is often given as a dryness index PET / P [e.g. 4, 29, 58] with range [0,∞ >. We adopt a moisture index [102, 319] instead that expresses the same information but on a bounded interval [-1,1] and thus fits our criteria better. We have chosen to use the term "seasonality of aridity" over "precipitation seasonality" as is used in Berghuijs et al. [29] and Addor et al. [4]. The precipitation seasonality metric is based on an expression of local P and PET time

45

series as sinusoidal functions and finding the difference between the timing of the P and PET peak. This approach was originally developed in the context of snow modelling [347] and thus assumes that PET follows a distinct summer-winter pattern due to temperature seasonality. This assumption is appropriate in the context of the US but less so towards the equator. Furthermore, precipitation seasonality only confers information about timing and not relative volumes of P and PET. Therefore, we have opted to use the within-year range of our monthly moisture index as a seasonality metric instead. This metric conveys information about the possible states of water availability a location can go through, which is relevant information on a global scale (although it has its own limitations, see section 2.5.3).

2.5.3   On study limitations

This study has several limitations which can be improved upon in later work. First, we investigate the relation between climate and streamflow patterns by comparing averaged monthly climate values over a 30-year period and median daily streamflow values. This approach smooths out outliers in both climate and streamflow data, but ignores interannual variability. The results shown in this work form a good basis to investigate interannual variability from. Second, the number of catchments could be increased. The GRDC catchments were selected for their global nature and availability of daily flow records, but this still leads to an under representation of African and Asian river systems. Additionally, to keep as many catchments for the analysis as possible, we have not set an upper bound to catchment size (provided that the catchment boundaries are known). It is possible that in large catchments not every part of the catchment contributes equally to overall runoff. The catchment-average climate that we use might thus not be representative of the climate in the runoff generating part of large catchments. A larger database of river basins would allow us to restrict the analysis to smaller catchments where this issue is unlikely to play a role. Third, the seasonality of aridity metric ($I_{m,r}$, eq. 2.3) can be improved. The metric measures the range between the lowest and highest monthly aridity value but this range is not necessarily symmetrical around the average aridity (I_m, eq. 2.2) value. There is thus a certain amount of non-uniqueness for each combination of I_m and $I_{m,r}$ values. For example, $I_m = 0$ and $I_{m,r} = 1$ can theoretically mean "this location is on average neither arid nor wet, but reaches a very arid state at some point during the year", "it is neither arid nor wet on average, but has a large water-surplus at some point" and everything in between. Extremely asymmetrical occurrences are unlikely though, because this would require nearly balanced precipitation and potential evapotranspiration all-year round, apart from a single extremely dry/rainy month. The impact of this effect is currently hard to judge but

might be investigated through an increased number of catchments. Another limitation of the I_m and $I_{m,r}$ indices is that (unlike the sine curve approach of Milly, [219]) they do not allow us to reconstruct the monthly times series of climate. Other choices of climate indices could lead to improvements, but the results already look promising: by comparing just three simple indices, we are able to locate catchments with similar seasonal flow patterns and flow regimes. The climate index values can be used to define a quantitative measure of "climatic similarity" between catchments in an easier, more succinct way than is possible with earlier climate classification schemes.

2.6 Conclusions

Hydrology needs its own structured way to quantify climates, acknowledging that climates vary gradually on a global scale, that distinct boundaries between climate classes do not represent reality well, and that climate descriptors should explicitly including those climate aspects that drive changes in hydrologic regimes. Until now, climate classification in hydrology has either used classification schemes from other disciplines (e.g. the Köppen-Geiger scheme) or used ad hoc methods (e.g. a within-study selection of metrics such as aridity or streamflow elasticity). In this work, causal factors (climate) and streamflow response are intentionally separated, meaning that the classification scheme presented here is based on only climatic information and can be evaluated with independent streamflow data. We define the hydro-climate on a global scale, using three dimensionless indices that describe each location's aridity, the seasonal changes in aridity and the fraction of precipitation that occurs as snowfall. Using 1103 catchments, we show that typical streamflow regimes and streamflow signature values correlate strongly with the local hydro-climate. Gradual spatial changes in climatic conditions are accompanied by gradual changes in flow regimes. In a climate classification context, using these three indices is a better way to identify hydrologically similar catchments than the Köppen-Geiger classification. This is partly because the Köppen-Geiger scheme is not hydrologically based and does not capture the hydrologically relevant nuances of colder climates properly, and partly because the Köppen-Geiger scheme uses discrete climate classes. The gradual changes in climatic and streamflow conditions are not adequately captured using discrete classes. Instead, it is more useful to view the global hydro-climate as a continuous spectrum on which every catchment is located. Regionalisation of typical streamflow patterns and streamflow signature values tends to be better when a small number of climatically similar basins (i.e. close together in the climate space described by our climate indices) is used instead of donors chosen based on either Köppen-Geiger or climate cluster grouping. Using the work shown here, a catchment's climate can be described with three simple numbers,

which allows easier knowledge transfer between catchments and can form the basis of a catchment classification method.

CHAPTER

3

GLOBAL PRECIPITATION MODALITY

Author's note: this chapter addresses an open question concerning the hydrologic climate classification presented in Chapter 2. The default interpretation of our climate seasonality metric $I_{m,r}$ is an assumption of one wet season and one dry season per year. In regions with bimodal rainfall regimes (two wet and dry seasons per year), this assumption would be incorrect. Because the within-year distribution of precipitation can influence within-year streamflow patterns and water availability, this might be an important factor to consider in any hydrologic classification study. However, information on precipitation modality is scattered across a large number of regional studies. We attempted to addresses the two-part question: "(1) How common and where are regions with more than a single rainfall season per year on a global scale and (2) should this be accounted for in hydrologically-informed classification schemes?"

We identified that two rainfall seasons per year occur on a small, but not insignificant fraction of the land surface. However, we could not locate streamflow data from regions with strong bimodal rainfall seasonality and thus could not investigate the relationship between bimodal rainfall and streamflow patterns. We can therefore answer the first part of our research question but must leave the second part for future work. Our investigation into the global occurrence of bimodal rainfall regimes is published as a stand-alone paper in the International Journal of Climatology, where it builds logically on earlier work published in the same journal [30]. The two papers deal with within-year climate patterns only and the relationship with streamflow is not discussed in either. This chapter presents our paper in its peer-reviewed form with only minor alterations relating to spelling and grammar. Its relevance to the thesis as a whole is further discussed in Chapter 6.

Abstract

Global precipitation patterns lead to differences in seasonal distributions of rainfall between locations, in the form of alternating dry and wet seasons. Many locations experience a single wet and dry season per year, but some studies report the occurrence of two wet and dry seasons per year. This bimodal rainfall pattern is commonly associated with locations within the tropics but is reported outside the tropics as well. However, this information is fragmented and studies of bimodality are mainly restricted to monthly rainfall totals. Here we use a gridded global data set and simple harmonic analysis to provide a systematic overview of global bimodal rainfall and rain-day frequency. We find a good agreement between the various regional studies concerning bimodal precipitation and our global overview, showing that bimodal rainfall occurs on approximately 7% of the global land surface. In the tropics, regions of bimodal rainfall totals (P) and regions of bimodal rain-day

frequency (N) tend to overlap due to the presence of dry seasons that have zero precipitation. Outside the tropics P and N are more independent which leads to complex within-year patterns of precipitation intensity. A secondary outcome of our results is an improved low-dimensional global parametrisation of monthly rainfall regimes. Our results provide the first gridded global overview of bimodal rainfall patterns and show the usefulness of simple mathematical approaches for detecting patterns in large data sets.

3.1 Introduction

Within-year climate patterns are a driving force behind environmental response in many locations. For example, the seasonal distribution of annual rainfall is important for large scale water resources planning and climate change impact assessment, because the seasonal availability of water influences society's need for reservoirs, flood defences, irrigation schemes, ecosystem management plans, etc. Seasonal precipitation patterns are complicated because they result from the interaction of large-scale climate factors with local topography [e.g. 67, 139, 260, 308]. Precipitation patterns vary from non-existent in arid regions, to having several alternating dry and wet seasons. We define this as a location's modality (unimodal being one wet and dry season per year; bimodal meaning two of each). This study investigates the global occurrence of bimodal precipitation (P) patterns and the accompanying number of rainfall events (N, days per month with $P \geq 0.1mm$, hereafter referred to as "rain-day frequency").

Early attempts to characterize rainfall seasonality rely on scattered station observations and include the use of a seasonality index in Africa, Australia and the US [238] and harmonic analysis on a global scale [146]. The seasonality index is unable to detect the differences between unimodal and bimodal regimes, but harmonic analysis can make this distinction. Later regional studies document bimodal precipitation patterns at a higher level of detail, here summarized by geographic region. The scattered nature of information makes it difficult to assess if, and unlikely that, this literature review covers all global occurrences of bimodal rainfall regimes. The tropics are relatively well-documented, although studies use different time periods, data sets, definitions and methods. Most observed rainfall patterns here are not bimodal, but either unimodal, constant or non-existent. Bimodal precipitation patterns outside the tropics seem similarly rare, although information on these regions is even more fragmented. We did not locate any global study showing regions of bimodal rain-day frequency in the literature, though this too has occasionally been documented in local studies [e.g. 202, 294]. Note that different authors assign different meanings to 'bimodal'. The term is either used to refer to alternating wetter and drier seasons or to the occurrence of two rainfall peaks within a single wet season. Additional

51

ambiguity stems from the lack of consistency in the definition of 'wet' and 'dry' seasons. For example, the Pacific slope of the mountains of Panama experiences "a rainy season from late April-early May to late November, and a dry season in December-April, during which 2 [precipitation] maxima are witnessed in June and October" [260]. These precipitation peaks are reportedly 5 months apart but are considered to belong to a single rainfall season. The distinction between wet and dry seasons tends to be a relative assessment per location, depending on a study's purpose and method. In the short literature review in the following section, the term 'bimodal' is used to refer to alternating wetter and drier seasons (as reported in the cited publications). Two rainfall peaks within a single wet season are mentioned as such.

Bimodal precipitation in Latin and South America: Bimodal monthly precipitation patterns in Meso- and tropical South America are limited to the western Andean cordillera, central Colombia, the Venezuelan Andes, the Eastern Guyana Highlands [260] and certain Guyana coastal regions [56]. The remaining regions have chiefly unimodal precipitation patterns [e.g. 14, 234, 260]. Precipitation over Mexico and Central America follows a dry winter, wet summer pattern. The summer season has a marked drop in precipitation during July and August known as the midsummer drought, but these months are still significantly wetter than the dry winter months [201, 338]. This is also the case for Caribbean islands north of approximately 17^oN [14]. Hsu & Wallace [146] also detect bimodal seasonality in central and north-eastern Argentina. **Africa:** Herrmann and Mohr [139] investigate the African continent and find bimodal precipitation patterns in east Somalia, Ethiopia, Kenya and South-Sudan, confirmed by Owiti and Zhu [246]. Bimodal precipitation also occurs in coastal Côte d'Ivoire, Ghana and central Madagascar, along with dual peaks in a single wet season in adjacent regions [139]. Regions with one wet season with two peaks also occur scattered around the equator [139]. Soliman [308] reports bimodal rainfall in southern Egypt. **Asia:** The tropical area of India is chiefly unimodal with a rainy season from July to September [247], although the south-eastern regions receive their main rains from October to December [68]. Western Sri Lanka has bimodal precipitation [203]. Chang et al. [67] find mainly unimodal patterns in South-East Asia. Exceptions are the Malaysian peninsula, North-East Borneo and certain small islands on the equator. **Oceania:** Precipitation patterns in northern Australia are unimodal [259]. **Extratropics:** Examples of bimodal precipitation regimes outside the tropics include western Arizona in the USA [138], the Pyrenees region in southern Europe [266], central Egypt [308] and the Indus delta in Pakistan [264]. Sayemuzzaman and Jha [277] report examples of major-minor wet seasons in coastal North Carolina. Keables [168] reports two wet peaks during summer in the Upper Midwest USA, within a single wet season. Tang and Reiter [313] further mention complex multi-modal (up to four annual peaks) rainfall patterns

over the North American plateau, stretching from the Canadian Rocky Mountains south to the Mexican border. Herrmann and Mohr [139] report bimodal rainfall in select regions in South Africa, Botswana, Namibia and Mozambique, in addition to scattered regions with two rainfall peaks in a single wet season. **Global:** the findings of Hsu & Wallace [146] are generally confirmed by the regional studies cited above.

A variety of methods are used to determine whether bimodal rainfall occurs, including visual assessment of time series [e.g. 14], harmonic analysis [e.g. 67, 246] and rule-based data analysis [e.g. 139], depending on the size of the study area. For large study areas, simplified representations of global climate are often used to reduce data dimensionality. To this effect, harmonic analysis has been used to describe precipitation seasonality in the USA [107] and the tropics [67, 119, 246]. Several other studies use sinusoidal curves to describe monthly climate variables on a regional [e.g. 219, 259] or global scale [30]. These studies calibrate a single harmonic with a 1-year period and assess how accurately this periodic function describes the original data. This approach has the advantage of compactly summarizing a climate variable's seasonality in three numbers with a clear physical interpretation: the mean value (corresponding to the monthly mean), the amplitude (describing the seasonality) and the phase (timing of the annual maximum). The studies that use sine curves apply these with a period of 12 months, giving a single peak per calendar year. Milly [219], and Berghuijs and Woods [30] suggested that a period of 6 months might be more appropriate for regions that experience bimodal rainfall patterns, but the effectiveness of this approach is so far unknown.

This paper builds on the framework presented in Berghuijs and Woods [30] to create a systematic and objective overview of global bimodal precipitation (P) and rain-day frequency patterns (N), by comparing independent, truncated sinusoidal functions with 12-month and 6-month periods in their capability to reproduce observed monthly P and N. When a curve with a single peak is more accurate than a curve with two annual maxima, we define the local climate as unimodal (i.e. having one drier and one wetter season). Similarly, when the curve with two peaks is more accurate than the curve with a single peak, we define the local climate as bimodal. Using the relative accuracy of two simple mathematical descriptions to define modality, we can avoid relying on subjective definitions of dry and wet seasons and avoid the use of empirical rules-based analysis.

3.2 Data

This study uses precipitation (P) and rain-day frequency (N) data directly obtained from the CRU TS v3.23 dataset [137], which contains monthly climate values for both variables for the period 1901-2014 on the planet's land areas (except Antarctica). It is based on

station observations combined with synthetic climatology and interpolated to a 0.5^o latitude/longitude resolution. The study period is defined as 1984-2014, averaged into a typical year (e.g. the average January is based on the 30 monthly January values from years 1984-2014). This balances data quality (number of observations on which the gridded data is based declines sharply further away from the present [137]), the influence of interannual variability (due to using a relatively long period) and the influence of climate change (by using the most recent data).

We use four additional data sets of varying resolution to assess the robustness of our conclusions with respect to uncertainty in the climate data: Global Precipitation Climatology Centre (GPCC, 1984-2013, 0.5^ox0.5^o resolution [283]), Global Precipitation Climatology Project (GPCP, 1984-2014, 2.5^ox2.5^o resolution [5]), Terrestrial Precipitation: 1900-2014 Gridded Monthly Time Series (TP, 1984-2014, 0.5^ox0.5^o resolution [207]) and Multi-Source Weighted-Ensemble Precipitation (MSWEP, 1984-2014, daily data aggregated to monthly scale, 0.25^ox0.25^o resolution [25]).

3.3 Methods

This study uses sinusoidal functions to approximate monthly averaged precipitation (P) and rain-day frequency (N) data of all global land cells on a 0.5^o latitude/longitude resolution. Each cell's P data is approximated with two independent sine curves, with a 12-month period (one annual peak) and 6-month period (two annual peaks) respectively. We quantify each curve's accuracy with an objective function defined as the average monthly error expressed as a percentage of the mean [30]. We assume that the sine curve with the lowest objective function value is the more appropriate description of the grid cell's climate, i.e. if the sine curve with period 6 months has a lower error than the 12-month period sine, we assume that the rainfall is bimodal. We apply the same procedure to the rain-day frequency data. A measure of the robustness of our conclusions follows from applying the sine-fitting method described above to 29 overlapping 3-year time windows in the 1984-2014 range and quantifying for how many periods each cell is classified as bimodal.

3.3.1 Sinusoidal functions

We use truncated sinusoidal functions that cannot reach physically unrealistic negative values to approximate monthly climate data [30]:

$$\hat{P}(t) = max\left(0, \bar{P}\left(1 + C_r(\delta_P) + \delta_P sin\left(2\pi\frac{t+3-s_P}{\tau}\right)\right)\right) \tag{3.1}$$

$$C_r(\delta_P) = -0.001\delta_P^4 + 0.026\delta_P^3 - 0.245\delta_P^2 + 0.2432\delta_P - 0.038 \tag{3.2}$$

$\hat{P}(t)$ is approximated monthly precipitation [mm], $\bar{P}$ the observed mean precipitation [mm], $C_r(\delta_P)$ an empirically derived correction factor which ensures that the mean of the truncated curve $\hat{P}(t)$ stays equal to $\bar{P}$, δ_P the fraction seasonal precipitation amplitude [-] and s_P the precipitation peak phase shift [months], which is shifted by an additional +3 months. The peak of a curve with $s_p = 0$ thus occurs in January, making non-zero s_p values easier to interpret. τ is the curve's period [months]. The functions for rain-day frequency N (number of days per month with $P \geq 0.1mm$) are defined similarly:

$$\hat{N}(t) = max\left(0, \bar{N}\left(1 + C_r(\delta_N) + \delta_N sin\left(2\pi\frac{t+3-s_N}{\tau}\right)\right)\right) \tag{3.3}$$

$$C_r(\delta_N) = -0.001\delta_N^4 + 0.026\delta_N^3 - 0.245\delta_N^2 + 0.2432\delta_N - 0.038 \tag{3.4}$$

These formulations necessarily mean that both peaks in the bimodal curve have the same amplitude. This limits the potential accuracy of our formulation to describe the data in locations where the climate is asymmetrical (e.g. when two annual dry seasons are of unequal length or severity). However, this limitation is required to keep the degrees of freedom the same between unimodal and bimodal formulations. Without this requirement, the accuracy metrics of both curves cannot be compared in a fair and meaningful way.

3.3.2 Calibration and evaluation

Parameters expressing the mean of the sine curve ($\bar{P}$, $\bar{N}$) are calculated directly as the mean of the data. Parameters δ (seasonal amplitude) and s (peak shift) are obtained through calibration by minimizing objective functions (eq. 3.1 and 3.3 respectively) using a multi-start Nelder-Mead simplex search method [188, 228]. The parameters are calibrated twice, once with seasonal length $\tau = 12$ [months] (from now called "unimodal sine") and once with $\tau = 6$ [months] ("bimodal sine").

This study concerns the typical seasonal pattern of climate variables. Therefore, the objective functions express the average error per month as a percentage of the mean [30]. This gives an overall impression of the sine curve's accuracy rather than focussing on a specific aspect such as getting the highest or lowest month captured accurately:

$$X_P = \frac{1}{12} \sum_{t=1}^{t=12} \frac{|\hat{P}(t) - P(t)|}{\bar{P}} \tag{3.5}$$

$$X_N = \frac{1}{12} \sum_{t=1}^{t=12} \frac{|\hat{N}(t) - N(t)|}{\bar{N}} \tag{3.6}$$

X_P shows the average simulated precipitation error [-] relative to the mean precipitation $\bar{P}$. $\hat{P}(t)$ are estimates of data P(t) obtained using sine curves. The objective function for N is analogous to P. We compare the accuracy of the calibrated unimodal and bimodal curves through simple subtraction (e.g. $\Delta X_P = X_{P,\tau=12} - X_{P,\tau=6}$) and interpret positive values to indicate the presence of a bimodal climate regime, because the bimodal curve better fits the data. These findings are compared to literature to judge the suitability of our method. We quantify our findings for both climate variables, assess the sensitivity of our conclusions to the choice of data period and comment on possible improvements to our scheme.

3.4 Results

Sine curves describing mean monthly precipitation totals (P, eq. 3.1) and rain-day frequency (N, eq. 3.3) are calibrated to fit observed monthly data using objective functions that minimise the average monthly error of each sine curve (eq. 3.5, 3.6 respectively). Using a seasonal duration $\tau = 12$ (one peak per calendar year), the average global P error is 0.18 [-], median 0.15 [-], with standard deviation 0.10 [-] which is comparable to the findings in Berghuijs and Woods [30]. The average global N error is 0.12 [-], median 0.097 [-], with standard deviation 0.098 [-]. In both cases larger errors occur mostly in locations with low annual precipitation (P < 240 mm/year, N < 30 days/year) or locations with hyperseasonal regimes (i.e. where the seasonal range in P or N is larger than the mean). Changing the seasonal duration to $\tau = 6$ (two peaks per calendar year) results in improved accuracy of P approximations for 7.9% and 6.7% land cells in and outside the tropics respectively. Increased accuracy for N with $\tau = 6$ is obtained for 8.5% of tropical and 8.9% of extratropical cells respectively (Table 3.1).

Figure 3.1a and Figure 3.1b shows the best objective function value obtained for P and N respectively, selecting the lowest option from unimodal and bimodal curves. Large errors occur mainly in drier areas where small absolute errors can be large relative to the mean. Figure 3.1c and 3.1d show the differences between unimodal and bimodal objective functions. A seasonal duration $\tau = 12$ is the more accurate option for most cells, but distinct areas with bimodal P and N are visible (areas with positive values, here bounded in red).

Table 3.1: Fraction of global grid cells where each combination of unimodal and bimodal P and N is most appropriate. Bimodal precipitation is slightly more common in the tropics than outside. Storm frequency shows no significant differences inside and outside the tropics.

		Global P			Tropics P			Non-tropics P		
		$\tau = 12$	$\tau = 6$	Σ	$\tau = 12$	$\tau = 6$	Σ	$\tau = 12$	$\tau = 6$	Σ
	$\tau = 12$	0.87	0.039	0.91	0.89	0.021	0.91	0.87	0.043	0.91
N	$\tau = 6$	0.055	0.032	0.087	0.026	0.059	0.085	0.065	0.024	0.089
	Σ	0.93	0.071		0.92	0.079		0.94	0.067	

Figure 3.1e and 3.1f show how often each cell is classified as bimodal when the average climate of overlapping 3-year sub-periods is used, rather than the average climate of the full 1984-2014 period. Darker areas show more stable bimodality while light shading indicates that a bimodal sine outperforms the unimodal sine in only a fraction of the 29 sub periods. It is therefore likely that in the darker areas the seasonal pattern is indeed bimodal. In the lighter areas conclusions about the climate's modality are more sensitive to the choice of data period, possibly because the seasonal signal is occasionally obscured by interannual variability.

Bimodal rain and rain-day frequency areas tend to overlap in the tropics (5.9% of cells, Table 3.1). Alternating rain and no-rain seasons is a hypothesis that can explain this overlap in the tropics, because this would lead to similar within-year distributions of both variables (non-zero N values being conditional on non-zero P values). Outside the tropics zero-rain months are less common and bimodal regions overlap less (2.4% of cells), indicating that the within-year distributions of both variables can be very different. Furthermore, the appropriateness of either climate description changes gradually in space. For locations further away from the equator, the difference between unimodal and bimodal sines is smaller than for locations close to the equator. Stronger climate seasonality in the tropics is a possible explanation, which can be investigated through the sine curve parameter values (Figure 3.2).

Figure 3.2e and 3.2f show the peak shift parameter in months away from January and are therefore represented on a circular colour scale. For areas with a seasonal length of 12 months, parameter values s_P and s_N of -6 and +6 months are equivalent. Within the black lines a seasonal length of 6 months is more accurate and here s_P and s_N values of -6, 0 and +6 months are equivalent. The locations where bimodal precipitation patterns occur cannot be explained through any obvious patterns in the sine curve parameter values, and the seasonal pattern is governed solely by the chosen seasonal length (τ) of the fitted curves.
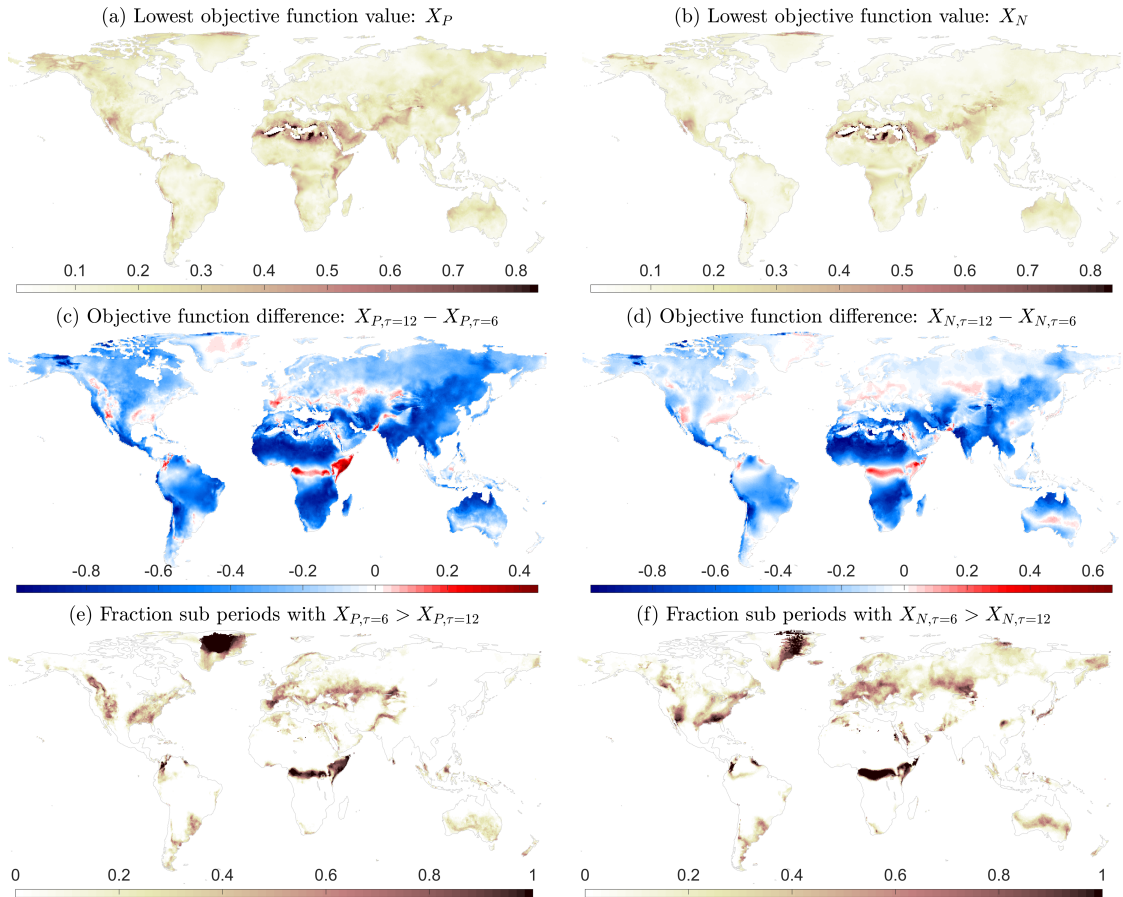
FIGURE 3.1. Calibration results from fitting sinusoidal functions to monthly averaged precipitation (P, left column) and rain-day frequency data (N, right column). Certain locations in the Sahara Desert show no objective function values as a mathematical artefact of there being no rainfall and no rain-days in the assessed period. (a, b) Lowest objective function value found for P and N respectively. (c, d) Difference between objective function values $X_{\tau=12}$ and $X_{\tau=6}$ for P and N respectively: lower (higher) values suggest that a seasonal length of 12 months (6 months) is more appropriate than the alternative. (e, f) Fraction of 3-year sub periods in the 1984-2014 data period in which a seasonal length of 6 months is more accurate than a seasonal length of 12 months.

However, the parameters do show that areas with larger objective function differences (Figure 3.1a, 3.1b) coincide with more extreme climates. These include arid and highly seasonal precipitation regimes in Africa (low $\bar{P}$, high δ_P), strongly seasonal P and N in India (high δ_P and δ_N values) and regions with a high number of rain events in Argentina

58

(a) Mean P parameter [mm/month]: $\bar{P}$

(b) Mean N parameter [days/month]: $\bar{N}$

(c) P seasonality parameter [-]: δ_P

(d) N seasonality parameter [-]: δ_N

(e) P timing parameter [month]: S_P

(f) N timing parameter [month]: S_N
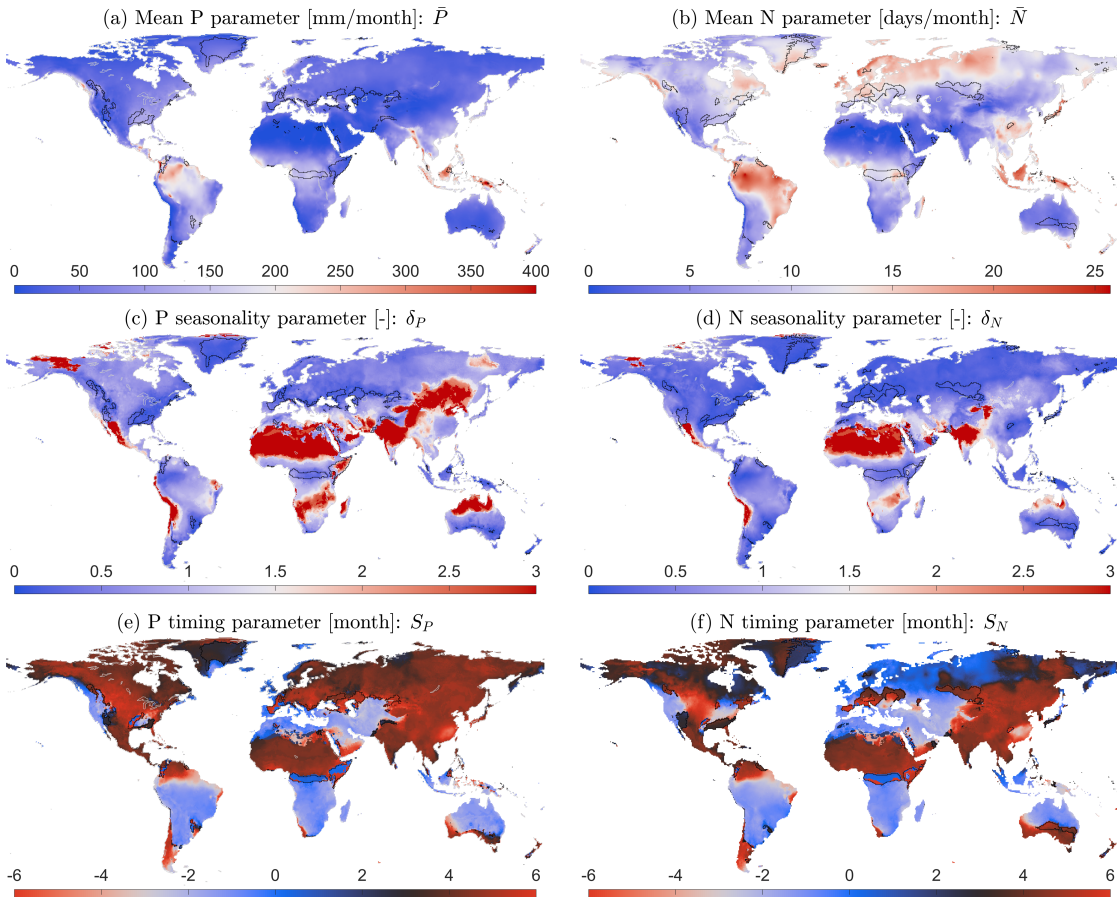
FIGURE 3.2. Parameter values for sine curves describing monthly average precipitation (P) and rain-day frequency (N), with areas where a seasonal length of 6 months is more appropriate than 12 months bounded in black. (a) Mean monthly P [mm], calculated from data; the scale is capped at 400mm to better highlight regional differences (0.06% of cells have values > 400mm/month). (c) P seasonality as a dimensionless multiplier of mean P; the scale is capped at 3 to better highlight regional differences (12.4% of cells have values > 3, predominantly in the extremely arid regions). (e) Timing of the P peak given in months away from January. (b) Mean monthly N [events/month], calculated from data. (d) N seasonality as a dimensionless multiplier of mean N, the scale is capped at 3 (5.8% of cells have values higher than 3, again predominantly in the very arid regions). (f) Timing of the N peak given in months away from January.

(high $\bar{N}$ values). Peak timing parameters s_P and s_N are similar for most regions in the Southern Hemisphere, meaning P and N distributions are roughly in phase and thus the

number of rainfall events goes up as monthly rainfall volumes increase. In the Northern Hemisphere P and N are more independent and the two distributions can be out of phase (e.g. Northern Europe). An increase in monthly rainfall totals thus does not necessarily lead to an increase in rainfall days, but might cause changes in the average rainfall per event.

3.5 Discussion

Previously, information about global bimodal precipitation was scattered across many different studies. This study uses gridded global climate data and simple mathematical analysis to create a systematic global overview of the location of bimodal precipitation regimes. To assess the sensitivity of our conclusions to data uncertainty, we have repeated the analysis with four additional precipitation datasets. Across all data sets, the same regions are consistently identified as having bimodal rainfall climates (Figure 3.3). Figure 3.3c is a straightforward comparison of the CRU TS and GPCC data sets, showing that results are consistent for the majority of grid cells. The tropical locations generally align well with investigated literature, except for a region of Africa centred on the equator. Herrmann and Mohr [139] classify this region as "humid" instead of bimodal. This definition of humid climates however depends on the minimum monthly rainfall in relation to the monthly temperature and not on the pattern of the monthly precipitation values, which has distinct wetter and drier (relative to the wet months) periods in this region. There are no other obvious instances within the tropics where our method indicates a bimodal rainfall regime but the available literature does not (deviations where our method indicates no bimodal regime but the literature does are discussed in a later section). Outside the tropics we see various cases where the bimodal sine curve is mathematically slightly more accurate than the unimodal sine (e.g. several locations in southern South America, large areas in the South-East USA, most of Greenland, certain Balkan areas and large regions of Southern Russia and Kazakhstan). However, the seasonal pattern in these areas is relatively asymmetrical and furthermore subject to strong interannual variations (as evidenced by Figure 3.1e, which shows that conclusions for these areas can change depending on the data period used). Combining the information about the accuracy of both curves with information about the sine curve parameters shows that in these regions the differences in accuracy between both curves are small ($|X_{\tau=12} - X_{\tau=6}| < 0.1$), but the accuracy of both is quite good ($X_P < 0.25$; compare figures 3.1a and 3.1c). From the parameter values we can deduce that these areas experience relatively high rainfall (figure 3.2a) and low seasonality (figure 3.2c). The seasonal pattern occurs on top of a high baseline rainfall, and the seasonal fluctuations in the pattern are thus relatively small compared to the

overall signal. For certain locations regional studies report bimodal regimes where one wet season has distinctly higher rainfall than the other wet season (i.e. Caribbean islands [14]; Malaysian peninsula and several Indonesian islands [67]; scattered regions on the African continent [139]). Our method shows no clear improvement of bimodal sine curves over unimodal sines for these locations because our very simple bimodal sine curves are unable to replicate this specific behaviour. These regions show in our results as areas where the accuracy of both sine functions is similar (figure 3.1c, d) and low (figure 3.1a, b), indicating that neither curve represents the seasonal pattern well (see also panels b and c in figure 3.5, where the asymmetry of the wet seasons causes the unimodal sine to be more accurate than the bimodal sine, but the objective function values show that both are inaccurate for this specific case). More complex mathematical formulations might have the required degrees of freedom to describe these regimes more accurately but will necessarily have to sacrifice parsimony for improved accuracy. The 2-month long midsummer drought in Mexico and Central America , and similar patterns in other regions such as equatorial Africa, do not show up as bimodal in our results. In these cases, the stronger signal is the single wet season as described by the unimodal curve, and the double peaks within this single season are relatively minor fluctuations.

Figure 3.4 gives visual examples of the extent to which sine curves can capture the within-year variability, by comparing the accuracy of unimodal and bimodal curves for P and N at five different percentiles of $X_{\tau=12} - X_{\tau=6}$. The leftmost plots show that for certain locations the bimodal sine curve gives only a marginal improvement over the unimodal sine and that for other cases neither sine curve is a very appropriate description of the local climate. The actual errors between simulated and observed values are quite low for these specific instances and the climate might be considered constant instead of uni- or bimodal. The rightmost plots are in the top 2.5% of improvements and show that a bimodal sine curve can be good approximation of seasonal patterns. The unimodal sine is unsuitable for these cases, being either a straight line (N) or completely missing the peak events in both magnitude and timing (P). The middle columns show examples where a bimodal sine is more appropriate than a unimodal sine, but unable to capture the within-year variability properly. These cases provide a visual example of bimodal sines correctly detecting a bimodal regime but lacking mathematical freedom to simulate the seasonal pattern accurately.

However, even if this method is a rough approximation of local climates, comparing the accuracy of unimodal and bimodal curves gives insight into how climates change gradually in space on a global scale. We can interpret the difference in accuracy (Figure 3.1c, 3.1d) from negative to positive as a unimodal regime being more likely, to a bimodal regime being more likely. This implies that a gradual change occurs between two extremes, with
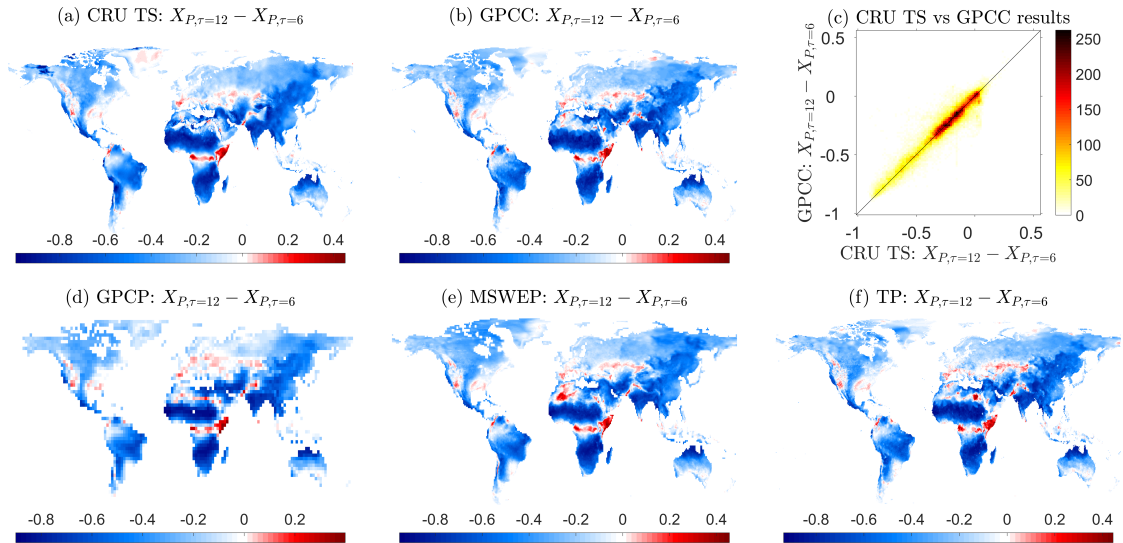
FIGURE 3.3. Stability of results under different precipitation data sets. Plot a, b, d, e and f show objective function differences between calibrated unimodal and bimodal sine curves for different data sets. (a) CRU TS data, 1984-2014, 0.5°x0.5° resolution, (b) GPCC data, 1984-2013, 0.5°x0.5° resolution, (c) overlap between CRU TS results and GPCC results, (d) GPCP data, 1984-2014, 2.5°x2.5° resolution, (e) MSWEP data, 1984-2014, 0.25°x0.25° resolution, (f) TP data, 1984-2014, 0.5°x0.5° resolution.

regimes in the middle that are neither clearly unimodal nor bimodal and/or where both mathematical formulations are inadequate. Figure 3.5 shows a longitudinal cross section of eastern equatorial Africa, where the climate gradually changes from asymmetrical bimodal (39^olon; the dry seasons have different lengths and the rainfall peaks have different heights) to more symmetrical bimodal (40.5^olon) to nearly unimodal rainfall (42^olon). For the first case, limitations in the bimodal sine curve formulation (fixed amplitude, 6-month period) make it unable to outperform the unimodal sine curve in objective function terms, even though the bimodal sine curve gives a better general idea of the rainfall seasonality. As the observed pattern shifts to a more symmetrical bimodal regime ($+40.5^o$ to $+41.5^o$lon) the bimodal sine clearly outperforms the unimodal sine. Further east the rainfall regime slowly shifts to being more unimodal (one of the precipitation peaks gradually becomes lower) and the unimodal sine becomes a more appropriate choice, as is reflected by its improved objective function value. Even though the asymmetry in bimodal regimes is not captured, this example shows that gradual changes in dominant climate patterns can be detected with this method.
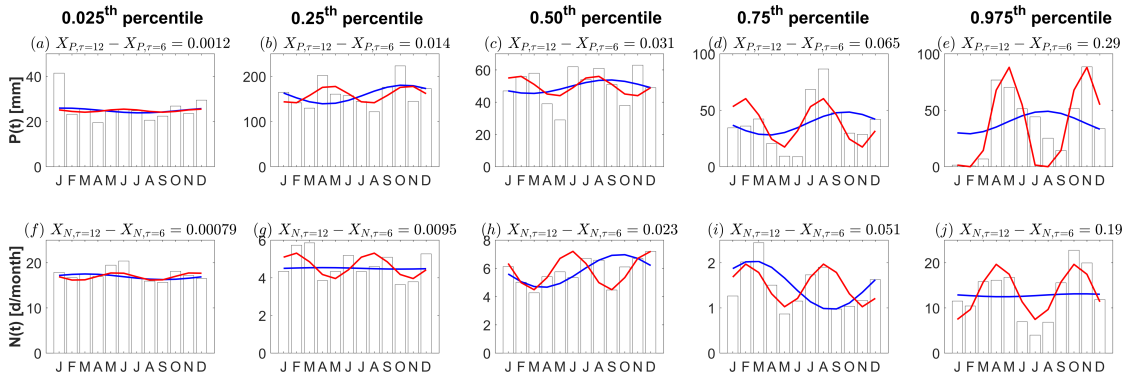
FIGURE 3.4. (a-e) 0.025^{th}, 0.25^{th}, 0.50^{th}, 0.75^{th} and 0.975^{th} percentile respectively of locations where a bimodal sine curve is a better approximation of local rainfall regime than a unimodal curve is. (f-j) Similar plots for local rain-day frequency patterns. The leftmost plots indicate regions where both curves have similar accuracy; bimodal curves become more accurate compared to unimodal curves in the columns to the right.

This work provides insight in the sequence of dry and wet seasons on a global scale, and the relative importance of these intra-annual variations compared to the baseline signal. Strong intra-annual precipitation variability can be seen in much of the Southern Hemisphere (Figure 3.1c, 3.1d), expressed as one of our mathematical formulations being much more accurate than the other. In these regions, the rain-day frequency pattern tends to be correlated with the precipitation (total) pattern, implying that the number of rain-days and precipitation total are conditional on one another. This is less so for much of the Northern Hemisphere, where the intra-annual precipitation variability is relatively small compared to the baseline (mean precipitation) and the rain-day frequency follows a more independent seasonal pattern. As a low-dimensional parametrisation of monthly rainfall regimes, these mathematical functions are an objective way to determine the strength of modality, they show spatial changes in climatic seasonality as a gradual change instead of a jump between categories and create the possibility to attach physical meaning to the functions' parameters. This make them a useful tool in any large-scale climate study. For example, these methods can be applied to investigate whether global rainfall regimes experience shifts from unimodal to bimodal, and whether this occurs as a result of natural variability or if trends can be found.
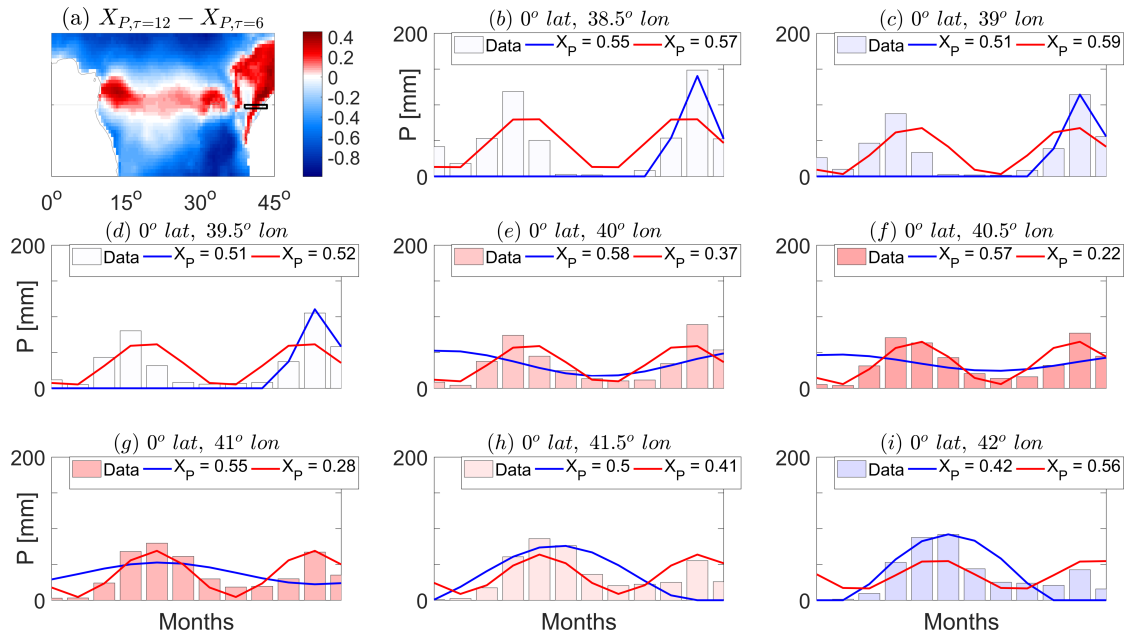
FIGURE 3.5. Longitudinal cross-section of East Africa, showing average precipitation regimes. (a) Objective function difference between calibrated unimodal and bimodal sine curves. The black box indicates the locations where the following plots zoom in on. (b-i) Average precipitation regime (bars), calibrated unimodal sine curve (blue) and bimodal sine curve (red). Objective function (X_P) values for both curves are given in the legend and data bars are coloured proportionally to which curve approximates the data more accurately. The images show a precipitation regime that gradually shifts from asymmetrical bimodal (b & c) to approximately symmetrical bimodal (f) to nearly unimodal (i).

3.6 Conclusions

This study provides a systematic global overview of where bimodal precipitation (P) regimes occur, along with information on bimodal rain-day frequency (N), based on gridded climate data. Before, this information was scattered throughout various regional studies using different data, methods and definitions. This work aims to introduce a measure of objectivity by using simple sinusoidal functions to describe monthly rainfall and rain-day frequency data, using formulations that allow for either one or two peaks per year. Comparing the average monthly error obtained by both formulations indicates whether a unimodal or bimodal climate regime is more likely for a given location. Globally, bimodal precipitation occurs in approximately 7.1% of grid cells, divided into 7.9% of tropical cells and 6.7%

of non-tropical cells. These results mostly confirm findings from regional studies, except one case due to definitions used in the regional study and several cases where the annual climate pattern is asymmetrical in either peak height or the respective length of dry and wet season(s), which is more complex than our simple formulations can describe. Overall, the spatial transition between unimodal and bimodal climates is gradual (e.g. north- and southwards of equatorial Africa and in northern Guyana), and our simple approach can detect these patterns. Sensitivity analysis shows that for areas in the transition zones between unimodal and bimodal (e.g. in north-eastern Kenya and northern Borneo), natural variability over a few years can cause the climate to switch back and forth between unimodal and bimodal. Within the tropics, the modality of rain-day frequency N tends to overlap with the modality of P, due to the occurrence of alternating dry and wet seasons with zero or low rainfall during the dry period. Outside the tropics, modality of P and N is more independent, which leads to complex changes in precipitation intensity during the year.

MULTI-MODEL COMPARISON FRAMEWORK

This chapter is currently under review as a model description paper to Geoscientific Model Development. Slight modifications have been made to better fit the general layout of this thesis. This chapter is the result of a collaboration with Dr. Murray Peel and Dr. Keirnan Fowler, who acted as co-supervisors during Wouter Knoben's visit to the University of Melbourne.

Citation: Knoben, W. J. M., Woods, R. A., & Freer, J. E. (2018). Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT) v1.0: an open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous space-state formulations. Geoscientific Model Development, under review. https://doi.org/10.5194/gmd-2018-332

Abstract

This paper presents the Modular Assessment of Rainfall-Runoff Models Toolbox (MAR-RMoT): a modular open-source toolbox containing documentation and model code for 46 existing conceptual hydrologic models. The toolbox is developed in Matlab and works with Octave. Models are implemented following several good practices for model development: definition of model equations (the mathematical model) is kept separate from the numerical methods used to solve these equations (the numerical model) to generate clean code that is easy to adjust and debug; the Implicit Euler time-stepping scheme is provided as the default option to numerically approximate each model's Ordinary Differential Equations in a more robust way than (common) Explicit schemes would; threshold equations

are smoothed to avoid discontinuities in the model's objective function space; and model equations are solved simultaneously, avoiding physically unrealistic sequential solving of fluxes. Generalized parameter ranges are provided to assist with model inter-comparison studies. In addition to this paper and its Supporting Materials, a User Manual is provided together with several workflow scripts that show basic example applications of the toolbox. The toolbox and documentation are available from https://github.com/wknoben/MARRMoT (DOI: 10.5281/zenodo.2482542). Our main scientific objective in developing this toolbox is to facilitate the inter-comparison of conceptual hydrological model structures which are in widespread use, in order to ultimately reduce the uncertainty in model structure selection.

4.1 Introduction

Rainfall-runoff modelling is useful to extrapolate our hydrologic understanding beyond measurement availability [38, 45]. We can challenge and improve our understanding of the way catchments function through model-based hypothesis testing [37, 74, 106, 173, 174] and simulate the impact of changes in climatic conditions and catchment characteristics such as land use [23, 100, 176, 249, 290, 333]. Many different modelling approaches are possible, ranging from lumped, empirical, deterministic bucket-style models to distributed, process-oriented, stochastic, 3D physics-based models [45]. Each of these approaches has its own advantages and drawbacks, concerning the level of spatial detail, amount of model 'realism' in terms of processes represented, input data requirements and computational time. The toolbox presented in this paper uses deterministic, spatially lumped bucket-style models, also referred to as conceptual hydrological models. Note that this definition of a conceptual model is different from the definition used by authors discussing the modelling process, where the conceptual model is a step between having a mental, perceptual model of a catchment and the collection of equations referred to as a mathematical/procedural model [e.g. 45, 72, 128].

Every application of a rainfall-runoff model is complicated by various aspects of uncertainty [e.g. 47, 248, 249]. Uncertainty is introduced during measurement of model input variables such as precipitation [e.g. 244] and temperature [e.g. 22] and derived variables such as potential evapotranspiration [e.g. 12, 242, 244]. Uncertainty is also present in measurements against which model output is compared, such as streamflow [e.g. 92, 213], water table depth [e.g. 114] and water quality [e.g. 214]. Values of model parameters can be uncertain due to dependency of 'optimal' parameter values on climatic conditions during model calibration [e.g. 80, 110], due to the choice of calibration algorithm [16] or due to the performance metric used [e.g. 96, 129]. Finally, the choice of model structure (i.e. the collection of equations and their internal connections that make up the model) itself is a source

of uncertainty [11, 80, 104, 106, 185, 325]. Currently, a wide variety of models are available. They may be different in spatial and temporal resolution, or include different processes, be deterministic or stochastic, might be based on top-down or bottom-up philosophies, or be different in some other way. This paper contributes to the investigation of model structure uncertainty of lumped, deterministic conceptual models. We hope to make progress towards answering a core question in hydrologic modelling: out of the overwhelming number of available models, which one is the most appropriate choice for a given catchment?

Conceptual models tend to have low data requirements (catchment-averaged forcing instead of spatially explicit) and are less computationally intensive than spatially explicit models. They are used in both scientific and operational settings [254]. A wide range of conceptual model structures exists, e.g. SACRAMENTO [60, 225], TOPMODEL [46], SIMHYD [70], the TANK model [310] and many more, but there is no clear basis to choose between the different models [45]. Models are different both in their internal structure (i.e. which storages are represented and how they are connected) and in their choice of flux equations (i.e. whether and how any given flux is quantified with a mathematical equation). Choosing the right model for a catchment where hydrological responses are measured is difficult because achieving a 'good' value on a performance metric is a necessary but not sufficient condition to determine whether a model produces the "right results for the right reasons" [173]. Different model structures can achieve superficially similar performance metrics, but might reach this point by wildly different internal dynamics [88, 121, 254]. Therefore, good simulation metrics do not necessarily tell us which model structure is more appropriate for this catchment. Choosing a suitable model structure when the catchment is ungauged is even more challenging. This model structure uncertainty is largely unquantified, even for existing models with a long legacy of 'successful' (often defined as having achieved a high value for some performance metric) applications. However, comparison of different models can be an expensive task if each model needs to be set up individually. Model inter-comparison studies are further complicated by the fact that documented computer code is unavailable for many model structures.

In recent years multi-model frameworks have received considerable attention. These provide a standardized framework in which several models are presented, or users can construct new models, or both. This reduces the time cost of a model comparison study, allows fair comparison of different model structures in a test case and allows the investigator to isolate choices in the model development process. Examples include the Modular Modelling System (MMS) [191], the Rainfall-Runoff Modelling Toolbox (RRMT) [331], the Framework for Understanding Structural Errors (FUSE) [78], a fuzzy model selection framework [20], SUPERFLEX [103, 161], the Catchment Modelling Framework (CMF) [184] and the Structure for Unifying Multiple Modelling Alternatives (SUMMA) [75, 76].

These frameworks are either limited to a small number of existing models (e.g. MMS, RRMT), use a pre-defined internal organization of stores (FUSE), consist of generic model elements (i.e. stores, fluxes and lags) that are not easily recognizable as existing models (e.g. CMF, SUPERFLEX), or are more physics-based and thus difficult to use with conceptual models (e.g. SUMMA).. Thus, despite these many existing frameworks, there is a need for a new framework that provides a user-friendly, standardized way to construct and compare existing, widely-used conceptual models, without constraining the allowed model architecture a priori.

This paper introduces the Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT) to fill a gap in the current selection of multi-model frameworks. MARRMoT provides an open-source, easy-to-use, expandable framework that currently includes 46 different conceptual model formulations. This provides all the benefits of a multi-model framework: models are constructed in a modular fashion from separate flux equations, which allows easy modification of provided models and expansion of the framework with new models or fluxes; good practices for numerical model solving are implemented as standard options; and all MARRMoT models require and provide standardized inputs and outputs. The large number of models in the framework facilitates studies that lead to more generalizable conclusions about model and/or catchment functioning. This work also provides a pragmatic overview of the wide variety of different flux equations and model structures that are currently used, facilitating studies and discussion beyond direct model comparison. Due to the code being open source, transparency and repeatability of research is encouraged, additions to the framework are possible, and the community can find and correct any mistakes. Finally, MARRMoT is provided with extensive documentation about the models included, the conversion of flux equations to computer code, recommendations for generalized parameter ranges for model sensitivity analysis and/or calibration, a User Manual explaining framework set-up, functioning and use, and several example workflow scripts that allow use of the framework even with minimal programming experience.

4.2   MARRMoT design considerations

MARRMoT takes inspiration from earlier modular frameworks (e.g. FUSE [78], FLEX [103]) and uses modular code with individual flux equations as the basic building blocks. Multi-model frameworks benefit from modular implementation because this simplifies programming of the framework and makes it easier to (i) re-use components of a model in a different context, including cases where the same basic equation is used by multiple models; and (ii) add new options to the framework [78]. MARRMoT follows several other good practices for model development which are briefly described in the following sub-sections.

4.2.1 Separation of model equations and equation solving

First, MARRMoT uses a distinct separation of model equations (ODEs) and the numerical approach used to solve these equations. In the theoretical process of developing a new hydrological model, the modeller ideally goes through several distinct steps [e.g. 45, 72, 128]. To start, the modeller develops a mental, perceptual model of catchment behaviour based on observations and/or other knowledge (i.e. expert opinion). Next, this model is simplified into an abstraction that shows the connection of the most important fluxes and storages (also termed a conceptual model, but this is a distinctly different meaning than when applied to a bucket-type hydrologic model). These relations are then formalized as Ordinary Differential Equations and their constitutive functions in a mathematical model. Finally, creating computer code to solve these equations sequentially as a time series is done with the procedural model. In practice however, these stages are often not distinct and tend to overlap [e.g. 164], a process referred to as "ad hoc" modelling. Overlap of the mathematical and procedural model can lead to altered model behaviour and difficulty with parameter estimation [72, 160, 164]. A clear separation between model equations and the code used to solve those equations gives computer code that is easier to understand and update with new time-stepping schemes or flux equations, relative to code where the model equations are interwoven with the numerical scheme.

4.2.2 Robust numerical approximation of model equations

Second, MARRMoT gives the possibility to choose a numerical method to approximate the ODEs in discrete time steps. Currently, a fixed-step Implicit Euler method is recommended as default, and an Explicit Euler method is provided for result matching with previous studies. Many implementations of hydrologic models use the Explicit Euler method to approximate storage changes [285, 297]. The Explicit Euler method relies on storage values at the start of a time step to estimate flux sizes in the current time step: FLUX(t) = f(STORE(t-1)). This method is easy to implement and fast to compute, but has several disadvantages: it has low accuracy and only conditional stability, which can lead to large numerical errors and amplification of such errors under certain conditions [72, 160, 285]. Implicit methods such as Implicit Euler instead rely on an iterative procedure that relates flux size to storage at the end of a time step: FLUX(t) = f(STORE(t)). These methods require more intensive iterative computation, but avoid the aforementioned issues even when implemented with fixed time step sizes [167, 285]. Higher-order numerical approximation methods are currently not provided in MARRMoT but can be included in a straightforward manner. Note that fixed time step size refers to the use of a single time step size throughout a simulation (e.g. hourly, daily), and does not prescribe the time step size.

4.2.3   Smoothing of threshold discontinuities in model equations

Third, MARRMoT removes threshold discontinuities in model equations through logistic smoothing [78, 163]. Hydrologic processes are often characterized by thresholds, e.g. snowmelt starts when a certain temperature is exceeded, and saturation excess flow occurs when the soil is saturated. Introducing threshold behaviour into hydrologic models leads to discontinuities in the model's objective function, which can complicate parameter estimation when small changes in parameter values may lead to large changes in objective function value or in the gradient thereof [163]. Smoothing model equations avoids these discontinuities but also involves a fundamental change to the model equations. Kavetski and Kuczera [163] recommend logistic functions to smooth threshold equations that closely resemble the original threshold function but are continuous throughout the function's domain. MARRMoT smooths storage-based thresholds with a logistic function [78]:

$$Q_o = Q_{in}\left(1 - \Phi(S, S_{max}, \rho_S, \epsilon)\right) \tag{4.1}$$

$$\Phi(S, S_{max}, \rho_S, \epsilon) = \frac{1}{1 + e^{\frac{S - S_{max} - \omega\epsilon}{\omega}}} \tag{4.2}$$

Where Q_o and Q_{in} are flux output and input respectively and $\Phi(..)$ the smoothing operator. S and S_{max} are current and maximum storage respectively, ω represents the degree of smoothing according to $\omega = \rho_S * S_{max}$, and ϵ is a coefficient that ensures that S does not exceed S_{max}. ρ_S and ϵ can be specified by the user, or used with default values of 0.01 and 5.00 respectively [78]. Temperature-based thresholds are smoothed with a different logistic function [163]:

$$P_s = P\Phi(T, T_0, \rho_T) \tag{4.3}$$

$$\Phi(T, T_t, \rho_T) = \frac{1}{1 + e^{\frac{T - T_0}{\rho_T}}} \tag{4.4}$$

Where P_s is precipitation as snow, P incoming precipitation and $\Phi(..)$ the smoothing operator. T and T_0 are the current and threshold temperatures respectively, and ρ_T is the smoothing parameter with default value 0.01.

4.2.4   Simultaneous solving of model equations

Fourth, MARRMoT solves all model equations simultaneously rather than sequentially. Operator-splitting (OS) numerical approximations integrate fluxes sequentially and can be useful in cases such as large systems of partial differential equations, where computational

speed would otherwise be a limiting factor [103]. Sequential calculation of model fluxes is common practice in many hydrologic models (e.g. SACRAMENTO and GR4J) but this approach assumes that fluxes occur in a pre-determined order. It is preferable to integrate model fluxes simultaneously to avoid "physically unsatisfying assumption[s]" [103, 271]. MARRMoT follows this recommendation, barring certain cases where the model is divided into two distinct parts due to a delay function, in which case simultaneous solving of the first and second part of the model is impossible.

4.3 MARRMoT

MARRMoT provides Matlab code for 46 conceptual models following the model development practices outlined in Section 4.2. This section provides a summary of the framework because it is infeasible to discuss every individual model here. References to the Appendices guide the interested reader to a more in-depth discussion of each model and its implementation in MARRMoT. In addition to this chapter, the MARRMoT documentation includes the following:

- Appendix B.2 - Model descriptions. This section contains descriptions of all 46 models in a standardized format. Each description includes a short introduction to the model, a list of parameters, a model schematic and a discussion of the ODEs and constitutive functions that describe the model's storage changes and fluxes.

- Appendix B.3 - Flux equation code. This section contains an overview of the 105 different flux equations used in MARRMoT, and their implementation as computer code.

- Appendix B.4 - Unit Hydrograph overview. This section contains an overview of the 7 different Unit Hydrograph routing schemes used in MARRMoT.

- Appendix B.5 - Parameter ranges. This section contains an overview of recommended parameter ranges for the 46 models based on published literature about hydrologic process and model application studies. The ranges are standardized across models, so that similar processes use similar parameter ranges. Use of these parameter ranges is optional.

- User Manual (available from the MARRMoT Github page): This document helps a user set up MARRMoT for use in either Matlab or Octave, outlines the inner workings of the standardized models, provides several workflow examples and provides examples on how to create a new flux equation or model.

4.3.1 General MARRMoT outline

Figure 4.1 shows the set-up of the MARRMoT framework and what the framework requires (i.e. data, model options, etc.) and provides for a given modelling study. Each model has its own separate model function, which contains both the numerical implementation of the model (i.e. the ODEs and fluxes that make up this model, as given in Appendix B.2, B.3 and B.4) and the necessary code to handle user input, run the model to produce a time series and generate output. The user is expected to provide the following inputs: time series of climate variables, initial values for each model store, choice of numerical integration method and settings for Matlab solvers, and values for each model parameter. Note that the solver selection relates to time-stepping numerics, not parameter selection and optimisation. Optionally, MARRMoT's provided parameter range guidance (Appendix B.5) can inform the choice of parameter values. Parameter ranges have been standardized as much as possible across all models, such that similar processes use the same range of possible parameter values across models (e.g. this ensures that all models that have an interception component with a maximum capacity can use the same range, 0-5mm, for their respective interception capacity parameter). Each model generates a time series of total simulated flow and total simulated evaporation as default output. Optionally, users can request variables with time series of storages and internal fluxes, as well as a summary of the main water balance components. The User Manual provides several workflow examples that showcase possible uses of MARRMoT: the examples cover (i) application of a single model, with a single parameter set to a single catchment, (ii) random parameter sampling from provided parameter ranges for a single model, (iii) application of three different models to a single catchment, and (iv) calibration of a single parameter set for a single model.

The basic building blocks inside each model function are flux functions. Each flux function describes a single flux, for example evaporation from an interception store, water exchange between two soil moisture stores or baseflow from groundwater. Flux functions are kept separate from the model functions, and each model calls several flux functions as needed. This allows for consistency across models (if errors are present in any flux function, at least they are the same in all models), easy implementation of new flux equations and facilitation of studies that are specifically interested in differences between various mathematical equations that all represent the same flux or process. The inputs required, and output returned by each flux function varies. Appendix B.3 provides a full overview of the mathematical functions used to represent fluxes in each model description, relevant constraints, numerical implementation of each flux in MARRMoT and a list of models that use each flux function). Various models use a Unit Hydrograph approach to delay flows
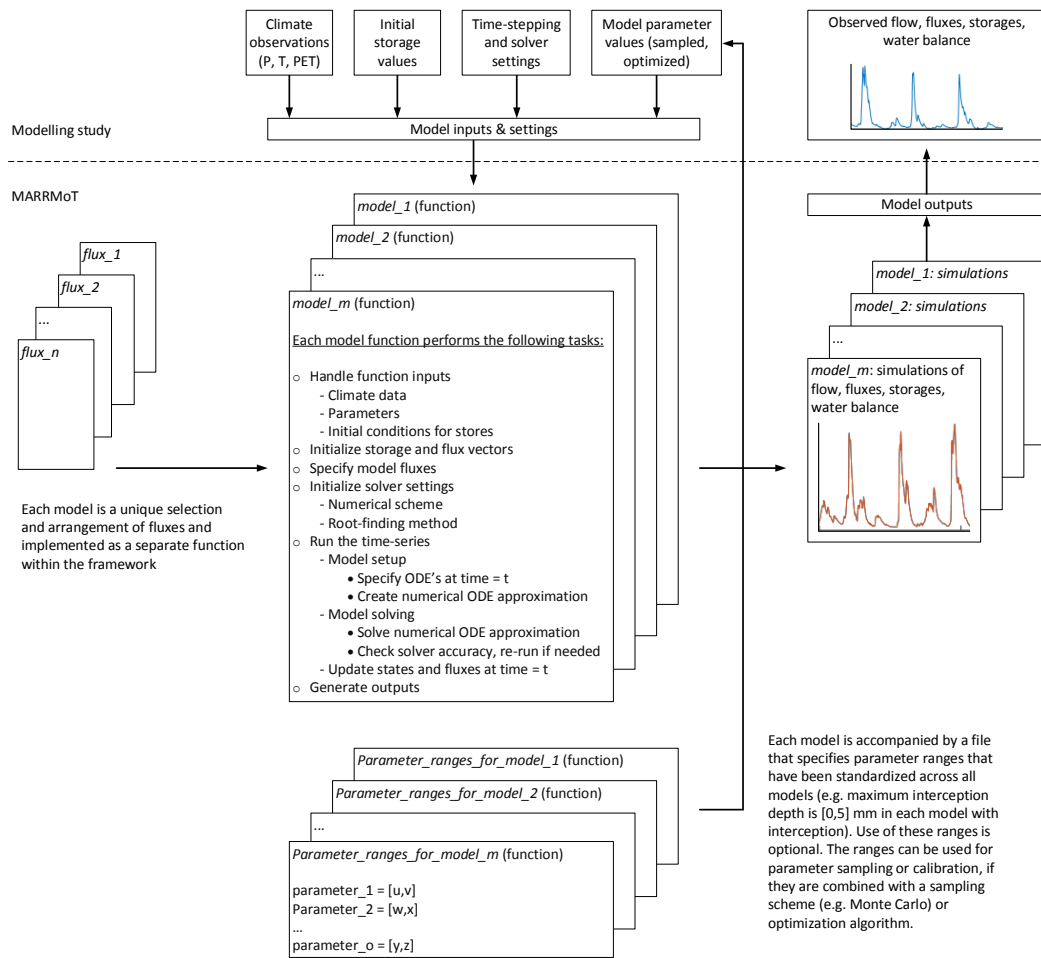
FIGURE 4.1. Schematic overview of the MARRMoT framework. MARRMoT provides 46 conceptual models implemented in a standardized way (part below the dotted line). Each model is a unique collection and arrangement of fluxes, but the coding set-up of each model is the same. Inputs required to run a model are time series of climate variables, settings for the numerical approximation scheme, values for the model parameters (which can optionally be sampled or optimized using provided, standardized ranges), and initial conditions for each model store. The model returns time series of simulated flow, fluxes and storages and a summary of the simulated water balance.

within the model and/or simulate flow routing. Appendix B.4 provides a full overview of Unit Hydrographs currently implemented in MARRMoT.

4.3.2 Summary of included models

Table 4.1 shows which models are currently implemented in MARRMoT and the main reference(s) for each. Some of the models have a long history of application, others are part of model comparison or development studies. MARRMoT development was not guided by a specific modelling objective (e.g. droughts, floods) and the current selection of model structures mainly aims for variety in the range of model structures. The User Manual provides guidance on changing and expanding the framework and, due to its open nature, these additions can be shared with the wider community. Each model is internally different from the others, either through using different configurations of stores and their connections, or through using different flux equations, or both. Models with sequential numbering (e.g. mopex1, mopex2) are part of the same study and tend to be similar but more elaborate as the number increases. Detailed model descriptions can be found in Appendix B.2.

Figure 4.2 provides a summarized overview of the model differences, expressed through the number of stores, number of parameters and hydrological processes represented. Models use between 1 and 8 stores, and between 1 and 23 parameters. The number of parameters tends to increase with the number of stores, but exceptions exist. Most models' stores are used to track moisture availability (i.e. across all models 162 stores are used, 155 of which track moisture availability); deficit stores are much rarer (i.e. only 7 out of 162 stores are used to track moisture deficit). Soil moisture storage is the most commonly modelled concept, occurring in every model. Routing stores (e.g. "fast flow routing") are included in 18 models, groundwater stores in 13 models, snow storage in 12, interception in 10, unit hydrograph routing also in 10, surface depression storage in 2 and channel storage in 1 model. However, these numbers should not be seen as representative of all conceptual models, because our model overview is necessarily incomplete and some of our models are part of model development studies (where a model is modified until satisfactory performance is obtained). These studies skew the number of stores in certain categories.

Table 4.1: MARRMoT models and function names

ID	Name	Main reference(s)	MARRMoT function
1	Traditional bucket model	[158]	m_01_collie1_1p_1s
2	Wetland, FLEX-Topo	[274]	m_02_wetland_4p_1s
3	Unnamed	[158]	m_03_collie2_4p_1s
4	Unnamed	[19]	m_04_newzealand1_6p_1s
5	IHACRES	[84, 199]	m_05_ihacres_6p_1s
6	Unnamed	[95]	m_06_alpine1_4p_2s
7	GR4J	[255, 271]	m_07_gr4j_4p_2s
8	Unnamed	[20]	m_08_us1_5p_2s
9	Unnamed	[309]	m_09_susannah1_6p_2s
10	Unnamed	[309]	m_10_susannah2_6p_2s
11	Unnamed	[158]	m_11_collie3_6p_2s
12	Unnamed	[95]	m_12_alpine2_6p_2s
13	Hillslope, FLEX-Topo	[274]	m_13_hillslope_7p_2s
14	TOPMODEL	[50, 78]	m_14_topmodel_7p_2s
15	Plateau, FLEX-Topo	[274]	m_15_plateau_8p_2s
16	Unnamed	[19]	m_16_newzealand2_8p_2s
17	Penman drying curve	[251, 331]	m_17_penman_4p_3s
18	SIMHYD	[70]	m_18_simhyd_7p_3s
19	Unnamed	[101]	m_19_australia_8p_3s
20	GSFB	[224, 350]	m_20_gsfb_8p_3s
21	FLEX-B	[106]	m_21_flexb_9p_3s
22	VIC	[78, 194]	m_22_vic_10p_3s
23	LASCAM	[303]	m_23_lascam_24p_3s
24	Unnamed	[349]	m_24_mopex1_5p_4s
25	TCM	[222]	m_25_tcm_6p_4s
26	FLEX-I	[106]	m_26_flexi_10p_4s
27	TANK model	[310]	m_27_tank_12p_4s
28	XINANJIANG	[351]	m_28_xinanjiang_12p_4s
29	HyMOD	[57, 330]	m_29_hymod_5p_5s
30	Unnamed	[349]	m_30_mopex2_7p_5s
31	Unnamed	[349]	m_31_mopex3_8p_5s
32	Unnamed	[349]	m_32_mopex4_10p_5s
33	SACRAMENTO	[60, 225]	m_33_sacramento_11p_5s
34	FLEX-IS	[106, 233]	m_34_flexis_12p_5s
35	Unnamed	[349]	m_35_mopex5_12p_5s
36	MODHYDROLOG	[69, 71]	m_36_modhydrolog_15p_5s
37	HBV-96	[197]	m_37_hbv_15p_5s
38	TANK model - SMA	[310]	m_38_tank2_16p_5s
39	MCRM	[222]	m_39_mcrm_16p_5s
40	SMAR	[236, 312]	m_40_smar_8p_6s
41	NAM	[232]	m_41_nam_10p_6s
42	HYCYMODEL	[115]	m_42_hycymodel_12p_6s
43	GSM-SOCONT	[281]	m_43_gsmsocont_12p_6s
44	ECHO	[282]	m_44_echo_16p_6s
45	PRMS	[192, 205]	m_45_prms_18p_7s
46	CLASSIC	[85]	m_46_classic_12p_8s
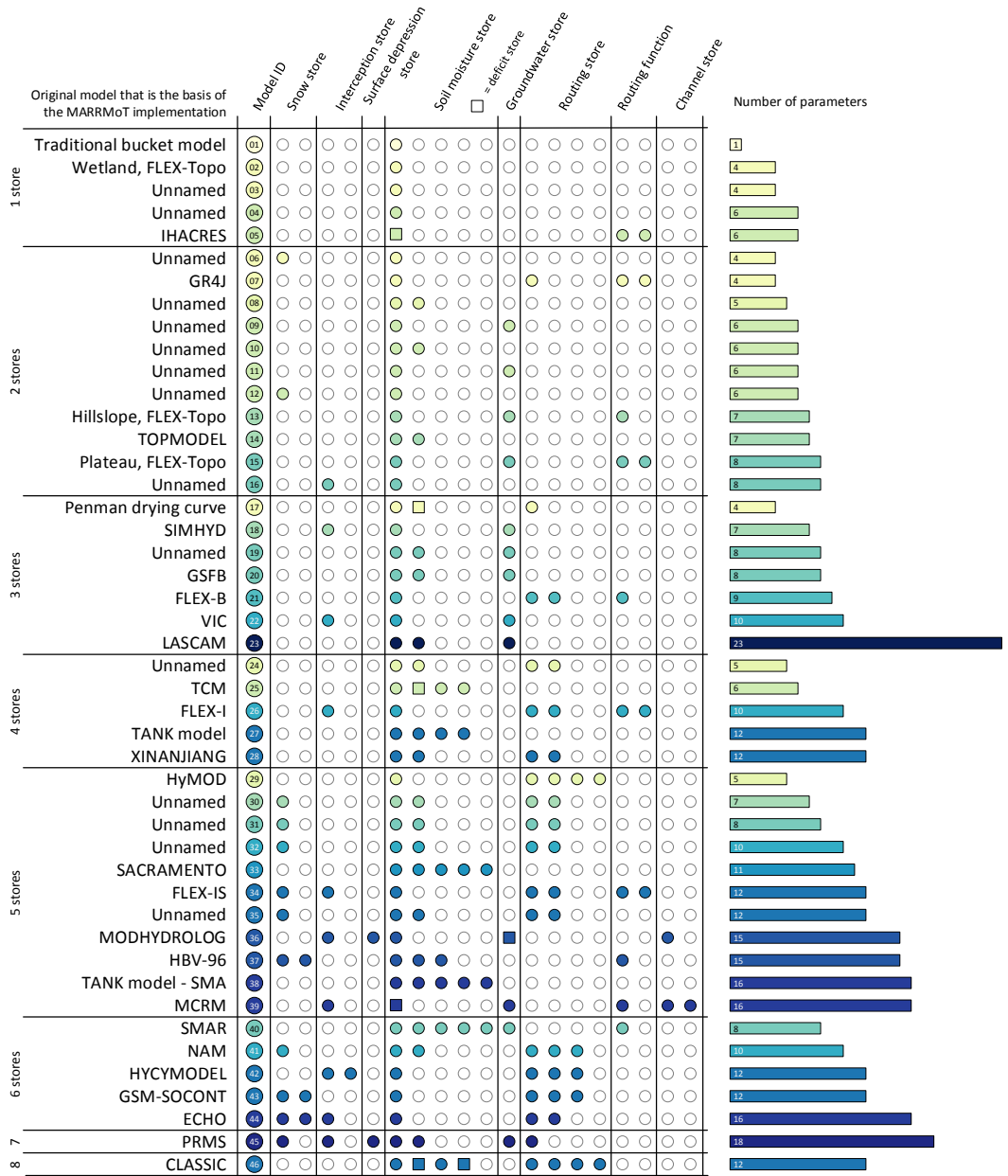
FIGURE 4.2. Overview of MARRMoT models. Models are sorted vertically by number of stores (1 at the top, 8 at the bottom). The columns show broad categories of hydrologic process that can be represented by a model. Coloured circles indicate the model has a store dedicated to the representation of this hydrological process (squares indicate a deficit store). The bar plot on the right shows each model's number of parameters. Colouring refers to the number of parameters.

4.4 46 model application test case

To demonstrate the potential of the framework, we calibrated all 46 MARRMoT models to flow observations at Hickory Creek near Brownstown, Illinois (USGS ID: 05592575). This catchment was randomly selected from the CAMELS data set [4]. The catchment is small with an area of approximately 115 km^2, located at 176 m.a.s.l. at latitude 38.9^o. It has a strong seasonal cycle with temperatures varying between -20°C in extreme winters, up to nearly 30°C in summers. Average annual rainfall is approximately 1117mm, 6.4% of which occurs as snowfall. The runoff ratio is around 29% of precipitation. The flow regime is flashy (baseflow index is 0.18) and ephemeral (no flow is observed 18% of the time), High flows (95th percentile flow is 3.7mm/d) are more common in winter and spring, while low flows (5th percentile flow is 0mm/d) are more common in summer and autumn. Soils are a mixture of silt (60%), clay (24%) and sand (16%).

PET input was estimated using climate data included in CAMELS and the Priestley-Taylor method [261]. Model calibration uses the time period 1989-1998, model evaluation uses the period 1999-2009. Initial states are found by iteratively running each model with data from the year 1989, until model states reach an equilibrium. The calibration algorithm is the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [136], using the Kling-Gupta Efficiency [129] as the objective function. CMA-ES optimizes a single parameter set per model using MARRMoT's provided parameter ranges. Note that parameter optimization and sampling are currently not part of the provided tools but connecting MARRMoT to various calibration algorithms or Monte Carlo sampling strategies is straightforward (the User Manual provides several basic workflow examples).

Figure 4.3a shows KGE values during calibration and evaluation for each model. Each result is coloured to indicate the number of calibrated parameters. The number of model parameters seems unrelated to model performance and several models with higher numbers of parameters are outperformed by the simplest 1-parameter bucket model. After analysing the components present in most successful models (not shown), we can speculate that a saturation excess mechanism is key to achieve satisfactory calibration efficiency values in this catchment, and that this catchment's flashy behaviour could be related to rainfall events on soil with low available storage.

Figure 4.3b shows values for two common hydrologic signatures, calculated for time series of simulated flow by each model (blue/yellow dots, shade showing the KGE value during calibration) and for observations (red dot). These signatures are calculated for the calibration period. There is significant scatter around the observed signature values and models with "good" calibration efficiency (darker shades) are not necessarily closer to observed signature values than models with lower calibration performance. From this
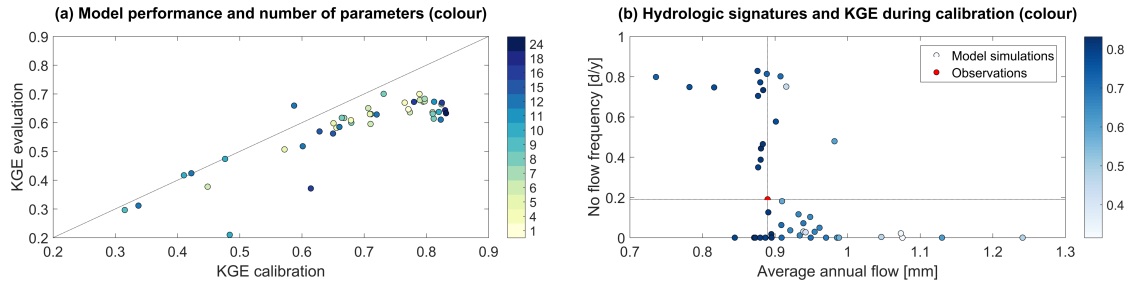
FIGURE 4.3. Example of MARRMoT application to Hickory Creek near Brownstown (USA). (a) model performance during calibration (1989-1998) and evaluation (1999-2009) periods. Each dot represents a single model and is coloured according to the model's number of calibrated parameters. (b) Comparison of simulated average flow and no-flow frequency signature values and observed values for those signatures (red dot bisected with lines).

we can conclude that even though certain model structures can achieve "high" values for a given objective function, there is no guarantee that the simulated flow series have the same statistical properties as the observed time series the models were calibrated against. Furthermore, this shows that a saturation-excess model can achieve high efficiency values, but that the full hydrologic behaviour in this catchment is likely more nuanced than a single runoff generation mechanism.

This test case highlights the power of multi-model comparison frameworks: from two simple plots we have deduced a plausible important runoff mechanism in this catchment, found that this mechanism alone cannot satisfactorily explain the catchment's hydrologic behaviour, and that a higher number of model parameters does not necessarily result in more realistic or better performing models. Further investigation of the model structures and their performance could lead us to more insights about hydrologic behaviour and inter-model differences, but that is beyond the scope of this test case.

4.5 MARRMoT considerations

4.5.1 Reliance on imperfect methods

MARRMoT uses built-in Matlab root-finding methods to solve the ODE approximations on every time step. Currently, *fzero* is the default option for models with one store and *fsolve* is the default in multi-store models. *lsqnonlin* is used as a slower but more robust alternative if the former methods are not sufficiently accurate (compared to a user-specified accuracy tolerance). In most cases, this set-up performs within acceptable bounds of accuracy.

However, for special cases (e.g. very small maximum storage values), the root-finding method might return solutions that are outside the bounds of expected model behaviour (e.g. storages values below 0, storages higher than their maximum capacity or complex numbers), even if "realistic" solutions also exist. Additional constraints must be introduced into the flux equations to prevent this behaviour, because in a large-sample study these issues are difficult to troubleshoot if they occur during the sampling of several thousands of combinations of models and catchments. This involves a fundamental change to model equations necessitated by the use of these solvers. More robust solvers such as *lsqnonlin* allow specification of bounds to the solution space but are less computationally efficient. The current trade-off favours constraints implemented into the fluxes and default use of faster root-finding methods over the more elegant, but much slower, solution provided by *lsqnonlin*. Further optimization of the root-finding methods is considered outside the scope of this version of MARRMoT.

4.5.2 Speed versus readability

Several considerations during MARRMoT design have been heavily influenced by readability and user-friendliness over computational efficiency. Implementing fluxes as anonymous functions rather than regular functions leads to reduced computational speed but increased clarity of the code.

Matlab was chosen out of similar concerns. Fortran or similar compiled languages would grant significant speed-ups but reduce user-friendliness.

4.5.3 Correspondence between MARRMoT and original publications

During MARRMoT development, we have tried to stay close to the original publications that introduced the models. Differences are unavoidable however, due to our criteria of creating a uniform framework. Most changes have to do with spatial discretisation, where we reduced the level of detail in a model to make all 46 models lumped.

For certain models (e.g. SACRAMENTO [60, 225]) model code and numerical implementation are so interwoven that far-reaching changes were required to make these models fit into this generalized framework. For all models, it is likely that the use of the default Implicit Euler scheme will provide different results to previous studies that use the (much more common) Explicit Euler scheme. Furthermore, the smoothing of model equations will also cause differences to arise with previous studies. We strongly recommend readers to compare the original publication of each model with the version given in this toolbox, to place results from the MARRMoT models in a proper context of earlier work with these models.

4.5.4 Parameter optimization and sampling

MARRMoT provides model code and recommended parameter ranges but does not include any parameter optimisation, parameter sampling or sensitivity analysis methods. This is a conscious choice because these methods continue to be developed and keeping a latest, state-of-the-art version of each packaged in the MARRMoT distribution is infeasible. We refer the reader to e.g. Arsenault et. al. [16] for a recent discussion of various optimization methods, to e.g. Beven and Binley [44] for a recent discussion of GLUE-based uncertainty analysis and to Pianosi et. al. [257] for a recent publication of an open-source sensitivity analysis toolbox. Application of any of these methods with MARRMoT models is straightforward. The User Manual provides workflow examples for parameter sampling and parameter calibration, which can be used as a starting point to integrate parameter optimization, sampling or sensitivity analysis methods.

4.5.5 Possible extensions

Lists of contemporary relevant hydrologic models are hard to come by. Such a list would always be incomplete because new models and model variants continue to be developed. As such, there is no reason to assume that the current 46 models in MARRMoT showcase all possible lumped conceptual hydrologic models. Likewise, although MARRMoT includes a wide variety of flux equations, this list should not be assumed to be complete. The MARRMoT User Manual therefore provides detailed guidance on creating new model and flux functions, and the code's location and licensing on Github allows these new models to be shared freely. Extensions to the framework are thus possible and encouraged.

Currently lacking in the code is the possibility to use adaptive time stepping. Fixed-step Implicit Euler approximations are sufficiently accurate for most applications [72, 160, 285] but adaptive time-stepping can provide additional benefits [78, 161, 285]. Our initial assessment is that it would be relatively straightforward to replace the current fixed-step time-stepping implementation with adaptive time-stepping (see e.g. Clark and Kavetski [72] for further reading on adaptive time-stepping).

4.6 Code availability and dependencies

MARRMoT is provided under the terms of the GNU General Public License version 3.0. MARRMoT code and User Manual can be downloaded from https://github.com/wknoben/MARRMoT (DOI: 10.5281/zenodo.2482542). Additional documentation can be found in the Supplementary Materials to this paper. MARRMoT has been developed on Matlab version 9.2.0.538062 (R2017a), with the Optimization Toolbox

Version 7.6 (R2017a). The Octave distribution has been tested with Octave 4.4.1 and requires the "optim" package. See the User Manual for some detail regarding running MARRMoT in Octave.

4.7 Conclusions

This paper introduces the Modular Assessment of Rainfall-Runoff Models Toolbox (MAR-RMoT). This modelling framework is based on a review of conceptual hydrologic models. Across these models, over 100 different flux equations and 7 different Unit Hydrographs (UHs) are used. These are implemented as separate functions and each model draws from this library to select the fluxes and UHs it needs. This results in standardized implementations of 46 unique, lumped model structures. The framework is implemented in Matlab, can be used in Octave, and is provided as open source software (https://github.com/wknoben/MARRMoT; DOI: 10.5281/zenodo.2482542). Requirements for running a model are simple: (i) time series of precipitation, potential evapotranspiration and optionally temperature, (ii) initial storage values, (iii) settings that specify the numerical integration method (currently provided are Implicit Euler (recommended) and Explicit Euler) and Matlab solver behaviour, and (iv) values for the model parameters (these can be sampled or optimized from parameter ranges provided as part of MARRMoT). MARRMoT comes with documentation that describes (i) each model and its equations, (ii) the conversion from model equations to computer code, (iii) the implementation of 7 different types of Unit Hydrographs, and (iv) the references used to inform standardized parameter ranges,. The User Manual provides guidance on navigating the Matlab functions in which each model is implemented, several examples of how the framework can be used (with workflow scripts that show the Matlab code required for these analyses), information on how to create new models or flux functions, and several small modifications that can speed up the model code by disabling certain output messages from Matlab's built-in solvers. The main purpose of MARRMoT is to enable multi-model comparison studies and objective testing of model hypotheses. Additional benefits can be gained from the framework's documentation, which provides an easy-to-navigate comparison of 46 unique conceptual hydrologic models. MARRMoT is provided to the community in the hopes that it will be useful and to encourage a growing trend of open and reproducible science.

LARGE-SAMPLE MODEL COMPARISON

This chapter is based on a draft manuscript that will be submitted as a research article to Water Resources Research.

Citation: Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A. & Woods, R. A. (in prep). Exploring conceptual model structure uncertainty through a large-sample approach: Comparative analysis of 36 models across 559 catchments. To be submitted to Water Resources Research.

Abstract

Choosing an appropriate model is a critical aspect of any modelling study. Previous studies have shown that the choice of model structure, i.e. a model's selection of states and fluxes and the equations used to describe them, has an important impact on model performance. This work investigates differences in performance of 36 lumped conceptual model structures, used for daily streamflow simulation in 559 catchments across the United States. Model performance is quantified using three different objective functions based on the Kling-Gupta Efficiency that focus on high flows, low flows and a combination of high and low flows respectively. Each model is calibrated for every combination of catchment and objective function with a Covariance Matrix Adaptation Evolution Strategy (CMA-ES), using a 10-year calibration period. Model performance is evaluated using a separate 10-year period. We find that adequate model performance can be achieved in most catchments for each objective function, but that no single model is capable of this. Instead, nearly every model is both one of the best choices in certain catchments and one of the worst in

other catchments. In most catchments, several models achieve performance very similar to that of the best model for that catchment, resulting in high levels of model structure uncertainty. This is especially apparent when an objective function is used that focusses on high flows, in which case models obtain very similar efficiency values. With objective functions that focus on low and combined flows, larger differences between model structures are seen and the better model structures can be more clearly identified. We find no relation between model performance during calibration or evaluation periods and number of model parameters, nor is there evidence of increased risk of overfitting for models with more parameters. However, we do find that certain structures seem better suited to certain objective functions and we formulate several hypotheses to investigate why this might be the case. We attempt to relate model performance to catchment attributes such as climate, geology, topography, soil type and streamflow signatures and report that the clearest link is between model performance and streamflow signatures. This indicates that certain models are inherently better suited for certain flow regimes, but the relation between flow regimes and catchment attributes, and the relationship between catchment attributes and model performance, remains elusive. Due to the intensive computational demands inherent to large-sample studies such as this, calibration results of all models (parameters, simulated model storages and fluxes, and obtained efficiency values, as well as source code) will be made available for further investigation once calibration of the final 10 models completes.

5.1 Introduction

There is an ongoing debate in hydrology whether a "one model fits all" approach is realistic, based on the assumption that the fundamental hydrological processes are the same everywhere [e.g 103, 198, 255, 273]. The assumption that this is true has led to development of various hydrological models which are now applied across a wide range of catchments (see e.g. discussion of the GR4J model in [103] and consider more recent applications of this model in 142 catchments in the US [245]). This assumption is contrasted at the coarser catchment scale by the concept of "uniqueness of place" [34]; the idea that in a practical sense every catchment is unique because there are limits to our understanding of fundamental processes and the availability of sufficiently detailed measurements. As a result of this uniqueness, many hydrological models have been developed that all aim to represent the dominant processes in a given catchment [e.g. 298] but, importantly, these models tend to include *different hydrologic processes* or use *different equations* for the same processes. In theory, we should be able to use a single model based on fundamental hydrologic principles, but in practice there are many different models available that all represent a certain view of what constitutes these fundamental hydrologic processes (often

informed by the hydrologic conditions in the catchment(s) where the model was originally developed). Choosing an appropriate model out of all possible options is critical if the model is to be used for prediction, especially if the prediction conditions (e.g. expected climatic forcing) are different from the conditions used to develop the model and/or calibrate any model parameters [173].

Within the sub-set known as conceptual models (i.e. bucket-type models) that is the focus of this thesis, some models employ a low-parameter, process-aggregated approach (e.g. GR4J: 4 parameters, fluxes represent aggregated results of all possible processes [255]). Others follow a process-explicit approach (e.g. MODHYDROLOG: 15 parameters, fluxes represent e.g. interception, infiltration, surface storage, groundwater-channel exchanges [69, 71]). Many models are somewhere in between (e.g. VIC: 10 parameters, fluxes represent interception but also an aggregated runoff contributing area [194]). The choice of model structure to use is one of the main sources of uncertainty in a modelling study [e.g. 11, 80, 104, 106, 185, 325] but between-model differences are currently not well understood [65, 128].

Large-sample modelling studies can increase our understanding of model functioning and the need for them has been discussed often [e.g. 11, 82, 111, 144, 190, 198, 254, 293, 299]. The consensus seems to be that assessing a single model's performance under a wide range of different conditions can either lead to increased understanding through model failure or improve our confidence in the model's abilities. Comparing the performance of different models for a given catchment can lead to increased understanding of hydrologic similarity and between-model differences. Studies that follow these approaches usually compromise on either the number of models or the number of catchments. For example, Franchini and Pacciani [111] uses 1 catchment and 7 models; Euser et al. [98] works with 2 catchments and 11 models; Seiller et al. [293] uses 2 catchments and 20 models; Nijzink et al. [233] uses 3 catchments and 4 models; Lidén and Harlin [195] uses 4 catchments and 1 model; De Boer-Euser et al. [88] works with 5 catchments and 8 models; Krueger et al. [185] uses 6 catchments and 72 models; and Bell et al. [28] and Moore and Bell [222] work with 9 catchments and 10 models. This is a logical result of large-sample studies that need to balance "depth with breadth" [130], in order to keep analysis and visualization of results manageable. There are studies that use both a larger model and larger catchment sample (e.g. Perrin et al. [254] uses 19 models in 429 catchments and Van Esse et al. [325] uses 13 models in 237 catchments). However, especially in these larger samples, investigating every interesting individual case is infeasible and instead lessons must be learned from emergent patterns across the full sample [144].

Large-sample emergent patterns can provide unique insights in how well models function across a variety of different catchment types and provide insight in hydrologic

similarity between different places. Ideally, these findings are generalisable, because this allows the modelling of ungauged catchments through regionalisation approaches. This requires (i) an appropriate model choice for the gauged catchment, and (ii) a thorough understanding of what makes catchments hydrologically similar and thus how appropriate it is to transpose a model from a gauged to ungauged catchment. However, as mentioned earlier, finding an appropriate model for a given catchment is not straightforward, nor is accurately defining hydrologic similarity. Hydrology currently lacks a complete catchment classification scheme [209, 334] but certain aspects of hydrologic similarity are well studied. It is well known that climate plays a dominant role in determining streamflow statistics across large scales [29, 58, 179, 219] and climatic conditions have been linked to conceptual model performance as well [e.g. 86, 195, 218, 325, 326]. Less is known about the influence of catchment characteristics such as geology, vegetation and topography on streamflow regimes and by extension on model performance. Large-sample results [3, 241] suggest that catchment attributes are of lesser importance than climatic conditions in determining certain hydrological signatures, but small-scale results suggest that knowledge of local catchment attributes is critical for model development and improvement [105, 190], and that hydrologic behaviour might vary in response to climatic and landscape gradients [62]. Recent development of large-scale hydro-climatic data sets [4, 8] makes it possible to investigate the relation between model performance and catchment characteristics on a large scale.

Given the current incomplete knowledge on between-model differences and the availability of large-sample data sets, an opportunity exists to study similarities and differences in the behaviour of multiple different models across a wide variety of places, and investigate whether emergent patterns in the relationship between catchment structure and model structure exist. The aim of this chapter is to explore model performance and model structure uncertainty in a wide range of different catchments and relate these findings to catchment attributes such as climate, topography and streamflow signatures. We simulate streamflow in 559 non-human impacted catchments across the United States, using 36 unique model structures from the multi-model framework developed in Chapter 4. Models are used in a lumped (catchment-averaged) fashion. On a per-catchment basis, we quantify model structure uncertainty in terms of different KGE metrics. On a per-model basis, we investigate whether certain models are inherently better suited for certain performance metrics (e.g. with a focus on high flow simulation) or certain regions (e.g. in terms of climate or catchment attributes). The results presented here contribute to answering several open research questions posed by Clark et al. [78]:

- What are the differences between model structures when using different evaluation

criteria?

- Is the trade-off between objective functions the same for different model structures?

- Are some models inherently better than others at reproducing certain parts of the hydrograph?

- Is it possible to identify model structures that are best suited for different climates or catchment types?

- Are relations between model performance and model structure consistent across basins?

5.2 Data and models

5.2.1 CAMELS catchment data

This study uses the CAMELS data set [4], which provides time series of meteorological variables and streamflow [229], and tables with catchment attributes for 671 catchments in the contiguous United States. Following recommendations for practical application [4, 229], we perform several basic data checks and remove those catchments with large (>10%) discrepancies between catchment area as used for averaging of the meteorological time series and area as published by the USGS [323] or the higher resolution GAGES II data set (provided as part of the CAMELS data set). We use preliminary screening [e.g. 206] to remove those catchments that fall outside the energy-limit and water-limit on the Budyko curve. This leaves 559 catchments for use in this study, distributed across the contiguous US (Figure 5.1). We use the years 1989-2009 in this study, because these are sufficiently long to calibrate and evaluate the models, without being impacted by too many missing streamflow observations [4, 229]. Streamflow records are complete during this period for 546 catchments. For the remaining 13 catchments, days with missing streamflow values are ignored during the calculation of objective function values. Missing values account for 0.013%, 0.25%, 0.41%, 1.1%, 1.2%, 1.2%, 1.2%, 2.9%, 3.3%, 3.4%, 3.6%, 4.1% and 4.8% of all observations in these catchments.

The provided meteorological time series [229] contain daily values of precipitation [mm/d], snow water equivalent, [kg/m^2], minimum and maximum air temperature [^oC], duration of daylight period [sec/day], incident shortwave radiation [W/m^2] and daily average water vapour pressure [Pa]. Three different forcing products are provided, and for each product the time series are available as catchment-averaged values, separated by elevation band, and per hydrological response unit. This study investigates lumped models

89
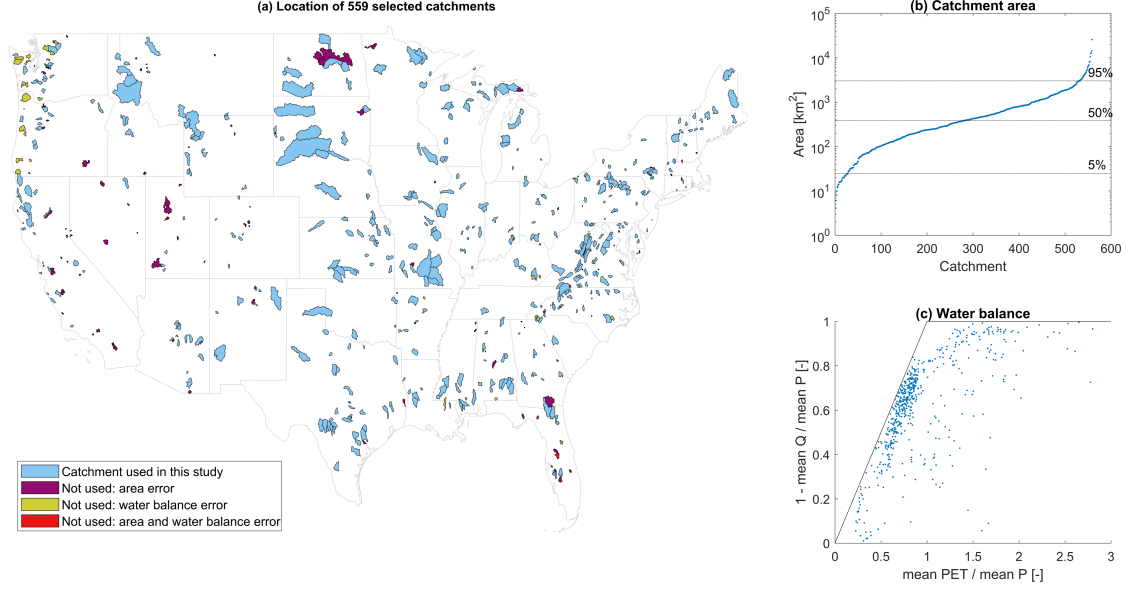
Figure 5.1: Selected catchments from CAMELS data set, after removing those catchments with uncertain area estimates (>10% differences between two geo-spatial data sets and the USGS reported value) and water balance errors. (a) Geographical location and reason for exclusion from this study. (b) Catchment area distribution of the 559 selected catchments. (c) Aridity index against 1 - runoff ratio for the 559 selected catchments.

(catchments are treated as a single entity) and thus uses catchment-averaged forcing data. We follow Newman et al. [229] and Addor et al. [4] in using the Daymet product, which is based on the highest spatial resolution of all three products (1km x 1km compared to 12km x 12km for Maurer and NLDAS products) and is more likely to provide accurate data for smaller catchments and locations with complex topography. Time series of daily potential evapotranspiration (PET) are estimated using the Priestley-Taylor method [261]:

$$PET = \frac{\alpha_{PT}}{\lambda} \frac{s(R_n - G)}{s + \gamma} \tag{5.1}$$

Where α_{PT} is the Priestley-Taylor coefficient $[-]$, λ is the latent heat of vaporization of water $[MJ/kg]$, s is the slope of the saturation vapor pressure-temperature relationship $[kPa/^oC]$, R_n is the net radiation $[MJ/m^2/d]$, G is the soil heat flux $[MJ/m^2/d]$, and γ is the psychometric constant $[kPa/^oC]$. We assume G to be zero and assume a reference crop albedo of 0.23 (short grass) in our calculations. λ, s, R_n and γ are calculated using the Daymet variables, the day of year, catchment-average latitude and elevation, and various equations for the constituent radiation terms [7, 212]. α_{PT} is set to 1.26, which is an average value for evaporation under saturated (energy-limited) conditions and thus appropriate for estimating potential evapotranspiration [261]. Negative estimates of PET

occasionally occur when outgoing longwave radiation exceeds net incoming shortwave radiation, and R_n thus reaches negative values. These negative values are corrected to zero.

Data for each catchment is divided into two 10-year periods covering 01-Jan 1989 to 31-Dec 1999 and 01-Jan 2000 to 31-Dec 2009 respectively. Average climate characteristics are approximately constant during these periods with the exception of regions with high mean precipitation ($\bar{P} \geq 5mm/d$, precipitation has decreased somewhat) and regions with low mean temperatures ($\bar{T} \leq 5^oC$, temperatures have increased). Estimated potential evapotranspiration rates are approximately constant between the two periods (Figure C.1).

5.2.2   MARRMoT modelling framework

This study uses the Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT) v1.0 [178]. MARRMoT contains model code for 46 unique conceptual model structures within a uniform framework. This has the main advantage that the implementation of models and fluxes is consistent and any differences in simulation are thus solely due to differences in model structure (i.e. the choice of states and processes to include in the model, and the choice of equations to represent them). The MARRMoT models are all based on published literature and cover a wide range of possible structures, from the traditional 1-parameter overflow bucket model to structures with up to 8 stores or 23 parameters. The toolbox is provided with literature-based parameter ranges for each model. These aim to standardize the parameter ranges as much as possible, so that models have the same amount of parameter-freedom (e.g. in the case of interception capacity, all models that simulate the interception process use a range of 0-5 mm). These ranges are optional but no changes to them were deemed necessary. The differential equations that express each model's changes in storage(s) with time are numerically approximated with a fixed-step Implicit Euler method, which uses the same step size as the forcing data. This provides better accuracy and stability compared to the more common Explicit Euler method [167, 285].

Due to time constraints and the intensive computational demands of the Implicit Euler approximation method coupled with the large sample of catchments, not all MARRMoT models could be used in this work. At the time of writing, calibration has completed for 36 of the models. These results will be made publicly available for further analysis and will be updated as the remaining models finish calibration. Figure 5.2 provides an overview of the 36 models for which results are currently available and used in this work.
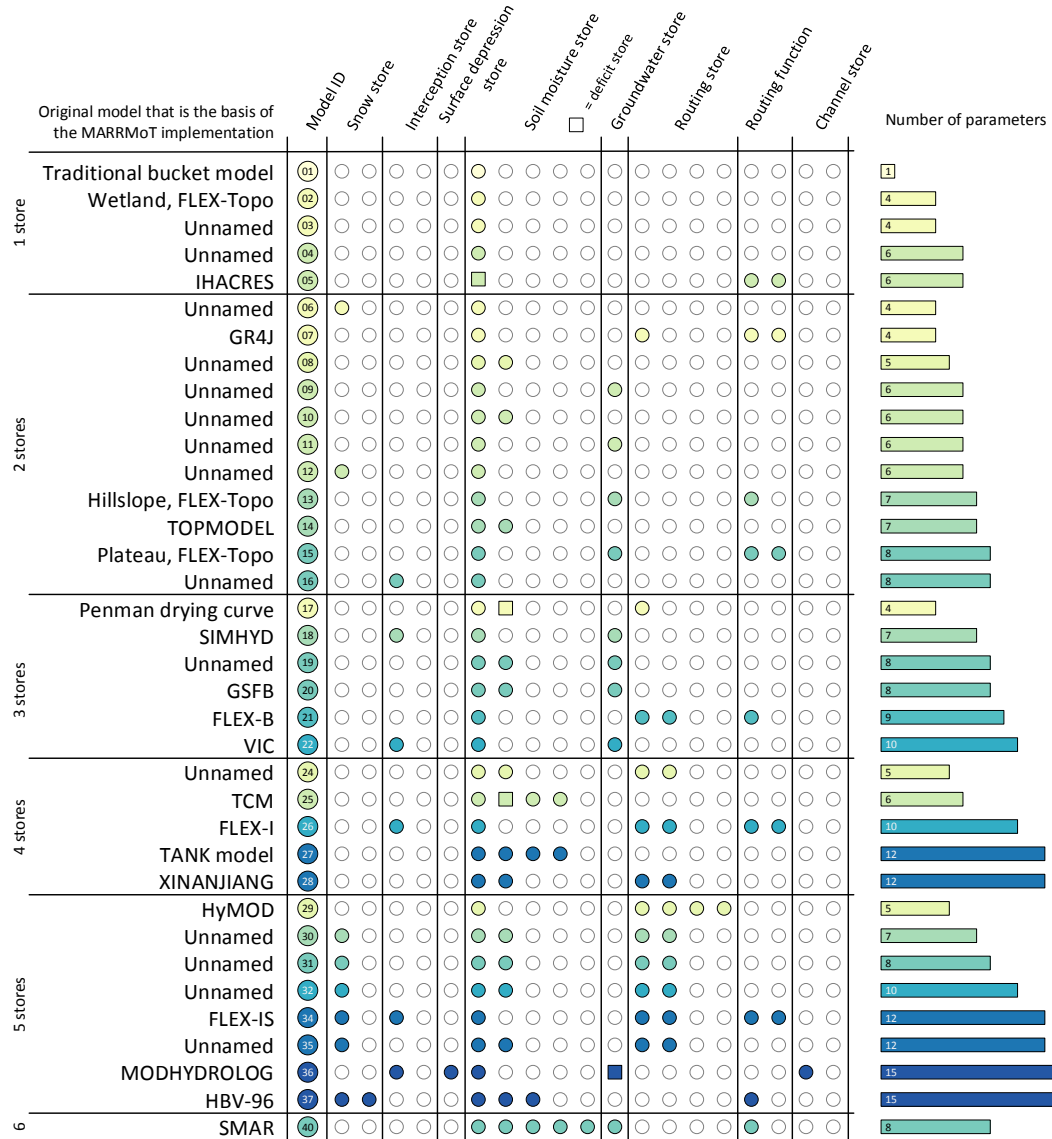
Figure 5.2: Summary of the 36 models used in this work (adapted from Knoben et al. [178], Figure 2). Models are sorted by the number of stores first (indicated in the left column) and by their number of parameters second (bars in right column). Model ID refers to a constant identifier that is used in the MARRMoT documentation and subsequent analysis in this work. The middle part of the figure gives an overview of processes each model's store(s) is intended to represent. Note that MARRMoT model implementations can deviate from the source models they are based on. See the MARRMoT documentation for details.

5.3 Method

5.3.1 Model set-up

This study uses the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [136] to optimize model parameters, using data from the period 1989-1998. Model performance is evaluated using data from 1999-2009. CMA-ES is a single-objective optimizer that compares favourably to various other methods for finding the global optimum of difficult functions and in rugged objective function landscapes [135, 136]. The algorithm has seen successful application in hydrology [e.g. 16, 110], as well as many other fields [134]. The algorithm is allowed to run until either the change in objective function or the change in parameter values drops below a threshold of 1E-3 (absolute change in either value), which we consider an acceptable compromise between accuracy and speed. We use three objective functions based on the Kling-Gupta Efficiency (KGE) [129]:

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2} \tag{5.2}$$

Where subscripts *obs* and *sim* refer to observed and simulated time series of flow respectively, r is the linear correlation coefficient between observed and simulated flow, σ denotes the standard deviation of flows and μ the mean of flows. We aim to compare model performance for a variety of flow conditions, therefore our choice of objective functions emphasizes higher flows, lower flows and a combination of both. The objective functions used are (O_1) the KGE calculated on time series of flow, KGE(Q), (O_2) the KGE of inverse flows, KGE(1/Q) (shown to be more appropriate than a log transform to emphasise low flows, [e.g. 262, 272] and (O_{12}) the mean of KGE(Q) and KGE(1/Q). For inverse flows, we avoid problems with zero flow values by adding a constant e to observed and simulated flows before the transformation is performed [262]. Following Pushpalatha et al. [262], e is set at 1% of the mean observed flow.

Model warm-up periods are used to reduce the impact of (uncertain) initial conditions on model performance. Recent studies have attempted to provide guidelines for warm-up period length in conceptual models [172, 263] but these studies are limited in number of models (1 and 2, respectively) and catchments (18 and 1 respectively) and therefore difficult to generalise to a large-sample study such as this. Instead of using a fixed number of warm-up days, we determine the initial storages in an iterative procedure by letting the model repeat year 1 of the data period until the stores reach an equilibrium for the 1st day of the year (<1% change in storage value(s) between runs). Storage values might not converge for certain parameter sets (e.g. when a store of unlimited depth has very low outflow), in which case the procedure is stopped after 50 iterations.

5.3.2 Analysis of results

Given the size of this sampling study, we must necessarily limit our analysis to a few key questions. The analysis is divided into three distinct parts. First, we investigate model structure uncertainty on a per-catchment basis. Next, we investigate inter-model differences using aggregated efficiency values. Finally, we synthesize our findings about catchments and models, and attempt to relate inter-model differences to structural uncertainty in the catchments. Each aspect is detailed in the following sections.

5.3.2.1 Quantification of model structure uncertainty

In the first part, we attempt to quantify conceptual model structure uncertainty on a per-catchment basis. There are no universally agreed metrics to quantify model structure uncertainty but using differences in objective function values is a common approach [e.g. 106, 141, 344]. Therefore, we define high model structure uncertainty as cases where several models achieve similar efficiency values, which makes it difficult to choose the most appropriate model and thus makes the structural uncertainty high. Cases where different models achieve very different efficiency values are examples of low structural uncertainty, because the choice of the most appropriate model is straightforward. To quantify model structure uncertainty, we report the maximum evaluation efficiency per catchment and the difference in efficiency values between the best model and the 10th percentile model (i.e. with 36 models we compare performance between the 1st and 32nd model). This gives some idea of how well the data in a catchment can be modelled by a variety of models and how large the variation of efficiency values in the model sample is, without being too sensitive to very poorly performing models. We also calculate Cumulative Distribution Functions (CDFs) of the number of models that fall within 0.01, 0.05, 0.10 and 0.25 KGE value [-] from the best model for each catchment. For comparison, we also give the number of models per catchment that outperform the standard (albeit low) benchmark that is the mean flow (i.e. $KGE = 1 - \sqrt{2}$, equivalent to Nash-Sutcliffe Efficiency (NSE) = 0). These metrics give better insight into the number of models that have similar performance for each catchment, compared to the first approach which shows variation in the model sample as a whole.

5.3.2.2 Model inter-comparison

In the second part, we investigate differences between individual models across the 559 catchments and 3 objective functions. Earlier studies have found that models with a higher number of parameters are at risk of being overparametrised, a scenario in which the

model has more degrees of freedom (i.e. parameters) than can be robustly estimated from the data [254]. This can lead to parameter equifinality and overcalibration. Parameter equifinality refers to cases where different parameter sets lead to the same objective function value during calibration. However, there is no guarantee that these parameter sets lead to similar objective functions values during evaluation. Overcalibration (also called overfitting) refers to cases where the model is calibrated to noise in the calibration data [40, 284, 295]. Both parameter equifinality during calibration and model overfitting can lead to a performance reduction during model evaluation, due to differences in parameter sensitivity and differences in data noise respectively (note that model performance can also improve during evaluation, but we focus mostly on the decline of model performance). Additionally, model performance might change as a result of changes in climatic forcing and/or catchment conditions between calibration and evaluation periods. Models with highly identifiable parameters are particularly at risk if the distribution of 'good' parameter values is slightly different between calibration and evaluation periods (i.e. a narrow range of parameter values leads to high efficiency values, but this narrow range is different between both periods). We investigate whether certain models are more consistent in terms of efficiency values between calibration and evaluation, and how the number of model parameters relates to these results. Additionally, we perform exploratory analysis aimed at finding out whether certain models are inherently better than others at reproducing different parts of the hydrograph (expressed through the three objective functions) and whether the trade-off between objective functions is the same for different model structures.

To investigate (1) whether models with more parameters are better at fitting calibration data than models with fewer parameters, (2) whether the number of parameters affects the stability of model performance and (3) whether certain models perform better for certain objective functions, we compare the distributions of model performance values of all models. We do this visually on a per-model basis, with histograms of KGE values during calibration, evaluation and the change between both periods. We then investigate whether this relates to the number of parameters in each model by quantifying the presence and statistical probability of trends in the median value of these distributions with a Mann-Kendall test [120, 169, 204]. Because both histograms and the Mann-Kendall test use aggregated metrics (respectively they use distributions and probabilities associated with median model performance), we lose information on model performance per catchment. Therefore we perform this assessment on model performance ranks as well, which allows us to investigate whether certain models are inherently better than others, and if these findings vary for different objective functions.

5.3.2.3 Relation between model structure performance and catchment attributes

The CAMELS dataset provides catchment attributes spread across six main categories: climate, geology, topography, soil, land cover and streamflow [4]. We attempt to use these descriptors to clarify the relation between model performance and catchment type. However, efficiency values cannot directly be compared between places [see e.g. 279, although this paper discusses Nash-Sutcliffe Efficiency, their points apply to KGE values as well]. We are thus limited to investigating the relative performance of each model compared to other models in our study by using model ranks instead of efficiency values. Furthermore, hydrology currently doesn't have a way to formally classify catchments [209, 334]. We therefore don't know which catchments are supposed to be similar and where we thus might expect similar performance for any given model. We thus compare model ranks directly to the catchment attributes given in the CAMELS data for want of a better alternative. We visually investigate the relationship between model ranking and catchment attributes on a per-model basis with scatter plots. We summarize and quantify these findings through Spearman rank correlations.

5.4 Results

5.4.1 Quantification of model structure uncertainty

Figure 5.3 shows the maximum achieved evaluation efficiency for each objective function in each catchment and the previously described measure of model structure uncertainty. Both the maximum efficiency values and structure uncertainty are subject to strong spatial organization (note the differences between both coasts and the central US), although exceptions to the pattern exist. In the majority of catchments, at least one model out of 36 is able to achieve reasonable objective function values (Figure 5.3a, 5.3c, 5.3e; note that this is not necessarily the same model everywhere) and both spatial patterns and KGE distributions are roughly similar for the three objective functions. Note that the definition of "reasonable" objective function values relies on a subjective assessment of adequacy associated with certain efficiency values (see e.g. [155] for a ranking of efficiency values from "unsatisfactory" ($NSE < 0.3$) to "very good" ($NSE \geq 0.7$) but the metric itself does not contain any inherent justification for these terms. We chose the term "reasonable" because in nearly every catchment at least one model improves on the traditional benchmark obtained from the mean flow, i.e. $KGE = 1 - \sqrt{2} \approx -0.41$ (functionally equivalent to NSE = 0). However, maximum evaluation efficiency varies considerably across the 559 catchments. For the high flow objective function, maximum efficiency (i.e. what the best model out of 36

achieves) ranges from -0.11 to 0.93. This range is -0.74 to 0.96 for the low flow objective function and -0.15 to 0.91 for the combined flow objective function. In geographical terms, maximum model performance tends to be low in the central US (plains areas east of the Rocky Mountains) and certain parts of the south west (Figure 5.3a, 5.3c, 5.3e). These areas share a tendency to be very arid (see Figure 3c in Addor et al. [4]).
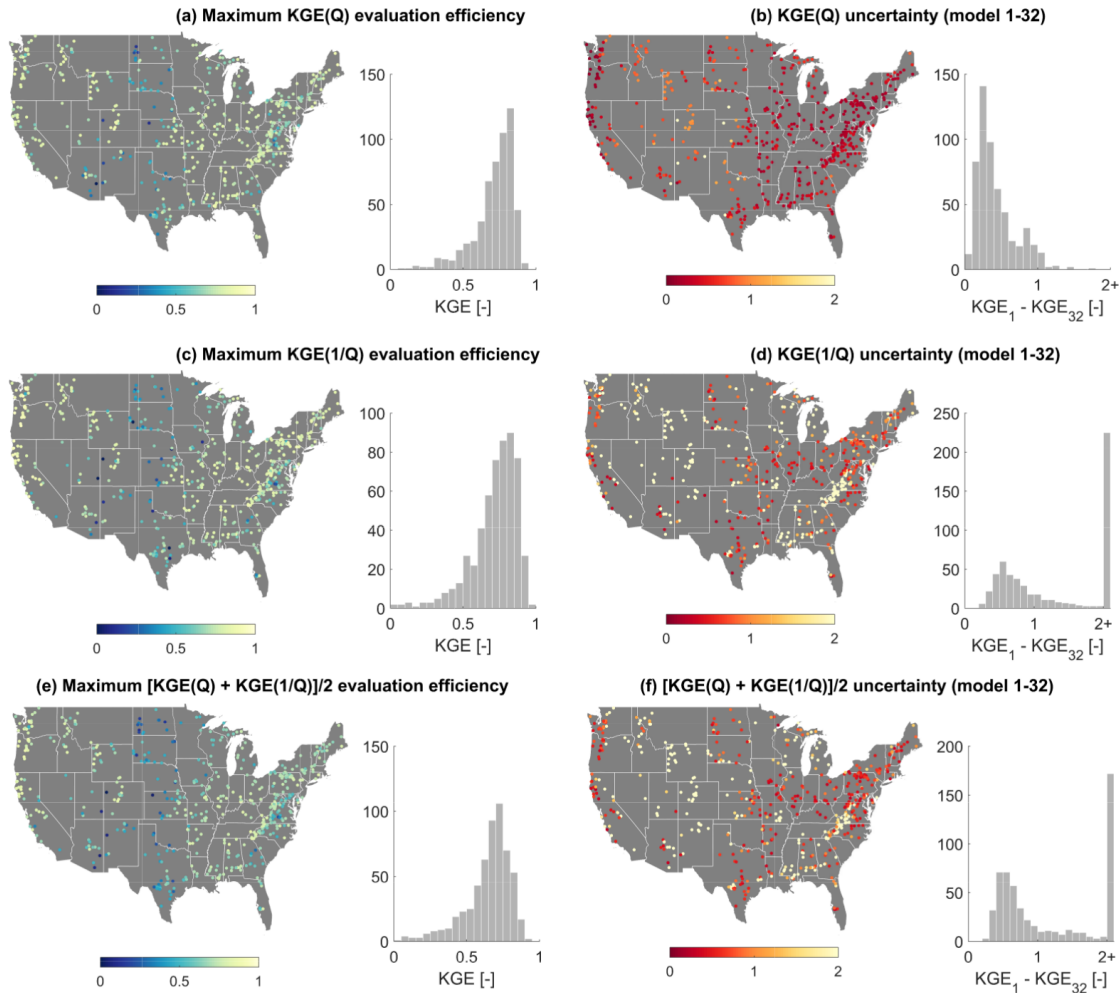
Figure 5.3: Overview of maximum model efficiency for three objective functions (left column) and a measure of model structure uncertainty defined as the difference in efficiency between the best (model 1) and the 10th percentile model (model 32). Note that axes are capped at [0,1] (figures a, c, e) and [0,2] (figures b, d, f) for clarity; efficiency values for 1/3/1 catchments are cut off in figures a/c/e, and lumped into the 2+ category for 7/225/172 catchments in figures b/d/f.

Large differences exist between the three objective functions in the extent of model structure uncertainty (Figure 5.3b, 5.3d, 5.3f). With the high flow objective function, model

structure uncertainty is relatively high in all catchments (i.e. models perform similarly, making it hard to distinguish between them). The difference between models that work well and those that work less well is much larger for the low flow and combined objective functions, suggesting that more models struggle with accurate representation of recessions and/or low flow periods. Interestingly, there is no obvious relation between maximum model performance and magnitude of model structure uncertainty. This indicates that in some regions all models perform equally well or poorly, irrespective of how high (in KGE terms) this performance is. Model structure uncertainty tends to be lower in the Rocky Mountains (Figure 5.3b, 5.3d, 5.3f, it is easy to distinguish the better and worse performing models here) and this can be explained by many models not having a snow routine. In case of the low (Figure 5.3d) and combined flow objectives (Figure 5.3f) differences between models are large outside the Rocky Mountains as well. There are no obvious physical characteristics shared between these catchments, but the streamflow regimes in these regions tend to be baseflow-dominated and non-flashy (analysis not shown for brevity). Many models struggle to achieve high KGE values for these regimes but Figures 5.3c and 5.3e show that this is not impossible, only that an appropriate model structure must be used.

Figure 5.4 shows summary plots of the number of models that fall within certain performance thresholds. Curves that stay closer to the bottom indicate that fewer models are 0.01/0.05/0.10/0.25 or less KGE value away from the best model in a given catchment. For example, in Figure 5.4a the orange line indicates that in approximately 350 catchments, no model is less than 0.01 KGE value away from the best model, while in the remaining 200 catchments at least 1 and up to 8 models have performance that is less than 0.01 KGE away from the best model for each catchment. These plots show that model structure equifinality can be a real issue (confirming findings by e.g. Perrin et al [254]): in approximately 200/150/100 catchments (high/low/combined flow objective) there are at least two and up to eight/seven/four model structures that have virtually identical performance during model evaluation (difference <0.01 KGE from the best model). Similarly, it is also clear that the KGE score obtained from mean runoff is not a particularly taxing benchmark, because the majority of model structures manage to beat this benchmark in nearly every catchment (interestingly though, there are several catchments where the mean flow is very difficult to beat, showing that the severity of this benchmark is catchment-dependent; this will be touched upon in Chapter 6). Both findings combined indicate that achieving efficiency values above the traditional mean runoff benchmark is by no means a guarantee of having used a catchment-appropriate model structure, nor that achieving a "high" efficiency value during evaluation (for a given definition of "high") means that the only appropriate model has been found. Without additional model falsification tests it is impossible to identify a single clearly most appropriate model in the vast majority of catchments studied here.
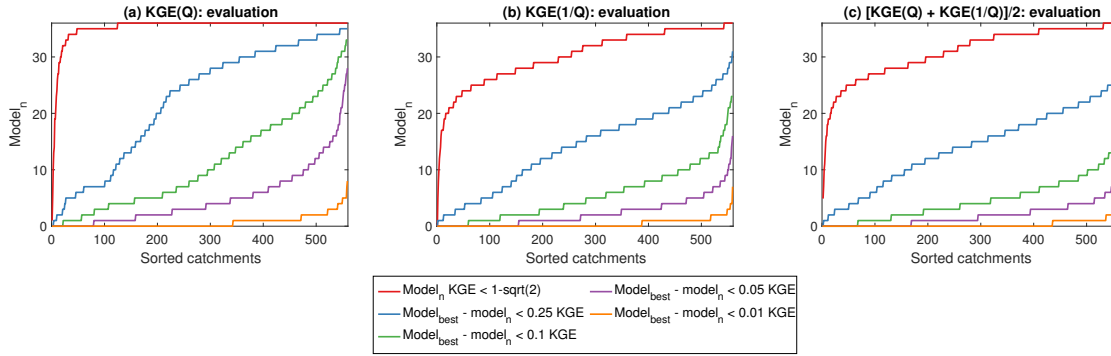
Figure 5.4: Summary plots showing the number of models that fall within certain performance thresholds. The red line shows the number of models in each catchment that outperform the traditional benchmark that is the mean runoff ($KGE = 1 - \sqrt{2} = -0.41$, analogous to NSE = 0). The other lines show the number of model structures that are within 0.01, 0.05, 0.10 and 0.25 KGE value of the best performing model for each catchment. Note that catchments are sorted independently for each CDF and that lines thus should not be compared on a per-catchment basis.

5.4.2 Model inter-comparison

The second research question, which focuses on differences between model structures, is answered in two parts: the first investigates the relationship between model performance and number of parameters, the second investigates the relative suitability of models for different objective functions.

5.4.2.1 Number of parameters does not adequately explain differences between models

Figure 5.5 summarizes the performance of all 36 models during high flow simulation (KGE(Q) objective function), sorted by the number of calibrated model parameters. Contrary to expectations, a higher number of model parameters does not necessarily lead to higher efficiency values during calibration (Figure 5.5a, darker shades indicate more parameters; see Figure 5.2 for more details about each model). In fact, several of the models with lower distributions of calibration efficiency ranges (e.g. m21, m26) have a relatively high number of free parameters (9 and 10 respectively). Both simpler (e.g. m07, 4 parameters) and more complex models (e.g. m35, 15 parameters) show higher distributions of efficiency values during calibration. Figure 5.6a shows the median model performance during calibration as well as aggregated median model performance per number of free parameters. A Mann-Kendall test shows that there is a tendency for models with a higher number of parameters to perform better during calibration (relatively large and positive

Z-value). However, this trend is not particularly significant ($p = 0.11$). Additionally, this trend is in a large part the result of the poor performance of the 1-parameter model. Removing this model from the analysis gives $Z = 0.94$ with $p = 0.35$.

Model performance during evaluation of the high flow metric (Figure 5.5b) shows a similar pattern: there are certainly differences between the ranges of efficiency values obtained by different models, but this seems unrelated to the number of parameters each model has. A Mann-Kendall test shows that models with a higher number of parameters tend to have better median evaluation values at a reasonably significant p-value ($Z = 1.97$, $p = 0.05$, Figure 5.6b). However, here too the poor performance of the 1-parameter model strongly affects this trend ($Z = 1.36$, $p = 0.18$ when m01 is removed from the analysis). Overall, evaluation efficiency ranges are somewhat lower than calibration ranges, which indicates either a change in catchment conditions that the models insufficiently account for (e.g. change in climatic forcing), or a degree of overcalibration (i.e. the models are calibrated to a certain amount of data noise). Analysis of each model's performance change from calibration to evaluation period (Figure 5.5c) shows that, whichever is the cause, all models suffer from this problem approximately equally. Figure 5.5c also shows that performance decline during evaluation is not a given, and in approximately a quarter of all catchments model performance instead increases during evaluation (note that these are not necessarily the same catchments for each model). The number of parameters does not seem to relate to the performance change any model experiences (Mann-Kendall results: $Z = -1.07$, $p = 0.28$, Figure 5.6c) and the number of parameters might thus not be a useful way to summarize model structures.

These conclusions are the same for the low flow ($KGE(1/Q)$) and combined flow ($\frac{1}{2}[KGE(Q) + KGE(1/Q)]$) objective functions (see Supporting Information, Figure C.2-C.5). However, models m21 and m26 that were highlighted as having low efficiency values during calibration with the high flow metric, achieve much better efficiency ranges when the models are calibrated using the low flow metric (Figure C.2a). This suggests that (certain aspects of) model structures directly affect how well or poorly the model performs for certain objective functions. Additionally, the histograms showing low flow performance (Figure C.2) are more flat and spread out than those showing high flow performance (Figure 5.5). This shows that model performance is not consistent between different objective functions. Adequate model performance for a given objective function is no guarantee of adequate performance for a different objective.
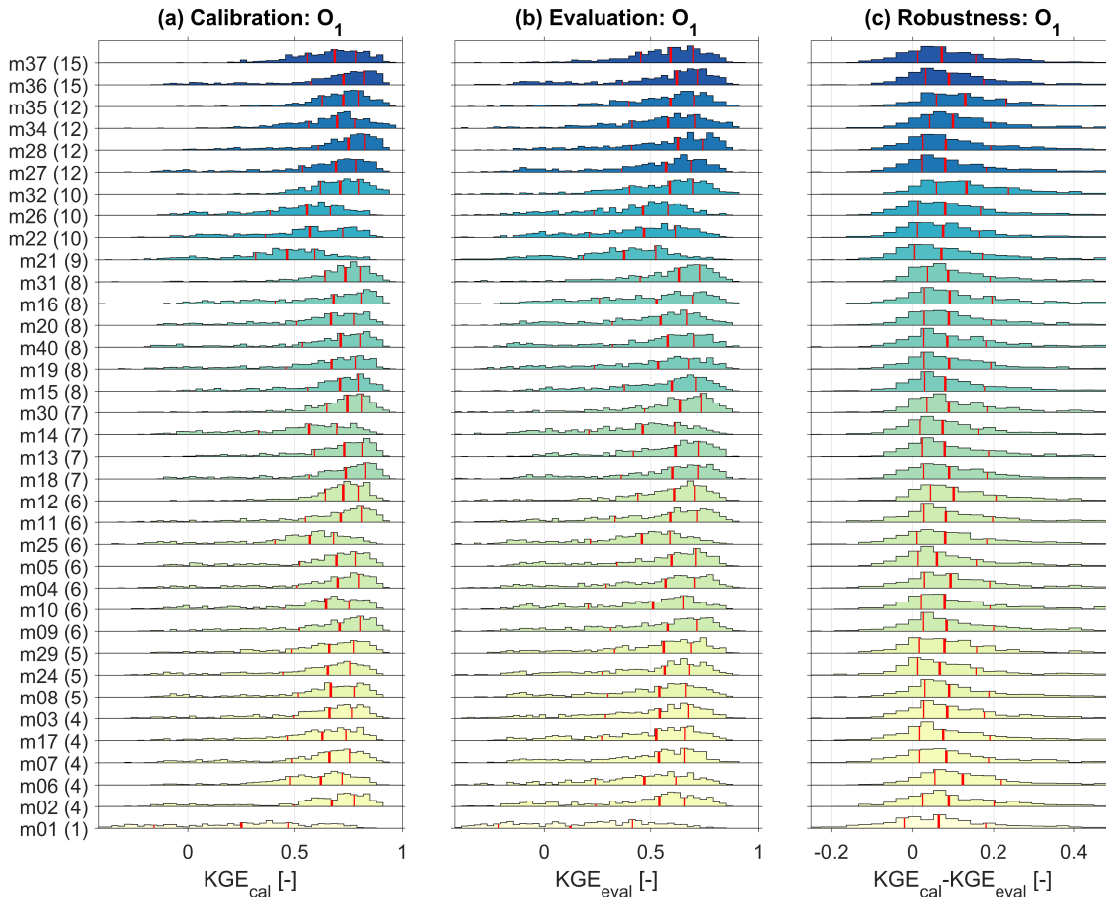
Figure 5.5: Summary of model performance during high flow simulation (KGE(Q)) in all 559 catchments. Models are sorted by number of parameters, with the model with the most parameters at the top (shown in brackets after each model ID). Coloring and model IDs match those used in Figure 5.2. Each histogram is based on 559 values. Red lines show the 25th, 50th (thick line) and 75th percentiles. Bin widths have been set at 0.03 KGE [-] through trial-and-error, settling on this value because it shows the distributions in a sufficient, but not overly high, level of detail. (a) Calibration efficiency. (b) Evaluation efficiency. (c) Difference in efficiency between calibration and evaluation period.

5.4.2.2   Model suitability for different objectives

We use model ranks to compare model performance on a per-catchment basis (Figure 5.7). For all models, the distribution of calibration ranks (Figure 5.7a) are similar to their distribution of evaluation ranks (Figure 5.7b) and there appears to be no relation between a model's number of parameters, whether it will rank high or low in either calibration or evaluation, and the model's change in relative rank between calibration and evaluation periods (Figure 5.7c). Some of the ranking distributions during calibration and evaluation
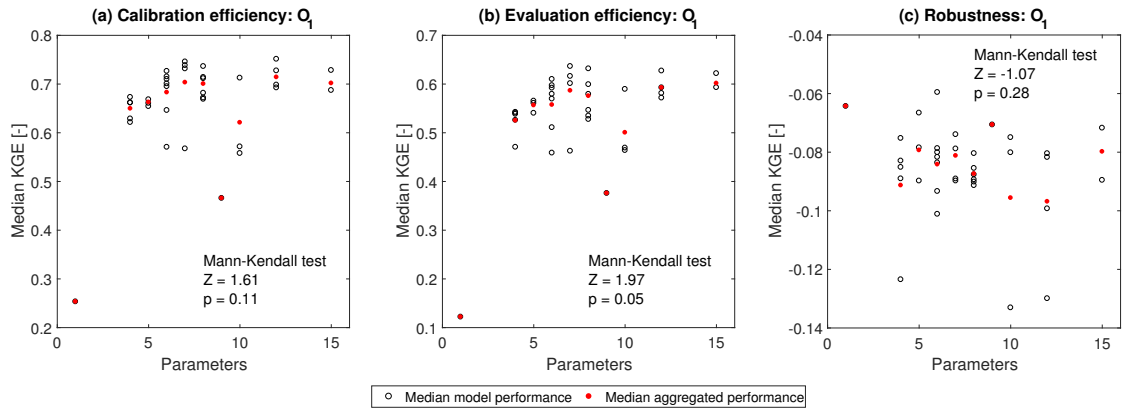
Figure 5.6: Overview of median model performance across 559 catchments (open circles) and median aggregated model performance per number of calibrated parameters (red dots) for the high flow ($KGE(Q)$) objective function. Statistics are shown for (a) the calibration period, (b) the evaluation period, and (c) the robustness of model performance given as $KGE_{val} - KGE_{cal}$. Mann-Kendall statistical tests using aggregated model performance show the relative strength and direction of trends (Z-value, negative values indicate downward trend whereas positive values indicate an upward trend) and the associated probability level (p).

cannot be satisfactorily summarized with the median, nor any other single value (e.g. the hourglass-shaped distributions of m37, m34, m12). This prevents the meaningful use of the Mann-Kendall test in these cases. The median does seem an appropriate metric to summarize the models' change in rankings (Figure 5.7c) and a Mann-Kendall test here confirms the lack of any trends related to the number of model parameters ($Z = 0.36$, $p = 0.72$). These findings also apply to the low flow objective function (Figure C.6; Mann-Kendall test on low flow rank change: $Z = -0.36$, $p = 0.72$) and to the combined flow objective function (Figure C.7; Mann-Kendall test on combined flow rank change: $Z = -0.18$, $p = 0.86$).

The model rankings (Figure 5.7, C.6, C.7) show that certain models have strong tendencies to perform well/poorly for certain objective functions (see e.g. m21, m22 and m26, that tend to rank low during high flow simulation but high during low flow simulation; compare Figure 5.7, C.6). Model ranks are summarized in Figure 5.8a, showing how often each model ranks in the top and bottom 3. Models that do not rank in the top or bottom 3 are indicated in green. The plot shows that no single model is clearly the best choice everywhere, because virtually all models appear in both the top and bottom 3 ranks for at least one of the 559 catchments. Exceptions are model m01 which does not rank in the top 3 for the low and combined flow objective functions; models m18 and m22 which do not rank in the bottom 3 for the combined flow objective function; and model m28, which

does not rank in the bottom 3 for the high and combined flow objective functions. That said, there are certainly models with more well-rounded performance than other models show. Figure 5.8b shows per objective function the 90th percentile model efficiency during evaluation. We can distinguish several model categories: (1) models that perform well on all three objective functions (i.e. the top-right corner with dark blue dots); (2) models that perform well for high flows but less so for low flows (i.e. models in the lower centre); (3) models that perform well for low flows but less so for high flows (i.e. models in the upper left corner); and (4) models that perform well on the two single objectives but lose some performance on the combined objective (i.e. models close to the 1:1 line with lighter colours). These patterns start to form around the 50th percentile performance and are relatively consistent up to the 95th percentile (Figure C.8, C.9).

It is noteworthy that model m28 is consistently one of the best, if not the best, performing model at very different percentiles (m28 is not the single best model everywhere though, as Figure 5.8a and C.9 show). The model has a unique feature not seen in any other model, namely a double parabolic curve that is used to represent the fraction of the catchment that contributes to free drainage [153]. We can speculate that this gives the model a unique capability that allows it to perform well in a wide variety of catchments. An obvious example of the impact of model structure choices is given by the model ranks that show an hourglass pattern (e.g. m37, m35, m34, m06; Figure 5.7a, 5.7b). These models all include a snow routine, which in certain catchments places them in top 10 models, whereas in less snow-dominated catchments their relative performance can vary. As another example, models m21, m26 and m22 show similar high performance on the low flow objective (Figure 5.8b; top left) and share a structural feature: all three include a store that represents the fraction of the catchment contributing to runoff which then drains into a linear reservoir. This construction is also seen in model m13, m28, m29 and m34, all of which show relatively high performance on the low flow objective function. However, while it is possible that this combination allows a model to perform well during low flow simulation, there are many other models without this particular arrangement of stores that also do well on this objective. There is thus a certain non-uniqueness in which arrangements of stores and flux equations allow a model to do well on these objectives. Furthermore, we can only use the results shown in this work as an opportunity to explore possible hypotheses about model behaviour. Our experimental set-up does not allow us to test them to either falsify or strengthen our ideas. Further investigation requires a more thorough examination of model differences and similarities, a detailed formulation of hypotheses and the construction of appropriate new models to perform these tests, which is beyond the scope of this work.
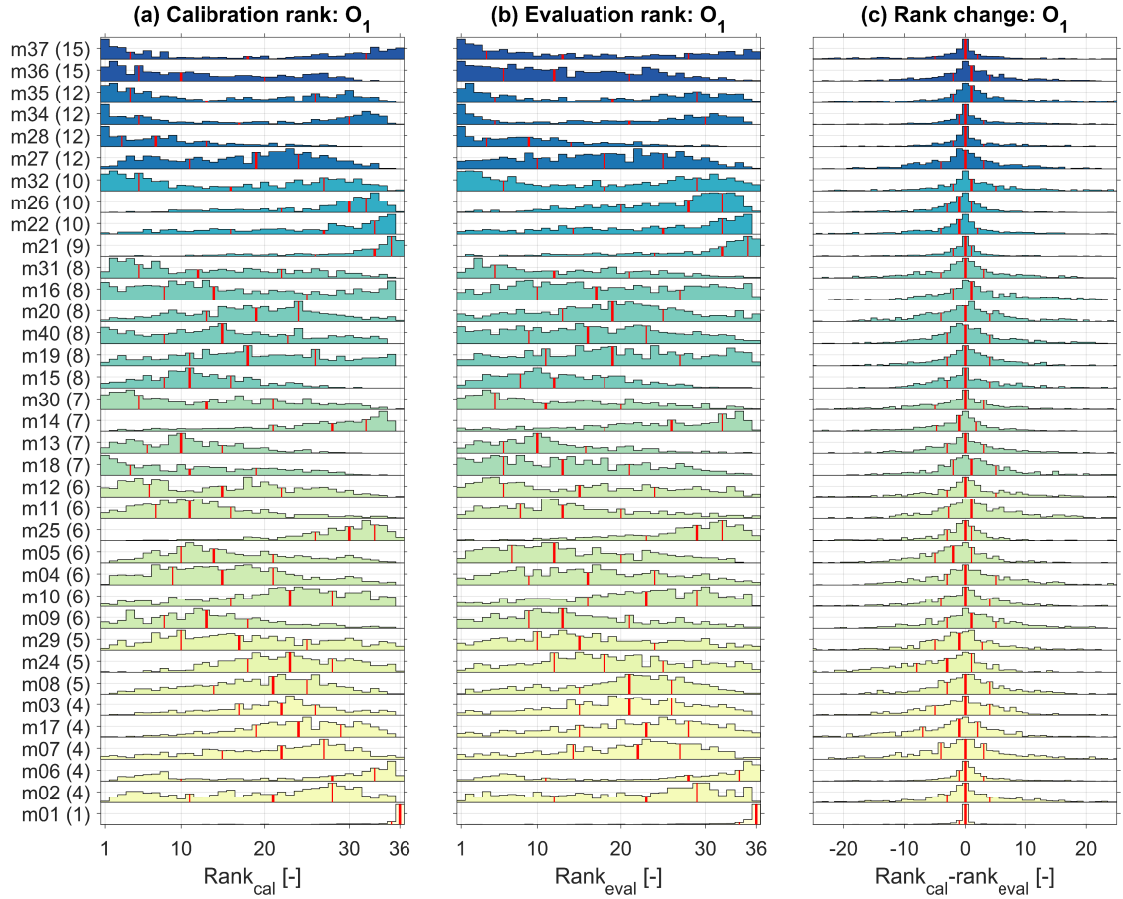
Figure 5.7: Summary of model ranking during high flow simulation ($KGE(Q)$) in all 559 catchments. Rank 1 is assigned to the model with the highest efficiency value in a given catchment. Models are sorted by number of parameters, with the model with the most parameters at the top (shown in brackets after each model ID). Colouring and model IDs match those in Figure 5.2. Each histogram is based on 559 values. Note that each histogram is scaled relative to its own maximum, meaning that absolute heights should not be compared between different rows. Red lines show the 25th, 50th (thick line) and 75th percentiles. (a) Calibration ranking. (b) Evaluation ranking. (c) Difference in ranking between calibration and evaluation period.
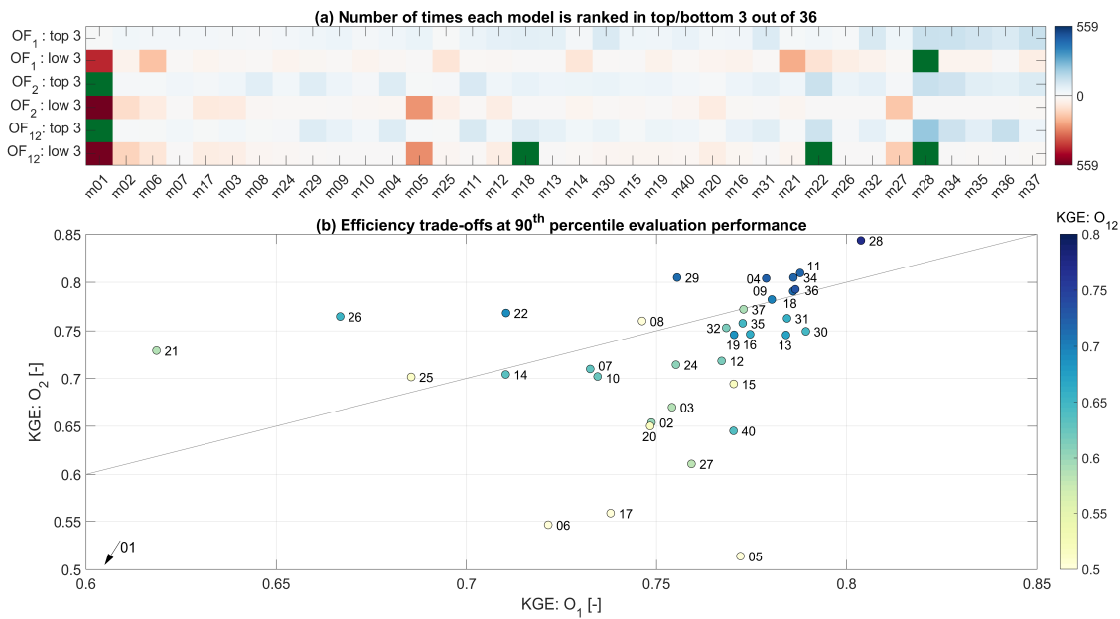
Figure 5.8: (a) Summary of model ranks during evaluation, showing the number of times each model is in the top 3 and bottom 3 for each objective function. Blue shades are used for top 3, red shades for bottom 3 counts. Zero counts are shown in green. (b) Model efficiency trade-off during evaluation of all three objective functions at the 90th percentile of model performance; i.e. the lowest model performance per model, per objective function in the top 10% of catchments. Model 01 falls outside the shown ranges, with $KGE : O_1 = 0.57$, $KGE : O_2 = -1.67$ and $KGE : O_{12} = -0.88$.

5.4.3   Relation between model structure performance and catchment attributes

We use scatter plots and correlation analysis to investigate the relationship between model ranks and the catchment attributes provided as part of the CAMELS data set. Scatter plots of model ranks for the high flow objective function (not shown for brevity) indicate that empirical relationships exist between model ranks and certain types of catchments. An obvious relation exists between models that include a snow component and catchments where a larger fraction of the annual precipitation occurs as snowfall. In these snow-dominated catchments, the snow models perform better than models without the capability to simulate snow accumulation and melt. Figure 5.9 summarizes the relationships between model ranks and catchment attributes using the Spearman rank correlation coefficient. This figure only includes catchments where less than 10% of annual precipitation occurs as snowfall. This clarifies any relations beyond snow, which are much harder to see in a correlation figure based on all catchments (Figure C.10). The strongest correlations for most models can be found with streamflow signatures. This suggests that certain models perform better or worse depending on the flow pattern they are attempting to simulate. Beyond snow, the link between model performance and climate conditions or catchment attributes is less well-defined. Using model ranks obtained during low flow evaluation (Figure C.11, C.12) and combined flow evaluation (Figure C.13, C.14) also shows that model ranks correlate most strongly with streamflow signature values, and much less so with catchment attributes used here. Generally speaking, model ranks during low flow simulation have lower correlations with catchment attributes (Figure C.11, C.12) than the ranks during high flow simulation (Figure 5.9, C.10). Correlations with catchment attributes are strongest for the combined flow objective (Figure C.13, C.14). It is possible that the multi-objective approach of the combined objective function puts greater demands on the model structure and thus relations between model performance and streamflow regimes become more pronounced.

Models can be relatively neatly grouped together based on their correlations with various attributes. In Figure 5.9 (high flow objective), models m17 to m22 seem to perform similarly well for extreme flow types, i.e. zero/low flows and high flows (for a description of these signatures, see Addor et al. [3]). Cursory examination of the model documentation shows that most of these models have a component that allows precipitation to bypass the soil moisture routine and directly contribute to runoff. Similarly, models m09, m04 and m11 share a general layout that consists of three different flow pathways (saturation excess, non-linear interflow and baseflow) coupled with a soil moisture threshold that must be exceeded before interflow occurs. These models seem to perform relatively better

in wetter catchments with occasional very high flows. Finally, models m21, m26, m28 and m34 perform relatively better in baseflow-dominated catchments without flashy streamflow behaviour. These particular models share a structural feature that consists of a soil moisture routine that simulates a variable contributing area, which then drains into a linear reservoir. This particular feature might also contribute to better relative performance of these models during low and combined flow simulation (Figure C.12, C.14). Note that these hypotheses were developed after the models were calibrated, evaluated, ranked relative to one another and grouped by their correlation with catchment attributes.

Figure 5.9: Spearman rank correlation between model ranks obtained during high flow objective function evaluation and CAMELS catchment attributes. Models are ranked from best to worst, with rank 1 indicating the best model. Only catchments where less than 10% of annual precipitation occurs as snowfall are included in this figure. A similar figure including all 559 catchments can be found in the Supporting Information (Figure C.10). Only correlations with p-value < 0.05 are shown. Marker size corresponds to the strength of the correlation, as does the intensity of the colours. Models are sorted manually in an attempt to place models with similar correlation patterns close together. Example interpretation: see the dark purple dot for the combination of model 01 and baseflow index. This indicates a strong positive correlation. Thus, as the baseflow index of catchments increases, model 01 tends to be ranked higher, i.e. worse. From a model perspective this is a sensible result, because model 01 only contains a saturation excess flow mechanism and no interflow or baseflow mechanism.

5.5 Discussion

5.5.1 Synthesis

Large-sample analysis such as this study can provide unique insights into our ability to model a wide variety of catchments and how models differ from one another in a practical sense. We show that a reasonable model can be found for nearly every catchment and objective function (Figure 5.3a, c, e), but that there are large differences in model structure uncertainty between different catchments and objective functions (Figure 5.3b, d, f). Structure uncertainty is high for high flow simulation (many models have similar performance to the best model in a given catchment, a comparable finding to the results of Perrin et al. [254]) and is much lower for the low flow and combined objectives (fewer models do well in any given catchment). Both maximum model performance and model structure uncertainty are subject to geographical organization, indicating that for certain types of catchments, achieving higher KGE values is difficult for many models.

Two catchment features stand out as strongly influencing whether a model will perform well for a given location. The first is the obvious requirement of a snow component in snow-dominated catchments (Figure C.10), although the complexity of this snow routine does not seem to influence performance much. Most models in our sample only include a very basic degree-day-model using a single temperature threshold for snowfall and snow melt. Model m37 uses a more complex snow routine that accounts for retaining and refreezing of liquid water in the snow pack, but that does not seem to give this model much of an edge over simpler alternatives (see e.g. Figure 5.5 and 5.7). The second feature relates to the variability of the streamflow regime. Differences between the models in our sample become apparent in catchments with higher BFI and Q5 values (relative to the other catchments). Models that simulate a variable contributing catchment area linked to a linear reservoir seem to do better for these flow regimes than other models in our sample (Figure 5.9). In contrast, these models do poorly in catchments where the flow regime is less constant and more prone to extremes. Generally, our results suggest that model structures strongly dictate which flow regimes a model is able to simulate properly.

The number of model parameters however has little to do with the range of performance each model achieves across the 559 catchments (Figure 5.5, C.2, C.3). Given that the models use the same data, same coding framework, same numerical implementation and same calibration algorithm, the only other source of differences is the model structure (i.e. the choice of processes and equations). Surprisingly, model performance change during evaluation seems unrelated to the number of parameters (Figure 5.5, C.2, C.3) and thus must be due to either differences in model structures or changes in climate or catchment

conditions between both periods. This contrasts with findings by Perrin et al. [254], who claim an inverse relation between model robustness (defined in their paper as the decrease of mean model performance between calibration and verification periods) and number of model parameters. This pattern is not visible in our sample (Figure 5.5, C.2, C.3) and not found by our use of statistical tests (Figure 5.6, C.4, C.5). While overfitting (i.e. performance loss in evaluation due to noise fitting during calibration, [e.g. 40, 284, 295]) is a clear issue with high-degree polynomials [p. 30, 123], it seems that these principles do not apply to our sample of conceptual hydrologic models and catchments. Our results suggest that how the parameters are used in a structure is more important to dictate a model's capabilities for a given objective function than how many parameters the model has.

The conclusions presented here are generalizations, and exceptions exist. Figure 5.8a makes clear that for virtually every combination of model and objective function, a catchment can be found where that model is one of the best choices out of our sample of 36 candidates. Equally, a different catchment will exist where that model is one of the worst possible choices (again confirming results from Perrin et al. [254]). This highlights a critical weakness of small-sample studies: conclusions are always conditional on the sample of models and catchments used and generalisation of findings should thus be done very carefully. Using larger samples reduces this risk, but it is currently still difficult to quantify exactly how representative any sample of models or catchments actually is. We will however attempt this in the following section.

5.5.2 How representative are our model and catchment sample?

One of the big challenges in hydrology is knowledge transfer from well-monitored catchments to ungauged ones [e.g. 54]. This relies on a measure of catchment similarity and a formal way of describing this similarity is so far unavailable [209, 334]. Therefore, we cannot easily say how well the catchments in this study represent the global distribution of different catchment types. Open-source data sets of catchment attributes continue to be developed [e.g. 4, 8] but we currently lack standardised metrics to quantify these attributes in a hydrologically meaningful way. Even though large-scale studies are increasingly common [e.g. 24], the lack of relevant high-quality global gridded catchment attribute data sets means that on a global scale the relationship between catchment attributes and streamflow regimes is still mostly unknown.

However, we can comment in the range of climatic conditions used in this work. Knoben et al. [179] defines the global hydro-climate with three dimensionless indices and shows that catchments that have similar values for these indices tend to have similar streamflow regimes as well (Chapter 2). The indices describe the annual average aridity (I_m), expressed
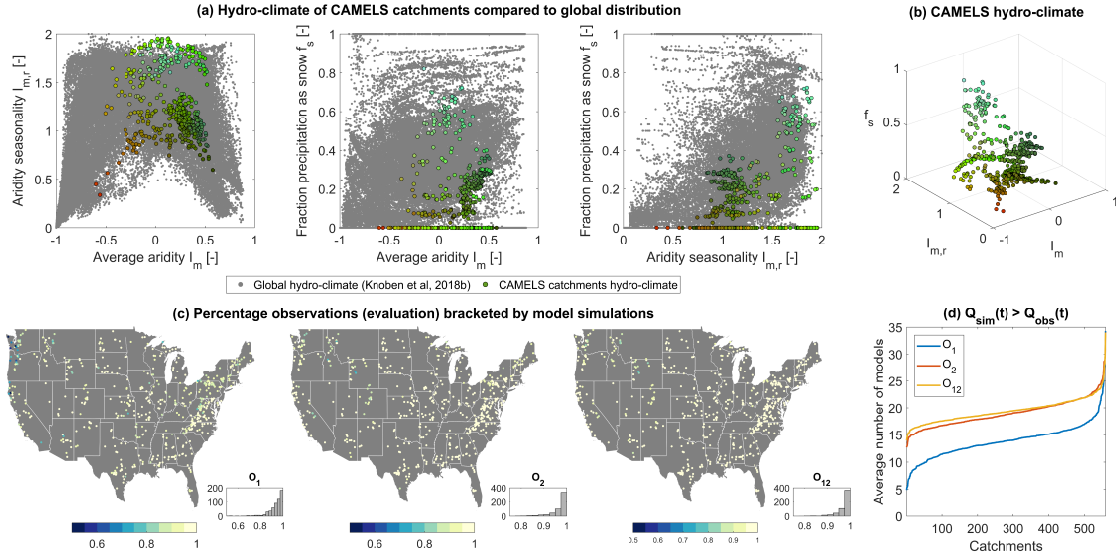
Figure 5.10: (a, b) Representativeness of the catchment sample and (c, d) model structures used in this study. (a) Comparison between the global hydro-climate (grey dots, [179, 180]) and the hydro-climate of the catchments used in this work (coloured dots), shown in 3 different 2D planes. The colour scheme replicates that of Chapter 2. (b) 3D visualization of the hydro-climate in catchments used in this work. (c) Percentages of streamflow observations in the model evaluation period that are bracketed by simulations from the model ensemble, where the models are calibrated using the high flow (O_1, $KGE(Q)$), low flow (O_2, $KGE(1/Q)$) and combined objective function (O_{12}, $\frac{1}{2}[KGE(Q) + KGE(1/Q)]$) respectively. (d) Summary graph showing whether models tend to underestimate or overestimate the observations per catchment. For each daily observation $Q_{obs}(t)$ in each catchment, we determined the number of model simulation $Q_{sim}(t)$ that are above this observation. We then averaged this number for the full evaluation period.

as a moisture index with range [-1,1] (-1 being completely arid, +1 having zero evaporation); the maximum range of month-to-month aridity ($I_{m,r}$) based on the monthly moisture index (range [0,2] with 0 indicating no month-to-month change in aridity and 2 indicating that the catchment at its driest is completely arid and at its wettest has precipitation without evaporation); and the fraction of precipitation that occurs as snowfall (f_s) based on monthly average temperature and rainfall (range [0,1] going from no snowfall to all precipitation as snow). Figure 5.10a shows the global distribution of these indices (grey dots) and where the catchments used in this study fall in the 3D-space given by these climate indices (coloured dots; CAMELS-only summary in Figure 5.10b). The catchments cover a wide range of climatic conditions, but certain areas are under-represented. There are few catchments on either end of the aridity scale ($I_m \leq -0.5$ and $I_m \geq 0.5$). High seasonality climates are well represented ($I_{m,r} \geq 1$) but low seasonality climates less so ($I_{m,r} < 0.5$). On the snow-axis,

111

moderately snow-influenced catchments ($f_s < 0.4$) are fairly well represented. The CAMELS catchments used here also include snow-dominated catchments ($f_s > 0.4$) but these tend to fall in the high seasonality climatic region as well. Therefore our work can be considered applicable to temperate catchments, but care should be taken when extrapolating our findings to climates with more extreme aridity values (e.g. deserts, tropical rain forests), to regions with less seasonally varied aridity values (e.g. climatic transition zones on the edges of deserts and rain forest), and to places with a less pronounced summer-winter temperature cycle (e.g. taiga).

Ideally, model simulations in an ensemble bracket the observations because this indicates that a sufficiently varied selection of models has been used [78]. Figure 5.10c shows that for the vast majority of catchments most of the observations fall within the range of model simulations during evaluation. However, during evaluation of the high flow objective (O_1) observations in several catchment on the west coast are not bracketed by model simulations, suggesting that models are not sufficiently diverse or the presence of bias in the forcing data (we have rejected several other catchments in this area due to water balance errors, see Figure 5.1). During evaluation of the other two objectives observations in these catchments are bracketed by model simulation, suggesting that in these places the problem lies in the calibration procedure and not in a lack of diversity in the models. In contrast, there are several catchments where only approximately 80% of observations are bracketed by model simulations regardless of objective function (for example, two catchments in western Wyoming and one catchment in western Montana). Further investigation of such individual cases is required to determine whether the current selection of model structures all lack a critical process or whether this is an artefact of the calibration procedure or data quality (for an example framework that can be used to distinguish between model structure deficiencies and calibration artefacts, see Fowler et al. [109]).

Figure 5.10d shows for each catchment how many models on average simulate streamflow higher than observations. By definition, the remaining models simulate streamflow lower than observations. The model ensemble seems relatively unbiased during low and combined flow simulation, with approximately half of the models simulating streamflow that exceeds observations. High flow simulation results indicate a more biased model ensemble, showing that in certain catchments the majority of models underestimate observations (bottom left) while in other catchments the majority of models overestimate observations (top right). Additional analysis (not shown for brevity) shows that the catchments where model ensemble seems biased (either towards underestimation or overestimation of flows) are predominantly arid ($I_m < -0.25$). The tendency to underestimation of flows is caused by the high flow objective function, because with a different objective function the

ensemble does bracket observations in these catchments. This is the case in catchments where periods with very low flow (Q < 1 mm/d) are alternated with several medium (Q > 10 mm/d) to very high (Q > 30 mm/d) flow events. In these cases most models simulate no flow instead of very low flow, because the calibration procedure focuses on the peaks only. With the low flow objective function more attention is given to these low flows and the models tend to bracket the observations better. Cases where most models tend to overestimate flows occur in the most ephemeral catchments, where long periods of zero flow are alternated with flood peaks that are both very high (up to 90 mm/d) and short (lasting around 1 day). Many models struggle to replicate this extremely flashy behaviour and it seems likely that either our chosen time-step size of 1 day provides insufficient temporal resolution to simulate the flow response for these catchments, or that our model ensemble lacks the appropriate structures that could simulate these flow regimes.

5.5.3   The need to select an appropriate hydrological model

It is clear that selecting the most appropriate model for any catchment is difficult, if not impossible, within the current study set up. Performance ranges of the different models are too similar to easily distinguish between them. This is especially the case when the high flow objective function is used, but model structure uncertainty is high for the low and combined objective function as well. It has been long known that simple models (here meaning models with fewer calibrated parameters) can compete with more complex (i.e. parameter-heavy) models in terms of model performance [e.g. 152, 254]. Overparametrisation and the inability to properly identify parameter values through calibration to streamflow data are often cited as a reason for this [e.g. 32, 186]. As a consequence, only models with a few parameters can be calibrated properly using just streamflow data (properly meaning that a unique optimal parameter set can be found). However, such simple models cannot contain all potentially relevant hydrologic processes, because this would require more parameters than can be identified from streamflow data alone. Yet, such detailed models are required if predictions under changing conditions are to be made [173]. This leads to a dilemma succinctly stated in Kuczera and Mroczkowski [186]: "A simple model cannot be relied upon to make meaningful extrapolative predictions, whereas a complex model may have the potential but because of information constraints may be unable to realize it".

Parsimonious models have distinct advantages when streamflow estimates are required under stationary conditions; model set up, calibration and use will be fast, and data requirements are low [254]. However, when a choice must be made between several equally plausible model structures for prediction under changing conditions (i.e. when it is

unknown which model is the most appropriate hypothesis about catchment behaviour), there is no justification for selecting the model with the fewest calibration parameters as the preferable alternative [239]. The impact of the choice of model structure can be significant: in a climate change impact assessment across 605 catchments in the United States, three equally plausible model structures (HBV, VIC and Sacramento-SMA) do not even predicted the same sign of future changes in 26% of the catchments [217]. Therefore, as many authors have noted, more in-depth tests are needed to distinguish between different models and make a catchment-appropriate choice. Different model structures should be considered as alternative working hypotheses [e.g. 73], which one should then attempt to falsify by subjecting them to various tests (for a recent commentary on model falsification, see Beven [42]). Our work shows for a wide selection of models and across a wide variety of catchments that these procedures must become the norm. Relying only on aggregated efficiency metrics to determine whether a model has 'adequate performance', without considering the performance of alternative model structures for a given catchment is simply insufficient.

5.5.4 Study limitations

5.5.4.1 Aggregated metrics, models and data

Our results suggest that model performance is dictated by the model structure and that certain structures, or certain arrangements of stores and fluxes, are better suited for certain types of flow simulation. This can potentially be of use in selecting appropriate models for a given study purpose and improving model functioning in a diagnostic setting. Still, our analyses are based on general performance metrics that aggregate model performance into a single efficiency score. Higher resolution diagnostics such as seasonal or time-step based performance metrics [e.g. 83] might provide insight into why there is also considerable overlap in aggregated model performance across our sample (see Figure 5.5).

Similarly, we use lumped models, catchment-averaged daily forcing data and average catchment attributes. Our approach does not account for any spatial heterogeneity in climatic inputs and spatial heterogeneity is only included in a basic way in those models that simulate a variable contributing area. We also assume that mean catchment attributes are a representative measure of the catchment as a whole. As mentioned, large-sample studies must necessarily sacrifice level of detail to facilitate the increased sample size (compare to for example model development in an experimental catchment) and lessons must be learned from emergent patterns across the large sample. The assumptions made in this work are not uncommon in hydrology: lumped models are often used with catchment-averaged, daily data, and spatially explicit descriptors of catchment attributes are mostly

unavailable. However, this lack of spatial explicitness makes it challenging to get clear answers to our questions, and much work remains to be done.

5.5.4.2 Data and parameter uncertainty

This work focusses on model structure uncertainty and leaves both data and parameter uncertainty unaccounted for. We have chosen to ignore parameter uncertainty due to the large number of models and catchments involved. Accounting for parameter uncertainty through for example GLUE-like analysis [47] would have increased computational times and the amount of data to analyse and visualise beyond feasible limits. With respect to parameter uncertainty, we see no clear evidence of model overfitting in our current approach but have no information on parameter identifiability. It is possible that cases exist where parameters are poorly identifiable during calibration, but that significant performance differences exist between such parameter sets during model evaluation [109]. We have treated data uncertainty as a constant for all models for similar reasons. Properly accounting for input data uncertainty is difficult, because these errors tend to vary in time. Accounting for non-stationary errors in rainfall data is possible but computationally intensive. Uncertainty in model evaluation data tends to be more stable in time, but also requires additional sampling (for a recent review of data uncertainty in hydrology, see McMillan et al. [216]). Our approach is constrained by a need to limit computational times, but ignoring data uncertainty can force the calibration procedure to compensate for errors in the measured rainfall-runoff relationship and not accounting for parameter uncertainty ignores the possible failure of the calibration procedure to select the most appropriate parameter set out of several sets with similar performance during calibration. However, our results suggest that models can be divided into groups, where models within each group are similarly suited towards particular flow regimes. These groups can be used to select a small number of promising models within our ensemble. Reducing the number of models reduces computational demands and allows investigation of parameter uncertainty and identifiability, as well as data uncertainty on a subset of models that are representative of the larger model sample.

5.5.4.3 Relating model performance to catchment attributes

We were unable to satisfactorily relate model performance to catchment attributes. In our opinion, efficiency values should not be used directly because equally high KGE values are not equally hard to achieve in different catchments [e.g. 279] and baseline performance is not the same across catchments (Figure 5.4). This makes it difficult to quantify how good model performance really is and better model benchmarks are needed. Lacking

these, we used model ranks to relate relative model performance to catchment attributes. These findings seem to indicate that certain model structures perform better or worse depending on the streamflow regime they attempt to simulate, but the relation between model performance and catchment attributes is inconclusive. The relationship between catchment attributes and streamflow signatures is currently not well-understood [3] either. This might mean that we are not using the right catchment attributes, or that different arrangements of catchment attributes lead to similar flow patterns and thus obscure any prevalent correlations, or that catchment-averaged values for catchment attributes lack the required level of detail to be of use. More work is required on understanding the differences and similarities between catchments, in terms of catchment attributes, streamflow regimes and model performance.

5.5.4.4 Generating and testing model hypotheses

Hypothesis-based model testing has been advocated for some time [e.g. 34, 42, 73, 104, 173, 256] but few studies so far have put this approach into practice. Our results give rise to several hypotheses about conceptual model behaviour (see Section 5.4.3), but again note that these hypotheses were made after we calibrated, evaluated, ranked and grouped the models. Strict testing of the hypotheses is thus necessary (i.e. comparing the performance of two model structures that are identical apart from the element under consideration under a wide variety of conditions) before these ideas can be used to guide model development. Testing these is not possible in the current study framework but hopefully these results can inform future studies on this subject. Additionally, the current results include two subsets of models that have very similar structures and can be used to investigate the impact of single model decision on a large scale. Models m22, m26 and m34 are all based on the FLEX framework [103] and can be used to assess the impact of adding an interception routine and a snow routine respectively. Models m24, m30, m31, m32 and m35 are part of the same study [349] and can similarly be used to isolate the impact of specific modelling decisions.

5.5.5 Fostering further work

This study presents selected results from a large-sampling study, using 36 models for streamflow simulation in 559 different catchments, with each model being calibrated three times for three different objective functions. In this work we trade "breadth for depth" [130] to explore a variety of questions. As a result, we have made several concessions to the study method and have raised as many questions as we hoped to answer. However, the computational demands of a large-sample studies such as this can be high. To facilitate

further research, calibration results of all models (parameters, simulated model storages and fluxes and obtained efficiency values) will be made available through an open-access repository once calibration of all models is completed. The CAMELS dataset and MARRMoT modelling toolbox are also freely available and can be found through their respective references.

5.6 Conclusions

We calibrated 36 lumped conceptual models for streamflow simulation in 559 catchments across the United States, using three different formulations of the Kling-Gupta Efficiency as a performance metric. The metrics focus on high, low and combined flow simulation respectively. Model calibration and evaluation were done using two different 10-year periods. All results are based on model performance during evaluation, unless noted otherwise. Here, we return to the five questions posed in the introduction and summarize what this work contributes to answering them. Model structure uncertainty is highest (i.e. it is hard to distinguish between models in term of performance) during high flow simulation, but is still significant during low and combined flow simulation (Figure 5.3). For the majority of catchments, there is little practical difference (< 0.05 KGE difference) between the best performing model and up to 27/16/13 other models in the sample for high/low/combined flow simulation (Figure 5.4). Histograms of model performance (Figure 5.5) show that distributions of model performance are remarkably similar during high flow simulation, and this holds to a slightly lesser extent for low flow simulation (Figure C.2). The largest differences in model performance can be seen during combined flow simulation (Figure C.3). Contrary to expectations, Mann-Kendall statistical tests (Figure 5.6) indicate that the number of model parameters does not relate to the distributions of model performance values during calibration or evaluation, nor to the change in model performance between both periods. Histograms of model ranking relative to one another (Figure 5.7, C.6, C.7) suggest that certain model structures are inherently better than others at reproducing certain parts of the hydrograph. This can also be seen in summary plots of absolute KGE values (Figure 5.8b, C.8, C.9) which suggest clear differences in the trade-off between objective functions for different models. However, it is impossible to find a single best model structure, because virtually every model ranks in both the top 3 and bottom 3 of models for at least a few of the 559 catchments (Figure 5.8a). We struggled to identify which model structures are the best suited for certain climates or catchment types, and instead found the strongest relation between relative model rank and streamflow signatures (Figure 5.9, C.10-C.14). This reinforces the idea that certain model structures are better suited for simulation of certain parts of the hydrograph, as various

aspects of the flow regime change in importance across the catchment sample. Scatter plots of model ranks against 52 catchment attributes (not shown for brevity) confirm significant scatter in these relations. Our catchment sample covers a significant range, but is not representative of all possible hydro-climates (Figure 5.10a, b). Similarly, our model ensemble seems relatively unbiased in the majority of catchments (Figure 5.10c, d) but there is some evidence that we lack appropriate model structures (in terms of missing processes or too crude spatial or temporal resolution) for high flow simulation in flash-flood-dominated arid catchments. Given our catchment-averaged approach to model use, data and analysis, more detailed investigation of between-model differences is needed and care should be taken when applying our findings to future modelling efforts that extend beyond the limits of our approach.

CONCLUSIONS AND OUTLOOK

The work in this thesis all contributes to improving our ability to perform large-sample model comparison studies. This started with investigating the global hydro-climate, so that we might quantify how representative any given sample of catchment types is. We followed this with a detailed investigation of conceptual model structures and the development of a novel model comparison coding framework. We then used the framework for a large-sample model comparison study which led to new insights about the differences between model structures and therefore can inform future model development work. From the start, it was obvious that addressing all concerns that currently complicate large-sample studies in a single thesis is infeasible. Our goal has been to contribute to, rather than provide, an answer to the question "what is the most appropriate model for a given objective?". We've taken several steps along the road to this answer but much work remains to be done. To facilitate this, all our results are, or will be, publicly available. Now we can only hope that this groundwork will enable and inspire further research.

Chapter 2 Hydrologic climate classification
`https://dx.doi.org/10.5523/bris.16ctquxqxk46h2v61gz7drcdz3`
Chapter 3 Global precipitation modality
`https://dx.doi.org/10.5523/bris.2ynd0zj7oqd1t24t6vhhh83exn`
Chapter 4 Multi-model comparison framework
`https://github.com/wknoben/MARRMoT`
Chapter 5 Model inter-comparison study
`to be decided`

6.1 Chapter summaries

Chapter 2: Hydrologic climate classification In order to select representative samples in a given study, we first need to have some idea of the population as a whole. This means hydrology needs an overview of the global variety of catchments, with respect to climatic forcing and local attributes such as topography, geology, etc. High quality global data sets of catchment attributes are currently unavailable but gridded global data does exist for several climatic attributes. However, hydrology has no formal way to structure the climatic forcing that underlies hydrologic response. Various climate classification systems can be borrowed from other disciplines but these are based on different organizing principles than a hydrological classification might need. Chapter 2 presents a hydrologically informed way to quantify global climates, explicitly addressing the shortcomings in non-hydrological climate classifications. Causal factors (climate) and hydrologic response (streamflow) are separated, meaning that our classification scheme is based only on climatic information and can be evaluated with independent streamflow data. We use theoretical understanding of the causal relation between climatic forcing and hydrologic response to define six climate indices. Sensitivity analysis reduces this number to three dimensionless indices for any location. These describe annual aridity, aridity seasonality, and precipitation-as-snow. We use these indices to create several climate groups and define the membership degree of 1,103 catchments to each of the climate groups, based on each catchment's climate. Streamflow patterns within each group tend to be similar, and patterns tend to be different between groups. Visual comparison of flow regimes and Wilcoxon two-sample statistical tests on 16 streamflow signatures show that this index-based approach is more effective than the often-used Köppen-Geiger classification for grouping hydrologically similar catchments. Climate forcing exerts a strong control on typical hydrologic response and we show that at the global scale both change gradually in space. We argue that hydrologists should consider the hydro-climate as a continuous spectrum defined by the three climate indices on which all catchments are positioned and show examples of this in a regionalisation context.

Chapter 3: Global precipitation modality The new climate classification shows that seasonality of climate can be an important control on hydrological response. Many locations experience a single wet and dry season per year, but some studies report the occurrence of two wet and dry seasons per year and our hydrologic climate classification does not specifically account for this. However, different seasonal precipitation patterns lead to differences in seasonal streamflow distributions and this might be relevant during catchment classification. Multiple dry and wet seasons per calendar year (referred to as

bimodal precipitation in the specific case of two dry and two wet seasons) are commonly associated with locations within the tropics but are reported outside the tropics as well. However, this information is fragmented and studies of bimodality are mainly restricted to monthly rainfall totals. In Chapter 3, we use a gridded global data set and simple harmonic analysis to provide a systematic overview of global bimodal rainfall and rain-day frequency. We find a good agreement between the various regional studies concerning bimodal precipitation and our global overview, showing that bimodal rainfall occurs on approximately 7% of the global land surface. In the tropics, regions of bimodal rainfall totals (P) and regions of bimodal rain-day frequency (N) tend to overlap due to the presence of dry seasons that have zero precipitation. Outside the tropics P and N are more independent which leads to complex within-year patterns of precipitation intensity. A secondary outcome of our results is an improved low-dimensional global parametrization of monthly rainfall regimes. Our results provide the first gridded global overview of bimodal rainfall patterns and show the usefulness of simple mathematical approaches for detecting patterns in large data sets. In the context of global catchment classification, bimodal precipitation patterns are strongest in the Horn of Africa, equatorial Africa and various smaller, scattered regions in northern South America, Sri Lanka and southern Pakistan. However, because streamflow records from these areas were unavailable, we must content ourselves with noting that bimodal precipitation regimes exist without evaluating their impact on hydrologic regimes.

Chapter 4: Multi-model comparison framework In the next part of the thesis, we searched for a framework that facilitates the inter-comparison of conceptual hydrological model structures across a very wide range of climatic conditions. Our aim is to compare commonly used models and improve our understanding of where and why models behave similarly and differently, with the ultimate goal of reducing uncertainty in model structure selection. We found that existing model comparison frameworks do not fulfil our needs. In Chapter 4, we developed a new model comparison framework called the the Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT) that does work for our specific purposes. MARRMoT is a modular open-source toolbox that contains documentation and model code for 46 existing conceptual hydrologic models. The toolbox is developed in Matlab and works with Octave, so it can be used in an open-source computing environment. Models are implemented following several good practices in model development: definition of model equations (the mathematical model) is kept separate from the numerical methods used to solve these equations (the numerical model) which leads to clean code that is easy to adjust and debug; the Implicit Euler time-stepping scheme is provided as the default option to numerically approximate each model's Ordinary Differential Equations in a more robust way than (common) Explicit schemes would; threshold equations are

smoothed to avoid discontinuities in the model's objective function space; and the model equations are solved simultaneously, avoiding physically unrealistic sequential solving of fluxes. Generalized parameter ranges are provided to assist with model inter-comparison studies. To encourage use of the framework, we created a large body of supporting materials detailing (i) descriptions of the 46 models, (ii) descriptions of how each flux equation is translated into computer code, (iii) similar descriptions for all 7 routing methods used by the models, (iv) generalised parameter ranges with literature references to support these choices, and (v) an extensive User Manual accompanied by several workflow scripts that show basic example applications of the toolbox (i.e. running a model, parameter sampling, model structure comparison and model calibration).

Chapter 5: Model structure uncertainty In the last part of the thesis, we finally attempt to shed some light on the main research question of this work: "what is the most appropriate model for a given study objective?" Choosing an appropriate model is a critical aspect of any modelling study. Previous studies have shown that the choice of model structure, i.e. a model's selection of states and fluxes and the equations used to describe them, has an important impact on model performance and thus on the study's conclusions. Chapter 5 investigates the performance differences of 36 lumped conceptual model structures, used for daily streamflow simulation in 559 catchments in the United States, using the MARRMoT framework developed in Chapter 4. Calibration of the final 10 models is still ongoing, but time constraints prohibited their inclusion in this chapter. Model performance is quantified using three different objective functions based on the Kling-Gupta Efficiency that focus on high flows, low flows and a combination of high and low flows respectively. Each model is calibrated for every combination of catchment and objective function with a Covariance Matrix Adaptation Evolution Strategy (CMA-ES), using a 10-year calibration period. Model performance is evaluated using a separate 10-year period. We find that adequate model performance can be achieved in most catchments for each objective function, but that no single model is capable of this. Instead, nearly every model is both one of the best choices in certain catchments and one of the worst in other catchments. In most catchments, several models achieve performance very similar to that of the best model for that catchment, resulting in high levels of model structure uncertainty. This is especially apparent during high flow simulation. During low and combined flow simulation, larger differences between model structures are seen. We find no relation between model performance during calibration or evaluation periods and number of model parameters, nor is there evidence of increased risk of overfitting for models with more parameters. However, we do find that certain structures seem better suited to certain objective functions and we formulate several hypotheses to investigate why this might

be the case. We attempt to relate model performance to catchment attributes such as climate, geology, topography, soil type and streamflow signatures and report the clearest link between model performance and streamflow signatures. This suggests that certain models are inherently better suited for certain flow regimes, but the relation between flow regimes and catchment attributes, and the relationship between catchment attributes and model performance, remains elusive. There is enormous potential for follow-up studies, using the results from this thesis as groundwork.

6.2 Overarching remarks

Figure 6.1 returns to the ideal model-comparison study presented in Chapter 1 (Figure 1.2) and shows how the work in this thesis contributes to the overall challenge of doing better model comparison studies. As noted in Chapter 1, the perfect model comparison study is currently out of reach for a variety of reasons. We have identified (i) a lack of knowledge with respect to defining a representative sample of testing conditions, (ii) a lack of knowledge with respect to defining a representative sample of models, and (iii) a practical limitation with respect to accounting for data and parameter uncertainty in a large-sample model comparison. The work in this thesis represents progress in the first two aspects.

The hydrologic climate classification (Chapter 2) and our analysis of bimodal precipitation patterns (Chapter 3) have increased our understanding of the global hydro-climate and these findings impose limits on the possible ranges that the global hydro-climate can take. The precipitation seasonality work (Chapter 3) also directly builds upon an open question posed in earlier work [30, 219], namely whether precipitation modality should be included in low-dimensional parametrisations of global climatic conditions. Many studies have quantified the global climate [e.g. 30] and several have quantified global streamflow patterns [e.g. 24]. Local assessments of the relation between climate, catchment and streamflow are also common [e.g. 29, 219, 270, 346]. Our hydrologic climate classification goes beyond this and provides a global overview of the relation between climate and streamflow. This work can provide a stepping stone towards a full catchment classification, that also accounts for the impact of the catchment on streamflow regimes.

Like other authors [e.g. 254, 293], we could not select a representative sample of models. While theoretically we should be able to define the total model space perfectly, in practice there are too many models and there is too little understanding of what makes them different or similar. Our results in Chapter 5 however suggest that large-sample studies can reveal emergent patterns in model behaviour and that it might be possible to define groups of similar model structures based on these patterns (Figure 5.9). This provides an
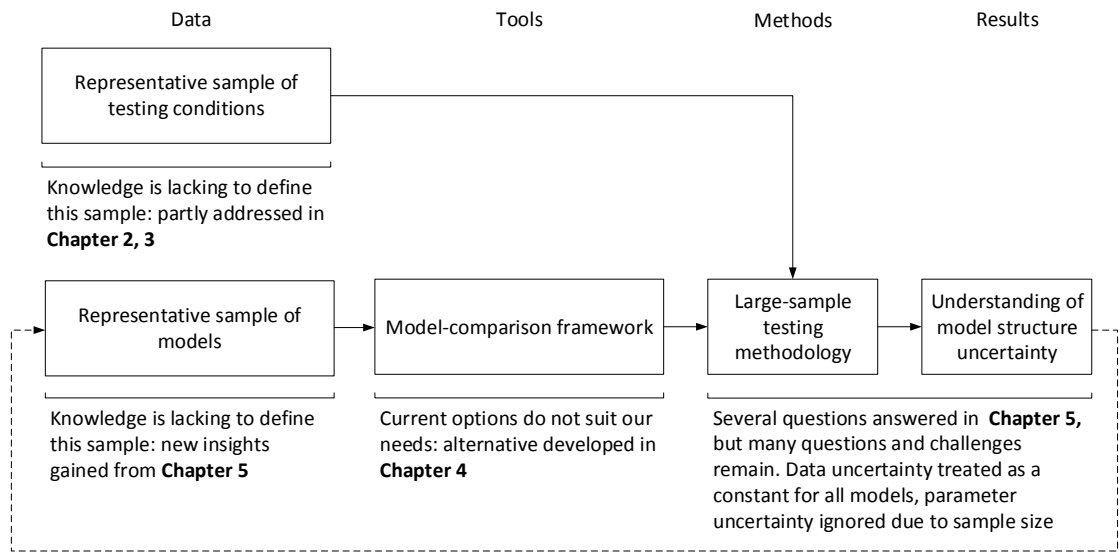
FIGURE 6.1. Overview of how the work in thesis contributes to improving our ability to perform large-sample model structure comparison studies. Because defining a representative sample of models is currently not possible, an iterative procedure might be adopted where results based on a large sample of models inform the selection of a representative model sample in future work.

interesting solution to the "representative model sample" problem: iterative approaches can teach us about the existence of groups in the total model space, where certain models with similar behaviour are found. One or a few members from each group can be selected as representative models and used in the next model comparison iteration. Reducing the number of models has the benefit that more in-depth analyses of inter-model differences become possible (e.g. one can go beyond aggregated metrics, which is a key limitation of our large-sample comparison in Chapter 5), as well as giving the opportunity to account for data and parameter uncertainty. It is important to note though, that different model groups emerge for different objective functions, and a representative model sample is thus conditional on the objective function used in the first iteration of the comparison. Our model-comparison framework presented in Chapter 4 can be a vehicle for such iterative analysis.

6.3 Outlook

This thesis leads to various opportunities for future research. This section outlines two interesting research directions in detail and provides a short list of other possible projects to conclude the section.

6.3.1 From climate to catchment classification

This thesis has made a few steps towards clarifying the link between climatic conditions and streamflow, but the relation of either with catchment attributes is still not quantified. Several steps can be taken to bring a hydrologic catchment classification closer.

In Chapter 2 we created a hydrologic climate classification, but did not account for precipitation modality. We show in Chapter 3 that two rainfall seasons do occur on a significant fraction of the land surface, but we lack streamflow data from regions where two rainfall seasons are common. Future work can focus on evaluating the impact of bimodal precipitation on streamflow regimes and extend our classification from Chapter 2 if that seems justified. This is of special interest in locations where bimodal precipitation occurs due to alternating rain and no-rain seasons (e.g. the Horn of Africa), because the impact on streamflow will be easier to see and more significant than in regions where bimodal precipitation occurs on top of high baseline rainfall (where the precipitation pattern is relatively less important compared to the overall higher precipitation volume, as is the case in parts of Equatorial Africa).

Similarly, the classification in Chapter 2 does not account for precipitation intensity because our data set lacked the temporal resolution for this (CRU-TS v2.32, monthly data [137]). Precipitation intensity interacts with a catchment's top layer to determine infiltration rate, ponding and overland flow. This could be an important factor for distinguishing catchments and future work can focus on clarifying these relations. The MSWEP dataset [25] provides 3-hourly gridded global precipitation data for the period 1979-2015 and might provide the resolution needed for such analysis.

Newly developed sets of catchment data can also be of use in this effort. CAMELS-US [4] and CAMELS-CL [8] include a wide variety of catchment attributes for 671 catchments in the US and 516 catchments in Chile, respectively. Similar data sets for the UK and Australia are being prepared [personal communication with G. Coxon and K. Fowler respectively] and all might prove useful for catchment classification. However, there are still various challenges to overcome. First, these data sets include attributes related to topography, geology, soil, vegetation, human influence, streamflow signatures, and others, but it is currently unknown which of these attributes are relevant (see e.g. Chapter 5

125

or [3]) and whether we should use these attributes as they are or convert them into more hydrologically meaningful numbers first (e.g. use soil depth in relation to annual precipitation instead of soil depth as an absolute value) [346]. Further difficulties arise from the fact that catchment attributes and time series of observations typically come from a range of different data sources, meaning that uncertainties and data quality can vary within the combined data set. It is also possible that average catchment values are not a hydrologically meaningful way to describe catchments, all of which can lead to difficulty with relating catchment attributes to flow regimes. It is also possible that different arrangements of catchment attributes lead to similar streamflow regimes, or that seemingly similar arrangements of attributes lead to different streamflow regimes [e.g. 243], which complicates finding clear relationships between the two. Second, streamflow signatures are an increasingly common way to summarise hydrologic regimes, but we don't understand them well enough [3], nor are we particularly good at selecting them [215], nor is quantifying their uncertainties [e.g. 340] common practice. Because of all the potential non-uniqueness of relationships and the inherent uncertainties of the data being used, unsupervised clustering approaches are probably not the best method for catchment classification, even though these approaches are relatively common [e.g. 187, 275, 276]. Instead, using hydrologic theory [e.g. 77] to create a conceptual understanding of these relationships and using clustering approaches only to test the validity of this understanding would be a much stronger basis for catchment classification. This approach has worked in Chapter 2 to clarify the relation between climate and streamflow and it is possible that with newly available data sets a similar approach could work for the climate-catchment-streamflow relationship.

6.3.2   The need for better benchmarks

In Chapter 4, we calibrated 36 models by optimising three different objective functions based on the Kling-Gupta Efficiency [KGE, 129]. Although there is guidance on what constitutes an acceptable efficiency value [223], existing standards tend to have been developed in specific environments and are not necessarily appropriate elsewhere. Our data set covers a wide variety of environments and the obtained efficiency values can not directly be compared across catchments [279]. Consider the hypothetical cases in Figure 6.2. In both rows, simulated streamflow has the same variation around the observations (2% and 10% of the mean in top and bottom row respectively). Figure 6.2d, 6.2e and 6.2f show cases where the model captures the seasonal pattern but fails to simulate the daily variation correctly. The KGE values in these three cases vary dramatically, due to large differences in the correlation component r. These KGE differences are directly related to
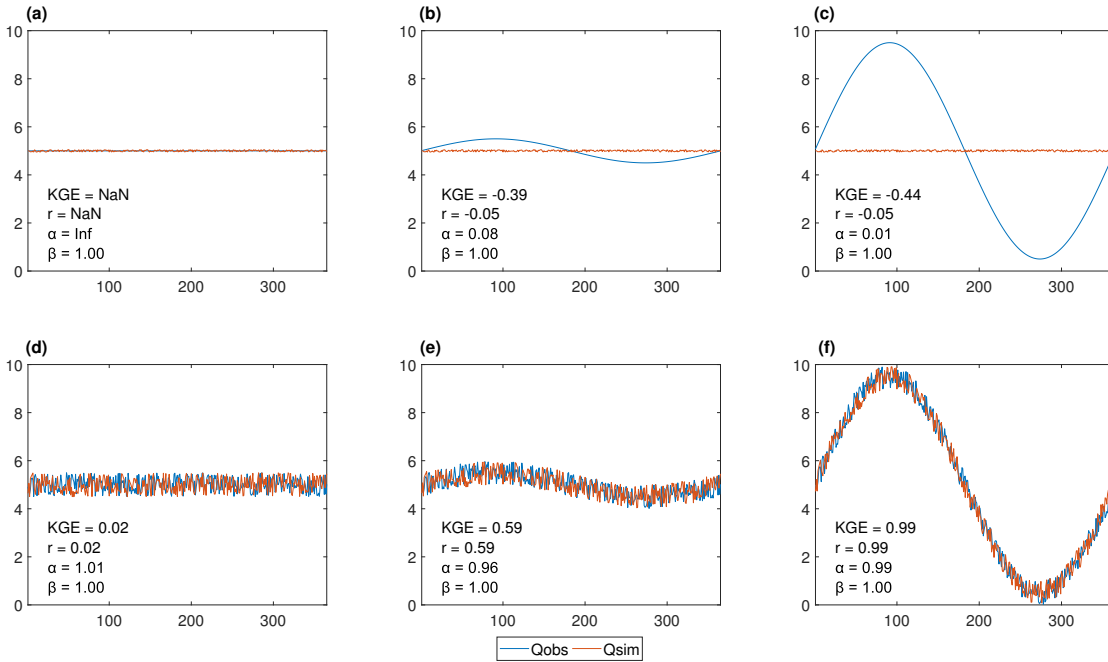
FIGURE 6.2. Kling-Gupta Efficiency calculations for a variety of synthetic stream-flow cases. r, α and β are the individual components of the KGE equation, expressing linear correlation, ratio of standard deviations and ratio of means respectively. KGE computations in (a) result in errors because the observations are the same value on each day. This returns in undefined correlation r and divide by zero error for ratio α.

the magnitude of the seasonal signal. We've shown in Chapter 2 that streamflow patterns vary geographically as a result of variations in climate. Wet and seasonal climates, and snow-dominated climates generally have higher and/or more defined streamflow peaks. These peaks are direct results from seasonality in the climatic forcing and any model that can thus transform these seasonal climatic patterns into seasonal streamflow patterns will be able to achieve high efficiency values in these catchments. The magnitude of the streamflow seasonality then dictates how high these efficiency values will be. In order to compare efficiency metrics across catchments with varying seasonal streamflow patterns, model efficiency values need *catchment-specific* context.

Benchmarks can provide catchment-specific context, by giving an indication of what base efficiency value any model is expected to improve upon. Traditionally, the mean flow is used as a comparison benchmark because this is an inherent feature of the Nash-Sutcliffe Efficiency and gives NSE = 0 [e.g. 129]. In KGE terms, the mean flow corresponds to

$KGE = 1 - \sqrt{2} \approx -0.41$ (Figure 6.2a-c, exact values depend on chance correlation between randomness in simulated flow and observations). This approach ignores that for certain places the mean flow can be quite representative of the typical flow regime (e.g. in locations with very low seasonal changes in the hydrograph, Figure 6.2a), whereas in other locations it is a very poor predictor of the typical flow regime (e.g. snow-dominated catchments, Figure 6.2c). This benchmark is thus not equally hard to beat everywhere, nor is it a particularly challenging benchmark in many cases (see Figure 5.4).

Calls for better benchmarks are not new [e.g. 288] but progress has been slow and with the increasing interest in large sample hydrology [130], there is a need for improved standardisation of model evaluation. Approaches so far can be divided into data-based and model-based benchmarks. The mean flow is a well-known data-based benchmark. The mean-calendar-day flow (e.g. a "typical streamflow year" created by averaging all January 1^{st}s, all January 2^{nd}s, etc) [279] can be more suitable in catchments with a distinct seasonal flow regime that is stable across years (in terms of volume and timing) because it captures this seasonality as part of the baseline upon which models are expected to improve. We will add the suggestion of a median-calendar-day flow benchmark, because typical flow years created from median daily observations have proved useful in Chapter 2. Use of the median smooths out year-to-year flow variations and is thus less prone to the presence of extremes in the data. Another option is to use the Streamflow Natural Variability [231] as a data-based benchmark. In contrast to data-based benchmarks, model-based benchmarks generally involve defining a simple model that more complex models are expected to improve upon. These can be simple expressions of the general rainfall-runoff relationship [279] or fully developed rainfall-runoff models such as SWBM [240], GR4J [253] or the Sacramento Soil Moisture Accounting model [230]. It is also possible to create model-based lower and upper benchmarks, by using a random and calibrated parameter set respectively [292]. The ultimate goal of any benchmark is to make clear "what [efficiency values] should be expected in a given catchment" [292], within the limits of what is supported by the quality of the observation data [83], and noting that high efficiency values do not necessarily constitute "realistic" model behaviour [289].

Model-based benchmarks give a clear idea of which model performance can be expected in any given catchment, but they require several (subjective) decisions before they can be applied. Questions such as "which model should we use as benchmark?", "how can we choose the 'simplest model' and what is an appropriate level of simplicity?", "which parameters should this model use?", "is this model an equally hard to be beat benchmark in different places?" need to be answered and justified. For a single-catchment study, a model-based benchmark gives a good idea of whether using a more complex model can be justified. In a large-sample study however, model-based benchmarks still suffer from a lack

of transferability. Because models are not equally good everywhere, we still need a way to interpret the benchmark model's efficiency values in terms of what is a reasonable value for any given catchment. Data-based approaches have the benefit of assuming nothing about the rainfall-runoff relationship, so they avoid subjective questions about model choice and parametrisation, but this is their main weakness as well. Because these benchmarks do not specify any rainfall runoff relationship, their usefulness decreases under non-stationary conditions or in locations with high interannual streamflow variability. In the first case, the benchmark regime will gradually become less representative of the new regime as conditions change. In the second case, the mean or median values are not a good representation of the actual variable flow regime.

In conclusion, there is an urgent need for an an overarching methodology that allows us to better compare model efficiency across places, but our current generation of benchmarks does not offer an obvious best answer. In multi-model studies, a data-based benchmark might be preferable because it avoids choosing a 'simplest' model which is necessarily subjective. In model improvement studies, a model-based benchmark might be preferable because it can indicate whether increasing model complexity is justified. Both approaches need further attention.

6.3.3 List of possible follow-up projects

- Evaluate the hydrologic climate classification in time: use the climate classification to relate long-term changes in hydro-climate to historical records of streamflow observations and see if our defined relationships hold.

- Evaluate the hydrologic climate classification in space: African and Asian river systems are under represented in the GRDC dataset we used [315]. A new dataset has become available [GSIM, 93, 127] with a wider global coverage that might be used to evaluate our classification. GSIM only includes streamflow signatures and does not provide continuous streamflow observations, so these tests must be limited to comparison with our findings on streamflow signatures (Figure 2.7).

- Expand MARRMoT: addition of adaptive time-stepping [e.g. 145] could both increase numerical performance and computational speed of all models.

- Expand MARRMoT: addition of new models that show step-wise increases in complexity [e.g. 106] will allow the framework to be used for multiple working hypothesis assessment [e.g. 73].

- Use MARRMoT: the need for hypothesis testing in hydrology has been discussed often [e.g. 34, 37, 42, 73, 77, 256, 273, 329] but is only occasionally applied in practice [e.g. 106, 309]. It is possible that the effort involved in creating alternative model structures is prohibitive. MARRMoT can be of use in making multi-model studies accessible for a wider group of researchers and facilitate further model-comparison studies.

- Investigate inter-model differences: in Chapter 5 we have calibrated a large number of models for several hundreds of catchments but only assessed model differences through aggregated efficiency metrics and model rankings. A next step can be to assess differences in simulated fluxes from models that perform similarly well. Each model generates at least time series of flow (Q) and evaporation (Ea). Quantifying differences between simulated fluxes [e.g. 1] can lead us to better understand the distribution of model structures in model space and model independence (i.e. answer the question "do models with similar efficiency values have similar output dynamics?"). This could potentially assist in selecting appropriately different models for ensemble prediction [344].

- Investigate model failure: our assessment of model adequacy and model failure in Chapter 5 has only involved aggregated performance metrics (i.e. the similarity of observed streamflow and model simulated streamflow is expressed as a single aggregated number). This provides a certain amount of insight about how closely simulations resemble observations, but it provides little insight into where models deviate from observations. Time-step based diagnostics can provide the necessary resolution to investigate whether models fail consistently under certain conditions and improve our understanding of where models go wrong and whether this is consistent across different model structures [83].

This appendix has been published as Supporting Information for a research article in Water Resources Research. Slight modifications have been made to better fit the general layout of this thesis. We gratefully acknowledge the assistance provided by professor Jonathan Rougier in creating the empirical Wilcoxon statistical test.

A.1 Introduction

These Supplementary materials are provided to support several arguments made in chapter 2 and to provide in-depth explanation of the methodology for reproducibility.

Section A.2 deals with preparation of GRDC streamflow data (available on request from http://www.bafg.de/GRDC/). This study uses a sub set known as "Climate Sensitive Stations Dataset (Pristine River Basins)" [316] downloaded on 16-05-2017. Outlined in section A.2 are the steps taken to determine which catchments for which no catchment boundary information is available should be excluded from further analysis (section A.2.1), results of manual quality assurance (section A.2.2 and table A.1) and length of the flow record for the catchments used in this study (section A.2.3).

Section A.3 provides details about the calculation of streamflow signatures. Section A.3.1 explains our modifications to the standard Wilcoxon test to create an empirical Wilcoxon test, which is used to estimate the statistical significance of differences between

weighted grouped streamflow signature values. Section A.3.2 shows the results of individual signature calculations.

Section A.4 shows the geographical spread of GRDC catchments, sorted by their main clusters.

A.2   GRDC catchment and streamflow data

Out of 1182 available stations in the GRDC Pristine Basins dataset, 64 are excluded because no boundary data is available, and the catchment area exceeds a size threshold (section A.2.1). Table A.1 shows the results of a manual quality assurance procedure, which leads us to exclude another 15 catchments whose data are implausible for pristine catchments (section A.2.2). Section A.2.3 shows the length of the flow record for the 1103 catchments that are used in this study.

A.2.1   Area threshold used to exclude GRDC catchments for which the catchment-averaged climate can not be calculated by using catchment boundaries

Information on catchment boundaries is not available for all catchments in the GRDC data set. If boundaries are available, we use this information to calculate a catchment-averaged climate. Where boundaries are not available, but the catchment is small, we can use the climate at location of the catchment's gauge as representative of the climate in the entire catchment. This section describes the procedure used to determine the catchment size threshold above which we exclude the catchment from use in this study. The GRDC data set provides the location of each catchment's gauge [latitude and longitude coordinates] and the size of the catchment [km^2].

First, we use the square root of catchment area to find the approximate catchment length (Figure A.1). The majority of catchments for which no boundary information is available, have an approximated length smaller than the length of 1 grid cell as used in the climate data, and for these no action is necessary. Next, we calculate the correlation length of our three climate indices in the latitude and longitude direction (Figure A.2). If a catchment's length is too large compared to the distance until which the climate indices are correlated, we remove the catchment from the data set. Some subjectivity is involved in choosing a threshold level. We need to balance the total number of catchments we can use (which favours keeping larger catchments) and ensuring climatic consistency within the catchment (which favours keeping only small catchments). We have chosen a threshold
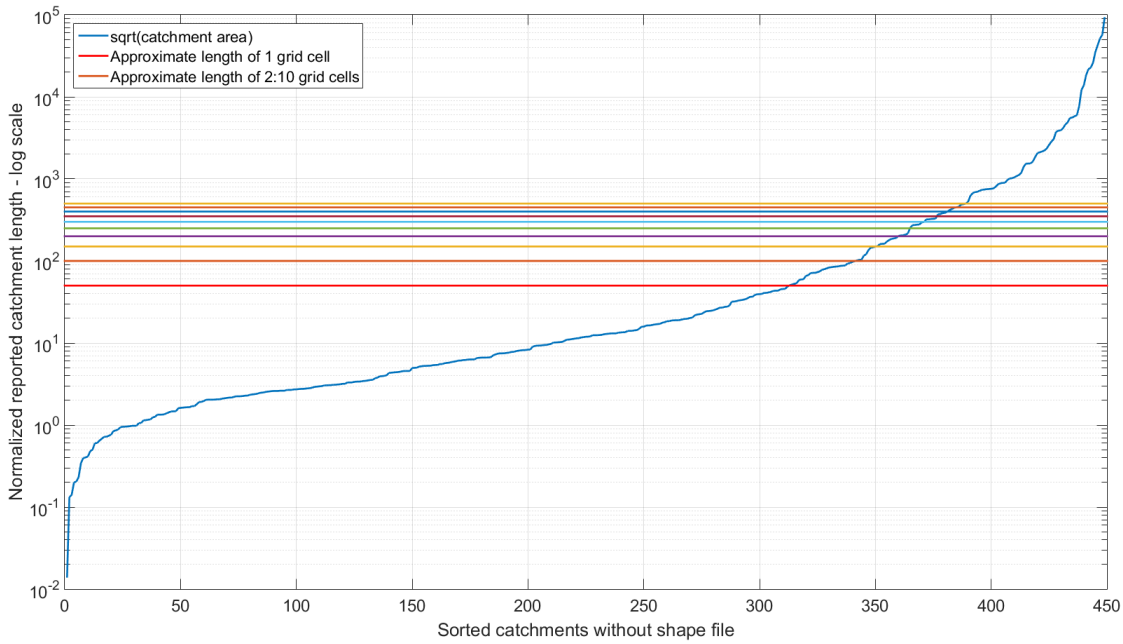
FIGURE A.1. GRDC Pristine Basins without information on catchment boundaries, sorted by approximated catchment length (square root of catchment area). Coloured lines indicated the approximated length of 1 to 10 grid cells as used in the CRU TS climate data. 312 GRDC Pristine Basins have areas smaller than 1 grid cell.

length of 3 grid cells as an appropriate middle ground. Catchments with an approximated length larger than 3 grid cells (approximately 150 km) are removed from further analysis.

A.2.2 Quality control of flow data

All 1182 Global Runoff Data Centre locations in the "pristine basins" data set have been visually inspected for data errors. Table A.1 shows 56 catchments that warranted further investigation and the result of this investigation. We exclude 5 stations due to doubtful data quality, another 5 stations due evidence of hydro power dam construction during the study period, 3 stations due to missing catchment area values and 2 stations due to highly implausible catchment area values. Two stations are suspected of consistent underestimation of flows during part of the time series, and data were adjusted to better fit the remainder of the time series (Figure A.4). For a further 19 stations we removed part of the time series due to measurement errors (e.g. inexplicable jumps in flow at the start of a new month/year, Figures A.5-A.9). Figure A.3 summarizes the results of the quality
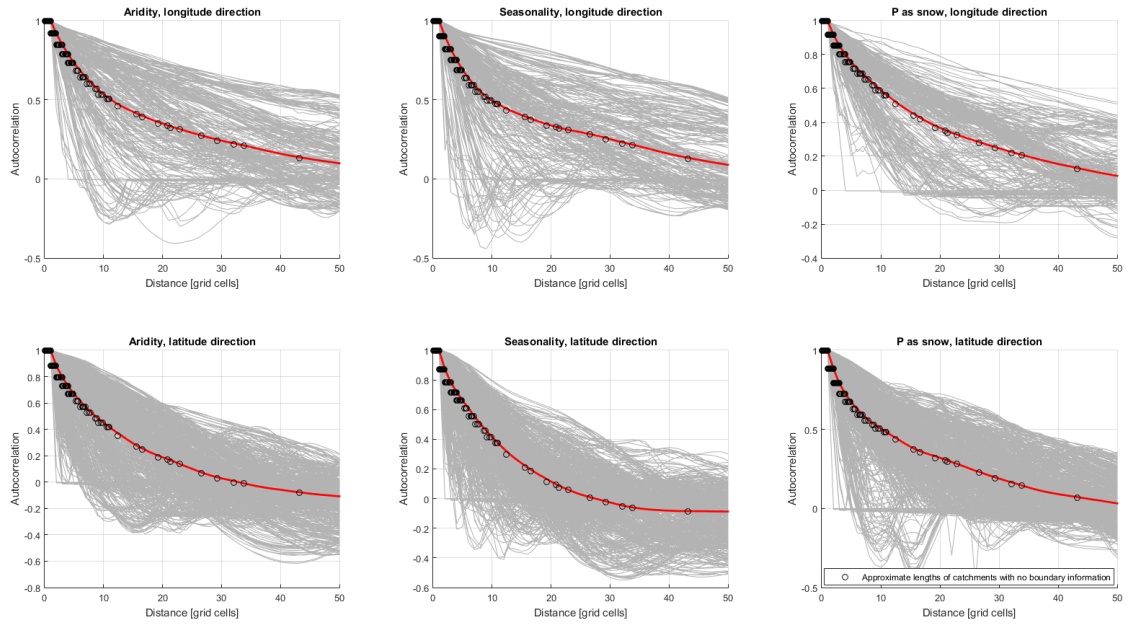
FIGURE A.2. Autocorrelation lengths in longitude (top) and latitude (bottom) direction for three climate indices: average annual aridity (left), aridity seasonality (middle) and fraction of precipitation as snow (right). Circles are the approximate lengths of all GRDC Pristine Basins for which no boundary information is available, matched up with the mean autocorrelation per grid cell distance (red line).

assurance procedure.
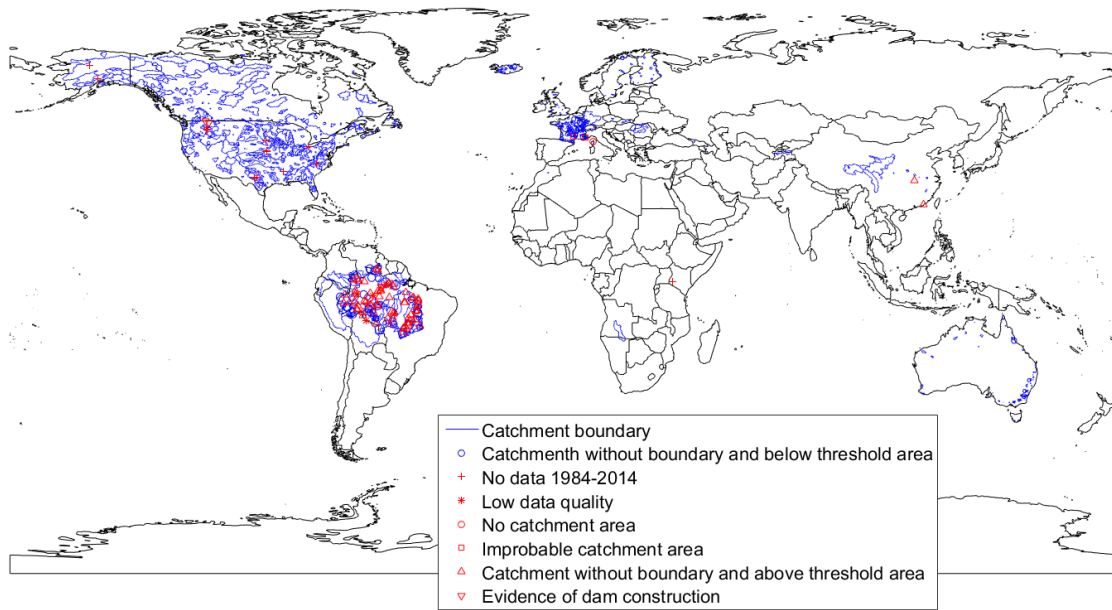
FIGURE A.3. Location and boundaries (where available) of GRDC Pristine Basins used in this study (blue) and Pristine Basins that are removed from further analysis for various reasons (red). Catchments for which no boundary data is available are used if their approximated length is smaller or equal to a climate correlation threshold and removed from the analysis if larger. The correlation threshold length is set at 3 times the size of a grid cell on which climate data is available, i.e. 0.5° (see section A.2.1).
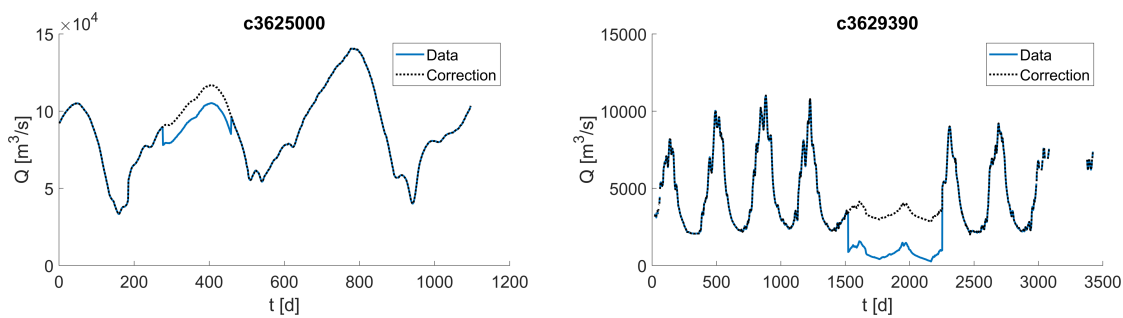
FIGURE A.4. Correction of flow underestimation for two GRDC catchments (see Table A.1 for details).
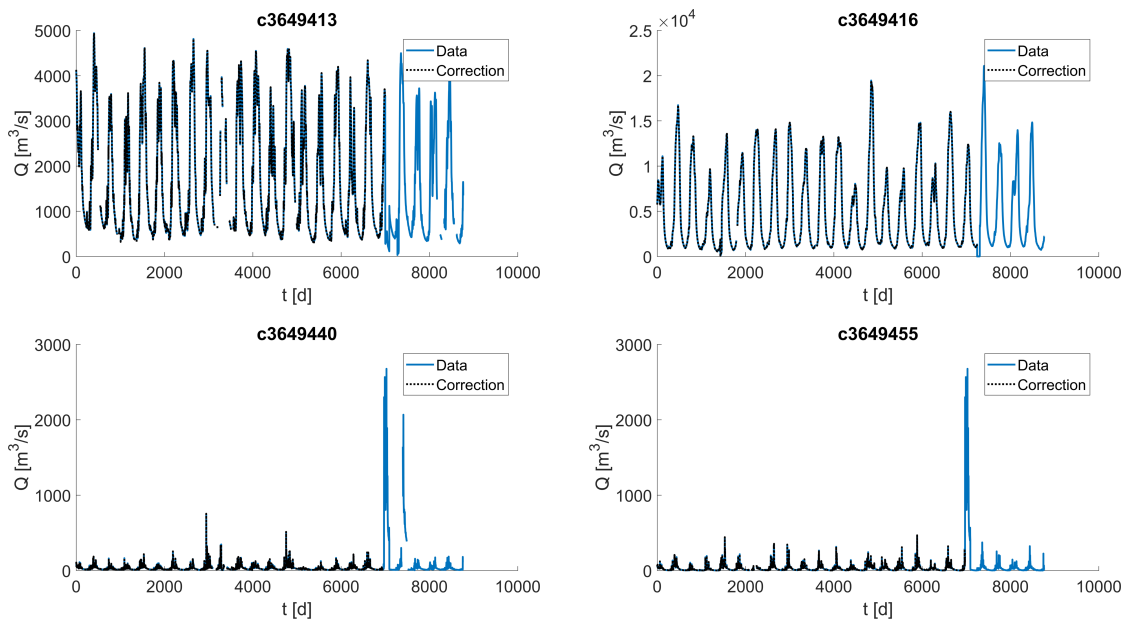
FIGURE A.5. Correction of flow records for GRDC catchments (see Table A.1 for details).
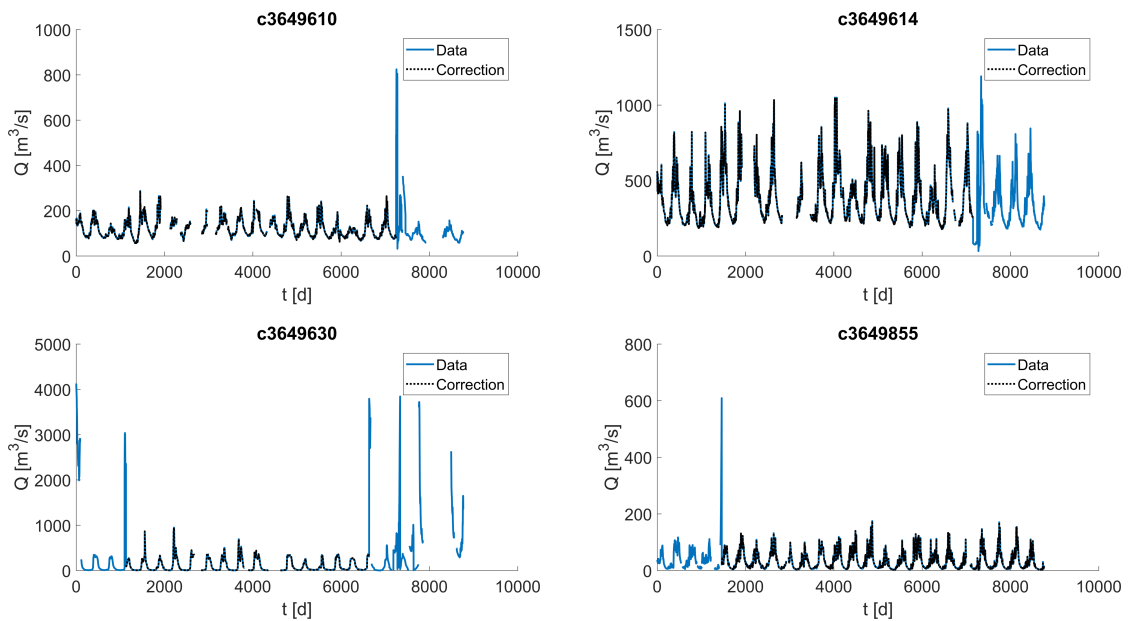
FIGURE A.6. Correction of flow records for GRDC catchments (see Table A.1 for details).
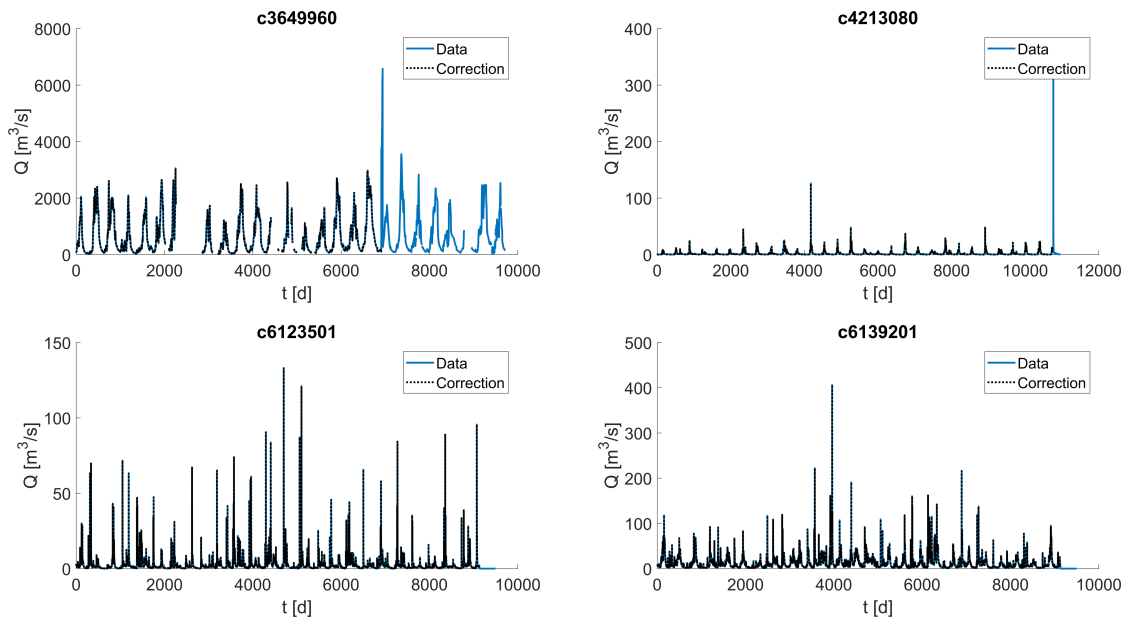
FIGURE A.7. Correction of flow records for GRDC catchments (see Table A.1 for details).
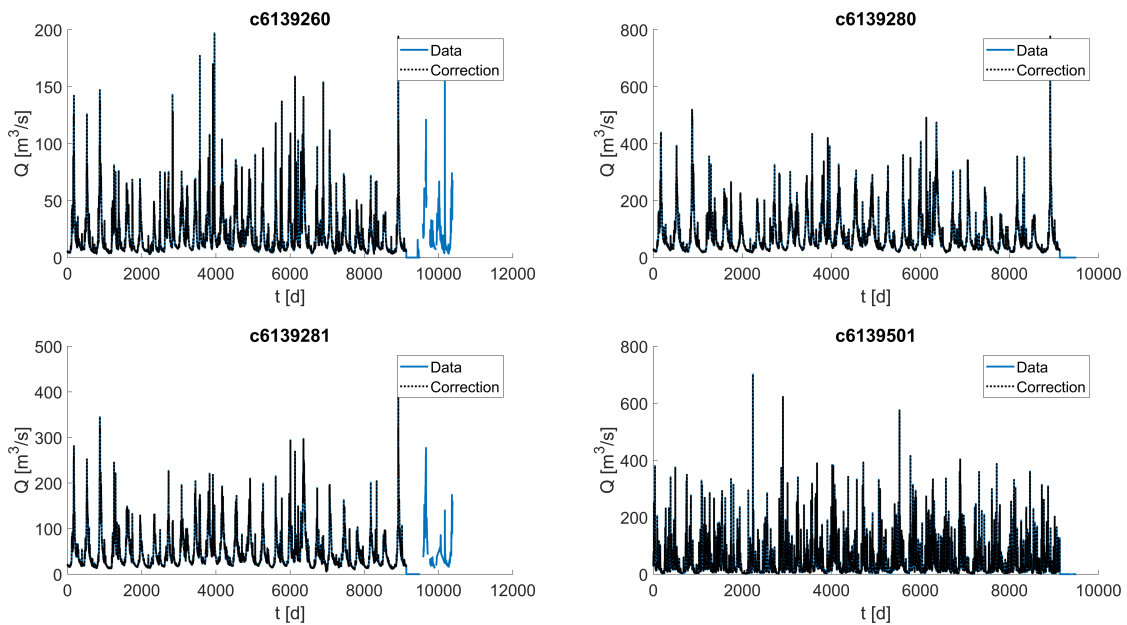
FIGURE A.8. Correction of flow records for GRDC catchments (see Table A.1 for details).
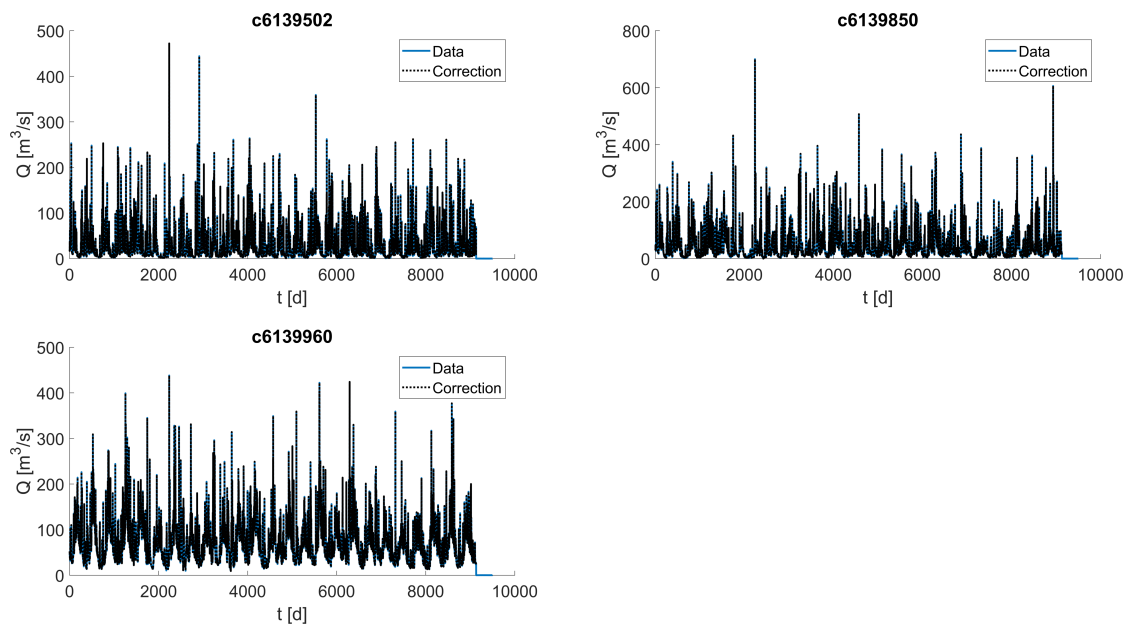
FIGURE A.9. Correction of flow records for GRDC catchments (see Table A.1 for details).

Table A.1: GRDC pristine basins that have been investigated further, based on visual inspection of data. Locations with extreme outlier-like peaks have been kept included if additional sources confirmed those peaks to be floods. Extreme peaks with no outside confirmation are treated as data errors. (Table continues on next page.)

Catchment ID	Excluded?	Reason
c3624250	Yes	Average flow during first 10 years is 400 m3/s, then a 2-year hiatus in measurements, average flow during last 10 years is 2000 m3/s
c3625000	No, corrected	Amazon. Days 277:458 are 0.6mm/d lower than rest of record
c3625310	Yes	Amazon, flows during middle 13 years are factor 5-10 lower than other years
c3627811	Yes	Bolivia. Very short time series with unexplained drops to zero
c3628300	Yes	Amazon. Unexplained increase in flow variability for middle 13 years
c3628400	Yes	Amazon. Flow regime changes drastically due to hydro power dam construction
c3628401	Yes	Amazon. Flow regime changes drastically due to hydro power dam construction
c3629390	No, corrected	Amazon. Days 1523:2253 are 1.2mm/d lower than rest of record
c3629800	Yes	Amazon. Unexplained flow regime changes in middle of data
c3649413	No, corrected	Brazil. Sudden drop in flow values, coinciding with start of new month. Possible procedure error. Removed 7000:end
c3649416	No, corrected	Brazil. Sudden drop in flow values, coinciding with start of new month. Possible procedure error. Removed days 7245:end
c3649440	No, corrected	Amazon. Sharp decrease in data quality towards end of series. Removed days 6491:end
c3649455	No, corrected	Amazon. Sharp decrease in data quality towards end of series. Also mentions dam construction around that period. Removed days 6972:end
c3649461	Yes	No catchment area
c3649465	Yes	No catchment area
c3649610	No, corrected	Amazon. Sharp decrease in data quality towards end of series. Removed days 7245:end
c3649614	No, corrected	Amazon. Sharp decrease in data quality towards end of series. Removed days 7153:end
c3649630	No, corrected	Amazon. Wonky data quality at beginning and end of series. Removed 1:1128 and days 6635:end
c3649855	No, corrected	Amazon. Sharp decrease in data quality towards beginning of series. Removed suspect data

Table A.1: Summary of GRDC flow data quality assurance - part 2. (Table continues on next page.)

Catchment ID	Excluded?	Reason
c3649960	No, corrected	Amazon river. Sharp decrease in data quality towards end of series. Removed days 6910:end
c4102450	No	Alaska. Shows 3 peaks that look like outliers in Oct-1986, Aug-2006 and Sep-2012. News reports confirm floods at those dates
c4115210	No	Washington State. Shows peak that looks like outlier in Feb-1996. News report confirms a flood.
c4115320	Yes	Montana. Subject to heavy dam construction. Streamflow record changes drastically for this site around 1970
c4115321	Yes	Idaho. Subject to heavy dam construction (Libby dam, 1972). Streamflow characteristics change during flow series
c4115322	Yes	Idaho. Subject to heavy dam construction (Libby dam, 1972). Streamflow characteristics change during flow series
c4118100	No	Nevada. Shows a peak that looks like an outlier. USGS fact sheet confirms a flood in Jan-1997
c4118105	No	Nevada. Shows a peak that looks like an outlier. USGS fact sheet confirms a flood in Jan-1997
c4119265	No	Idaho. Shows outlier-like peak. Various sources confirm a flood event in Jun-2008
c4123255	No	Ohio. Shows outlier-like peak. Various sources confirm extreme flood in Mar-1997
c4126850	No	Texas. Peaks are sudden and high. Typical ephemeral stream
c4146230	No	California. Outlier-like peaks. Typical ephemeral stream
c4146380	No	California. End of data looks higher than rest of series but no reason to assume errors
c4148070	No	Virginia, near NC. Outlier-like peak. No sources confirm flood but proximity to NC and matching flood dates imply hurricane-related flood here
c4148110	No	North Carolina. Outlier-like peak. Various sources confirm Sep-1999 flood due to hurricane
c4148125	No	North Carolina. Outlier-like peak. Various sources confirm Sep-1999 flood due to hurricane
c4148850	No	Florida. Low flow variability is high (quick changes), peaks look slower than normal rising limbs. No reason to assume errors, historical data from 1933-2017 looks the same
c4149405	No	Alabama. Outlier-like peak. News report confirms flood in May-2003

Table A.1: Summary of GRDC flow data quality assurance - part 3. (Table continues on next page.)

Catchment ID	Excluded?	Reason
c4149411	No	Mississippi.Data looks spiky but no reason to assume errors
c4149420	No	Florida. Outlier-like peak. Weather source confirms flood in Oct-1998
c4149510	No	Florida. Outlier-like peak. Weather source confirms flood in Oct-1998
c4150310	No	Texas. Outlier-like peak. News report confirms extreme flood in Oct-1998
c4207750	No	British Columbia. Outlier-like peak. News report confirms extreme flood in Oct-2003
c4213080	No, corrected	Alberta. Outlier-like peak near end of data. No confirmation found of extreme rain or flow, removed outlier
c5202057	No	New South Wales. Shows outlier-like peaks. No confirmation of extreme flow but confirmation of extreme rain in the approximate area
c5202065	No	New South Wales. Shows outlier-like peaks. No confirmation of extreme flow but confirmation of extreme rain in the approximate area
c6123501	No, corrected	France. Flows drop to (nearly) zero towards end of data, coinciding with start of new year. Possible procedure error. Removed 9133:end
c6125360	Yes	Catchment area too small, results in unrealistic flows
c6128702	Yes	Catchment area too small, results in unrealistic flows
c6139201	No, corrected	France. Flows drop to (nearly) zero towards end of data, coinciding with start of new year. Possible procedure error. Removed days 9133:end
c6139260	No, corrected	France. Flows drop to (nearly) zero towards end of data, coinciding with start of new year. Possible procedure error. Removed days 9133:end
c6139280	No, corrected	France. Flows drop to (nearly) zero towards end of data, coinciding with start of new year. Possible procedure error. Removed days 9133:end
c6139281	No, corrected	France. Flows drop to (nearly) zero towards end of data, coinciding with start of new year. Possible procedure error. Removed days 9133:end
c6139501	No, corrected	France. Flows drop to (nearly) zero towards end of data, coinciding with start of new year. Possible procedure error. Removed days 9133:end
c6139502	No, corrected	France. Flows drop to (nearly) zero towards end of data, coinciding with start of new year. Possible procedure error. Removed days 9133:end

Table A.1: Summary of GRDC flow data quality assurance - part 4.

Catchment ID	Excluded?	Reason
c6139850	No, corrected	France. Flows drop to (nearly) zero towards end of data, coinciding with start of new year. Possible procedure error. Removed days 9133:end
c6139960	No, corrected	France. Flows drop to (nearly) zero towards end of data, coinciding with start of new year. Possible procedure error. Removed days 9133:end
c6150300	Yes	No catchment area

A.2.3 Flow record length

Flow records within the GRDC Pristine Basins data set need to cover a minimum of 20 years, but these do not necessarily overlap with the study period of 1984-2014. Figure A.10 shows a histogram of the number of years within the period 1984-2014 available. 1041 catchments (94.3%) used in this study have data records longer than 20 years.
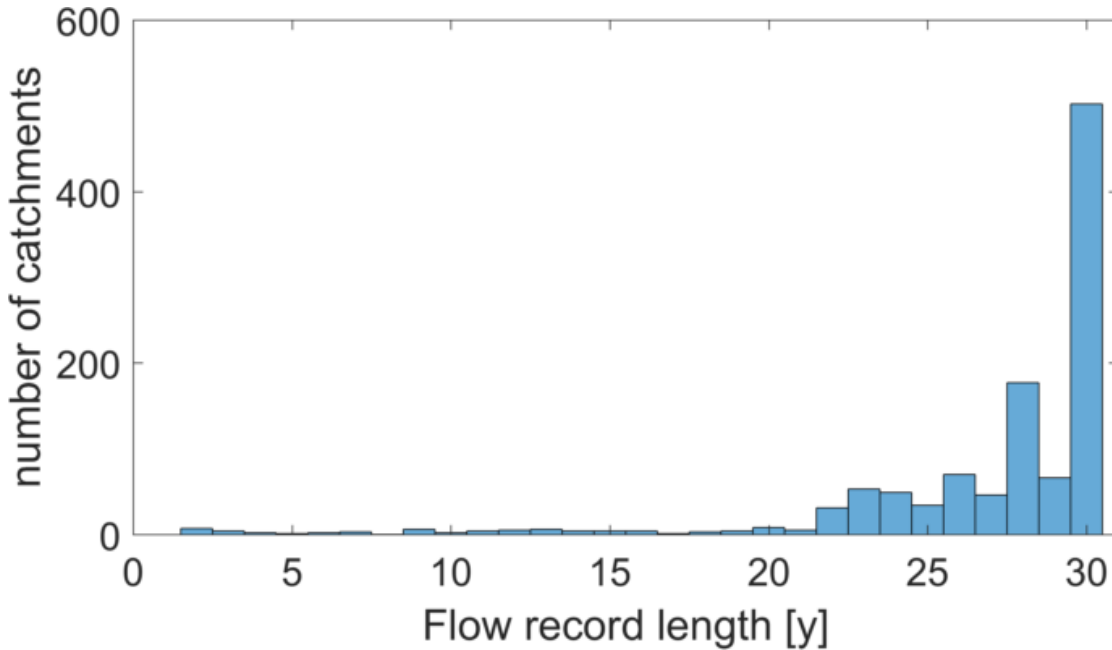
FIGURE A.10. Overview of record length of flow data for the catchments in the GRDC Pristine River basins data set.

A.3 Significance testing and signature values

This section gives details about the empirical Wilcoxon test used (section A.3.1), streamflow signature values for 16 signatures, based on catchment grouping by representative climate and by Köppen-Geiger class (section A.3.2) and the statistical difference in signature values per representative climate and Köppen-Geiger class (Section A.3.3).

A.3.1 Empirical Wilcoxon statistical test

Context

- We have calculated 16 signature values for 1103 catchments;

- We have determined the membership of each catchment to each of the 18 possible climate clusters;

- We want to know if there is a statistical difference between the signatures values that are associated with each cluster.

Problem

- We need a non-parametric method because we are unsure of the distributions that make up the signature values per climate cluster (i.e. we do not know if they are normal, so the t-test should not be used);

- The Wilcoxon rank test is a suitable test but can only be used in cases where the samples are unweighed (i.e. all samples belong for 100% to their respective class). The test thus needs to be adapted.

Wilcoxon procedure

- With the Wilcoxon test, we are testing the null hypothesis $\mu_1 = \mu_2$ against a suitable alternative [336], with μ_1 being the mean of all observations $x_{i=1:m}$ and μ_2 being the mean of all observations $y_{j=1:n}$;

- The unmodified Wilcoxon test consists of pair-wise comparison of all observations in x_i and y_j. If $x_i > y_j$, a new variable U_x is increased by 1, if $x_i < y_j$, U_y is increased by 1. After all pairs have been compared the p-value for obtaining a given $U = min(U_x, U_y)$ can be obtained from tables or otherwise, which gives a measure of the likelihood of obtaining a given U value under the null hypothesis. If the p-value is below a certain critical threshold, the null hypothesis can be rejected.

Solution

- Modify the Wilcoxon test to account for weighted observations and use bootstrapping to empirically determine the null distribution;

- In this case μ_1 refers to the mean of signatures values belonging to climate cluster 1, consisting of $x_{i=1:m}$ observations with accompanying weights $w_{x,i=1:m}$. Idem for μ_2. During pair-wise comparison, U_x is increased by $w_{x,i} * w_{y,j}$, if $x_i > y_j$. Otherwise, $w_{x,i} * w_{y,j}$ is added to U_y. $U = min(U_x, U_y)$ can now be tested against an empirical null distribution;

- To create the empirical null distribution, x_i and y_j are pooled together into a single sample, from which two new uniform random samples are drawn, with replacement,

equal in size to x_i and y_j. A U-value is then determined from pairwise comparison of both new samples. This process is repeated N times to form an empirical approximation of the null distribution. N is commonly set at 999 [87];

- Now the U-value of observations can be compared to the empirical distribution of U-values under the null hypothesis, using $p = \frac{r+1}{N+1}$, where N is the number of samples and r is the number of samples that have a U-value below the U-value calculated for the data [235].

Illustration of Wilcoxon test, adapted for fuzzy membership (Figure A.11)

- Distribution of weighted x_i and y_j as observed in the data. The y-axis is meaningless and only used to visualise x_i and y_j better. The U-value for these samples is 204.4;

- Uniform random sampling with replacement of x_i and y_j, from a pool made up of x_i and y_j. Visually, the samples are similar to one another, and very different from the pattern seen in Figure A.11a. The U-value is 697.7;

- The average U-value after N-samples levels out at $U \approx 840$ and $N = 999$ seems sufficient;

- Visually comparing the U-value calculated from data with the empirical U-value distribution shows that the null hypothesis can quite probably be rejected. The test statistic $p = \frac{r+1}{N+1} = \frac{0+1}{999+1} = 0.001$ confirms this.

A.3.2   Signature values

This section contains plots of the average signature value in each climate cluster (Figure A.12) and Köppen-Geiger class (Figure A.13). Signature values are calculated using daily flow data for each hydrological year available per catchment. This gives up to 30 annual values per catchment (depending on number of available hydrological years), from which the average annual signature value is calculated. Catchment membership to each climate cluster is used to create a weighted average signature value per climate cluster.

A.3.3   Statistical difference between climate groups/classes per streamflow signature

This section shows the results of significance testing per signature, with catchments grouped by representative climate (significance tested with empirical Wilcoxon test, Figure A.14) and Köppen-Geiger class (significance tested with regular Wilcoxon test, Figure A.15).
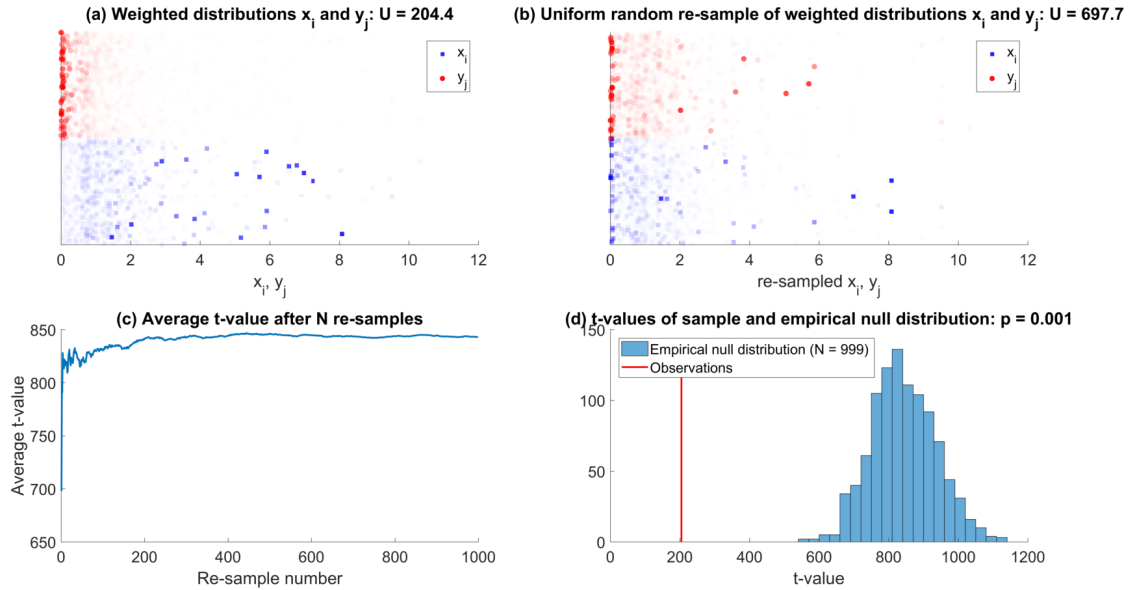
FIGURE A.11. Example of empirical Wilcoxon statistical test. (a) Distribution of observed values of the average_flow signature for representative climate 16 (x_i, very wet climate, low seasonality, no frost) and climate 1 (y_j, very dry climate, low seasonality, no frost). Data points correspond to all 1103 catchments with their transparency dependent on the weight with which they belong to each cluster. The low U-value indicates a large difference in ranks between both samples, which can also be seen from the concentration of high weight values for both climates. (b) Uniform random re-sampling of x_i and y_j, from a sample containing both. The re-sampled distributions are more similar (spread of height weights) and this leads to a higher U-value. (c) The average U-value of the re-sampled distribution after N samples. (d) The resulting distribution of re-sampled U-values (bars) versus the U-value of observations (red line). The observed U-value is very different from the re-sampled distribution, leading to a low empirical p-value and thus a high confidence that the H0 hypothesis $\mu_{xi} = \mu_{yj}$ should be rejected.

A.4 Geographical spread of GRDC catchments per main climate cluster

This section shows the geographical spread of GRDC catchments in the context of their main representative climate, showing that nearly all representative climates contain flow records from at least two continents (Figure A.16).
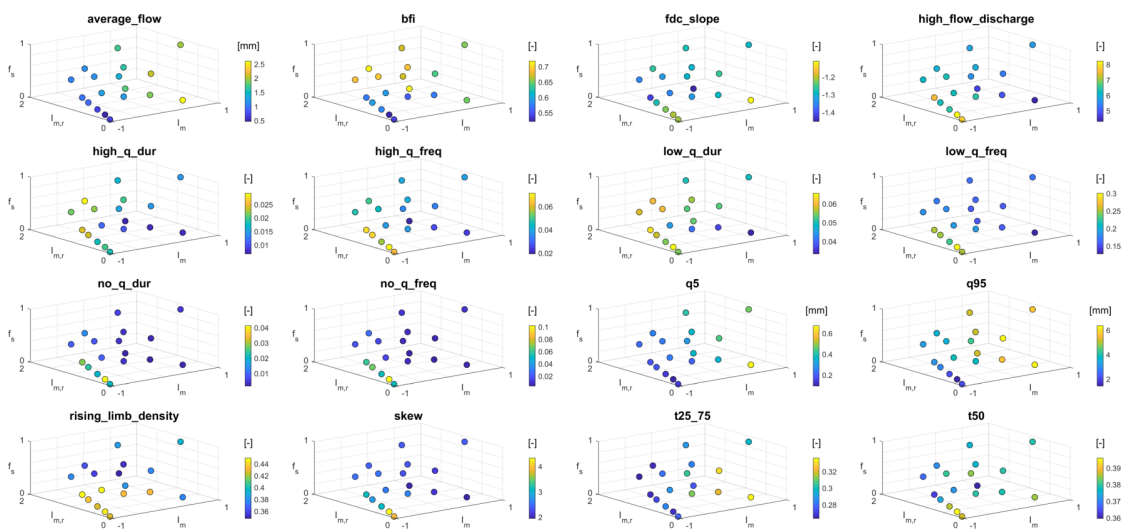
FIGURE A.12. Weighted mean signature values per climate cluster, plotted at the climate cluster centroid location. Values are the weighted mean of average annual signature values per catchment.
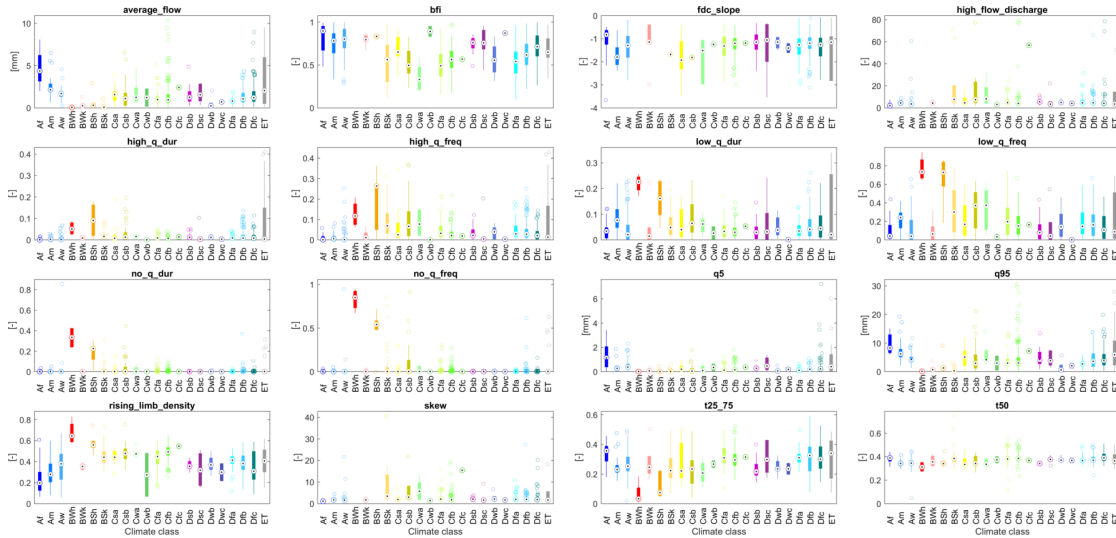
FIGURE A.13. Distribution of the mean of annual signature values per GRDC catchment within each Köppen-Geiger climate. Climate classes without flow values are not shown.
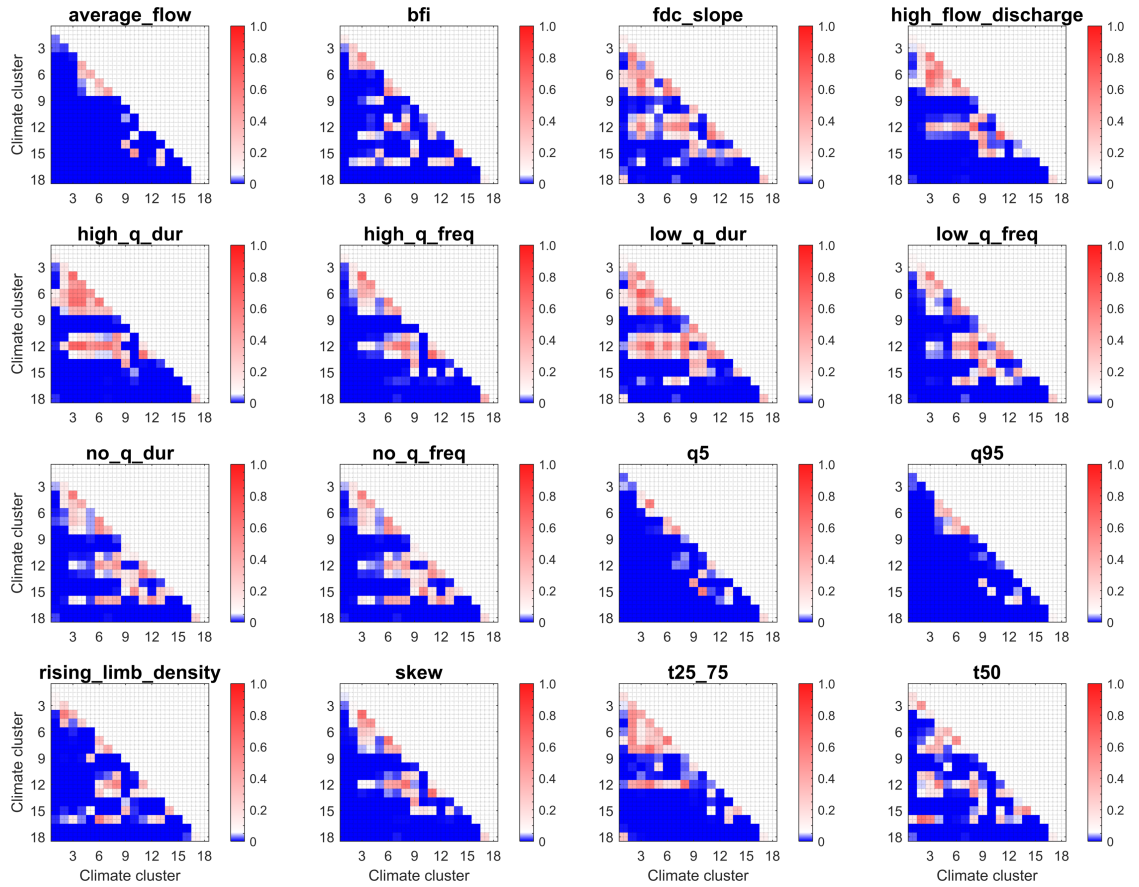
FIGURE A.14. Empirical p-values of differences between annual mean signatures values per climate cluster.
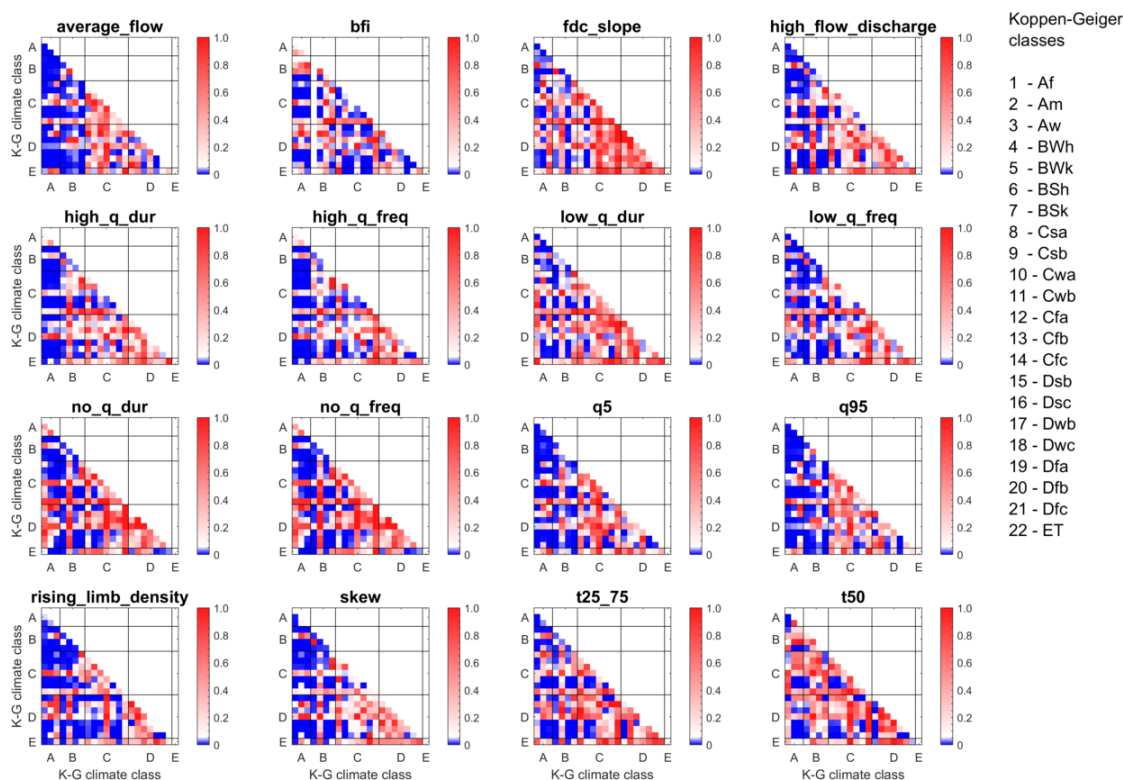
FIGURE A.15. p-values of differences between annual mean signatures values
        per Koppen-Geiger climate class
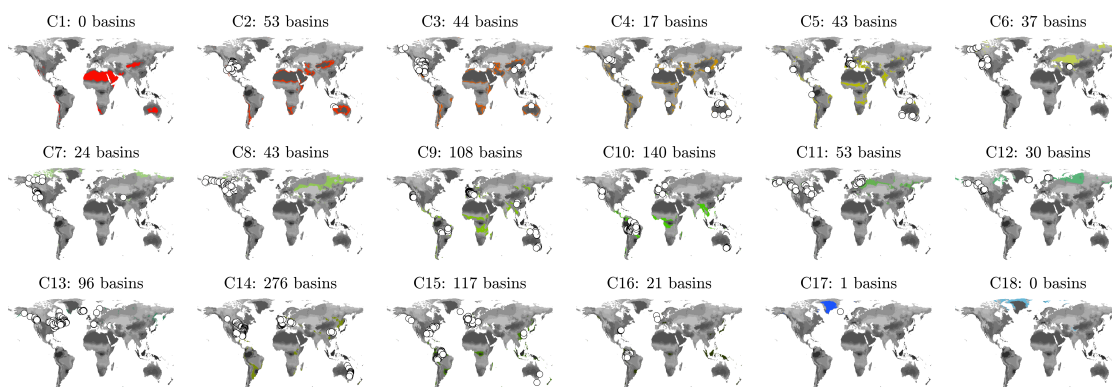
FIGURE A.16. GRDC catchments (white dots) per main climate cluster (cluster
        for which each catchment has the highest membership degree). Catchment-
        averaged climate is used to determine membership degrees where catchment
        boundary data is available (718 catchments), the climate at the outlet location
        is used in the remaining cases (385 catchments).

This appendix has been submitted as Supporting Materials for a Model description paper in Geoscientific Model Development. Slight modifications have been made to better fit the general layout of this thesis.

Citation: Knoben, W. J. M., Woods, R. A., & Freer, J. E. (2018). Supporting Material to "Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT) v1.0: an open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous space-state formulations". Geoscientific Model Development, under review. https://doi.org/10.5194/gmd-2018-332

B.1 Introduction

These Supporting Materials contain documentation for various parts of the MARRMoT software. Section B.2 contains model desciptions for the 46 conceptual models included in MARRMoT. Section B.3 shows how the constitutive functions of each model are translated into Matlab code, and which models use which of the resulting flux functions. Section B.4 shows how 7 different Unit Hydrograph approaches are coded in MARRMoT and which models use these. Section B.5 shows an overview of generalized parameter ranges for the 46 models.

B.2 Model descriptions

This section contains mathematical descriptions of all models that are included in the Modular Assessment of Rainfall-Runoff Models Toolbox v1.0 (MARRMoT). All descriptions follow the same layout (see the example model at the end of this section):

- Title: gives an informal name for the model structure followed by a unique ID;

- Introduction: gives a brief description of the model, including one or more original reference(s), the number of stores and parameters, a list containing parameter names and occasionally note-worthy deviations from the original model;

- Process list: a brief overview of the main processes the model is intended to represent;

- Figure: a wiring diagram that shows the names of model stores and fluxes;

- Matlab name section: gives the name of the file that contains Matlab code for this model;

- Model equations section: a mathematical description of the model. This uses Ordinary Differential Equations (ODEs) to describe the changes in model storage(s) and constitutive functions that detail how individual fluxes operate.

MARRMoT models intend to stay close to the original models they are based on but differences are unavoidable. We strongly recommend users to read the original paper cited for each model as well as our interpretation given in this document. In many cases, more than one version of a model exists, but these are not always easily distinguishable. There is a certain degree of model name equifinility, where a single name is used to refer to various different version of the same base model. A good example is TOPMODEL, of which many variants exist based around the initial concept of topographic indices. MARRMoT models tend to be based on older rather than newer publications for any given model (to stay close to the "intended" model by the original author(s)) but our selection has been pragmatic to achieve greater variety in the available fluxes and model structures in MARRMoT. The description of each model lists the papers that form the basis of the MARRMoT version of that model.

MARRMoT is set up to work with arbitrary user-defined time step sizes for climate input data. For consistency of parameter values across different time step sizes, the internal dynamics of each model are specified using the base units $[mm]$ and $[d]$. The temporal resolution of climate data is converted to $[mm/d]$ within each model, and model output is converted back to the user-specified time step size. Internal fluxes in each MARRMoT

model use the base units and are in $[mm/d]$ and parameter values are specified in the base or derived units (e.g. $[d^{-1}]$ for time coefficients). These units are kept throughout this document.

The computational implementation of constitutive functions is given in section B.3 and Unit Hydrographs are specified in section B.4. Generalized parameter ranges for all models are given in section B.5.

Example model (model ID: nn)

The Example model (fig. B.1) is used in the MARRMoT User Manual to show how to create a new MARRMoT model from scratch [177]. It has 3 stores and 7 parameters (UZ_{max}, c_{rate}, p_{rate}, k_{lz}, α, k_g, d). The model aims to represent:

- Saturation excess from the upper zone;
- Two-way interaction between upper and lower zone through percolation and capillary rise;
- A split between fast subsurface flow and groundwater recharge from the lower zone;
- Slow runoff from the groundwater;
- Triangular routing of combined surface and subsurface flows.

File names

Model: m_nn_example_7p_3s

Parameter ranges: m_nn_example_7p_3s_parameter ranges

Model equations
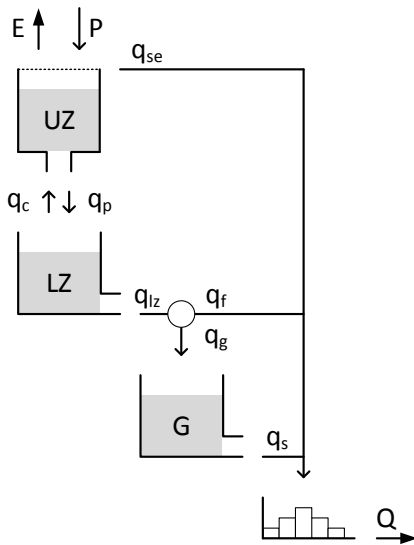
$$\frac{dUZ}{dt} = P + q_c - E - q_{se} - q_p \tag{B.1}$$

$$E = E_p * \frac{UZ}{UZ_{max}} \tag{B.2}$$

$$q_c = c_{rate}\left(1 - \frac{UZ}{UZ_{max}}\right) \tag{B.3}$$

$$q_{se} = \begin{cases} P, & \text{if } UZ = UZ_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.4}$$

$$q_p = p_{rate} \tag{B.5}$$

Figure B.1: Structure of the Example model

153

Where UZ [mm] is the current storage in the upper zone, refilled by precipitation P [mm/d] and capillary rise q_c [mm/d] and drained by evaporation E [mm/d], percolation q_p [mm/d] and saturation excess q_{se} [mm/d]. Evaporation occurs at the potential rate E_p scaled by the current storage in UZ compared to maximum storage UZ_{max} [mm]. Capillary rise occurs at a maximum rate c_{rate} [mm/d] if $UZ = 0$ and decreases linearly if not. Saturation excess flow only occurs when UZ is at maximum capacity. Percolation occurs at a constant rate p_{rate} [mm/d].

$$\frac{dLZ}{dt} = q_p - q_c - q_{lz} \tag{B.6}$$

$$q_{lz} = k_{lz} * LZ \tag{B.7}$$

Where LZ [mm] is the current storage in the lower zone, refilled by percolation q_p [mm/d] and drained by capillary rise q_c [mm/d] and outflow q_{lz} [mm/d]. Outflow has a linear relation with storage through time parameter k_{lz} [d^{-1}].

$$\frac{dG}{dt} = q_g - q_s \tag{B.8}$$

$$q_g = \alpha * q_{lz} \tag{B.9}$$

$$q_s = k_g * G \tag{B.10}$$

Where G [mm] is the current groundwater storage, refilled by recharge q_g [mm/d] and drained by slow flow q_s [mm/d]. Recharge is a fraction α [-] of outflow from the lower zone. Outflow has a linear relation with storage through time parameter k_g [d^{-1}]. Saturation excess q_{se}, interflow q_f and slow flow q_s are combined and routed with a triangular Unit Hydrograph with time base d [d] to give outflow Q [mm/d].

B.2.1   Collie River Basin 1 (model ID: 01)

The Collie River Basin 1 model (fig. B.2) is part of a top-down modelling exercise and is originally applied at the annual scale [158]. This is a classic bucket model. It has 1 store and 1 parameter (S_{max}). The model aims to represent:

- Evaporation from soil moisture;
- Saturation excess surface runoff.

B.2.1.1   File names

Model: m_01_collie1_1s_1p
Parameter ranges: m_01_collie1_1s_1p_parameter_ ranges

B.2.1.2   Model equations
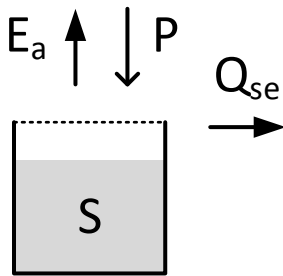
$$\frac{dS}{dt} = P - E_a - Q_{se} \tag{B.11}$$

$$Ea = \frac{S}{S_{max}} * Ep \tag{B.12}$$

$$Q_{se} = \begin{cases} P, & if\ S > S_{max} \\ 0, & otherwise \end{cases} \tag{B.13}$$

Figure B.2: Structure of the Collie River Basin 1 model

Where S [mm] is the current storage in the soil moisture and P the precipitation input [mm/d]. Actual evaporation E_a [mm/d] is estimated based on the current storage S, the maximum soil moisture storage S_{max} [mm], and the potential evapotranspiration E_p [mm/d]. Q_{se} [mm/d] is saturation excess overland flow.

B.2.2 Wetland model (model ID: 02)

The Wetland model (fig. B.3) is a conceptualization of the perceived dominant processes in a typical Western European wetland [274]. It belongs to a 3-part topography driven modelling exercise, together with a hillslope and plateau conceptualization. Each model is provided in isolation here, because they are well-suited for isolating specific model structure choices. It has 1 store and 4 parameters (D_w, $S_{w,max}$, β_w and K_w). The model aims to represent:

- Stylized interception by vegetation;
- Evaporation;
- Saturation excess runoff generated from a distribution of soil depths;
- A linear relation between storage and slow runoff.

B.2.2.1 File names

Model: m_02_wetland_4p_1s

Parameter ranges: m_02_wetland_4p_1s_parameter_ ranges

B.2.2.2 Model equations
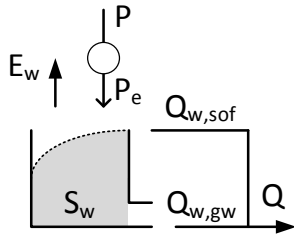
Figure B.3: Structure of the Wetland model

$$\frac{dS_w}{dt} = P_e - E_w - Q_{w,sof} - Q_{w,gw} \tag{B.14}$$

$$P_e = max(P - D_w, 0) \tag{B.15}$$

$$E_w = \begin{cases} E_p, & \text{if } S_w > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.16}$$

$$Q_{w,sof} = \left(1 - \left(1 - \frac{S_w}{S_{w,max}}\right)^{\beta_w}\right) * P_e \tag{B.17}$$

$$Q_{w,gw} = K_w * S_w \tag{B.18}$$

Where S_w is the current soil water storage [mm]. Incoming precipitation P [mm/d] is reduced by interception D_w [mm/d], which is assumed to evaporate before the next precipitation event. Evaporation from soil moisture E_w [mm/d] occurs at the potential rate E_p whenever possible. Saturation excess surface runoff $Q_{w,sof}$ [mm/d] depends on the fraction of the catchment that is currently saturated, expressed through parameters $S_{w,max}$ [mm] and β_w [-]. Groundwater flow $Q_{w,gw}$ [mm/d] depends linearly on current storage S_w through parameter K_w [d^{-1}]. Total flow:

$$Q = Q_{w,sof} + Q_{w,gw} \qquad\qquad \text{(B.19)}$$

B.2.3 Collie River Basin 2 (model ID: 03)

The Collie River Basin 2 model (fig. B.4) is part of a top-down modelling exercise and is originally applied at the monthly scale [158]. It has 1 store and 4 parameters (S_{max}, S_{fc}, a, M). The model aims to represent:

- Separate bare soil and vegetation evaporation;
- Saturation excess surface runoff;
- Subsurface runoff.

B.2.3.1 File names

Model: m_03_collie2_4p_2s
Parameter ranges: m_03_collie2_4p_2s_parameter_ ranges

B.2.3.2 Model equations
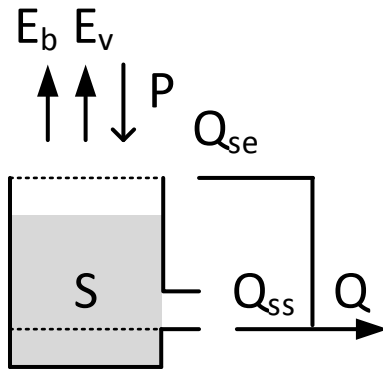
Figure B.4: Structure of the Collie River Basin 2 model

$$\frac{dS}{dt} = P - E_b - E_v - Q_{se} - Q_{ss} \tag{B.20}$$

$$Eb = \frac{S}{S_{max}}(1 - M) * Ep \tag{B.21}$$

$$Ev = \begin{cases} M * E_p, & if\ S > S_{fc} \\ \frac{S}{S_{fc}} * M * E_p, & otherwise \end{cases} \tag{B.22}$$

$$Q_{se} = \begin{cases} P, & if\ S > S_{max} \\ 0, & otherwise \end{cases} \tag{B.23}$$

$$Q_{ss} = \begin{cases} a * (S - Sfc), & if\ S > S_{fc} \\ 0, & otherwise \end{cases} \tag{B.24}$$

Where S [mm] is the current storage in the soil moisture and P [mm/d] the precipitation input. Actual evaporation is split between bare soil evaporation E_b [mm/d] and transpiration through vegetation E_v [mm/d], controlled through the forest fraction M [-]. The evaporation estimates are based on the current storage S, the potential evapotranspiration E_p [mm/d], maximum soil moisture storage S_{max} [mm] and field capacity S_{fc} [mm] respectively. Q_{se} [mm/d] is saturation excess overland flow. Q_{ss} [mm/d] is subsurface flow regulated by runoff coefficient a [d^{-1}]. Total flow:

$$Q = Q_{se} + Q_{ss} \tag{B.25}$$

B.2.4 New Zealand model v1 (model ID: 04)

The New Zealand model v1 (fig. B.5) is part of a top-down modelling exercise that focusses on several catchments in New Zealand [19]. It has 1 store and 6 parameters (S_{max}, S_{fc}, M, a, b and $t_{c,bf}$). The model aims to represent:

- Separate vegetation and bare soil evaporation;
- Saturation excess overland flow;
- Subsurface runoff when soil moisture exceeds field capacity;
- Baseflow.

B.2.4.1 File names

Model: m_04_newzealand1_6p_1s

Parameter ranges: m_04_newzealand1_6p_1s_parameter_ ranges

B.2.4.2 Model equations
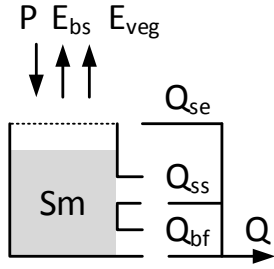
Figure B.5: Structure of the New Zealand model v1

$$\frac{dS_m}{dt} = P - E_{veg} - E_{bs} - Q_{se} - Q_{ss} - Q_{bf} \quad \text{(B.26)}$$

$$E_{veg} = \begin{cases} M * E_p, & \text{if } S > S_{fc} \\ \frac{S_m}{S_{fc}} * M * E_p, & \text{otherwise} \end{cases} \quad \text{(B.27)}$$

$$E_{bs} = \frac{S}{S_{max}} (1 - M) * E_p \quad \text{(B.28)}$$

$$Q_{se} = \begin{cases} P, & \text{if } S \geq S_{max} \\ 0, & \text{otherwise} \end{cases} \quad \text{(B.29)}$$

$$Q_{ss} = \begin{cases} \left(a * (S - S_{fc})\right)^b, & \text{if } S \geq S_{fc} \\ 0, & \text{otherwise} \end{cases} \quad \text{(B.30)}$$

$$Q_{bf} = t_{c,bf} * S \quad \text{(B.31)}$$

Where S_m [mm] is the current soil moisture storage which gets replenished through precipitation P [mm/d]. Evaporation through vegetation E_{veg} [mm/d] depens on the forest fraction M [-] and field capacity S_{fc} [-]. E_{bs} [mm/d] represents bare soil evaporation. When S exceeds the maximum storage S_{max} [mm], water leaves the model as saturation excess runoff Q_{se}. If S exceeds field capacity S_{fc} [mm], subsurface runoff Q_{ss} [mm/d] is generated controlled by time parameter a [d^{-1}] and nonlinearity parameter b [-]. Q_{bf} represents baseflow controlled by time scale parameter $t_{c,bf}$ [d^{-1}]. Total runoff Q_t [mm/d] is:

$$Q_t = Q_{se} + Q_{ss} + Q_{bf} \qquad \text{(B.32)}$$

B.2.5 IHACRES (model ID: 05)

The IHACRES model (fig. B.6) as implemented here is a modification of the original equations [84, 199, 350], which explicitly account for the various fluxes in a step-wise order. Furthermore, IHACRES usually uses temperature as a proxy for potential evapotranspiration (E_p). Here it uses estimated E_p directly to be consistent with other models. The equations for E_a and U are set up following Croke & Jakeman [84], with the non-linearity in U based on Ye et al [350]. THis version thus uses a catchment moisture deficit formulation, rather than a catchment wetness index. Littlewood et al [199] recommends the two parallel routing functions. The model has 1 *deficit* store and 6 parameters (lp, d, p, α, τ_q, τ_s). The model aims to represent:

- Catchment deficit build-up
- Slow and fast routing of effective precipitation.

B.2.5.1 File names

Model: m_05_ihacres_6p_1s
Parameter ranges: m_05_ihacres_6p_1s_parameter_ ranges

B.2.5.2 Model equations
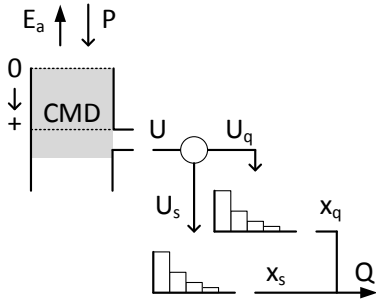
Figure B.6: Structure of the IHACRES model

$$\frac{dCMD}{dt} = -P + E_a + U \tag{B.33}$$

$$E_a = E_p * min\left(1, e^{2\left(1 - \frac{CMD}{lp}\right)}\right) \tag{B.34}$$

$$U = P\left(1 - min\left(1, \left(\frac{CMD}{d}\right)^p\right)\right) \tag{B.35}$$

$$U_q = \alpha * U \tag{B.36}$$

$$U_s = (1 - \alpha) * U \tag{B.37}$$

Where CMD is the current moisture deficit [mm], P [mm/d] the incoming precipitation that *reduces* the deficit, E_a [mm/d] evaporation that *increases* the deficit, and U [mm/d] the effective precipitation that occurs when the deficit is below a threshold d [mm].

Evaporation occurs at the potential rate E_p until the moisture deficit reaches wiliting point lp [mm], after which evaporation decreases exponentially with increasing deficit. Effective precipitation U equals incoming precipitation P when the deficit is zero, and

decreases as a linear fraction of P until moisture deficit is larger than a threshold d [mm], after which precipitation does not contribute to streamflow any longer. U is divided between fast and slow routing components based on fraction α [-]. Both routing schemes are exponentially decreasing over time with lags τ_q [d] and τ_s [d] respectively. Total flow:

$$Q = x_q + x_s \tag{B.38}$$

B.2.6   Alpine model v1 (model ID: 06)

The Alpine model v1 model (fig. B.7) is part of a top-down modelling exercise and represents a monthly water balance model [95]. It has 2 stores and 4 parameters (T_t, ddf, S_{max}, t_c). The model aims to represent:

- Snow accumulation and melt;
- Saturation excess overland flow;
- Linear subsurface runoff.

B.2.6.1   File names

Model: m_06_alpine1_4p_2s
Parameter ranges: m_06_alpine1_4p_2s_parameter_ ranges

B.2.6.2   Model equations
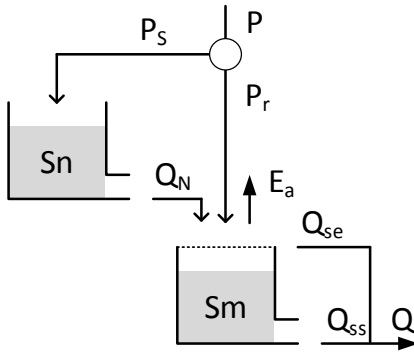
Figure B.7: Structure of the Alpine model v1

$$\frac{dSn}{dt} = P_s - Q_N \tag{B.39}$$

$$P_s = \begin{cases} P, & \text{if } T \le T_t \\ 0, & \text{otherwise} \end{cases} \tag{B.40}$$

$$Q_N = \begin{cases} ddf * (T - T_t), & \text{if } T \ge T_t \\ 0, & \text{otherwise} \end{cases} \tag{B.41}$$

Where S_N is the current snow storage [mm], P_s the precipitation that falls as snow [mm/d], Q_N snow melt [mm/d] based on a degree-day factor (ddf, [mm/°C/d]) and threshold temperature for snowfall and snowmelt (T_t, [°C]).

$$\frac{dS_m}{dt} = P_r + Q_N - E_a - Q_{se} - Q_{ss} \tag{B.42}$$

$$P_r = \begin{cases} P, & \text{if } T > T_t \\ 0, & \text{otherwise} \end{cases} \tag{B.43}$$

$$E_a = \begin{cases} E_p, & \text{if } S > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.44}$$

$$Q_{se} = \begin{cases} P_r + Q_N, & \text{if } S_m \geq S_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.45}$$

$$Q_{ss} = t_c * S_m \tag{B.46}$$

Where S_m [mm] is the current soil moisture storage, which is assumed to evaporate at the potential rate E_p [mm/d] when possible. When S_m exceeds the maximum storage S_{max} [mm], water leaves the model as saturation excess runoff Q_{se}. Q_{ss} represents subsurface flow controlled by time scale parameter t_c [d^{-1}]. Total runoff Q_t [mm/d] is:

$$Q_t = Q_{se} + Q_{ss} \tag{B.47}$$

B.2.7 GR4J (model ID: 07)

The GR4J model (fig. B.8) is originally developed with an explicit (operator-splitting) time-stepping scheme [255]. Recently a new version has been released that works with an implicit time-stepping scheme [271]. The implementation given here follows most of the equations from Santos et al [271], but uses the original Unit Hydrographs for flood routing given by Perrin et al [255]. It has 2 stores and 4 parameters (x_1, x_2, x_3, x_4). The model aims to represent:

- Implicit interception by vegetation, expressed as net precipitation or evaporation;
- Different time delays within the catchment expressed by two hydrographs;
- Water exchange with neighbouring catchments.

B.2.7.1 File names

Model: m_07_gr4j_4p_2s

Parameter ranges: m_07_gr4j_4p_2s_parameter_ ranges

B.2.7.2 Model equations
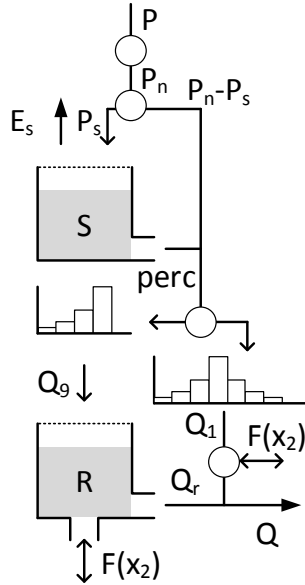
Figure B.8: Structure of the GR4J model

$$\frac{dS}{dt} = P_s - E_s - Perc \tag{B.48}$$

$$P_s = P_n * \left(1 - \left(\frac{S}{x1}\right)^2\right) \tag{B.49}$$

$$P_n = \begin{cases} P - Ep, & \text{if } P \geq Ep \\ 0, & \text{otherwise} \end{cases} \tag{B.50}$$

$$E_s = E_n * \left(2\frac{S}{x1} - \left(\frac{S}{x1}\right)^2\right) \tag{B.51}$$

$$E_n = \begin{cases} Ep - P, & \text{if } Ep > P \\ 0, & \text{otherwise} \end{cases} \tag{B.52}$$

$$Perc = \frac{x_1^{-4}}{4d} * \left(\frac{4}{9}\right)^{-4} S^5 \tag{B.53}$$

Where S is the current soil moisture storage [mm], P_s [mm/d] is the fraction of net precipitation P_n [mm/d] redirected to soil moisture, E_s [mm/d] is the fraction of net evaporation E_n [mm/d] subtracted from soil moisture, and $perc$ [mm/d] is percolation to deeper soil layers. Parameter x_1 [mm] is the maximum soil moisture storage.

Percolation *perc* and excess precipitation $P_n - P_s$ are divided into 90% groundwater flow, routed through a triangular routing scheme with time base x_4 [d], and 10% direct runoff, routed through a triangular routing scheme with time base $2x_4$ [d].

$$\frac{dR}{dt} = Q_9 + F(x_2) - Q_r \tag{B.54}$$

$$F(x_2) = x_2 * \left(\frac{R}{x_3}\right)^{3.5} \tag{B.55}$$

$$Q_r = \frac{x_3^{-4}}{4d} R^5 \tag{B.56}$$

Where R [mm] is the current storage in the routing store, $F(x_2)$ [mm/d] the catchment groundwater exchange, depending on exchange coefficient x_2 [mm/d] and the maximum routing capacity x_3 [mm], and Q_r [mm/d] routed flow. Total runoff Q_t [mm/d]:

$$Q_t = Q_r + max(Q_1 + F(x_2), 0) \tag{B.57}$$

167

B.2.8 United States model (model ID: 08)

The United States model (fig. B.9) is part of a multi-model comparison study using several catchments in the United States [20]. It has 2 stores and 5 parameters (α_{ei}, M, S_{max}, fc, α_{ss}). The model aims to represent:

- Interception as a percentage of precipitation;
- Separate unsaturated and saturated zones;
- Separate bare soil evaporation and vegetation transpiration;
- Saturation excess overland flow;
- Subsurface flow.

B.2.8.1 File names

Model: m_08_us1_5p_2s

Parameter ranges: m_08_us1_5p_2s_parameter_ ranges

B.2.8.2 Model equations
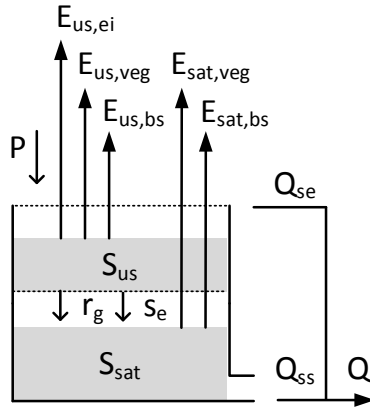
Figure B.9: Structure of the United States model

$$\frac{dS_{us}}{dt} = P - E_{us,ei} - E_{us,veg} - E_{us,bs} - r_g \tag{B.58}$$

$$E_{us,ei} = \alpha_{ei} * P \tag{B.59}$$

$$E_{us,veg} = \begin{cases} \frac{S_{us}}{S_{us}+S_{sat}} * M * E_p, & \text{if } S_{us} > S_{usfc} \\ \frac{S_{us}}{S_{us}+S_{sat}} * M * E_p * \frac{S_{us}}{S_{usfc}}, & \text{otherwise} \end{cases} \tag{B.60}$$

$$E_{us,bs} = \frac{S_{us}}{S_{us}+S_{sat}} * (1-M) * \frac{S_{us}}{S_{max}-S_{sat}} * E_p \tag{B.61}$$

$$r_g = \begin{cases} P, & \text{if } S_{us} > S_{usfc} \\ 0, & \text{otherwise} \end{cases} \tag{B.62}$$

$$S_e = \begin{cases} S_{us} - S_{usfc}, & if \ S_{us} > S_{usfc} \\ 0, & \text{otherwise} \end{cases} \tag{B.63}$$

$$S_{usfc} = fc * (S_{max} - S_{sat}) \tag{B.64}$$

Where S_{us} [mm] is the current storage in the unsaturated zone, $E_{us,ei}$ [mm/d] evaporation from interception, $E_{us,veg}$ [mm/d] transpiration through vegetation, $E_{us,bs}$ [mm/d]

bare soil evaporation and r_g [mm/d] drainage to the saturated zone. Interception evaporation relies on parameter α_{ei} [-], representing the fraction of precipitation P that is intercepted. The implicit assumption is that this evaporates before the next precipitation event. Transpiration uses forest fraction M [-], potential evapotranspiration E_p [mm/d] and the estimated field capacity S_{usfc} through parameter fc [-]. Bare soil evaporation relies also on the maximum soil moisture storage S_{max} [mm].

$$\frac{dS_{sat}}{dt} = r_g - E_{sat,veg} - E_{sat,bs} - Q_{se} - Q_{ss} \tag{B.65}$$

$$E_{sat,veg} = \frac{S_{sat}}{S_{max}} * M * E_p \tag{B.66}$$

$$E_{sat,bs} = \frac{S_{sat}}{S_{max}} * (1 - M) * E_p \tag{B.67}$$

$$Q_{se} = \begin{cases} r_g, & \text{if } S_{us} \geq S_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.68}$$

$$Q_{ss} = \alpha_{ss} * S_{sat} \tag{B.69}$$

Where S_{sat} [mm] is the current storage in the saturated zone, $E_{sat,veg}$ [mm/d] transpiration through vegetation, $E_{sat,bs}$ [mm/d] bare soil evaporation, Q_{se} [mm/d] saturation excess overland flow and Q_{ss} [mm/d] subsurface flow. Subsurface flow uses time parameter α_{ss} [d^{-1}] Total flow:

$$Q = Q_{se} + Q_{ss} \tag{B.70}$$

B.2.9   Susannah Brook model v1-5 (model ID: 09)

The Susannah Brook model v1-5 (fig. B.10) is part of a top-down modelling exercise designed to use auxiliary data [309]. It has 2 stores and 6 parameters (S_b, S_{fc}, M, a, b and r). The model aims to represent:

- Evaporation from soil and transpiration from vegetation;
- Saturation excess and non-linear subsurface flow;
- Groundwater recharge and baseflow.

B.2.9.1   File names

Model: m_09_susannah1_6p_2s
Parameter ranges: m_09_susannah1_6p_2s_parameter_ ranges

B.2.9.2   Model equations

Figure B.10: Structure of the Susannah Brook model v1-5
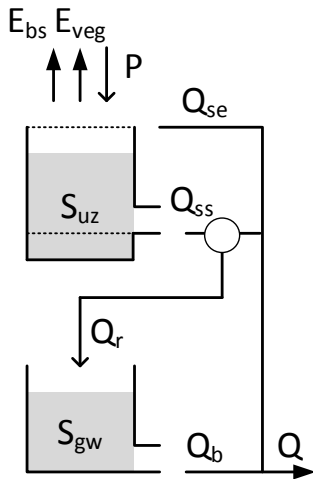
$$\frac{dS_{uz}}{dt} = P - E_{bs} - E_{veg} - Q_{se} - Q_{ss} \tag{B.71}$$

$$E_{bs} = \frac{S}{S_b}(1-M)E_p \tag{B.72}$$

$$E_{veg} = \begin{cases} M * E_p, & \text{if } S > S_{fc} \\ \frac{S}{S_{fc}} M * E_p, & \text{otherwise} \end{cases} \tag{B.73}$$

$$Q_{se} = \begin{cases} P, & \text{if } S \geq S_b \\ 0, & \text{otherwise} \end{cases} \tag{B.74}$$

$$Q_{ss} = \begin{cases} \left(\frac{S-S_{fc}}{a}\right)^{\frac{1}{b}}, & \text{if } S > S_{fc} \\ 0, & \text{otherwise} \end{cases} \tag{B.75}$$

Where S_{uz} is current storage in the upper zone [mm]. P [mm/d] is the precipitation input. E_{bs} is bare soil evaporation [mm/d] based on soil depth S_b [mm] and forest fraction M [-]. E_{veg} is transpiration from vegetation, using the wilting point S_{fc} [mm] and forest fraction M. Q_{se} is saturation excess flow [mm/d]. Q_{ss} is non-linear subsurface flow, using the wilting point S_{fc} as a threshold for flow generation and two flow parameters a [d] and b [-]. Q_r is groundwater recharge [mm/d].

$$\frac{DS_{gw}}{dt} = Q_r - Q_b \tag{B.76}$$

$$Q_r = r * Q_{ss} \tag{B.77}$$

$$Q_b = \left(\frac{1}{a}S_{gw}\right)^{\frac{1}{b}} \tag{B.78}$$

Where S_{gw} is the groundwater storage [mm], and Q_b the baseflow flux [mm/d]. Total flow [mm]:

$$Q = Q_{se} + (Q_{ss} - Q_r) + Q_b \tag{B.79}$$

B.2.10 Susannah Brook model v2 (model ID: 10)

The Susannah Brook model v2 model (fig. B.11) is part of a top-down modelling exercise designed to use auxiliary data [309]. It has 2 stores and 6 parameters (S_b, ϕ, fc, r, c, d). For consistency with other model formulations, S_b is is used as a parameter, instead of being broken down into its constitutive parts D and ϕ. The model aims to represent:

- Separation of saturated zone and a variable-size unsaturated zone;
- Evaporation from unsaturated and saturated zones;
- Saturation excess and non-linear subsurface flow;
- Deep groundwater recharge.

B.2.10.1 File names

Model: m_10_susannah2_6p_2s
Parameter ranges: m_10_susannah2_6p_2s_parameter_ ranges

B.2.10.2 Model equations
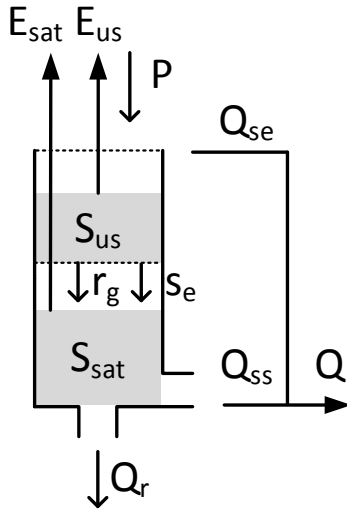
Figure B.11: Structure of the Susannah Brook v2 model

$$\frac{dS_{us}}{dt} = P - E_{us} - r_g - S_e \tag{B.80}$$

$$E_{us} = \frac{S_{us}}{S_b} * E_p \tag{B.81}$$

$$S_b = D * \phi \tag{B.82}$$

$$r_g = \begin{cases} P, & if\ S_{us} > S_{usfc} \\ 0, & \text{otherwise} \end{cases} \tag{B.83}$$

$$S_e = \begin{cases} S_{us} - S_{usfc}, & if\ S_{us} > S_{usfc} \\ 0, & \text{otherwise} \end{cases} \tag{B.84}$$

$$S_{usfc} = (S_b - S_{sat}) * \frac{fc}{\phi} \tag{B.85}$$

Where S_{us} is the current storage in the unsaturated store [mm], P the current precipitation [mm], S_b [mm] the maximum storage of the soil profile, based on the soil depth D [mm] and the porosity ϕ [-]. r_g is drainage from the unsaturated store to the saturated store [mm], based on the variable field capacity S_{usfc} [mm]. S_{usfc} is based on the current storage on the saturated zone S_{sat}

172

[mm], the maximum soil moisture storage S_b [mm], the field capacity fc [-] and the porosity ϕ [-]. S_e [mm] is the storage excess, resulting from a decrease of S_{usfc} that leads to more water being stored in the unsaturated zone than should be possible.

$$\frac{dS_{sat}}{dt} = r_g - E_{sat} - Q_{SE} - Q_{SS} - Q_R \tag{B.86}$$

$$E_{sat} = \frac{S_{sat}}{S_b} * E_p \tag{B.87}$$

$$Q_{SE} = \begin{cases} r_g + S_e, & \text{if } S_{sat} > S_b \\ 0, & \text{otherwise} \end{cases} \tag{B.88}$$

$$Q_{SS} = (1 - r) * c * (S_{sat})^d \tag{B.89}$$

$$Q_R = r * c * (S_{sat})^d \tag{B.90}$$

Where S_{sat} is the current storage in the saturated zone [mm], E_{sat} is the evaporation from the saturated zone [mm], Q_{SE} saturation excess runoff [mm] that occurs when the saturated zone reaches maximum capacity S_b [mm], Q_{SS} is subsurface flow [mm] and Q_R is recharge of deep groundwater [mm]. Both Q_{SS} and Q_R are based on the dimensionless fraction r and subsurface flow constants c [d^{-1}] and d [-]. Total runoff is the sum of Q_{SE} and Q_{SS}:

$$Q = Q_{SE} + Q_{SS} \tag{B.91}$$

B.2.11 Collie River Basin 3 (model ID: 11)

The Collie River Basin 3 model (fig. B.12) is part of a top-down modelling exercise and is originally applied at the daily scale [158]. It has 2 stores and 6 parameters (S_{max}, S_{fc}, a, M, b, λ). The model aims to represent:

- Separate bare soil and vegetation evaporation;
- Saturation excess surface runoff;
- Non-linear subsurface runoff;
- Non-linear groundwater runoff.

B.2.11.1 File names

Model: m_11_collie3_6p_2s

Parameter ranges: m_11_collie3_6p_2s_parameter_ ranges

B.2.11.2 Model equations
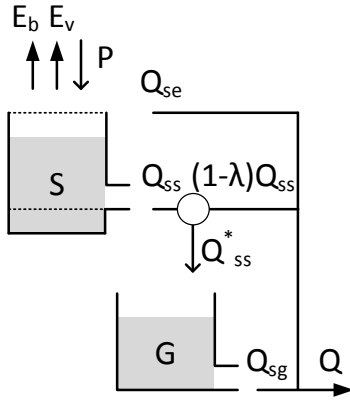
Figure B.12: Structure of the Collie River Basin 3 model

$$\frac{dS}{dt} = P - E_b - E_v - Q_{se} - Q_{ss} \tag{B.92}$$

$$Eb = \frac{S}{S_{max}}(1 - M) * Ep \tag{B.93}$$

$$Ev = \begin{cases} M * E_p, & if\ S > S_{fc} \\ \frac{S}{S_{fc}} * M * E_p, & otherwise \end{cases} \tag{B.94}$$

$$Q_{se} = \begin{cases} P, & if\ S > S_{max} \\ 0, & otherwise \end{cases} \tag{B.95}$$

$$Q_{ss} = \begin{cases} \left(a * (S - Sfc)\right)^b, & if\ S > S_{fc} \\ 0, & otherwise \end{cases} \tag{B.96}$$

Where S [mm] is the current storage in the soil moisture and P the precipitation input [mm/d]. Actual evaporation is split between bare soil evaporation E_b [mm/d] and transpiration through vegetation E_v [mm/d], controlled through the forest fraction M. The evaporation estimates are based on the current storage S, the potential evapotranspiration E_p [mm/d] and the maximum soil moisture storage S_{max} [mm], and field capacity S_{fc}

[mm] respectively. Q_{se} [mm/d] is saturation excess overland flow. Q_{ss} [mm/d] is non-linear subsurface flow regulated by runoff coefficients a [d^{-1}] and b [-].

$$\frac{dG}{dt} = Q_{ss}^* - Qsg \tag{B.97}$$

$$Q_{ss}^* = \lambda * Q_{ss} \tag{B.98}$$

$$Q_{sg} = (a * G)^b \tag{B.99}$$

Where G [mm] is groundwater storage. Q_{ss}^* mm/d] is the fraction of Q_{ss} directed to groundwater. Q_{sg} [mm/d] is non-linear groundwater flow that relies on the same parameters as subsurface flow uses. Total runoff:

$$Q = Q_{se} + (1 - \lambda) * Q_{ss} + Q_{sg} \tag{B.100}$$

B.2.12 Alpine model v2 (model ID: 12)

The Alpine model v2 (fig. B.13) is part of a top-down modelling exercise and represents a daily water balance model [95]. It has 2 stores and 6 parameters (T_t, ddf, S_{max}, C_{fc}, $t_{c,in}$, $t_{c,bf}$). The model aims to represent:

- Snow accumulation and melt;
- Saturation excess overland flow;
- Linear subsurface runoff.

B.2.12.1 File names

Model: m_12_alpine2_6p_2s
Parameter ranges: m_12_alpine2_6p_2s_parameter_ ranges

B.2.12.2 Model equations

Figure B.13: Structure of the Alpine model v1
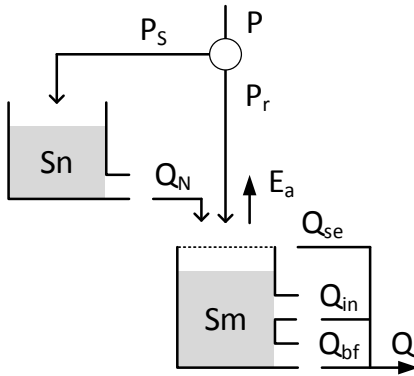
$$\frac{dSn}{dt} = P_s - Q_N \tag{B.101}$$

$$P_s = \begin{cases} P, & \text{if } T \le T_t \\ 0, & \text{otherwise} \end{cases} \tag{B.102}$$

$$Q_N = \begin{cases} ddf * (T - T_t), & \text{if } T \ge T_t \\ 0, & \text{otherwise} \end{cases} \tag{B.103}$$

Where S_N is the current snow storage [mm], P_s the precipitation that falls as snow [mm/d], Q_N snow melt [mm/d] based on a degree-day factor (ddf, [mm/°C/d]) and threshold temperature for snowfall and snowmelt (T_t, [°C]).

$$\frac{dS}{dt} = P_r + Q_N - E_a - Q_{se} - Q_{in} - Q_{bf} \tag{B.104}$$

$$P_r = \begin{cases} P, & \text{if } T > TT \\ 0, & \text{otherwise} \end{cases} \tag{B.105}$$

$$E_a = \begin{cases} E_p, & \text{if } S > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.106}$$

$$Q_{se} = \begin{cases} P_r + Q_N, & \text{if } S \geq S_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.107}$$

$$Q_{in} = \begin{cases} t_{c,in} * (S - S_{fc}), & \text{if } S > S_{fc} \\ 0, & \text{otherwise} \end{cases} \tag{B.108}$$

$$Q_{bf} = t_{c,bf} * S \tag{B.109}$$

Where S [mm] is the current soil moisture storage, which is assumed to evaporate at the potential rate E_p [mm/d] when possible. When S exceeds the maximum storage S_{max} [mm], water leaves the model as saturation excess runoff Q_{se}. If S exceeds field capacity S_{fc} [mm], interflow Q_{in} [mm/d] is generated controlled by time parameter $t_{c,in}$ [d^{-1}]. Q_{bf} represents baseflow controlled by time scale parameter $t_{c,bf}$ [d^{-1}]. Total runoff Q_t [mm/d] is:

$$Q_t = Q_{se} + Q_{in} + Q_{bf} \tag{B.110}$$

B.2.13 Hillslope model (model ID: 13)

The Hillslope model (fig. B.14) is a conceptualization of the perceived dominant processes in a typical Western European hillslope [274]. It belongs to a 3-part topography driven modelling exercise, together with a wetland and plateau conceptualization. Each model is provided in isolation here, because they are well-suited for isolating specific model structure choices. It has 2 store and 7 parameters (D_w, $S_{h,max}$, β_h, a, T_h, C and K_h). The model aims to represent:

- Stylized interception by vegetation;
- Evaporation;
- Separation between rapid subsurface flow and groundwater recharge;
- Capillary rise and linear relation runoff from groundwater.

B.2.13.1 File names

Model: m_13_hillslope_7p_2s

Parameter ranges: m_13_hillslope_7p_2s_parameter_ ranges

B.2.13.2 Model equations

Figure B.14: Structure of the Hillslope model
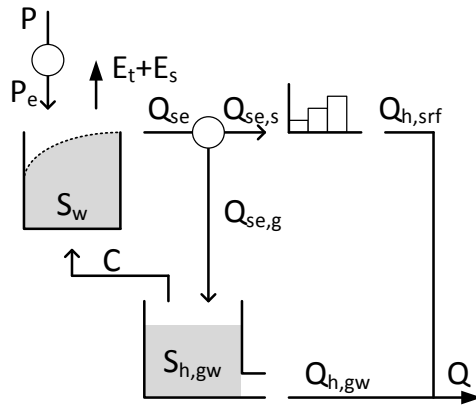
$$\frac{dS_w}{dt} = P_e + C - (E_t + E_s) - Q_{se} \quad \text{(B.111)}$$

$$P_e = max(P - D_h, 0) \quad \text{(B.112)}$$

$$C = c. \quad \text{(B.113)}$$

$$E_t + E_s = \begin{cases} E_p, & \text{if } S_w > 0 \\ 0, & \text{otherwise} \end{cases} \quad \text{(B.114)}$$

$$Q_{se} = \left(1 - \left(1 - \frac{S_h}{S_{h,max}}\right)^{\beta_h}\right) * P_e \quad \text{(B.115)}$$

$$\text{(B.116)}$$

Where S_w is the current soil water storage [mm]. Incoming precipitation P [mm/d] is reduced by interception D_h [mm/d], which is assumed to evaporate before the next precipitation event. C is capillary rise from groundwater [mm/d], given as a constant rate. Evaporation from soil moisture $E_t + E_s$ [mm/d] occurs at the potential rate E_p whenever possible. Storage excess surface runoff

Q_{se} [mm/d] depends on the fraction of the catchment that is currently saturated, expressed through parameters $S_{h,max}$ [mm] and β_h [-].

$$\frac{dS_{h,gw}}{dt} = Q_{se,g} - C - Q_{h,gw} \tag{B.117}$$

$$Q_{se,g} = (1 - a) * Q_{se} \tag{B.118}$$

$$Q_{h,gw} = K_h * S_{h,gw} \tag{B.119}$$

Where $S_{h,gw}$ is current groundwater storage [mm]. $Q_{se,g}$ is the groundwater fraction of storage excess flow Q_{se} [mm/d], with $Q_{se,s}$ as its complementary part. a is the parameter controlling this division [-]. Groundwater flow $Q_{h,gw}$ [mm/d] depends linearly on current storage $S_{h,gw}$ through parameter K_h [d^{-1}]. Total flow Q_t is the sum of $Q_{h,gw}$ and $Q_{h,srf}$, the latter of which is $Q_{se,s}$ lagged over T_h days.

B.2.14   TOPMODEL (model ID: 14)

The TOPMODEL (fig. B.15) is originally a semi-distributed model that relies on topographic information [48]. The model(ling concept) has undergone many revisions and significant differences can be seen between various publications. The version presented here is mostly based on Beven et al [50], with several necessary simplifications. Following Clark et al [78], the model is simplified to a lumped model (removing the distributed routing component) and all parameters are calibrated. This means the distribution of topographic index values that characterizes TOPMODEL are estimated using a shifted 2-parameter gamma distribution instead of being based on DEM data [78, 300]. For simplicity of the evaporation calculations, the root zone store and unsaturated zone store are combined into a single threshold store with identical functionality to the original 2-store concept. The model has 2 stores and 7 parameters ($S_{UZ,max}$, S_t, K_d, q_0, f, χ, ϕ). The model aims to represent:

- Variable saturated area with direct runoff from the saturated part;
- Infiltration and saturation excess flow;
- Leakage to, and non-linear baseflow from, a deficit store.

B.2.14.1   File names

Model: m_14_topmodel_7p_2s
Parameter ranges: m_14_topmodel_7p_2s_parameter_ ranges

B.2.14.2   Model equations
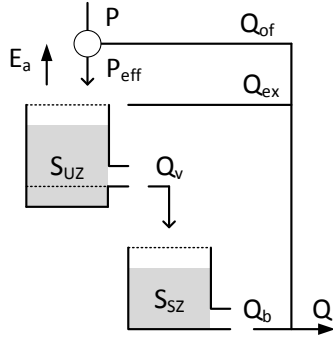
Figure B.15: Structure of the TOPMODEL

$$\frac{dS_{UZ}}{dt} = P_{eff} - Q_{ex} - E_a - Q_v \tag{B.120}$$

$$P_{eff} = P - Q_{of} = P - A_C * P \tag{B.121}$$

$$Q_{ex} = \begin{cases} P_{eff}, & \text{if } S_{UZ} = S_{uz,max} \\ 0, & \text{otherwise} \end{cases} \tag{B.122}$$

$$E_a = \begin{cases} E_p, & \text{if } S_{UZ} > S_t * S_{UZ,max} \\ \frac{S_{UZ}}{S_t * S_{UZ,max}} * E_p, & \text{otherwise} \end{cases} \tag{B.123}$$

$$Q_v = \begin{cases} k_d \frac{S_{UZ} - S_t * S_{UZ,max}}{S_{UZ,max}(1 - S_t)}, & \text{if } S_{UZ} > S_t * S_{UZ,max} \\ 0, & \text{otherwise} \end{cases} \tag{B.124}$$

Where S_{UZ} [mm] is the current storage in the combined unsaturated zone and root zone, with S_t [-] (fraction of $S_{UZ,max}$) indicating the boundary between the two and being the threshold above which drainage to the saturated zone can occur. P_{eff} [mm/d] is the fraction of precipitation that does not fall on the saturated area A_c [-], E_a [mm/d] is evaporation that occurs at the potential rate for the unsaturated zone and scaled linearly with storage in the root zone, Q_{ex} [mm/d] is overflow when the bucket reaches maximum capacity $S_{UZ,max}$ [mm], and Q_v [mm/d] is drainage to the saturated zone, depending on time parameter k_d [d^{-1}] and the relative storage in the unsaturated zone compared to the current deficit in the saturated zone.

$$\frac{dS_{SZ}}{dt} = -Q_v + Q_b \tag{B.125}$$

$$Q_b = q_0 * e^{-f*S_{SZ}} \tag{B.126}$$

Where S_{SZ} [mm] is the current storage *deficit* in the saturated zone store, which is increased by baseflow Q_b [mm/d] and decreased by drainage Q_v. Q_b relies on saturated flow rate q_0 [mm/d], parameter f [mm^{-1}] and current deficit S_{SZ}. Total flow:

$$Q = Q_{of} + Q_{ex} + Q_b \tag{B.127}$$

$$Q_{of} = A_c * P \tag{B.128}$$

The saturated area A_c is calculated as follows. First, the within-catchment distribution of topographic index values is estimated with a shifted 2-parameter gamma distribution [78, 300]:

$$f(\zeta) = \begin{cases} \frac{1}{\chi\Gamma(\phi)}\left(\frac{\zeta-\mu}{\chi}\right)^{\phi-1} exp\left(-\frac{\zeta-\mu}{\chi}\right), & \text{if } \zeta > \mu \\ 0, & \text{otherwise} \end{cases} \tag{B.129}$$

Where Γ is the gamma function and χ, ϕ and μ are parameters of the gamma distribution. Following Clark et al [78], μ is fixed at $\mu = 3$ and χ and ϕ are calibration parameters. ζ represents the topographic index $ln(a/tan\beta)$ with mean value $\lambda = \chi\phi + \mu$. Saturated area A_c is computed as the fraction of the catchment that is above a deficit-dependent critical value ζ_{crit}:

$$A_c = \int_{\zeta_{crit}}^{\infty} f(\zeta)d\zeta \tag{B.130}$$

$$\zeta_{crit} = f * S_{SZ} + \lambda \tag{B.131}$$

B.2.15 Plateau model (model ID: 15)

The Plateau model (fig. B.16) is a conceptualization of the perceived dominant processes in a typical Western European plateau [274]. It belongs to a 3-part topography driven modelling exercise, together with a wetland and hillslope conceptualization. Each model is provided in isolation here, because they are well-suited for isolating specific model structure choices. It has 2 stores and 8 parameters (F_{max}, D_p, $S_{u,max}$, lp, p, T_p, C and K_p). The model aims to represent:

- Stylized interception by vegetation;
- Evaporation controlled by a wilting point and moisture constrained transpiration;
- Separation between infiltration and infiltration excess flow;
- Capillary rise and linear relation runoff from groundwater.

B.2.15.1 File names

Model: m_15_plateau_8p_2s

Parameter ranges: m_15_plateau_8p_2s_parameter_ ranges

B.2.15.2 Model equations
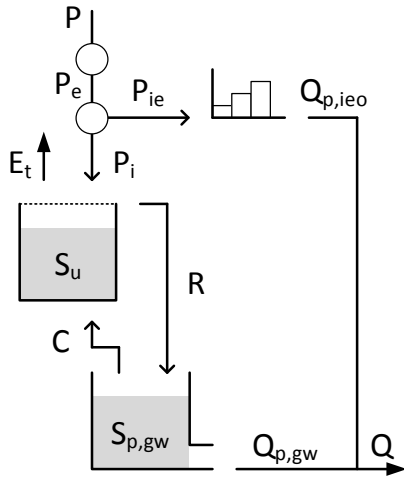
Figure B.16: Structure of the Plateau model

$$\frac{dS_u}{dt} = P_i + C - E_t - R \tag{B.132}$$

$$P_i = min(P_e, F_{max}) \tag{B.133}$$

$$= min\left(max(P - D_p, 0), F_{max}\right) \tag{B.134}$$

$$C = c. \tag{B.135}$$

$$E_t = E_p * max\left(p\frac{S_u - S_{wp}}{S_{u,max} - S_{wp}}, 0\right) \tag{B.136}$$

$$R = \begin{cases} P_i + C, & \text{if } S_u = S_{u,max} \\ 0, & \text{otherwise} \end{cases} \tag{B.137}$$

Where S_u is the current soil water storage [mm]. Incoming precipitation P [mm/d] is reduced by interception D_p [mm/d], which is assumed to evaporate before the next precipitation event. P_e is further divided into infiltration P_i [mm/d] based on the maximum infiltration rate F_{max} [mm/d] and infiltration excess $P_{ie} = P_e - P_i$ [mm/d]. C is capillary

rise from ground water [mm/d], given as a constant rate.

Evaporation from soil moisture E_t [mm/d] occurs at the potential rate E_p when S_u is above the wilting point S_{wp} [mm] (here defined as $S_{wp} = lp * S_{u,max}$) and is further constrained by coefficient p [-], which is between 0 and 1. Storage excess R [mm/d] flows into the groundwater.

$$\frac{dS_{p,gw}}{dt} = R - C - Q_{p,gw} \tag{B.138}$$

$$Q_{p,gw} = K_p * S_{p,gw} \tag{B.139}$$

Where $S_{p,gw}$ is current groundwater storage [mm]. Groundwater flow $Q_{p,gw}$ [mm/d] depends linearly on current storage $S_{p,gw}$ through parameter K_p [d^{-1}]. Total flow Q_t is the sum of $Q_{p,gw}$ and $Q_{p,ieo}$, the latter of which is P_{ie} lagged over T_p days.

B.2.16 New Zealand model v2 (model ID: 16)

The New Zealand model v2 (fig. B.17) is part of a top-down modelling exercise that focusses on several catchments in New Zealand [19]. It has 2 stores and 8 parameters (I_{max}, S_{max}, S_{fc}, M, a, b and $t_{c,bf}$, d). The model aims to represent:

- Interception by vegetation;
- Separate vegetation and bare soil evaporation;
- Saturation excess overland flow;
- Subsurface runoff when soil moisture exceeds field capacity;
- Baseflow;
- Flow routing.

B.2.16.1 File names

Model: m_16_newzealand2_8p_2s

Parameter ranges: m_16_newzealand2_8p_2s_parameter_ ranges

B.2.16.2 Model equations
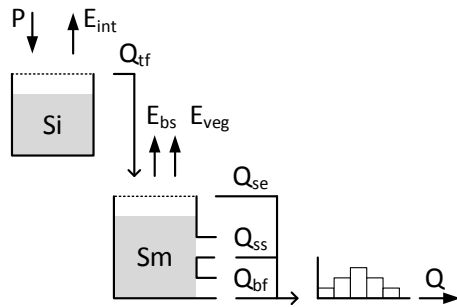
Figure B.17: Structure of the New Zealand model v1

$$\frac{dS_i}{dt} = P - E_{int} - Q_{tf} \tag{B.140}$$

$$E_{int} = Ep \tag{B.141}$$

$$Q_{tf} = \begin{cases} P, & \text{if } S_i \geq I_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.142}$$

Where S_i [mm] is the current interception storage which gets replenished through daily precipitation P [mm/d]. Intercepted water is assumed to evaporate (E_{int} [mm/d]) at the potential rate E_p [mm/d] when possible. Q_{tf} [mm/d] represents throughfall towards soil moisture.

$$\frac{dS_m}{dt} = Q_{tf} - E_{veg} - E_{bs} - Q_{se} - Q_{ss} - Q_{bf} \tag{B.143}$$

$$E_{veg} = \begin{cases} M * E_p, & \text{if } S > S_{fc} \\ \frac{S_m}{S_{fc}} * M * E_p, & \text{otherwise} \end{cases} \tag{B.144}$$

$$E_{bs} = \frac{S}{S_{max}}(1 - M) * E_p \tag{B.145}$$

$$Q_{se} = \begin{cases} P, & \text{if } S \geq S_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.146}$$

$$Q_{ss} = \begin{cases} \left(a * (S - S_{fc})\right)^b, & \text{if } S \geq S_{fc} \\ 0, & \text{otherwise} \end{cases} \tag{B.147}$$

$$Q_{bf} = t_{c,bf} * S \tag{B.148}$$

Where S_m [mm] is the current soil moisture storage which gets replenished through daily precipitation P [mm/d]. Evaporation through vegetation E_{veg} [mm/d] depens on the forest fraction M [-] and field capacity S_{fc} [-]. E_{bs} [mm/d] represents bare soil evaporation. When S exceeds the maximum storage S_{max} [mm], water leaves the model as saturation excess runoff Q_{se}. If S exceeds field capacity S_{fc} [mm], subsurface runoff Q_{ss} [mm/d] is generated controlled by time parameter a [d^{-1}] and nonlinearity parameter b [-]. Q_{bf} represents baseflow controlled by time scale parameter $t_{c,bf}$ [d^{-1}]. Total runoff Q_t [mm/d] is:

$$Q_t = Q_{se} + Q_{ss} + Q_{bf} \tag{B.149}$$

Total flow is delayed by a triangular routing scheme controlled by time parameter d [d].

B.2.17 Penman model (model ID: 17)

The Penman model (fig. B.18) is based on the drying curve concept described by Penman [251, 331]. It has 3 stores and 4 parameters (S_{max}, ϕ, α, k_1). The model aims to represent:

- Moisture accumulation and evaporation from the root zone;
- Bypass of excess moisture to the stream;
- Deficit-based groundwater accounting;
- Linear flow routing.

B.2.17.1 File names

Model: m_17_penman_4p_3s

Parameter ranges: m_17_penman_4p_3s_parameter_ ranges

B.2.17.2 Model equations
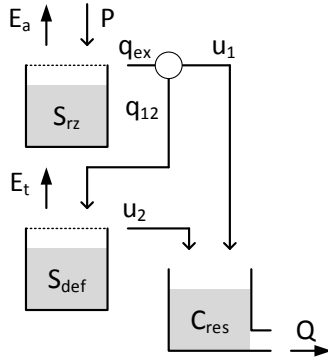
Figure B.18: Structure of the Penman model

$$\frac{dS_{rz}}{dt} = P - E_a - Q_{ex} \tag{B.150}$$

$$E_a = \begin{cases} E_p, & \text{if } S_{rz} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.151}$$

$$P_{ex} = \begin{cases} P, & \text{if } S_{rz} = S_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.152}$$

Where S_{rz} [mm] is the current storage in the root zone, refilled by precipitation P [mm/d] and drained by evaporation E_a [mm/d] and moisture excess q_{ex} [mm/d]. E_a occurs at the potential rate E_p [mm/d] whenever possible. q_{ex} occurs only when the store is at maximum capacity S_{max} [mm].

$$\frac{dS_{def}}{dt} = E_t + u_2 - q_{12} \tag{B.153}$$

$$E_t = \begin{cases} \gamma * E_p, & \text{if } S_{rz} = 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.154}$$

$$u_2 = \begin{cases} q_{12}, & \text{if } S_{def} = 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.155}$$

$$q_{12} = (1 - \phi) * q_{ex} \tag{B.156}$$

Where S_{def} [mm] is the current moisture *deficit*, which is increased by evaporation E_t [mm/d] and reduced by inflow q_{12} [mm/d]. E_t occurs only when the upper store S_{rz} is empty and at a fraction γ [-] of E_p. Inflow q_{12} is the fraction $(1-\phi)$ [-] of q_{ex} that does not bypass the lower soil layer. Saturation excess u_2 [mm/d] occurs only when there is zero deficit.

$$\frac{dC_{res}}{dt} = u_1 + u_2 - Q \tag{B.157}$$

$$Q = k_1 * C_{res} \tag{B.158}$$

Where C_{res} [mm] is the current storage in the routing reservoir, increased by u_1 and u_2, and drained by runoff Q [mm/d]. Q has a linear relationship with storage through time scale parameter k_1 [d^{-1}].

B.2.18   SIMHYD (model ID: 18)

The SIMHYD model (fig. B.19) is a simplified version of MODHYDROLOG, originally developped for use in Australia [70]. It has 3 stores and 7 parameters (INSC, COEFF, SQ, SMSC, SUB, CRAK, K). The model aims to represent:

- Interception by vegetation;
- Infiltration and infiltration excess flow;
- Preferential groundwater recharge, interflow and saturation excess flow;
- Groundwater recharge resulting from filling up of soil moisture storage capacity;
- Slow flow from groundwater.

B.2.18.1   File names

Model: m_18_simhyd_7p_3s

Parameter ranges: m_18_simhyd_7p_3s_parameter_ ranges

B.2.18.2   Model equations

Figure B.19: Structure of the SIMHYD model
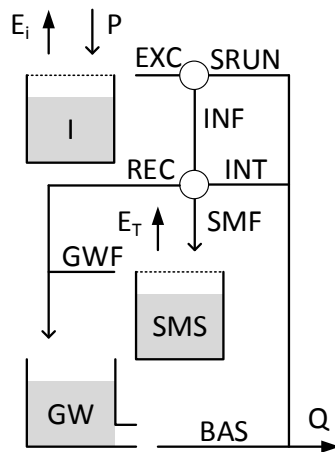
$$\frac{dI}{dt} = P - E_i - EXC \tag{B.159}$$

$$E_i = \begin{cases} E_p, & \text{if } I > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.160}$$

$$EXC = \begin{cases} P, & \text{if } I = INSC \\ 0, & \text{otherwise} \end{cases} \tag{B.161}$$

  Where I is the current interception storage [mm], P precipitation [mm/d], E_i the evaporation from the interception store [mm/d] and EXC the excess rainfall [mm/d]). Evaporation is assumed to occur at the potential rate when possible. When I exceeds the maximum interception capacity $INSC$ [mm], water is routed to the rest of the model as excess precipitation EXC.

$$\frac{dSMS}{dt} = SMF - E_T - GWF \tag{B.162}$$

$$SMF = INF - INT - REC \tag{B.163}$$

$$INF = min\left(COEFF * exp\left(\frac{-SQ * SMS}{SMSC}\right), EXC\right) \tag{B.164}$$

$$INT = SUB * \frac{SMS}{SMSC} * INF \tag{B.165}$$

$$REC = CRAK * \frac{SMS}{SMSC} * (INF - INT) \tag{B.166}$$

$$E_T = min\left(10 * \frac{SMS}{SMSC}, PET\right) \tag{B.167}$$

$$GWF = \begin{cases} SMF, & \text{if } SMS = SMSC \\ 0, & \text{otherwise} \end{cases} \tag{B.168}$$

Where SMS is the current storage in the soil moisture store [mm]. INF is total infiltration [mm/d] from excess precipitation, based on maximum infiltration loss parameter COEFF [-], the infiltration loss exponent SQ [-] and the ratio between current soil misture storage SMS and the maximum soil moisture capacity SMSC [mm]. INT represents interflow and saturation excess flow [mm/d], using a constant of proportionality SUB [-]. REC is preferential recharge of groundwater [mm/d] based on another constant of proportionality CRAK [-]. SMF is flow into soil moisture storage [mm/d]. E_T evaporation from the soil moisture that occurs at the potential rate when possible [mm/d], and GWF the flow to the groundwater store [mm/d]:

$$\frac{dGW}{dt} = REC + GWF - BAS \tag{B.169}$$

$$BAS = K * GW \tag{B.170}$$

Where GW is the current storage [mm] in the groundwater reservoir. Outflow *BAS* [mm/d] from the reservoir has a linear relation with storage through the linear recession parameter K [d^{-1}]. Total outflow Q_t [mm/d] is the sum of three parts:

$$Q_t = SRUN + INT + BAS \tag{B.171}$$

$$SRUN = EXC - INF \tag{B.172}$$

B.2.19 Australia model (model ID: 19)

The Australia model (fig. B.20) is part of a top-down modelling exercise designed to use auxiliary data [101]. Some adjustments were made to the evaporation equations: these were originally separated between vegetation and bare soil evaporation, scaled between the unsaturated and saturated zone. This has been simplified to separation between unsaturated and saturated evaporation only. The model has 3 stores and 8 parameters (S_b, ϕ, fc, α_{SS}, β_{SS}, K_{deep}, α_{BF}, β_{BF}). For consistency with other model formulations, S_b is is used as a parameter, instead of being broken down into its constitutive parts D and ϕ. The model aims to represent:

- Separation of saturated zone and a variable-size unsaturated zone;
- Evaporation from unsaturated and saturated zones;
- Saturation excess and non-linear subsurface flow;
- Deep groundwater recharge and baseflow.

B.2.19.1 File names

Model: m_19_australia_8p_3s

Parameter ranges: m_19_australia_8p_3s_parameter_ ranges

B.2.19.2 Model equations
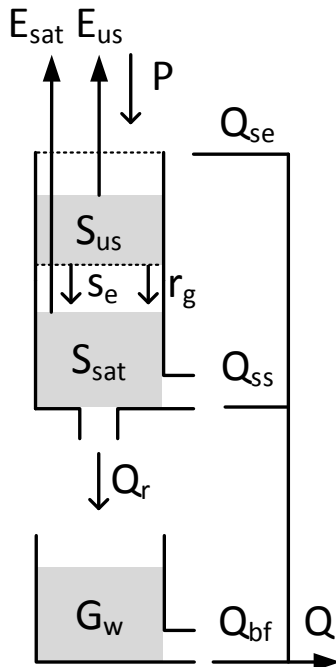
$$\frac{dS_{us}}{dt} = P - E_{us} - r_g - s_e \tag{B.173}$$

$$E_{us} = \frac{S_{us}}{S_b} * E_p \tag{B.174}$$

$$S_b = D * \phi \tag{B.175}$$

$$r_g = \begin{cases} P, & if\ S_{us} > S_{usfc} \\ 0, & \text{otherwise} \end{cases} \tag{B.176}$$

$$s_e = \begin{cases} S_{us} - S_{usfc}, & if\ S_{us} > S_{usfc} \\ 0, & \text{otherwise} \end{cases} \tag{B.177}$$

$$S_{usfc} = (S_b - S_{sat}) * \frac{fc}{\phi} \tag{B.178}$$

Where S_{us} is the current storage in the unsaturated store [mm], P the current precipitation [mm/d], S_b [mm] the maximum storage of the soil profile, based on the

Figure B.20: Structure of the Australia model

soil depth D [mm] and the porosity ϕ [-]. r_g [mm/d] is drainage from the unsaturated store to the saturated store, based on the variable field capacity S_{usfc} [mm]. S_{usfc} is based on the current storage on the saturated zone S_{sat} [mm], the maximum soil moisture storage S_b [mm], the field capacity fc [-] and the porosity ϕ [-]. s_e [mm/d] is the storage excess, resulting from a decrease of S_{usfc} that leads to more water being stored in the unsaturated zone than should be possible.

$$\frac{dS_{sat}}{dt} = r_g - E_{sat} - Q_{SE} - Q_{SS} - Q_R \tag{B.179}$$

$$E_{sat} = \frac{S_{sat}}{S_b} * E_p \tag{B.180}$$

$$Q_{SE} = \begin{cases} r_g + S_e, & \text{if } S_{sat} > S_b \\ 0, & \text{otherwise} \end{cases} \tag{B.181}$$

$$Q_{SS} = \alpha_{SS} * (S_{sat})^{\beta_{SS}} \tag{B.182}$$

$$Q_R = K_{deep} * S_{sat} \tag{B.183}$$

Where S_{sat} is the current storage in the saturated zone [mm], E_{sat} is the evaporation from the saturated zone [mm], Q_{SE} saturation excess runoff [mm/d] that occurs when the saturated zone reaches maximum capacity S_b [mm], Q_{SS} is subsurface flow [mm/d] and Q_R is recharge of deep groundwater [mm/d]. Both Q_{SS} and Q_R are based on the dimensionless fraction r and subsurface flow constants c [d^{-1}] and d [-].

$$\frac{dG_w}{dt} = Q_R - Q_{BF} \tag{B.184}$$

$$Q_{BF} = \alpha_{BF} * (G_w)^{\beta_{BF}} \tag{B.185}$$

$$\tag{B.186}$$

Where G_w is the current groundwater storage [mm] and Q_{BF} baseflow, dependent on parameters α_{BF} [d^{-1}] and β_{BF} [-]. Total runoff is the sum of QSE, Q_{SS} and Q_{BF}:

$$Q = Q_{SE} + Q_{SS} + Q_{BF} \tag{B.187}$$

B.2.20 Generalized Surface inFiltration Baseflow model (model ID: 20)

The GSFB model (fig. B.21) is originally developed for use in Australian ephemeral catchments [224, 350]. It has 3 stores and 8 parameters (C. NDC, S_{max}, E_{max}, F_{rate}, B, DPF, SDR_{max}). The model aims to represent:

- Saturation excess surface runoff;
- Threshold-based infiltration;
- Threshold-based baseflow;
- Deep percolation and water rise to meet evaporation demand.

B.2.20.1 File names

Model: m_20_gsfb_8p_3s

Parameter ranges: m_20_gsfb_8p_3s_parameter_ ranges

B.2.20.2 Model equations
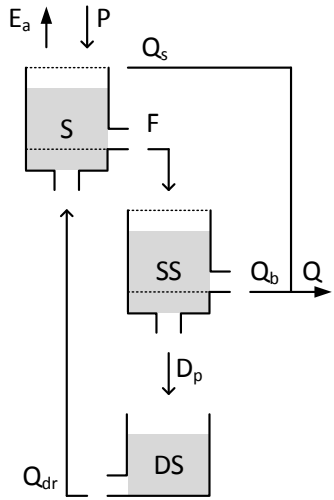
Figure B.21: Structure of the GSFB model

$$\frac{dS}{dt} = P + Q_{dr} - E_a - Q_s - F \tag{B.188}$$

$$Q_{dr} = \begin{cases} C * DS * \left(1 - \frac{S}{NDC * S_{max}}\right), & \text{if } S \leq NDC * S_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.189}$$

$$E_a = \begin{cases} E_p, & \text{if } S > NDC * S_{max} \\ min\left(E_p, E_{max} \frac{S}{NDC * S_{max}}\right), & \text{otherwise} \end{cases} \tag{B.190}$$

$$Q_s = \begin{cases} P, & \text{if } S = S_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.191}$$

$$F = \begin{cases} F_{rate}, & \text{if } S > NDC * S_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.192}$$

Where S [mm] is the current storage in the upper zone, refilled by precipitation P [mm/d] and recharge from deep groundwater Q_{dr} [mm/d]. The store is drained by evaporation E_a [mm/d], surface runoff Q_s [mm/d] and infiltration F [mm/d]. E_a occurs

at the potential rate E_p $[mm/d]$ if the store is above a threshold capacity given as the fraction NDC [-] of maximum storage S_{max} [mm]. Evaporation occurs at a reduced rate scaled by maximum evaporation rate E_{max} $[mm/d]$ if the store is below this threshold. Q_s occurs only if the store is at maximum capacity S_{max}. F occurs at a constant rate F_{rate} if the store is above threshold $NDC * S_{max}$. Recharge from deep percolation only occurs if the store is below threshold capacity $NDC * S_{max}$ and uses time parameter C $[d^{-1}]$ and current deep storage DS [mm].

$$\frac{dSS}{dt} = F - Q_b - D_p \tag{B.193}$$

$$Q_b = \begin{cases} B * DPF * (SS - SDR_{max}), & \text{if } SS > SDR_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.194}$$

$$D_p = (1 - B) * DPF * SS \tag{B.195}$$

Where SS [mm] is the current storage in the subsurface store, refilled by infiltration F and drained by baseflow Q_b $[mm/d]$ and deep percolation D_p $[mm/d]$. Outflow from this store is given as a function of storage DS and time coefficient DPF $[d^{-1}]$. A fraction $1 - B$ [-] of this outflow is deep percolation D_p. The remaining fraction B [-] is baseflow Q_b, provided the store is above threshold SDR_{max} [mm].

$$\frac{dDS}{dt} = D_p - Q_{dr} \tag{B.196}$$

$$\tag{B.197}$$

Where DS [mm] is the current storage in the deep store, refilled by a deep percolation D_p and drained by recharge to the upper store Q_{dr}. Total flow:

$$Q_t = Q_s + Q_b \tag{B.198}$$

B.2.21 Flex-B (model ID: 21)

The Flex-B model (fig. B.22) is the basis of a model development study [106]. It has 3 stores and 9 parameters (UR_{max}, β, D, $Perc_{max}$, L_p, $N_{lag,f}$, $N_{lag,s}$, K_f, K_s). The model aims to represent:

- Infiltration and saturation excess flow based on a distribution of different soil depths;
- A split between fast saturation excess flow and preferential recharge to a slow store;
- Percolation from the unsaturated zone to a slow runoff store.

B.2.21.1 File names

Model: m_21_flexb_9p_3s

Parameter ranges: m_21_flexb_9p_3s_parameter_ ranges

B.2.21.2 Model equations
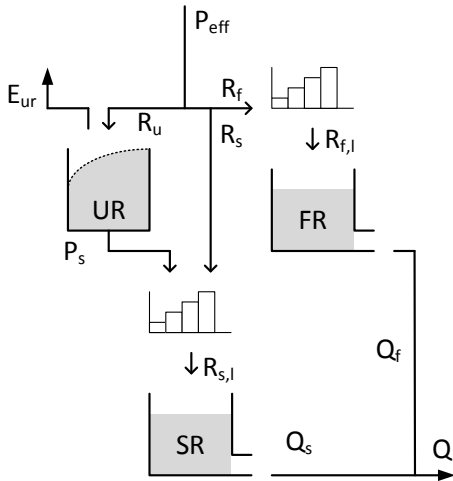
Figure B.22: Structure of the Flex-B model

$$\frac{dUR}{dt} = R_u - E_{ur} - R_p \tag{B.199}$$

$$R_U = (1 - C_r) * P_{eff} \tag{B.200}$$

$$C_r = \left[1 + exp\left(\frac{-UR/UR_{max} + 1/2}{\beta} \right) \right]^{-1} \tag{B.201}$$

$$E_{ur} = E_p * min\left(1, \frac{UR}{UR_{max}} \frac{1}{L_p} \right) \tag{B.202}$$

$$P_s = Perc_{max} * \frac{UR}{UR_{max}} \tag{B.203}$$

Where UR is the current storage in the unsaturated zone [mm]. R_u [mm/d] is the inflow into UR based on its current storage compared to maximum storage UR_{max} [mm] and a shape distribution parameter β [-].

E_{ur} the evaporation [mm/d] from UR which follows a linear relation between current and maximum storage until a threshold L_p [-] is exceeded. P_s is the percolation from UR to the slow reservoir SR [mm/d], based on a maximum percolation rate $Perc_{max}$ [mm], relative to the fraction of current storage and maximum storage. P_{eff} is routed towards the unsaturated zone based on Cr, with the remainder being divided into preferential recharge R_s [mm/d] and fast runoff R_f [mm/d]:

$$R_s = (P_{eff} - R_u) * D \tag{B.204}$$

$$R_f = (P_{eff} - R_u) * (1 - D) \tag{B.205}$$

Where R_s and R_f are the flows [mm/d] to the slow and fast runoff reservoir respectively, based on runoff partitioning coefficient D [-]. Both are lagged by linearly increasing triangular transformation functions with parameters $N_{lag,s}$ [d] and $N_{lag,f}$ [d] respectively. Percolation R_p is added to R_s before the transformation to $R_{s,l}$ occurs.

$$\frac{dFR}{dt} = R_{f,l} - Q_f \tag{B.206}$$

$$Q_f = K_f * FR \tag{B.207}$$

Where FR is the current storage [mm] in the fast flow reservoir. Outflow Q_f [mm/d] from the reservoir has a linear relation with storage through time scale parameter K_f [d^{-1}].

$$\frac{dSR}{dt} = R_{s,l} - Q_s \tag{B.208}$$

$$Q_s = K_s * SR \tag{B.209}$$

Where SR is the current storage [mm] in the slow flow reservoir. Outflow Q_s [mm/d] from the reservoir has a linear relation with storage through time scale parameter K_s [d^{-1}]. Total outflow Q [mm/d]:

$$Q = Q_f + Q_s \tag{B.210}$$

B.2.22 Variable Infiltration Capacity (VIC) model (model ID: 22)

The VIC model (fig. B.23) is originally developed for use with General Circulation Models and uses latent and sensible heat fluxes to determine the rainfall-runoff relationship [194]. For consistency with other models in this framework, we use a conceptualized version based in part of the VIC implementation in Clark et al [78]. In addition, the original Leaf-Area-Index-based interception capacity is replaced with a sinusoidal curve-based approximation of interception capacity. The model has 3 stores and 10 parameters ($\bar{I}$, I_δ, I_s, $S_{sm,max}$, b, k_1, c_1, $S_{gw,max}$, k_2, c_2). The model aims to represent:

- Time-varying interception by vegetation;
- Variable infiltration and saturation excess flow;
- Interflow and baseflow from a deeper groundwater layer.

B.2.22.1 File names

Model: m_22_vic_10p_3s
Parameter ranges: m_22_vic_10p_3s_parameter_ ranges

B.2.22.2 Model equations
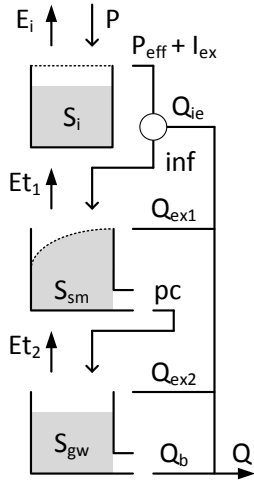
Figure B.23: Structure of the VIC model

$$\frac{dS_i}{dt} = P - E_i - P_{eff} - I_{ex} \tag{B.211}$$

$$E_i = \frac{S_i}{I_{max}} * E_p \tag{B.212}$$

$$I_{max} = \bar{I}(1 + I_\delta * sin(2\pi(t + I_s))) \tag{B.213}$$

$$P_{eff} = \begin{cases} P, & \text{if } S_i = I_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.214}$$

$$I_{ex} = max(S_i - I_{max}) \tag{B.215}$$

Where S_i [mm] is the current interception storage, refilled by precipitation P [mm/d] and drained by evaporation E_i [mm/d] and interception excess flows P_{eff}[mm/d] and I_{ex} [mm/d]. E_i decreases linearly with storage, based on maximum storage I_{max} [mm]. I_{max} is determined using the mean interception $\bar{I}$ [mm], fractional seasonal interception change I_δ [-] and time shift I_s [-]. It is implicitly assumed that 1 sinusoidal period corresponds with a growing season of 1 year. P_{eff} is effective rainfall when the store is at maximum capacity. I_{ex} is

an auxiliary flux used when a change in storage size result in current storage S_i exceeding I_{max}.

$$\frac{dS_{sm}}{dt} = inf - Et_1 - Q_{ex1} - pc \tag{B.216}$$

$$inf = (P_{eff} + I_{ex}) - Q_{ie} \tag{B.217}$$

$$Q_{ie} = (P_{eff} + I_{ex}) * \left(1 - \left(1 - \frac{S_{sm}}{S_{sm,max}}\right)^b\right) \tag{B.218}$$

$$E_{t1} = \frac{S_{sm}}{S_{sm,max}} * (E_p - E_i) \tag{B.219}$$

$$Q_{ex1} = \begin{cases} inf, & \text{if } S_{sm} = S_{sm,max} \\ 0, & \text{otherwise} \end{cases} \tag{B.220}$$

$$pc = k_1 * \left(\frac{S_{sm}}{S_{sm,max}}\right)^{c_1} \tag{B.221}$$

Where S_{sm} [mm] is the current soil moisture storage, refilled by infiltration inf [mm/d], and drained by evapotranspiration Et_1 [mm/d], storage excess Q_{ex1} [mm/d] and percolation pc [mm/d]. inf relies on the value of infiltration excess Q_{ie}, which is calculated using the maximum soil moisture storage $S_{sm,max}$ [mm] and shape parameter b [-]. Et_1 scales linearly with current storage. Q_{ex1} equals inf when the store is at maximum capacity. pc has a potentially non-linear relationship with current storage through time parameter k_1 [d^{-1}] and shape parameter c_1.

$$\frac{dS_{gw}}{dt} = pc - Et_2 - Q_{ex2} - Q_b \tag{B.222}$$

$$Et_2 = \frac{S_{gw}}{S_{gw,max}} * (E_p - E_i - Et_1) \tag{B.223}$$

$$Q_{ex2} = \begin{cases} pc, & \text{if } S_{gw} = S_{gw,max} \\ 0, & \text{otherwise} \end{cases} \tag{B.224}$$

$$Q_b = k_2 * \left(\frac{S_{gw}}{S_{gw,max}}\right)^{c_2} \tag{B.225}$$

Where S_{gw} [mm] is the current groundwater storage, refilled through percolation pc [mm/d] and drained by evapotranspiration Et_2 [mm/d], excess flow Q_{ex2} [mm/d] and baseflow Q_b [mm/d]. Et_2 is scaled linearly with current storage based on maximum storage $S_{gw,max}$ [mm]. Q_{ex2} equals pc when the store is at maximum capacity. Q_b has a potentially non-linear relationship with current storage through time parameter k_2 and shape parameter c_2. Total outflow:

$$Q_t = Q_{ie} + Q_{ex1} + Q_{ex2} + Q_b \qquad \text{(B.226)}$$

B.2.23   Large-scale catchment water and salt balance model element (model ID: 23)

The large-scale catchment water and salt balance model (LASCAM) (fig. B.24) is part of a study that investigates soil water and salt concentration before and after forest clearing [302]. It is a semi-distributed model made up of individual elements, such as described below. The model presented here simulates the water balance only (salt is ignored). It has 3 stores and 24 parameters (α_f, β_f, B_{max}, F_{max}, α_c, β_c, A_{min}, A_{max}, α_{ss}, β_{ss}, c, α_g, β_g, γ_f, δ_f, t_d, α_b, β_b, γ_a, δ_a, α_a, β_a, γ_b, δ_b). The model aims to represent:

- Stylized interception;
- Saturation and infiltration excess surface runoff;
- An inner layout representing near-stream saturated storage, deep saturated storage and medium-depth unsaturated storage;
- Subsurface saturation and infiltration excess flow to the near-stream store;
- Percolation to and capillary rise from groundwater.

B.2.23.1   File names

Model: m_23_lascam_24p_3s
Parameter ranges: m_23_lascam_24p_3s_parameter_ ranges

B.2.23.2   Model equations
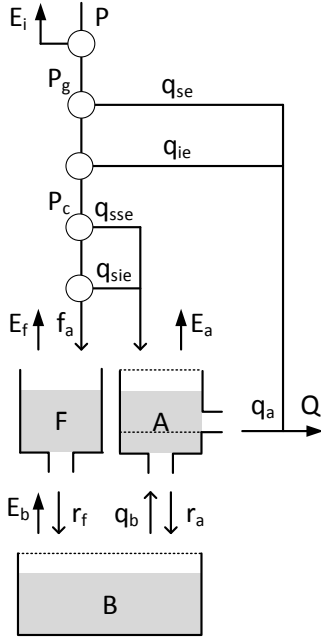
Figure B.24: Structure of the LASCAM model

$$\frac{dF}{dt} = f_a - E_f - r_f \tag{B.227}$$

$$f_a = min\left(P_c * max\left(1, \frac{1-\phi_{ss}}{1-\phi_c}\right), f_{ss}^*\right) \tag{B.228}$$

$$f_{ss}^* = \alpha_f\left(1 - \frac{B}{B_{max}}\right)\left(\frac{F}{F_{max}}\right)^{-\beta_f} \tag{B.229}$$

$$\phi_c = \begin{cases} \alpha_c\left(\frac{A-A_{min}}{A_{max}-A_{min}}\right)^{\beta_c}, & \text{if } A > A_{min} \\ 0, & \text{otherwise} \end{cases} \tag{B.230}$$

$$\phi_{ss} = \begin{cases} \alpha_{ss}\left(\frac{A-A_{min}}{A_{max}-A_{min}}\right)^{\beta_{ss}}, & \text{if } A > A_{min} \\ 0, & \text{otherwise} \end{cases} \tag{B.231}$$

$$P_c = min\left(P_g - q_{se}, f_s^*\right) \tag{B.232}$$

$$f_s^* = c \tag{B.233}$$

$$q_{se} = \phi_c * P_g \tag{B.234}$$

$$P_g = max\left(\alpha_g + \beta_g * P, 0\right) \tag{B.235}$$

$$E_f = \gamma_f * E_p\left(\frac{F}{F_{max}}\right)^{\delta_f} \tag{B.236}$$

$$r_f = t_d * F \tag{B.237}$$

Where F [mm] is the current storage in the unsaturated infiltration store, which controls the amount of subsurface runoff generated on the boundary of a more permeable top layer (store A) with a less permeable bottom layer (store F). F is refilled by actual infiltration f_a [mm/d], and drained by recharge r_f [mm/d] and evaporation E_b [mm/d]. f_a depends on the actual infiltration rate P_c [mm/d], the fraction saturated catchment area ϕ_{ss} [-], the fraction variable area contributing to overland flow ϕ_c [-] and a catchment-scale infiltration capacity f_{ss}^* [mm/d]. f_{ss}^* depends on a scaling parameter α_f [mm/d], the relative storage in groundwater B/B_{max}, the relative infiltration volume in the catchment F/F_{max} and non-linearity parameter β_t [-]. B_{max} [mm] and F_{max} [mm] are storage scaling parameters [-]. ϕ_c uses the minimum contributing area A_{min} [mm], maximum contributing area A_{max} [mm] and shape parameters α_c [-] and β_c [-] to control the shape of this distribution. ϕ_{ss} takes a similar shape as ϕ_c, using parameters α_{ss} [-] and β_{ss} [-]. P_c is the lesser of throughfall rate P_g [mm/d] minus saturation excess q_{se} [mm/d], and the catchment infiltration capacity f_s^* [mm/d]. f_s^* is assumed to have a constant rate c [mm/d]. q_{se} is determined as that part of throughfall P_g that falls on the variable contributing catchment area given by ϕ_c. P_g is determined as a fixed interception rate α_g [mm/d] and a fractional interception β_g [-]. Evaporation E_f uses the potential rate E_p [mm/d] scaled by

the relative storage in F and two shape parameters γ_f [-] and δ_f [-]. Recharge r_f [mm/d] has a linear relation with storage through time parameter t_d [d^{-1}].

$$\frac{dA}{dt} = q_{sse} + q_{sie} + q_b - E_a - q_a - r_a \tag{B.238}$$

$$q_{sse} = \frac{\phi_{ss} - \phi_c}{1 - \phi_c} P_c \tag{B.239}$$

$$q_{sie} = max\left(P_c * \frac{1 - \phi_{ss}}{1 - \phi_c} - f_{ss}^*, 0\right) \tag{B.240}$$

$$q_b = \beta_b \left(exp\left(\alpha_b \frac{B}{B_{max}}\right) - 1\right) \tag{B.241}$$

$$E_a = \phi_c * E_p + \gamma_a * E_p \left(\frac{A}{A_{max}}\right)^{\delta_a} \tag{B.242}$$

$$q_a = \begin{cases} \alpha_a \left(\frac{A - A_{min}}{A_{max} - A_{min}}\right)^{\beta_a}, & \text{if } A > A_{min} \\ 0, & \text{otherwise} \end{cases} \tag{B.243}$$

$$r_a = \phi_{ss} * f_{ss}^* \tag{B.244}$$

Where A [mm] is the current storage in the more permeable upper zone (above less permeable lower zone F), refilled by sub-surface saturation excess q_{sse} [mm/d], sub-surface infiltration excess q_{sie} [mm/d] and discharge from groundwater q_b [mm/d]. The store is drained by evaporation E_a, subsurface stormflow q_a [mm/d] and recharge r_a [mm/d]. Flow from store B, q_b, decreases exponentially as the store dries out, controlled by parameters β_b and α_b. Evaporation E_a occurs at the potential rate E_p from the variable saturated area ϕ_c and additionally at a rate scaled by the relative storage in A and two shape parameters γ_a [-] and $delta_a$ [-]/ Recharge r_a is a function of the saturated subsurface area ϕ_{ss} and the subsurface infiltration rate f_{ss}^*.

$$\frac{dB}{dt} = r_f + r_a - E_b - q_b \tag{B.245}$$

$$E_b = \gamma_b * E_p \left(\frac{B}{B_{max}}\right)^{\delta_b} \tag{B.246}$$

Where B [mm] is the current storage in the deep layers, refilled by recharge from stores A (r_a) and F (r_f), and drained by evaporation E_b and groundwater discharge q_b. E_b uses the potential rate E_p scaled by the relative storage in B and two shape parameters γ_b [-] and δ_b [-]. Total flow:

$$Q_t = q_{se} + q_{ie} + q_a \tag{B.247}$$

$$q_{ie} = P_g - q_{se} - P_c \tag{B.248}$$

Where q_{ie} $[mm/d]$ is infiltration excess on the surface.

B.2.24 MOPEX-1 (model ID: 24)

The MOPEX-1 model (fig. B.25) is part of a model improvement study that investigates the relationship between dominant processes and model structures for 197 catchments in the MOPEX database [349]. It has 4 stores and 5 parameters (S_{b1}, t_w, t_u, S_e, t_c). The model aims to represent:

- Saturation excess flow;
- Infiltration to deeper soil layers;
- A split between fast and slow runoff.

B.2.24.1 File names

Model: m_24_mopex1_5p_4s

Parameter ranges: m_24_mopex1_5p_4s_parameter_ ranges

B.2.24.2 Model equations

Figure B.25: Structure of the MOPEX-1 model
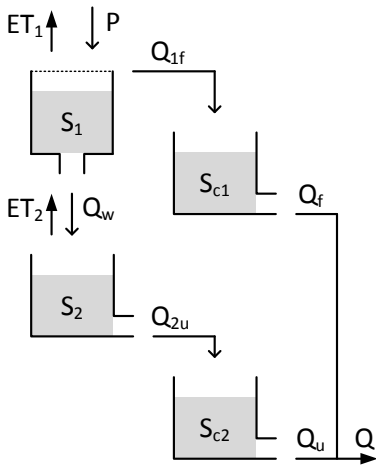
$$\frac{dS_1}{dt} = P - ET_1 - Q_{1f} - Q_w \tag{B.249}$$

$$ET_1 = \frac{S_1}{S_{b1}} * Ep \tag{B.250}$$

$$Q_{1f} = \begin{cases} P, & \text{if } S_1 \geq S_{b1} \\ 0, & \text{otherwise} \end{cases} \tag{B.251}$$

$$Q_w = t_w * S_1 \tag{B.252}$$

Where S_1 [mm] is the current storage in soil moisture and P precipitation [mm/d]. Evaporation ET_1 [mm/d] depends linearly on current soil moisture, maximum soil moisture S_{b1} [mm] and potential evapotransporation E_p [mm/d]. Saturation excess flow Q_{1f} [mm/d] occurs when the soil moisture bucket exceeds its maximum capacity. Infiltration to deeper groundwater Q_w [mm/d] depends on current soil moisture and time parameter t_w [d^{-1}].

$$\frac{dS_2}{dt} = Q_w - ET_2 - Q_{2u} \tag{B.253}$$

$$ET_2 = \frac{S_2}{S_e} * Ep \tag{B.254}$$

$$Q_{2u} = t_u * S_2 \tag{B.255}$$

Where S_2 [mm] is the current groundwater storage, refilled by infiltration from S_1. Evaporation ET_2 [mm/d] depends linearly on current groundwater and root zone storage capacity S_e [mm]. Leakage to the slow runoff store Q_{2u} [mm/d] depends on current groundwater level and time parameter t_u [d^{-1}].

$$\frac{dS_{c1}}{dt} = Q_{1f} - Q_f \tag{B.256}$$

$$Q_f = t_c * S_{c1} \tag{B.257}$$

Where S_{c1} [mm] is current storage in the fast flow routing reservoir, refilled by Q_{1f}. Routed flow Q_f depends on the mean residence time parameter t_c [d^{-1}].

$$\frac{dS_{c2}}{dt} = Q_{2u} - Q_u \tag{B.258}$$

$$Q_u = t_c * S_{c2} \tag{B.259}$$

Where S_{c2} [mm] is current storage in the slow flow routing reservoir, refilled by Q_{2u}. Routed flow Q_u depends on the mean residence time parameter t_c [d^{-1}]. Total simulated flow Q_t [mm/d]:

$$Q_t = Q_f + Q_u \tag{B.260}$$

B.2.25 Thames Catchment Model (model ID: 25)

The Thames Catchment Model (TCM) model (fig. B.26) is originally intended to be used in zones with similar surface characteristics, rather than catchments as a whole [222]. It has 4 stores and 6 parameters (ϕ, rc, γ, k_1, c_a, k_2). The model aims to represent:

- Effective rainfall before infiltration;
- Preferential recharge;
- Catchment drying through prolonged soil moisture depletion;
- Groundwater abstraction;
- Non-linear groundwater flow.

B.2.25.1 File names

Model: m_25_tcm_6p_4s
Parameter ranges: m_25_tcm_6p_4s_parameter_ ranges

B.2.25.2 Model equations

Figure B.26: Structure of the TCM model
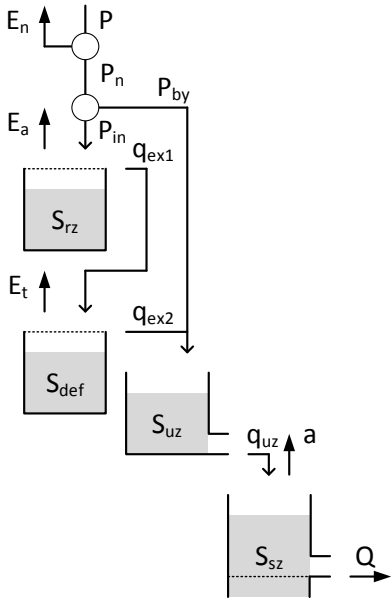
$$\frac{dS_{Rz}}{dt} = P_{in} - E_a - q_{ex1} \qquad (B.261)$$

$$P_{in} = (1 - \phi) * P_n \qquad (B.262)$$

$$P_n = max(P - E_p, 0) \qquad (B.263)$$

$$E_a = \begin{cases} E_p, & \text{if } S_{rz} > 0 \\ 0, & \text{otherwise} \end{cases} \qquad (B.264)$$

$$q_{ex1} = \begin{cases} P_{in}, & \text{if } S_{rz} > rc \\ 0, & \text{otherwise} \end{cases} \qquad (B.265)$$

Where S_{rz} [mm] is the current storage in the root zone, refilled by infiltrated precipitation P_{in} [mm/d], and drained by evaporation E_a [mm/d] and storage excess flow q_{ex1} [mm/d]. P_{in} is the fraction $(1 - \phi)$ [-] of net precipitation P_n [mm/d] that is not preferential recharge. P_n is the difference between precipitation P [mm/d] and potential evapotranspiration E_p [mm/d] per time step. E_a occurs at the net potential rate whenever possible. q_{ex1}

205

occurs only when the store is at maximum capacity rc [mm].

$$\frac{dS_{def}}{dt} = E_t + q_{ex2} - q_{ex1} \tag{B.266}$$

$$E_t = \begin{cases} \gamma * E_p, & \text{if } S_{rz} = 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.267}$$

$$q_{ex2} = \begin{cases} q_{ex1}, & \text{if } S_{def} = 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.268}$$

Where S_{def} [mm] is the current storage in the soil moisture *deficit* store. The deficit is increased by evaporation E_t [mm/d] and percolation q_{ex2} [mm/d]. The deficit is decreased by overflow from the upper store q_{ex1}. E_t only occurs when the upper zone is empty and at a fraction γ [-] of E_p. q_{ex2} only occurs when the deficit is zero.

$$\frac{dS_{uz}}{dt} = P_{by} + q_{ex2} - q_{uz} \tag{B.269}$$

$$P_{by} = \phi * P_n \tag{B.270}$$

$$q_{uz} = k_1 * S_{uz} \tag{B.271}$$

Where S_{uz} is the current storage in the unsaturated zone, refilled by preferential recharge P_{by} [mm/d] and percolation q_{ex2} [mm/d], and drained by groundwater flow q_{uz} [mm/d]. P_{by} is a fraction ϕ [-] of P_n. q_{uz} has a linear relation with storage through time parameter k_1 [d^{-1}].

$$\frac{dS_{sz}}{dt} = q_{uz} - a - Q \tag{B.272}$$

$$a = c_a \tag{B.273}$$

$$Q = \begin{cases} k_2 * S_{sz}^2, & \text{if } S_{sz} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.274}$$

Where S_{sz} [mm] is the current storage in the saturated zone, refilled by groundwater flow q_{uz} [mm/d] and drained by abstractions a [mm/d] and outflow Q [mm/d]. a occurs at a constant rate c_a [mm/d]. Abstractions can draw down the aquifer below the runoff generating threshold. Q has a quadratic relation with storage through parameter k_2 [$mm^{-1}d^{-1}$].

B.2.26 Flex-I (model ID: 26)

The Flex-I model (fig. B.27) is the part of a model development exercise [106]. It has 4 stores and 10 parameters (I_{max}, UR_{max}, β, D, $Perc_{max}$, L_p, $N_{lag,f}$, $N_{lag,s}$, K_f, K_s). The model aims to represent:

- Interception by vegetation;
- Infiltration and saturation excess flow based on a distribution of different soil depths;
- A split between fast saturation excess flow and preferential recharge to a slow store;
- Percolation from the unsaturated zone to a slow runoff store.

B.2.26.1 File names

Model: m_26_flexi_10p_4s

Parameter ranges: m_26_flexi_10p_4s_parameter_ ranges

B.2.26.2 Model equations
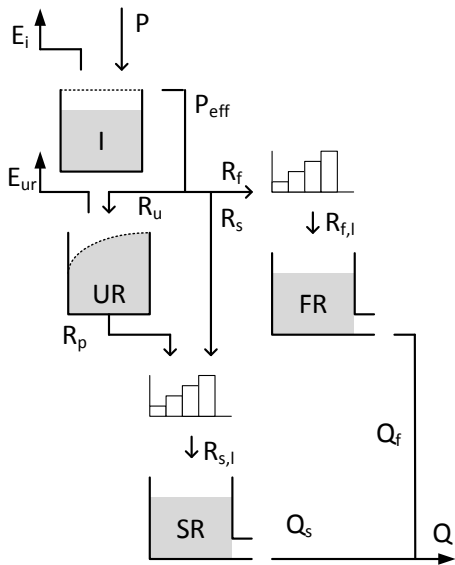
Figure B.27: Structure of the Flex-I model

$$\frac{dI}{dt} = P - E_i - P_{eff} \tag{B.275}$$

$$E_i = \begin{cases} Ep & \text{, if } I > 0 \\ 0 & \text{, otherwise} \end{cases} \tag{B.276}$$

$$P_{eff} = \begin{cases} P & \text{, if } I \geq I_{max} \\ 0 & \text{, otherwise} \end{cases} \tag{B.277}$$

Where I is the current interception storage [mm], P [mm/d] incoming precipitation, E_i [mm/d] evaporation from the interception store and P_{eff} [mm/d] interception excess routed to soil moisture. Evaporation occurs at the potential rate E_p [mm/d] whenever possible. Interception excess occurs when the interception store exceeds its maximum capacity I_{max} [mm].

$$\frac{dUR}{dt} = R_u - E_{ur} - R_p \tag{B.278}$$

$$R_U = (1 - C_r) * P_{eff} \tag{B.279}$$

$$C_r = \left[1 + exp\left(\frac{-UR/UR_{max} + 1/2}{\beta}\right)\right]^{-1} \tag{B.280}$$

$$E_{ur} = E_p * min\left(1, \frac{UR}{UR_{max}} \frac{1}{L_p}\right) \tag{B.281}$$

$$P_s = Perc_{max} * \frac{-UR}{UR_{max}} \tag{B.282}$$

Where UR is the current storage in the unsaturated zone [mm]. R_u [mm/d] is the inflow into UR based on its current storage compared to maximum storage UR_{max} [mm] and a shape distribution parameter β [-]. E_{ur} the evaporation [mm/d] from UR which follows a linear relation between current and maximum storage until a threshold L_p [-] is exceeded. P_s is the percolation from UR to the slow reservoir SR [mm/d], based on a maximum percolation rate $Perc_{max}$ [mm], relative to the fraction of current storage and maximum storage. P_{eff} is routed towards the unsaturated zone based on Cr, with the remainder being divided into preferential recharge R_s [mm/d] and fast runoff R_f [mm/d]:

$$R_s = (P_{eff} - R_u) * D \tag{B.283}$$

$$R_f = (P_{eff} - R_u) * (1 - D) \tag{B.284}$$

Where R_s and R_f are the flows [mm/d] to the slow and fast runoff reservoir respectively, based on runoff partitioning coefficient D [-]. Both are lagged by linearly increasing triangular transformation functions with parameters $N_{lag,s}$ [d] and $N_{lag,f}$ [d] respectively, that give the number of days over which R_s and R_f need to be transformed. Percolation R_p is added to R_s before the transformation to $R_{s,l}$ occurs.

$$\frac{dFR}{dt} = R_{f,l} - Q_f \tag{B.285}$$

$$Q_f = K_f * FR \tag{B.286}$$

Where FR is the current storage [mm] in the fast flow reservoir. Outflow Q_f [mm/d] from the reservoir has a linear relation with storage through time scale parameter K_f [d^{-1}].

$$\frac{dSR}{dt} = R_{s,l} - Q_s \tag{B.287}$$

$$Q_s = K_s * SR \tag{B.288}$$

Where SR is the current storage [mm] in the slow flow reservoir. Outflow Q_s [mm/d] from the reservoir has a linear relation with storage through time scale parameter K_s [d^{-1}]. Total outflow Q [mm/d]:

$$Q = Q_f + Q_s \qquad\qquad (\text{B.289})$$

B.2.27 Tank model (model ID: 27)

The Tank Model (fig. B.28) is originally developed for use constantly saturated soils in Japan [310]. It has 4 stores and 12 parameters (A_0, A_1, A_2, t_1, t_2, B_0, B_1, t_3, C_0, C_1, t_4, D_1). The model aims to represent:

- Runoff on increasing time scales with depth.

B.2.27.1 File names

Model: m_27_tank_12p_4s

Parameter ranges: m_27_tank_12p_4s_parameter_ ranges

B.2.27.2 Model equations
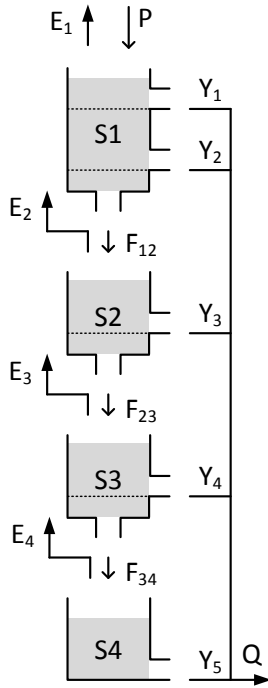
Figure B.28: Structure of the Tank Model

$$\frac{dS_1}{dt} = P - E_1 - F_{12} - Y_2 - Y_1 \tag{B.290}$$

$$E_1 = \begin{cases} Ep, & \text{if } S_1 > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.291}$$

$$F_{12} = A_0 * S_1 \tag{B.292}$$

$$Y_2 = \begin{cases} A_2 * (S_1 - t_2), & \text{if } S_1 > t_2 \\ 0, & \text{otherwise} \end{cases} \tag{B.293}$$

$$Y_1 = \begin{cases} A_1 * (S_1 - t_1), & \text{if } S_1 > t_1 \\ 0, & \text{otherwise} \end{cases} \tag{B.294}$$

Where S_1 [mm] is the current storage in the upper zone, refilled by precipitation P [mm/d] and drained by evaporation E_1 [mm/d], drainage F_{12} [mm/d] and surface runoff Y_1 [mm/d] and Y_2 [mm/d]. E_1 occurs at the potential rate E_p [mm/d] if water is available. Drainage to the intermediate layer has a linear relationship with storage through time scale parameter A_0 [d^{-1}]. Surface runoff Y_2 and Y_1 occur when S_1 is above thresholds t_2 [mm] and t_1 [mm] respectively. Both are linear relationships through time parameters A_2 [d^{-1}] and A_1 [d^{-1}] respectively.

$$\frac{dS_2}{dt} = F_{12} - E_2 - F_{23} - Y_3 \tag{B.295}$$

$$E_2 = \begin{cases} Ep, & \text{if } S_1 = 0 \ \& \ S_2 > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.296}$$

$$F_{23} = B_0 * S_2 \tag{B.297}$$

$$Y_3 = \begin{cases} B_1 * (S_2 - t_3), & \text{if } S_2 > t_3 \\ 0, & \text{otherwise} \end{cases} \tag{B.298}$$

Where S_2 [mm] is the current storage in the intermediate zone, refilled by drainage F_{12} from the upper zone and drained by evaporation E_2 [mm/d], drainage F_{23} [mm/d] and intermediate discharge Y_3 [mm/d]. E_2 occurs at the potential rate E_p if water is available and the upper zone is empty. Drainage to the third layer F_{23} has a linear relationship with storage through time scale parameter B_0 [d^{-1}]. Intermediate runoff Y_3 occurs when S_2 is above threshold t_3 [mm] and has a linear relationship with storage through time scale parameter B_1 [d^{-1}].

$$\frac{dS_3}{dt} = F_{23} - E_3 - F_{34} - Y_4 \tag{B.299}$$

$$E_3 = \begin{cases} Ep, & \text{if } S_1 = 0 \ \& \ S_2 = 0 \ \& \ S_3 > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.300}$$

$$F_{34} = C_0 * S_3 \tag{B.301}$$

$$Y_4 = \begin{cases} C_1 * (S_3 - t_4), & \text{if } S_3 > t_4 \\ 0, & \text{otherwise} \end{cases} \tag{B.302}$$

Where S_3 [mm] is the current storage in the sub-base zone, refilled by drainage F_{23} from the intermediate zone and drained by evaporation E_3 [mm/d], drainage F_{34} [mm/d] and sub-base discharge Y_4 [mm/d]. E_3 occurs at the potential rate E_p if water is available and the upper zones are empty. Drainage to the fourth layer F_{34} has a linear relationship with storage through time scale parameter C_0 [d^{-1}]. Sub-base runoff Y_4 occurs when S_3 is above threshold t_4 [mm] and has a linear relationship with storage through time scale parameter C_1 [d^{-1}].

$$\frac{dS_4}{dt} = F_{34} - E_4 - Y_5 \tag{B.303}$$

$$E_4 = \begin{cases} Ep, & \text{if } S_1 = 0 \ \& \ S_2 = 0 \ \& \ S_3 = 0 \ \& \ S_4 > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.304}$$

$$Y_5 = D_1 * S_4 \tag{B.305}$$

Where S_4 [mm] is the current storage in the base layer, refilled by drainage F_{34} from the sub-base zone and drained by evaporation E_4 [mm/d] and baseflow Y_5 [mm/d]. E_4 occurs at the potential rate E_p if water is available and the upper zones are empty. Baseflow Y_5 has a linear relationship with storage through time scale parameter D_1 [d^{-1}]. Total runoff:

$$Q_t = Y_1 + Y_2 + Y_3 + Y_4 + Y_5 \tag{B.306}$$

B.2.28 Xinanjiang model (model ID: 28)

The Xinanjiang model (fig. B.29) is originally intended for use in humid or semi-humid regions in China [351]. The model uses a variable contributing area to simulate runoff. The version presented here uses a double parabolic curve to simulate tension water capacities within the catchment [153], instead of the original single parabolic curve. The model has 4 stores and 12 parameters (A_{im}, a, b, W_{max}, LM, c, S_{max}, Ex, k_I, k_G, c_I, c_G). The model aims to represent:

- Runoff from impervious areas;
- Variable distribution of tension water storage capacities in the catchment;
- Variable contributing area of free water storages;
- Direct surface runoff from the contributing free area;
- Delayed interflow and baseflow from the contributing free area.

B.2.28.1 File names

Model: m_28_xinanjiang_12p_4s

Parameter ranges: m_28_xinanjiang_12p_4s_parameter_ ranges

B.2.28.2 Model equations
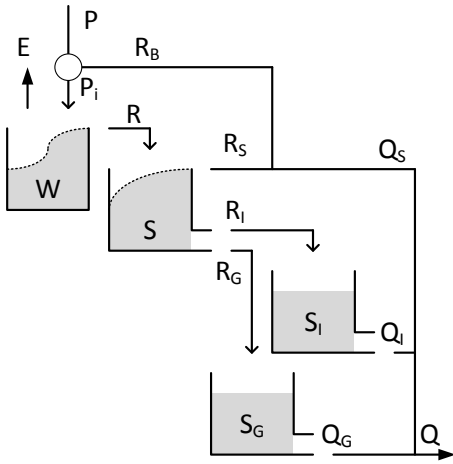
$$\frac{dW}{dt} = P_i - E - R \tag{B.307}$$

$$P_i = (1 - A_{im}) * P \tag{B.308}$$

$$R = \begin{cases} P_i * \left[(0.5 - a)^{1-b} \left(\frac{W}{W_{max}} \right)^b \right], & \text{if } \frac{W}{W_{max}} \leq 0.5 - a \\ P_i * \left[1 - (0.5 + a)^{1-b} \left(1 - \frac{W}{W_{max}} \right)^b \right], & \text{otherwise} \end{cases} \tag{B.309}$$

$$E = \begin{cases} E_p, & \text{if } W > LM \\ \frac{W}{LM} E_p, & \text{if } c * LM \geq W \leq LM \\ c * E_p, & \text{otherwise} \end{cases} \tag{B.310}$$

Figure B.29: Structure of the Xinan-jiang model

Where W [mm] is the current tension water storage, refilled by a infiltration P_i [mm/d] and drained by evaporation E [mm/d] and runoff R [mm/d]. P_i is the fraction of pre-

cipitation P [mm/d] that does not fall on impervious area A_{im} [-]. Runoff generation R uses a double parabolic curve to determine the fraction of catchment area that is at full tension storage and thus can contribute to runoff generation. This curve relies on shape parameters a [-] and b [-], and maximum tension water storage W_{max} [mm]. Evaporation rate E declines as tension water storage decreases. Evaporation occurs at the potential rate E_p [mm/d] if storage W is above threshold LM [mm], and reduces linearly below that up to a second threshold $c * LM$ [-]*[mm]. Below this threshold evaporation occurs at a constant rate $c * E_p$.

$$\frac{dS}{dt} = R - R_S - R_I - R_G \tag{B.311}$$

$$R_S = R * \left(1 - \left(1 - \frac{S}{S_{max}}\right)^{Ex}\right) \tag{B.312}$$

$$R_I = k_I * S * \left(1 - \left(1 - \frac{S}{S_{max}}\right)^{Ex}\right) \tag{B.313}$$

$$R_G = k_G * S * \left(1 - \left(1 - \frac{S}{S_{max}}\right)^{Ex}\right) \tag{B.314}$$

Where S [mm] is the current storage of free water, refilled by runoff R from filled tension water areas, and drained by surface runoff R_S [mm/d], interflow R_I [mm/d] and baseflow R_G [mm/d]. All runoff components rely on a parabolic equation to simulate variable contributing areas of the catchment, dependent on maximum free water storage S_{max} [mm] and shape parameter Ex [-]. R_I also uses a time coeficient k_I [d^{-1}]. R_G uses a time coeficient k_G [d^{-1}].

$$\frac{dS_I}{dt} = R_I - Q_I \tag{B.315}$$

$$Q_I = c_I * S_I \tag{B.316}$$

Where S_I [mm] is the current storage in the interflow routing reservoir, filled by interflow from free water R_I and drained by delayed interflow Q_I [mm/d]. Q_I uses a time coefficient c_I [d^{-1}].

$$\frac{dS_G}{dt} = R_G - Q_G \tag{B.317}$$

$$Q_G = c_G * S_G \tag{B.318}$$

Where S_G [mm] is the current storage in the baseflow routing reservoir, filled by baseflow from free water R_G and drained by delayed baseflow Q_G [mm/d]. Q_G uses a time coefficient c_G [d^{-1}]. Total flow depends on four separate runoff components:

$$Q_t = Q_S + Q_I + Q_G \tag{B.319}$$

$$Q_S = R_S + R_B \tag{B.320}$$

$$R_B = A_{im} * P \tag{B.321}$$

Where R_B [mm/d] is direct rainoff generated by precipitation P [mm/d] on the fraction impervious area A_{im} [-].

B.2.29 HyMOD (model ID: 29)

The HyMOD model (fig. B.30) combines a PDM-like soil moisture routine [e.g. 221]) with a Nash cascade of three linear reservoirs that simulates fast flow and a single linear reservoir intended to simulate slow flow [57, 330]. Although the model was originally intended as a flexible structure where the user defines which processes to include, this study includes only a single version that is commonly used. It has 5 parameters (S_{max}, b, a, k_f and k_s) and 5 stores. The model aims to represent:

- Different soil depths throughout the catchment;
- Separation of flow into fast and slow flow.

B.2.29.1 File names

Model: m_29_hymod_5p_5s
Parameter ranges: m_29_hymod_5p_5s_parameter_ ranges

B.2.29.2 Model equations
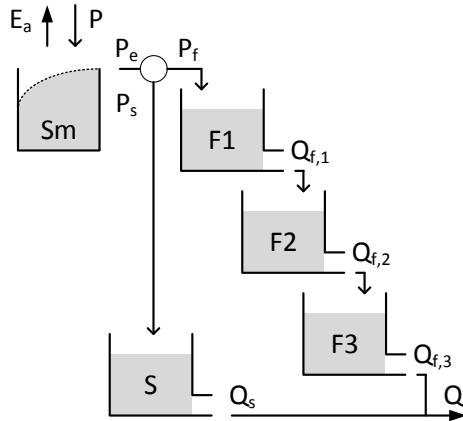
Figure B.30: Structure of the HyMOD model

$$\frac{dSm}{dt} = P - E_a - P_e \tag{B.322}$$

$$E_a = \frac{Sm}{S_{max}} * E_p \tag{B.323}$$

$$P_e = \left(1 - \left(1 - \frac{S}{S_{max}}\right)^b\right) * P \tag{B.324}$$

Where Sm is the current storage in Sm [mm], S_{max} [mm] is the maximum storage in Sm, E_a and E_p the actual and potential evapotranspiration respectively [mm/d] and b is the soil depth distribution parameter [-]. P [mm/d] is the precipitation input.

$$\frac{dF_1}{dt} = P_f - Q_{f,1} \tag{B.325}$$

$$P_f = a * P_e \tag{B.326}$$

$$Q_{f,1} = k_f * S_{f,1} \tag{B.327}$$

Where F_1 is the current storage in store F_1 [mm], a the fraction of P_e that flows into the fast stores and k_f the runoff coefficient of the fast stores. Stores F_2 and F_3 take the outflow

of the previous store as input ($Q_{f,1}$ and $Q_{f,2}$ respectively) and generate outflow analogous to the equations above.

$$\frac{dS}{dt} = P_s - Q_s \tag{B.328}$$

$$P_s = (1-a) * P_e \tag{B.329}$$

$$Q_s = k_s * S \tag{B.330}$$

Where S is the current storage in store S [mm], $1-a$ [-] the fraction of P_e that flows into the slow store and k_s the runoff coefficient of the slow store. Total outflow:

$$Q_t = Q_s + Q_{f,3} \tag{B.331}$$

B.2.30 MOPEX-2 (model ID: 30)

The MOPEX-2 model (fig. B.31) is part of a model improvement study that investigates the relationship between dominant processes and model structures for 197 catchments in the MOPEX database [349]. It has 5 stores and 7 parameters ($T_{crit}, ddf, S_{b1}, t_w, t_u, S_e, t_c$). The model aims to represent:

- Snow accumulation and melt;
- Saturation excess flow;
- Infiltration to deeper soil layers;
- A split between fast and slow runoff.

B.2.30.1 File names

Model: m_30_mopex2_7p_5s

Parameter ranges: m_30_mopex2_7p_5s_parameter_ ranges

B.2.30.2 Model equations

Figure B.31: Structure of the MOPEX-2 model
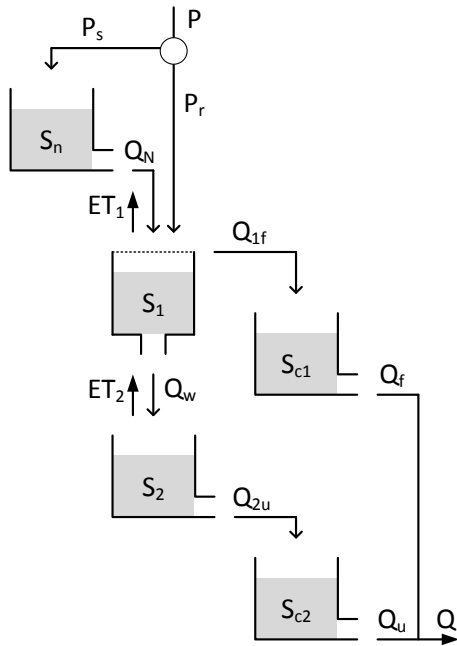
$$\frac{dS_n}{dt} = P_s - Q_n \tag{B.332}$$

$$P_s = \begin{cases} P, & \text{if } T \le T_{crit} \\ 0, & \text{otherwise} \end{cases} \tag{B.333}$$

$$Q_n = \begin{cases} ddf * (T - T_{crit}), & \text{if } T > T_{crit} \\ 0, & \text{otherwise} \end{cases} \tag{B.334}$$

Where S_n [mm] is the current snow pack. Precipitation occurs as snowfall P_s [mm/d] when current temperature T [^oC] is below threshold T_{crit} [^oC]. Snowmelt Q_N [mm/d] occurs when the temperature rises above the threshold temperature and relies in the degree-day factor dd [$mm/^oC/d$].

218

$$\frac{dS_1}{dt} = P_r - ET_1 - Q_{1f} - Q_w \tag{B.335}$$

$$P_r = \begin{cases} P, & \text{if } T > T_{crit} \\ 0, & \text{otherwise} \end{cases} \tag{B.336}$$

$$ET_1 = \frac{S_1}{S_{b1}} * Ep \tag{B.337}$$

$$Q_{1f} = \begin{cases} P, & \text{if } S_1 \geq S_{b1} \\ 0, & \text{otherwise} \end{cases} \tag{B.338}$$

$$Q_w = t_w * S_1 \tag{B.339}$$

Where S_1 [mm] is the current storage in soil moisture and P_r precipitation as rain [mm/d]. Evaporation ET_1 [mm/d] depends linearly on current soil moisture, maximum soil moisture S_{b1} [mm] and potential evapotranspiration E_p [mm/d]. Saturation excess flow Q_{1f} [mm/d] occurs when the soil moisture bucket exceeds its maximum capacity. Infiltration to deeper groundwater Q_w [mm/d] depends on current soil moisture and time parameter t_w [d^{-1}].

$$\frac{dS_2}{dt} = Q_w - ET_2 - Q_{2u} \tag{B.340}$$

$$ET_2 = \frac{S_2}{S_e} * Ep \tag{B.341}$$

$$Q_{2u} = t_u * S_2 \tag{B.342}$$

Where S_2 [mm] is the current groundwater storage, refilled by infiltration from S_1. Evaporation ET_2 [mm/d] depends linearly on current groundwater and root zone storage capacity S_e [mm]. Leakage to the slow runoff store Q_{2u} [mm/d] depends on current groundwater level and time parameter t_u [d^{-1}].

$$\frac{dS_{c1}}{dt} = Q_{1f} - Q_f \tag{B.343}$$

$$Q_f = t_c * S_{c1} \tag{B.344}$$

Where S_{c1} [mm] is current storage in the fast flow routing reservoir, refilled by Q_{1f}. Routed flow Q_f depends on the mean residence time parameter t_c [d^{-1}].

$$\frac{dS_{c2}}{dt} = Q_{2u} - Q_u \tag{B.345}$$

$$Q_u = t_c * S_{c2} \tag{B.346}$$

Where S_{c2} [mm] is current storage in the slow flow routing reservoir, refilled by Q_{2u}. Routed flow Q_u depends on the mean residence time parameter t_c [d^{-1}]. Total simulated flow Q_t [mm/d]:

$$Q_t = Q_f + Q_u \tag{B.347}$$

B.2.31 MOPEX-3 (model ID: 31)

The MOPEX-3 model (fig. B.32) is part of a model improvement study that investigates the relationship between dominant processes and model structures for 197 catchments in the MOPEX database [349]. It has 5 stores and 8 parameters (T_{crit}, ddf, S_{b1}, t_w, S_{b2}, t_u, S_e, t_c). The model aims to represent:

- Snow accumulation and melt;
- Saturation excess flow;
- Infiltration to deeper soil layers;
- Subsurface-influenced fast flow;
- A split between fast and slow runoff.

B.2.31.1 File names

Model: m_31_mopex3_8p_5s

Parameter ranges: m_31_mopex3_8p_5s_parameter_ ranges

B.2.31.2 Model equations
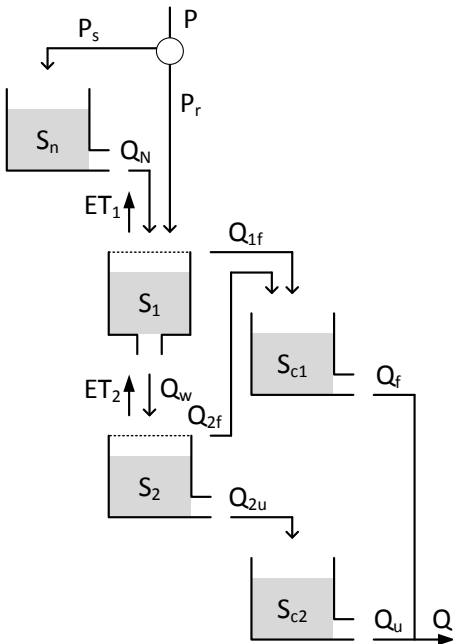
$$\frac{dS_n}{dt} = P_s - Q_n \tag{B.348}$$

$$P_s = \begin{cases} P, & \text{if } T \le T_{crit} \\ 0, & \text{otherwise} \end{cases} \tag{B.349}$$

$$Q_n = \begin{cases} ddf * (T - T_{crit}), & \text{if } T > T_{crit} \\ 0, & \text{otherwise} \end{cases} \tag{B.350}$$

Where S_n [mm] is the current snow pack. Precipitation occurs as snowfall P_s [mm/d] when current temperature T [^oC] is below threshold T_{crit} [^oC]. Snowmelt Q_N [mm/d] occurs when the temperature rises above the threshold temperature and relies in the degree-day factor dd [$mm/^oC/d$].

Figure B.32: Structure of the MOPEX-3 model

$$\frac{dS_1}{dt} = P_r - ET_1 - Q_{1f} - Q_w \tag{B.351}$$

$$P_r = \begin{cases} P, & \text{if } T > T_{crit} \\ 0, & \text{otherwise} \end{cases} \tag{B.352}$$

$$ET_1 = \frac{S_1}{S_{b1}} * Ep \tag{B.353}$$

$$Q_{1f} = \begin{cases} P, & \text{if } S_1 \geq S_{b1} \\ 0, & \text{otherwise} \end{cases} \tag{B.354}$$

$$Q_w = t_w * S_1 \tag{B.355}$$

Where S_1 [mm] is the current storage in soil moisture and P_r precipitation as rain [mm/d]. Evaporation ET_1 [mm/d] depends linearly on current soil moisture, maximum soil moisture S_{b1} [mm] and potential evapotranspiration E_p [mm/d]. Saturation excess flow Q_{1f} [mm/d] occurs when the soil moisture bucket exceeds its maximum capacity. Infiltration to deeper groundwater Q_w [mm/d] depends on current soil moisture and time parameter t_w [d^{-1}].

$$\frac{dS_2}{dt} = Q_w - ET_2 - Q_{2u} - Q_{2f} \tag{B.356}$$

$$ET_2 = \frac{S_2}{S_e} * Ep \tag{B.357}$$

$$Q_{2u} = t_u * S_2 \tag{B.358}$$

$$Q_{2f} = \begin{cases} Q_w, & \text{if } S_2 \geq S_{b2} \\ 0, & \text{otherwise} \end{cases} \tag{B.359}$$

Where S_2 [mm] is the current groundwater storage, refilled by infiltration from S_1. Evaporation ET_2 [mm/d] depends linearly on current groundwater and root zone storage capacity S_e [mm]. Leakage to the slow runoff store Q_{2u} [mm/d] depends on current groundwater level and time parameter t_u [d^{-1}]. When the store reaches maximum capacity S_{b2} [mm], excess flow Q_{2f} [mm/d] is routed towards the fast response routing store.

$$\frac{dS_{c1}}{dt} = Q_{1f} + Q_{2f} - Q_f \tag{B.360}$$

$$Q_f = t_c * S_{c1} \tag{B.361}$$

Where S_{c1} [mm] is current storage in the fast flow routing reservoir, refilled by Q_{1f} and Q_{2f}. Routed flow Q_f depends on the mean residence time parameter t_c [d^{-1}].

$$\frac{dS_{c2}}{dt} = Q_{2u} - Q_u \tag{B.362}$$

$$Q_u = t_c * S_{c2} \tag{B.363}$$

Where S_{c2} [mm] is current storage in the slow flow routing reservoir, refilled by Q_{2u}. Routed flow Q_u depends on the mean residence time parameter t_c [d^{-1}]. Total simulated flow Q_t [mm/d]:

$$Q_t = Q_f + Q_u \tag{B.364}$$

B.2.32 MOPEX-4 (model ID: 32)

The MOPEX-4 model (fig. B.33) is part of a model improvement study that investigates the relationship between dominant processes and model structures for 197 catchments in the MOPEX database [349]. It has 5 stores and 10 parameters (T_{crit}, ddf, S_{b1}, t_w, I_α, I_s, S_{b2}, t_u, S_e, t_c). The original model relies on observations of Leaf Area Index and a calibrated interception fraction. Liang et al [194] show typical Leaf Area Index time series, and a sinusoidal function is a reasonable approximation of this. Therefore, the model is slightly modified to use a calibrated sinusoidal function, so that the data input requirements for MOPEX-4 are consistent with other models. The model aims to represent:

- Snow accumulation and melt;
- Time-varying interception;
- Saturation excess flow;
- Infiltration to deeper soil layers;
- A split between fast and slow runoff.

B.2.32.1 File names

Model: m_32_mopex4_10p_5s

Parameter ranges: m_32_mopex4_10p_5s_parameter_ ranges

B.2.32.2 Model equations
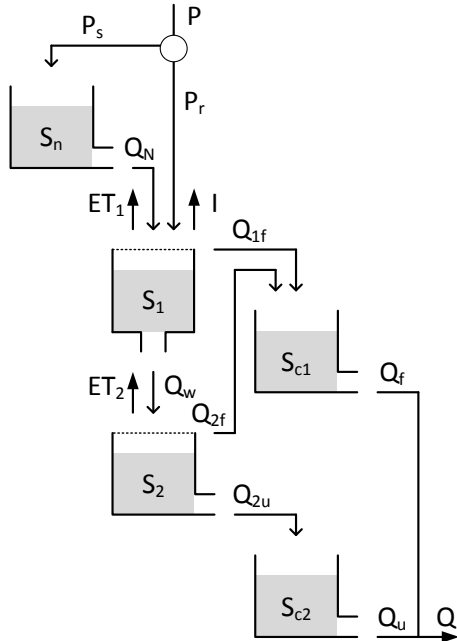
$$\frac{dS_n}{dt} = P_s - Q_n \tag{B.365}$$

$$P_s = \begin{cases} P, & \text{if } T \leq T_{crit} \\ 0, & \text{otherwise} \end{cases} \tag{B.366}$$

$$Q_n = \begin{cases} ddf * (T - T_{crit}), & \text{if } T > T_{crit} \\ 0, & \text{otherwise} \end{cases} \tag{B.367}$$

Where S_n [mm] is the current snow pack. Precipitation occurs as snowfall P_s [mm/d] when current temperature T [^oC] is below threshold T_{crit} [^oC]. Snowmelt Q_N [mm/d] occurs when the temperature rises above the threshold temperature and relies in the degree-day factor dd [$mm/^oC/d$].

Figure B.33: Structure of the MOPEX-4 model

$$\frac{dS_1}{dt} = P_r - ET_1 - I - Q_{1f} - Q_w \tag{B.368}$$

$$P_r = \begin{cases} P, & \text{if } T > T_{crit} \\ 0, & \text{otherwise} \end{cases} \tag{B.369}$$

$$ET_1 = \frac{S_1}{S_{b1}} * Ep \tag{B.370}$$

$$I = max\left(0, I_\alpha + (1 - I_\alpha)cos\left(2\pi \frac{t - I_s}{t_{max}}\right)\right) * P_r \tag{B.371}$$

$$Q_{1f} = \begin{cases} P, & \text{if } S_1 \geq S_{b1} \\ 0, & \text{otherwise} \end{cases} \tag{B.372}$$

$$Q_w = t_w * S_1 \tag{B.373}$$

Where S_1 [mm] is the current storage in soil moisture and P_r precipitation as rain [mm/d]. Evaporation ET_1 [mm/d] depends linearly on current soil moisture, maximum soil moisture S_{b1} [mm] and potential evapotranspiration E_p [mm/d]. Interception I [mm/d] depends on the mean intercepted fraction I_α [-], the maximum Leaf Area Index timing I_s [d] and the length of the seasonal cycle t_{max} [d] (usually set at 365 days). Saturation excess flow Q_{1f} [mm/d] occurs when the soil moisture bucket exceeds its maximum capacity. Infiltration to deeper groundwater Q_w [mm/d] depends on current soil moisture and time parameter t_w [d^{-1}].

$$\frac{dS_2}{dt} = Q_w - ET_2 - Q_{2u} - Q_{2f} \tag{B.374}$$

$$ET_2 = \frac{S_2}{S_e} * Ep \tag{B.375}$$

$$Q_{2u} = t_u * S_2 \tag{B.376}$$

$$Q_{2f} = \begin{cases} Q_w, & \text{if } S_2 \geq S_{b2} \\ 0, & \text{otherwise} \end{cases} \tag{B.377}$$

Where S_2 [mm] is the current groundwater storage, refilled by infiltration from S_1. Evaporation ET_2 [mm/d] depends linearly on current groundwater and root zone storage capacity S_e [mm]. Leakage to the slow runoff store Q_{2u} [mm/d] depends on current groundwater level and time parameter t_u [d^{-1}]. When the store reaches maximum capacity S_{b2} [mm], excess flow Q_{2f} [mm/d] is routed towards the fast response routing store.

$$\frac{dS_{c1}}{dt} = Q_{1f} + Q_{2f} - Q_f \tag{B.378}$$

$$Q_f = t_c * S_{c1} \tag{B.379}$$

Where S_{c1} [mm] is current storage in the fast flow routing reservoir, refilled by Q_{1f} and Q_{2f}. Routed flow Q_f depends on the mean residence time parameter t_c [d^{-1}].

$$\frac{dS_{c2}}{dt} = Q_{2u} - Q_u \tag{B.380}$$

$$Q_u = t_c * S_{c2} \tag{B.381}$$

Where S_{c2} [mm] is current storage in the slow flow routing reservoir, refilled by Q_{2u}. Routed flow Q_u depends on the mean residence time parameter t_c [d^{-1}]. Total simulated flow Q_t [mm/d]:

$$Q_t = Q_f + Q_u \tag{B.382}$$

B.2.33 SACRAMENTO model (model ID: 33)

The SACRAMENTO model (fig. B.34) is part of an ongoing model development project by the National Weather Service, which started several decades ago [60, 225]. The documentation mentions a specific order of flux computations. For consistency with other models, here all fluxes are computed simultaneously. It has 5 stores and 13 parameters ($PCTIM$, $UZTWM$, $UZFWM$, k_{uz}, $PBASE$, $ZPERC$, $REXP$, $LZTWM$, $LZFWPM$, $LZFWSM$, $PFREE$, k_{lzp}, k_{lzs}). The model also uses several coefficients derived from the calibration parameters [182]: $PBASE$ and $ZPERC$. The model aims to represent:

- Impervious and direct runoff;
- Within soil division of water storage between tension and free water;
- Surface runoff, interflow and percolation to deeper soil layers;
- Multiple baseflow processes.

B.2.33.1 File names

Model: m_33_sacramento_11p_5s

Parameter ranges: m_33_sacramento_11p_5s_parameter_ ranges

B.2.33.2 Model equations
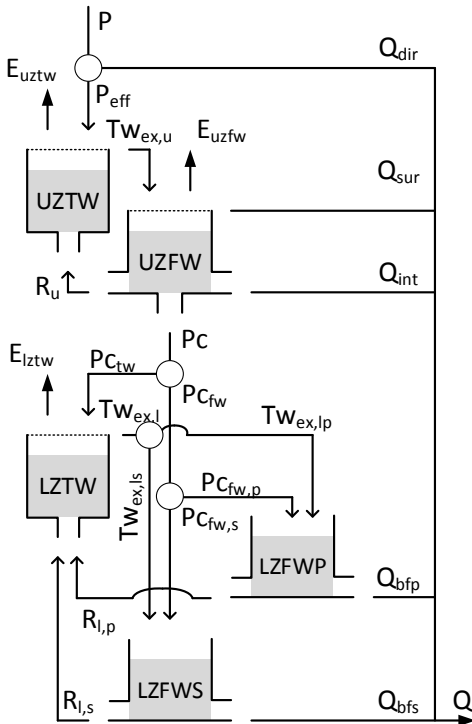
$$\frac{dUZTW}{dt} = P_{eff} + R_u - E_{uztw} - Tw_{ex,u} \qquad (B.383)$$

$$P_{eff} = (1 - PCTIM) * P \qquad (B.384)$$

$$R_u = \begin{cases} \dfrac{UZTWM * UZFW - UZFWM * UZTW}{UZTWM + UZFWM}, \\ \qquad \text{if } \dfrac{UZTW}{UZTWM} < \dfrac{UZFW}{UZFWM} \\ 0, \quad \text{otherwise} \end{cases}$$

$$(B.385)$$

$$E_{uztw} = \frac{UZTW}{UZTWM} * E_p \qquad (B.386)$$

$$Tw_{ex,u} = \begin{cases} P_{eff}, & \text{if } UZTW = UZTWM \\ 0, & \text{otherwise} \end{cases}$$

$$(B.387)$$

Where $UZTW$ [mm] is upper zone tension water, refilled by effective precipitation P_{eff} [mm/d]

Figure B.34: Structure of the SACRA-MENTO model

227

and redistribution of free water R_u [mm/d], and is drained by evaporation E_{uztw} [mm/d] and tension water excess $Tw_{ex,u}$ [mm/d]. P_{eff} is the fraction $(1-PCTIM)$ [-] of precipitation P that does not fall on impervious fraction $PCTIM$ [-]. R_u is only active when the relative deficit in tension water is greater than that in free water, and rebalances the available water in the upper zone. This uses the current storages, $UZTW$ and $UZFW$, and maximum storages, $UZTWM$ [mm] and $UZFWM$ [mm], of tension and free water stores respectively. Evaporation is determined with a linear relation between available, maximum upper zone tension storage and potential evapotranspiration E_p [mm/d]. $Tw_{ex,u}$ occurs only when the store is at maximum capacity.

$$\frac{dUZFW}{dt} = Tw_{ex,u} - E_{uzfw} - Q_{sur} - Q_{int} - Pc - R_u \tag{B.388}$$

$$E_{uzfw} = \begin{cases} E_p - E_{uztw}, & \text{if } UZFW > 0 \text{ \& } E_p > E_{uztw} \\ 0, & \text{otherwise} \end{cases} \tag{B.389}$$

$$Q_{sur} = \begin{cases} Tw_{ex,u}, & \text{if } UZFW = UZFWM \\ 0, & \text{otherwise} \end{cases} \tag{B.390}$$

$$Q_{int} = k_{uz} * UZFW \tag{B.391}$$

$$Pc = Pc_{demand} * \frac{UZFW}{UZFWM} \tag{B.392}$$

$$Pc_{demand} = PBASE * \left(1 + ZPERC * \left(\frac{\sum LZ_{deficiency}}{\sum LZ_{capacity}}\right)^{1+REXP}\right) \tag{B.393}$$

$$LZ_{deficiency} = [LZTWM - LZTW] + [LZFWPM - LZWFP] + [LZFWSM - LZFWS] \tag{B.394}$$

$$LZ_{capacity} = LZTWM + LZFWPM + LZFWSM \tag{B.395}$$

Where $UZFW$ [mm] is upper zone free water, refilled by excess water $Tw_{ex,u}$ that can not be stored as tension water, and drained by evaporation E_{uzfw} [mm/d], surface runoff Q_{sur} [mm/d], interflow Q_{int} [mm/d], and percolation to deeper groundwater Pc [mm/d]. Evaporative demand unmet by the upper tension water store is taken from upper free water storage at the potential rate. Q_{sur} occurs only when the store is at maximum capacity

$UZFWM$ [mm]. Q_{int} uses time coefficient k_{uz} [d^{-1}] to simulate interflow. Percolation Pc is calculated as a balance between the fraction water availability in upper zone free storage, and demand from the lower zone Pc_{demand}. The demand can be between a base percolation rate $PBASE$ [mm/d] and an upper limit of $ZPERC$ [-] times $PBASE$. This demand is scaled by the relative size of lower zone moisture deficiencies, expressed as the ratio between total deficiency and maximum lower zone storage. $LZTWM$ [mm], $LZFWP$ [mm], $LZFWS$ [mm] are the maximum capacity of the lower zone tension store, primary free water store and supplemental free water store respectively. The lower zone percolation demand is potentially non-linear through exponent $REGX$ [-]. $PBASE$ is calculated as $k_{lzp} * LZFWPM + K_{lzs} * LZFWSM$.

$$\frac{dLZTW}{dt} = Pc_{tw} + R_{l,p} + R_{l,s} - E_{lztw} - Tw_{ex,l} \tag{B.396}$$

$$Pc_{tw} = (1 - PFREE) * Pc \tag{B.397}$$

$$R_{l,p} = \begin{cases} LZFWPM * \frac{-LZTW(LZFWPM + LZFWSM) + LZTWM(LZWFP + LZWFS)}{(LZFWPM + LZFWSM)(LZTWM + LZFWPM + LZFWSM)}, \\ \qquad\qquad\qquad\qquad \text{if } \frac{LZTW}{LZTWM} < \frac{LZFWP + LZFWS}{LZFWPM + LZFWSM} \\ 0, \quad \text{otherwise} \end{cases}$$

$$\tag{B.398}$$

$$R_{l,s} = \begin{cases} LZFWSM * \frac{-LZTW(LZFWPM + LZFWSM) + LZTWM(LZWFP + LZWFS)}{(LZFWPM + LZFWSM)(LZTWM + LZFWPM + LZFWSM)}, \\ \qquad\qquad\qquad\qquad \text{if } \frac{LZTW}{LZTWM} < \frac{LZFWP + LZFWS}{LZFWPM + LZFWSM} \\ 0, \quad \text{otherwise} \end{cases}$$

$$\tag{B.399}$$

$$E_{lztw} = \begin{cases} \left(E_p - E_{uztw} - E_{uzfw}\right) * \frac{LZTW}{UZTWM + LZTWM}, & \text{if } LZTW > 0 \ \& \ E_p > (E_{uztw} + E_{uzfw}) \\ 0, & \text{otherwise} \end{cases}$$

$$\tag{B.400}$$

$$Tw_{ex,l} = \begin{cases} Pc_{tw}, & \text{if } LZTW = LZTWM \\ 0, & \text{otherwise} \end{cases} \tag{B.401}$$

Where $LZTW$ [mm] is lower zone tension water, refilled by percolation Pc_{tw} [mm/d] and drained by evaporation E_{lztw} [mm/d] and tension water excess $Tw_{ex,l}$ [mm/d]. Evaporative demand unmet b the upper zone can be satisfied from the lower zone tension water store, scaled by the current lower zone storage relative to total tension zone storage. Both $R_{l,p}$ and $R_{l,s}$ are only active when the relative deficit in tension water is greater than that in free water, and rebalances the available water in the lower zone. This uses the

current storages, $LZTW$, $LZFWP$ and $LZFWS$, and maximum storages, $LZTWM$ [mm], $LZFWPM$ [mm] and $LZFWSM$ [mm], of the tension and free water stores respectively. Pc_{tw} is the fraction $(1 - PFREE)$ [-] of percolation Pc that does not go into free storage. $Tw_{ex,l}$ occurs only when the store is at maximum capacity $LZTWM$ [mm].

$$\frac{dLZFWP}{dt} = Pc_{fw,p} + Tw_{ex,lp} - Q_{bfp} \tag{B.402}$$

$$Pc_{fw,p} = \left[\frac{LZFWPM - LZFWP}{LZFWPM \left(\frac{LZFWPM - LZFWP}{LZFWPM} + \frac{LZFWSM - LZFWS}{LZFWSM} \right)} \right] * (PFREE * Pc) \tag{B.403}$$

$$Tw_{ex,lp} = \left[\frac{LZFWPM - LZFWP}{LZFWPM \left(\frac{LZFWPM - LZFWP}{LZFWPM} + \frac{LZFWSM - LZFWS}{LZFWSM} \right)} \right] * Tw_{ex,l} \tag{B.404}$$

$$Q_{bfp} = k_{lzp} * LZFWP \tag{B.405}$$

Where $LZFWP$ [mm] is current storage in the primary lower zone free water store, refilled by excess tension water $TW_{ex,lp}$ [mm/d] and percolation $Pc_{fw,p}$ [mm/d] and drained by primary baseflow Q_{bfp} [mm/d]. Refilling of both lower zone free water stores (primary and supplemental) is divided between the two based on their relative, scaled moisture deficiency. Percolation from the upper zone $Pc_{fw,p}$ is scaled according to the relative current moisture deficit $\frac{LZFWPM - LZFWP}{LZFWM}$ compared to the total relative deficit in the lower free water stores $\left(\frac{LZFWPM - LZFWP}{LZFWPM} + \frac{LZFWSM - LZFWS}{LZFWSM} \right)$. $Tw_{ex,lp}$ is a similarly scaled part of $Tw_{ex,l}$. Q_{bfp} uses time parameter K_{lzp} [d^{-1}] to estimate primary baseflow.

$$\frac{dLZFWS}{dt} = Pc_{fw,s} + Tw_{ex,ls} - Q_{bfs} \tag{B.406}$$

$$Pc_{fw,s} = \left[\frac{LZFWSM - LZFWS}{LZFWSM \left(\frac{LZFWPM - LZFWP}{LZFWPM} + \frac{LZFWSM - LZFWS}{LZFWSM} \right)} \right] * (PFREE * Pc) \tag{B.407}$$

$$Tw_{ex,ls} = \left[\frac{LZFWSM - LZFWS}{LZFWSM \left(\frac{LZFWPM - LZFWP}{LZFWPM} + \frac{LZFWSM - LZFWS}{LZFWSM} \right)} \right] * Tw_{ex,l} \tag{B.408}$$

$$Q_{bfs} = k_{lzs} * LZFWS \tag{B.409}$$

Where $LZFWS$ [mm] is current storage in the supplemental free water lower zone store, refilled by excess tension water $TW_{ex,ls}$ [mm/d] and percolation $Pc_{fw,s}$ [mm/d], and drained by supplemental baseflow Q_{bfs} [mm/d]. $Pc_{fw,s}$ is determined based on relative

deficits in the lower zone free stores, as is $Tw_{ex,ls}$. Q_{bfs} uses time parameter K_{lzs} $[d^{-1}]$ to estimate supplementary baseflow. Total simulated outflow:

$$Q_t = Q_{dir} + Q_{sur} + Q_{int} + Q_{bfp} + Q_{bfs} \tag{B.410}$$

231

B.2.34 FLEX-IS (model ID: 34)

The FLEX-IS model (fig. B.35) is a combination of the FLEX-B model expanded with an interception (I) routine [106] and a snow (S) module [233]. It has 5 stores and 12 parameters ($TT, ddf, I_{max}, UR_{max}, \beta, L_p, Perc_{max}, D, N_{lag,f}, N_{lag,s}, K_f, K_s$). The model aims to represent:

- Snow accumulation and melt;
- Interception by vegetation;
- Infiltration and saturation excess flow based on a distribution of different soil depths;
- A split between fast saturation excess flow and preferential recharge to a slow store;
- Percolation from the unsaturated zone to a slow runoff store.

B.2.34.1 File names

Model: m_34_flexis_12p_5s

Parameter ranges: m_34_flexis_12p_5s_parameter_ ranges

B.2.34.2 Model equations

Figure B.35: Structure of the FLEX-IS model
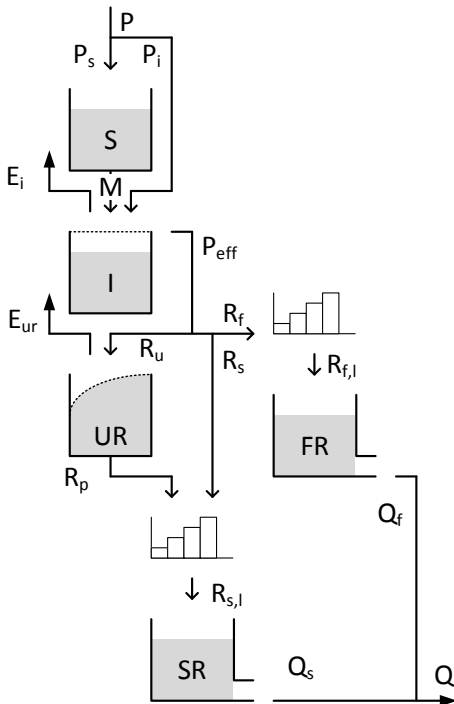
$$\frac{dS}{dt} = P_s - M \tag{B.411}$$

$$P_s = \begin{cases} P, & \text{if } T \le TT \\ 0, & \text{otherwise} \end{cases} \tag{B.412}$$

$$M = \begin{cases} ddf * (T - TT), & \text{if } T \ge TT \\ 0, & \text{otherwise} \end{cases} \tag{B.413}$$

Where S [mm] is the current snow storage, P_s the precipitation that falls as snow [mm/d], M the snowmelt [mm/d] based on a degree-day factor (ddf, [mm/°C/d]) and threshold temperature for snowfall and snowmelt (TT, [°C]).

$$\frac{dI}{dt} = P_I + M - E_I - P_{eff} \tag{B.414}$$

$$P_i = \begin{cases} P, & \text{if } T > TT \\ 0, & \text{otherwise} \end{cases} \tag{B.415}$$

$$E_i = \begin{cases} E_p, & \text{if } I > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.416}$$

$$P_{eff} = \begin{cases} P_i, & \text{if } I = I_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.417}$$

Where P_I [mm/d] is the incoming precipitation, I is the current interception storage [mm], which is assumed to evaporate (E_i [mm/d]) at the potential rate E_p [mm/d] when possible. When I exceeds the maximum interception storage I_{max} [mm], water is routed to the rest of the model as P_{eff} [mm/d].

$$\frac{dUR}{dt} = R_u - E_{ur} - R_p \tag{B.418}$$

$$R_u = (1 - \left[1 + exp\Big(\frac{-UR/UR_{max} + 1/2}{\beta}\Big)\right]^{-1}) * P_{eff} \tag{B.419}$$

$$E_{ur} = E_p * min\Big(1, \frac{UR}{UR_{max}}\frac{1}{L_p}\Big) \tag{B.420}$$

$$R_p = Perc_{max} * \frac{-UR}{UR_{max}} \tag{B.421}$$

Where UR is the current storage in the unsaturated zone [mm]. R_u [mm/d] is the inflow into UR based on its current storage compared to maximum storage UR_{max} [mm] and a shape distribution parameter β [-]. E_{ur} the evaporation [mm/d] from UR which follows a linear relation between current and maximum storage until a threshold L_p [-] is exceeded. R_p [mm/d] is the percolation from UR to the slow reservoir SR [mm], based on a maximum percolation rate $Perc_{max}$ [mm/d], relative to the fraction of current storage and maximum storage.

$$R_s = (P_{eff} - R_u) * D \tag{B.422}$$

$$R_f = (P_{eff} - R_u) * (1 - D) \tag{B.423}$$

Where R_s and R_f are the flows [mm/d] to the slow and fast runoff reservoir respectively, based on runoff partitioning coefficient D [-]. Both are lagged by linearly increasing

triangular transformation functions with parameters $N_{lag,s}$ and $N_{lag,f}$ respectively, that give the number of time steps over which R_s and R_f need to be transformed. R_p is added to R_s before the transformation occurs.

$$\frac{dFR}{dt} = R_{f,l} - Q_f \tag{B.424}$$

$$Q_f = K_f * FR \tag{B.425}$$

Where FR is the current storage [mm] in the fast flow reservoir. Outflow Q_f [mm/d] from the reservoir has a linear relation with storage through time scale parameter K_f [d^{-1}].

$$\frac{dSR}{dt} = R_{s,l} - Q_s \tag{B.426}$$

$$Q_s = K_s * SR \tag{B.427}$$

Where SR is the current storage [mm] in the slow flow reservoir. Outflow Q_s [mm/d] from the reservoir has a linear relation with storage through time scale parameter K_s [d^{-1}].

$$Q = Q_f + Q_s \tag{B.428}$$

Where Q [mm/d] is the total simulated flow as the sum of Q_s and Q_f.

B.2.35   MOPEX-5 (model ID: 35)

The MOPEX-5 model (fig. B.36) is part of a model improvement study that investigates the relationship between dominant processes and model structures for 197 catchments in the MOPEX database [349]. It has 5 stores and 12 parameters (T_{crit}, ddf, S_{b1}, t_w, I_a, I_s, T_{min}, T_{max}, S_{b2}, t_u, S_e, t_c). The original model relies on observations of Leaf Area Index and a calibrated interception fraction. Liang et al [194] show typical Leaf Area Index time series, and a sinusoidal function is a reasonable approximation of this. Therefore, the model is slightly modified to use a calibrated sinusoidal function, so that the data input requirements for MOPEX-5 are consistent with other models. The model aims to represent:

- Snow accumulation and melt;
- Time-varying interception and the impact of phenology on transpiration;
- Saturation excess flow;
- Infiltration to deeper soil layers;
- A split between fast and slow runoff.

B.2.35.1   File names

Model: m_35_mopex5_12p_5s

Parameter ranges: m_35_mopex5_12p_5s_parameter_ ranges

B.2.35.2   Model equations
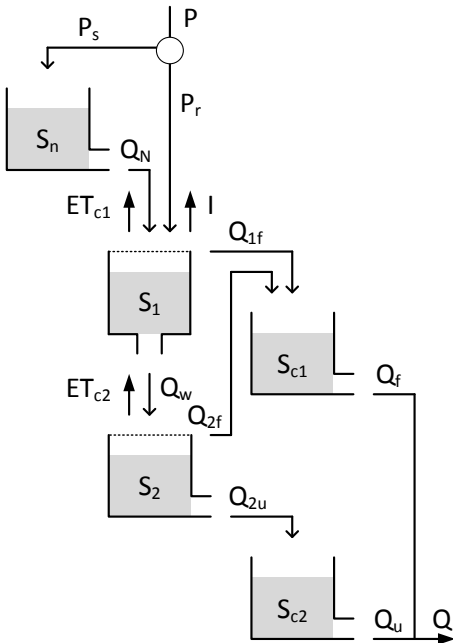
$$\frac{dS_n}{dt} = P_s - Q_n \tag{B.429}$$

$$P_s = \begin{cases} P, & \text{if } T \le T_{crit} \\ 0, & \text{otherwise} \end{cases} \tag{B.430}$$

$$Q_n = \begin{cases} ddf * (T - T_{crit}), & \text{if } T > T_{crit} \\ 0, & \text{otherwise} \end{cases} \tag{B.431}$$

Where S_n [mm] is the current snow pack. Precipitation occurs as snowfall P_s [mm/d] when current temperature T [^oC] is below threshold T_{crit} [^oC]. Snowmelt Q_N [mm/d] occurs when the temperature rises above the threshold temperature and relies in the degree-day factor dd [$mm/^oC/d$].

Figure   B.36:   Structure   of   the MOPEX-5 model

235

$$\frac{dS_1}{dt} = P_r - ET_1 - I - Q_{1f} - Q_w \tag{B.432}$$

$$P_r = \begin{cases} P, & \text{if } T > T_{crit} \\ 0, & \text{otherwise} \end{cases} \tag{B.433}$$

$$ET_{c1} = \frac{S_1}{S_{b1}} * Ep_c \tag{B.434}$$

$$I = max\left(0, I_\alpha + (1 - I_\alpha)sin\left(2\pi \frac{t + I_s}{365/d}\right)\right) \tag{B.435}$$

$$Q_{1f} = \begin{cases} P, & \text{if } S_1 \geq S_{b1} \\ 0, & \text{otherwise} \end{cases} \tag{B.436}$$

$$Q_w = t_w * S_1 \tag{B.437}$$

Where S_1 [mm] is the current storage in soil moisture and P_r precipitation as rain [mm/d]. Evaporation ET_1 [mm/d] depends linearly on current soil moisture, maximum soil moisture S_{b1} [mm] and phenology-corrected potential evapotranspiration:

$$Ep_c = Ep * GSI \tag{B.438}$$

$$GSI = \begin{cases} 0, & \text{if } T < T_{min} \\ \frac{T - T_{min}}{T_{max} - T_{min}}, & \text{if } T_{min} \geq T < T_{max} \\ 1, & \text{if } T \geq T_{max} \end{cases} \tag{B.439}$$

Where GSI is a growing season index based on parameters T_{min} [^oC] and T_{max} [^oC]. Interception I [mm/d] depends on the mean intercepted fraction I_α [-] and the maximum Leaf Area Index timing I_s [d]. Saturation excess flow Q_{1f} [mm/d] occurs when the soil moisture bucket exceeds its maximum capacity. Infiltration to deeper groundwater Q_w [mm/d] depends on current soil moisture and time parameter t_w [d^{-1}].

$$\frac{dS_2}{dt} = Q_w - ET_2 - Q_{2u} - Q_{2f} \tag{B.440}$$

$$ET_{c2} = \frac{S_2}{S_e} * Ep_c \tag{B.441}$$

$$Q_{2u} = t_u * S_2 \tag{B.442}$$

$$Q_{2f} = \begin{cases} Q_w, & \text{if } S_2 \geq S_{b2} \\ 0, & \text{otherwise} \end{cases} \tag{B.443}$$

Where S_2 [mm] is the current groundwater storage, refilled by infiltration from S_1. Evaporation ET_2 [mm/d] depends linearly on current groundwater and root zone storage capacity S_e [mm]. Leakage to the slow runoff store Q_{2u} [mm/d] depends on current groundwater level and time parameter t_u [d^{-1}]. When the store reaches maximum capacity S_{b2} [mm], excess flow Q_{2f} [mm/d] is routed towards the fast response routing store.

$$\frac{dS_{c1}}{dt} = Q_{1f} + Q_{2f} - Q_f \tag{B.444}$$

$$Q_f = t_c * S_{c1} \tag{B.445}$$

Where S_{c1} [mm] is current storage in the fast flow routing reservoir, refilled by Q_{1f} and Q_{2f}. Routed flow Q_f depends on the mean residence time parameter t_c [d^{-1}].

$$\frac{dS_{c2}}{dt} = Q_{2u} - Q_u \tag{B.446}$$

$$Q_u = t_c * S_{c2} \tag{B.447}$$

Where S_{c2} [mm] is current storage in the slow flow routing reservoir, refilled by Q_{2u}. Routed flow Q_u depends on the mean residence time parameter t_c [d^{-1}]. Total simulated flow Q_t [mm/d]:

$$Q_t = Q_f + Q_u \tag{B.448}$$

B.2.36 MODHYDROLOG (model ID: 36)

The MODHYDROLOG model (fig. B.37) is an elaborate groundwater recharge model, originally created for use in Australia [69, 71]. It has 5 stores (I, D, SMS, GW and CH) and 15 parameters (INSC, COEFF, SQ, SMSC, SUB, CRAK, EM, DSC, ADS, MD, VCOND, DLEV, k_1, k_2 and k_3). It originally includes a routing scheme that allows linking sub-basins together, which has been removed here. The model aims to represent:

- Interception by vegetation;
- Infiltration and infiltration excess flow;
- Depression storage and delayed infiltration;
- Preferential groundwater recharge, interflow and saturation excess flow;
- Groundwater recharge resulting from filling up of soil moisture storage capacity;
- Water exchange between shallow and deep aquifers;
- Water exchange between aquifer and river channel.

B.2.36.1 File names

Model: m_36_modhydrolog_15p_5s

Parameter ranges: m_36_modhydrolog_15p_5s_parameter_ ranges

B.2.36.2 Model equations
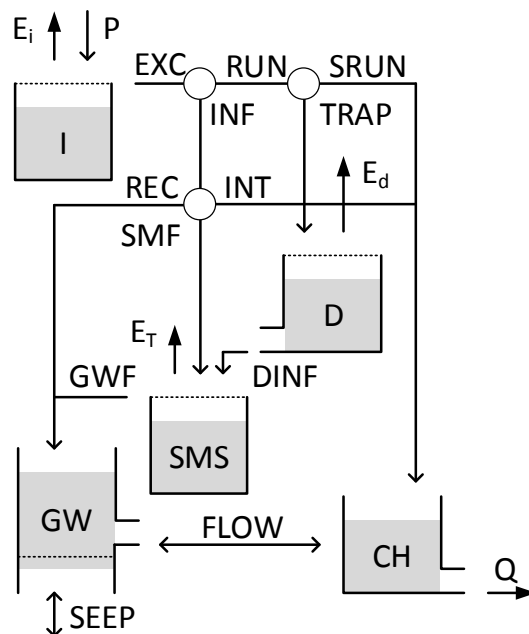
Figure B.37: Structure of the MODHY-DROLOG model

$$\frac{dI}{dt} = P - E_i - EXC \qquad (B.449)$$

$$E_i = \begin{cases} E_p, & \text{if } I > 0 \\ 0, & \text{otherwise} \end{cases} \qquad (B.450)$$

$$EXC = \begin{cases} P, & \text{if } I = INSC \\ 0, & \text{otherwise} \end{cases} \qquad (B.451)$$

Where I [mm] is the current interception storage, P the rainfall [mm/d], E_i the evaporation from the interception store [mm/d] and EXC the excess rainfall [mm/d]). Evaporation is assumed to occur at the potential rate E_p [mm/d] when possible. When I exceeds the maximum interception capacity $INSC$, water

is ruted to the rest of the model as excess precipitation EXC. The soil moisture store SMS is instrumental in dividing runoff between infiltration and surface flow:

$$\frac{dSMS}{dt} = SMF + DINF - E_T - GWF \tag{B.452}$$

$$SMF = INF - INT - REC \tag{B.453}$$

$$INF = min\left(COEFF * exp\left(\frac{-SQ * SMS}{SMSC}\right), EXC\right) \tag{B.454}$$

$$INT = SUB * \frac{SMS}{SMSC} * INF \tag{B.455}$$

$$REC = CRAK * \frac{SMS}{SMSC} * (INF - INT) \tag{B.456}$$

$$E_T = min\left(EM * \frac{SMS}{SMSC}, PET\right) \tag{B.457}$$

$$GWF = \begin{cases} SMF, & \text{if } SMS = SMSC \\ 0, & \text{otherwise} \end{cases} \tag{B.458}$$

Where SMS is the current storage in the soil moisture store [mm]. SMF [mm/d] and DINF [mm/d] are the infiltration and delayed infiltration respectively. INF is total infiltration [mm/d] from excess precipitation, based on maximum infiltration loss parameter COEFF [-], the infiltration loss exponent SQ and the ratio between current soil misture storage SMS [mm] and the maximum soil moisture capacity SMSC [mm]. INT represents interflow and saturation excess flow [mm/d], using a constant of proportionality SUB [-]. REC is preferential recharge of groundwater [mm/d] based on another constant of proportionality CRAK [-]. SMF is flow into soil moisture storage [mm/d]. E_T evaporation from the soil moisture that occurs at the potential rate when possible [mm/d], based on the maximum plant-controlled rate EM [mm/d]. GWF is the flow to the groundwater store [mm/d]:

$$\frac{dD}{dt} = TRAP - E_D - DINF \tag{B.459}$$

$$TRAP = ADS * exp\left(-MD\frac{D}{DSC - D}\right) * RUN \tag{B.460}$$

$$RUN = EXC - INF \tag{B.461}$$

$$E_D = \begin{cases} ADS * E_p, & \text{if } D > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.462}$$

$$DINF = \begin{cases} ADS * RATE, & \text{if } D > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.463}$$

$$RATE = COEFF * exp\left(-SQ\frac{SMS}{SMSC}\right) - INF - INT - REC \tag{B.464}$$

Where TRAP [mm/d] is the part of overland flow captured in the depression store [equation from 258]), E_D the evaporation from the depression store [mm/d], and DINF delayed infiltration to soil moisture [mm/d]. TRAP uses DSC as the maximum depression store capacity [mm], ADS as the fraction of land functioning as depression storage [-] and MD a depression storage parameter [-]. E_D relies on the potential evapotranspiration E_p. The grundwater store has no defined upper and lower boundary and instead fluctuates around a datum DLEV:

$$\frac{dGW}{dt} = REC + GWF - SEEP - FLOW \tag{B.465}$$

$$SEEP = VCOND * (GW - DLEV) \tag{B.466}$$

$$FLOW = \begin{cases} k_1 * |GW| + k_2 * (1 - exp(-k_3 * |GW|)), & \text{if } GW \geq 0 \\ -(k_1 * |GW| + k_2 * (1 - exp(-k_3 * |GW|))), & \text{if } GW < 0 \end{cases} \tag{B.467}$$

Where SEEP [mm/d] is the exchange with a deeper aquifer (can be negative or positive) and FLOW [mm/d] the exchange with the channel (can be negative or positive). VCOND is a leakage coefficient, DLEV a datum around which the groundwater level can fluctuate, and k_1, k_2 and k_3 are runoff coefficients. The channel store aggregates incoming fluxes and produces the total runoff Q_t [mm/d]:

$$\frac{dCH}{dt} = SRUN + INT + FLOW - Q \tag{B.468}$$

$$SRUN = RUN - TRAP \tag{B.469}$$

$$Q_t = \begin{cases} CH, & \text{if } CH > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.470}$$

B.2.37 HBV-96 (model ID: 37)

The HBV-96 model (fig. B.38) was originally developed for use in Sweden, but has been widely applied beyond its original region [197]. It can account for different land types (forest, open ground, lakes) but that distinction has been removed here. It has 5 stores and 15 parameters (TT, TTI, CFR, $CFMAX$, TTM, WHC, $CFLUX$, FC, LP, β, K_0, α, c, K_1, $MAXBAS$) parameters. The model aims to represent:

- Snow accumulation, melt and refreezing;
- Infiltration and capillary flow to, and evaporation from, soil moisture;
- A non-linear storage-runoff relationship from the upper runoff-generating zone;
- A linear storage-runoff relationship from the lower runoff-generating zone.

B.2.37.1 File names

Model: m_37_hbv_15p_5s
Parameter ranges: m_37_hbv_15p_5s_parameter_ ranges

B.2.37.2 Model equations
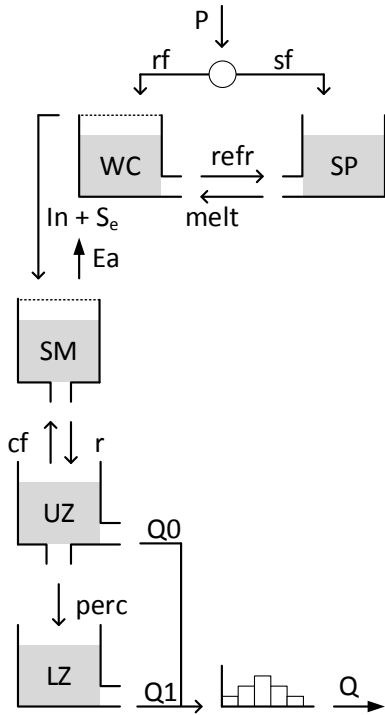
Figure B.38: Structure of the HBV-96 model

$$\frac{dSP}{dt} = sf + refr - melt \tag{B.471}$$

$$sf = \begin{cases} P, & \text{if } T \le TT - \frac{1}{2}TTI \\ P * \frac{TT + \frac{1}{2}TTI - T}{TTI}, & \text{otherwise} \\ 0, & \text{if } T \ge TT + \frac{1}{2}TTI \end{cases} \tag{B.472}$$

$$refr = \begin{cases} CFR * CFMAX * (TTM - T), & \text{if } T < TTM \\ 0, & \text{otherwise} \end{cases} \tag{B.473}$$

$$melt = \begin{cases} CFMAX * (T - TTM), & \text{if } T \ge TTM \\ 0, & \text{otherwise} \end{cases} \tag{B.474}$$

Where SP is the current snow storage [mm]. sf is precipitation that occurs as snowfall [mm/d] based on daily precipitation P [mm/d], threshold temperature for snowfall TT [°C] and the snowfall threshold interval length TTI [°C]. $refr$ [mm/d] is the refreezing of liquid snow if the current temperature T is below the melting threshold

242

TTM [°C], using a coefficient of refreezing CFR [-] and a degree-day factor CFMAX [mm/d/°C]. $melt$ represents snowmelt if the current temperature T is below the melting threshold TTM, using the degree-day factor CFMAX.

$$\frac{dWC}{dt} = rf + melt - refr - in - S_{excess} \tag{B.475}$$

$$rf = \begin{cases} 0, & \text{if } T \leq TT - \frac{1}{2}TTI \\ P * \frac{T - TT + \frac{1}{2}TTI}{TTI}, & \text{otherwise} \\ P, & \text{if } T \geq TT + \frac{1}{2}TTI \end{cases} \tag{B.476}$$

$$in = \begin{cases} rf + melt, & \text{if } WC \geq WHC * SP \\ 0, & \text{otherwise} \end{cases} \tag{B.477}$$

$$S_e = \begin{cases} WC - WHC * SP, & \text{if } WC \geq WHC * SP \\ 0, & \text{otherwise} \end{cases} \tag{B.478}$$

Where WC is the current liquid water content in the snow pack [mm], rf is the precipitation occurring as rain [mm/d] based on temperature threshold parameters TT and TTI, $refr$ is the refreezing flux, and in the infiltration to soil moisture [mm/d] that occurs when the water holding capacity of snow gets exceeded. S_{excess} [mm/d] represents excess stored water that is freed when the total possible storage of liquid water in the snow pack is reduced.

$$\frac{dSM}{dt} = (in + S_{excess}) + cf - E_a - r \tag{B.479}$$

$$cf = CFLUX * \left(1 - \frac{SM}{FC} \right) \tag{B.480}$$

$$E_a = \begin{cases} E_p, & \text{if } SM \geq LP * FC \\ E_p * \frac{SM}{LP*FC}, & \text{otherwise} \end{cases} \tag{B.481}$$

$$r = (in + S_{excess}) * \left(\frac{SM}{FC} \right)^{\beta} \tag{B.482}$$

Where SM is the current storage in soil moisture [mm], in the infiltration from the surface, cf the capillary rise [mm/d] from the unsaturated zone, E_a evaporation [mm/d] and r the flow to the upper zone [mm/d]. Capillary rise depends on the maximum rate CFLUX [mm/d], scaled by the available storage in soil moisture, expressed as the ration between current storage SM and maximum storage FC [mm]. Evaporation E_a occurs at

the potential rate E_p when current soil moisture is above the wilting point LP [mm], and is scaled linearly below that. Runoff r to the upper zone has a potentially non-linear relationship with infiltration in through parameter β [-].

$$\frac{dUZ}{dt} = r - cf - Q_0 - perc \tag{B.483}$$

$$Q_0 = K_0 * UZ^{(1+\alpha)} \tag{B.484}$$

$$perc = c. \tag{B.485}$$

Where UZ is the current storage [mm] in the upper zone. Outflow Q_0 [mm/d] from the reservoir has a non-linear relation with storage through time scale parameter K_0 [d^{-1}] and and α [-]. Percolation $perc$ [mm/d] to the lower zone is given as a constant rate c [mm/d]S.

$$\frac{dLZ}{dt} = perc - Q_1 \tag{B.486}$$

$$Q_1 = K_1 * LZ \tag{B.487}$$

Where LZ is the current storage [mm] in the lower zone. Outflow Q_1 [mm/d] from the reservoir has a linear relation with storage through time scale parameter K_1 [d^{-1}]. Total outflow is generated by summing Q_0 and Q_1 and applying a triangular transform based on lag parameter MAXBAS [d].

B.2.38 Tank Model - SMA (model ID: 38)

The Tank Model (fig. B.39) is originally developed for use in constantly saturated soils in Japan. This alternative Tank model - SMA (soil moisture accounting) version was developed for regions that are not continuously saturated [310]. This model is identical to the original tank model, but has an increased depth in the first store to represent primary soil moisture, and adds a new store to represent secondary soil moisture. It has 5 stores and 16 parameters (sm_1, sm_2, k_1, k_2, A_0, A_1, A_2, t_1, t_2, B_0, B_1, t_3, C_0, C_1, t_4, D_1). The model aims to represent:

- Runoff on increasing time scales with depth;
- Soil moisture storage;
- capillary rise to replenish soil moisture.

B.2.38.1 File names

Model: m_38_tank2_16p_5s
Parameter ranges: m_38_tank2_16p_5s_parameter_ ranges

B.2.38.2 Model equations
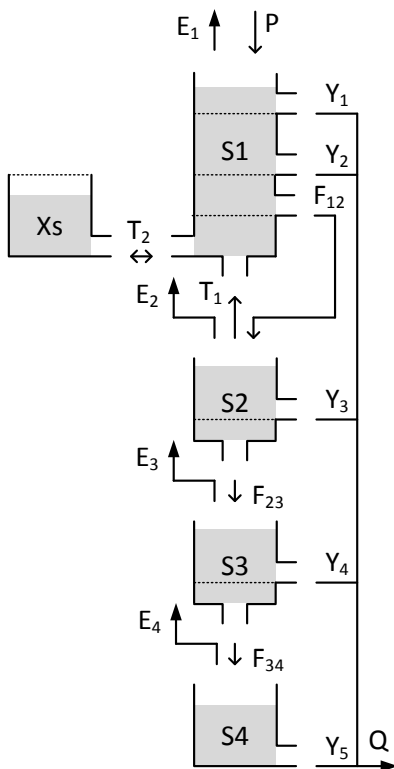
$$\frac{dS_1}{dt} = P + T_1 - T_2 - E_1 - F_{12} - Y_2 - Y_1 \tag{B.488}$$

$$T_1 = k_1 \left(1 - \frac{S_1}{sm_1}\right), \text{if } S_1 < sm_1 \tag{B.489}$$

$$T_2 = k_2 \left(\frac{min(S_1, sm_1)}{sm_1} - \frac{X_s}{sm_2}\right) \tag{B.490}$$

$$E_1 = \begin{cases} Ep, & \text{if } S_1 > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.491}$$

$$F_{12} = \begin{cases} A_0 * (S_1 - sm_1), & \text{if } S_1 > sm_1 \\ 0, & \text{otherwise} \end{cases} \tag{B.492}$$

$$Y_2 = \begin{cases} A_2 * (S_1 - t_2), & \text{if } S_1 > t_2 \\ 0, & \text{otherwise} \end{cases} \tag{B.493}$$

$$Y_1 = \begin{cases} A_1 * (S_1 - t_1), & \text{if } S_1 > t_1 \\ 0, & \text{otherwise} \end{cases} \tag{B.494}$$

Figure B.39: Structure of the Tank
Model - SMA

245

Where S_1 [mm] is the current storage in the upper zone, refilled by precipitation P [mm/d] and drained by evaporation E_1 [mm/d], drainage F_{12} [mm/d] and surface runoff Y_1 [mm/d] and Y_2 [mm/d]. If S_1 is below the soil moisture threshold sm_1 [mm], capillary rise T_1 [mm/d] from store S_2 can occur. Capillary rise has a base rate k_1 [mm/d] and decreases linearly as soil moisture S_1 nears sm_1. This store is connected to the secondary soil moisture store X_s through transfer flux

T_2 [mm/d]. This flux can work in either direction, based on a base rate k_2 [mm/d], the current storages S_1 [mm] and X_s [mm] and the maximum soil moistures storages sm_1 [mm] and sm_2 [mm]. Evaporation E_1 occurs at the potential rate E_p [mm/d] if water is available. Drainage to the intermediate layer has a linear relationship with storage through time scale parameter A_0 [d^{-1}]. Surface runoff Y_2 and Y_1 occur when S_1 is above thresholds t_2 [mm] and t_1 [mm] respectively. Both are linear relationships through time parameters A_2 [d^{-1}] and A_1 [d^{-1}] respectively.

$$\frac{dX_s}{dt} = T_2 \tag{B.495}$$

Where X_s [mm] is the current storage in the secondary soil moisture zone. This zone has a maximum capacity sm_2 [mm], used in the calculation of T_2. T_2 can be both positive and negative.

$$\frac{dS_2}{dt} = F_{12} - E_2 - T_1 - F_{23} - Y_3 \tag{B.496}$$

$$E_2 = \begin{cases} Ep, & \text{if } S_1 = 0 \ \& \ S_2 > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.497}$$

$$F_{23} = B_0 * S_2 \tag{B.498}$$

$$Y_3 = \begin{cases} B_1 * (S_2 - t_3), & \text{if } S_2 > t_3 \\ 0, & \text{otherwise} \end{cases} \tag{B.499}$$

Where S_2 [mm] is the current storage in the intermediate zone, refilled by drainage F_{12} from the upper zone and drained by evaporation E_2 [mm/d], drainage F_{23} [mm/d] and intermediate discharge Y_3 [mm/d]. E_2 occurs at the potential rate E_p if water is available and the upper zone is empty. Drainage to the third layer F_{23} has a linear relationship with

storage through time scale parameter B_0 [d^{-1}]. Intermediate runoff Y_3 occurs when S_2 is above threshold t_3 [mm] and has a linear relationship with storage through time scale parameter B_1 [d^{-1}].

$$\frac{dS_3}{dt} = F_{23} - E_3 - F_{34} - Y_4 \tag{B.500}$$

$$E_3 = \begin{cases} Ep, & \text{if } S_1 = 0 \ \& \ S_2 = 0 \ \& \ S_3 > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.501}$$

$$F_{34} = C_0 * S_3 \tag{B.502}$$

$$Y_4 = \begin{cases} C_1 * (S_3 - t_4), & \text{if } S_3 > t_4 \\ 0, & \text{otherwise} \end{cases} \tag{B.503}$$

Where S_3 [mm] is the current storage in the sub-base zone, refilled by drainage F_{23} from the intermediate zone and drained by evaporation E_3 [mm/d], drainage F_{34} [mm/d] and sub-base discharge Y_4 [mm/d]. E_3 occurs at the potential rate E_p if water is available and the upper zones are empty. Drainage to the fourth layer F_{34} has a linear relationship with storage through time scale parameter C_0 [d^{-1}]. Sub-base runoff Y_4 occurs when S_3 is above threshold t_4 [mm] and has a linear relationship with storage through time scale parameter C_1 [d^{-1}].

$$\frac{dS_4}{dt} = F_{34} - E_4 - Y_5 \tag{B.504}$$

$$E_4 = \begin{cases} Ep, & \text{if } S_1 = 0 \ \& \ S_2 = 0 \ \& \ S_3 = 0 \ \& \ S_4 > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.505}$$

$$Y_5 = D_1 * S_4 \tag{B.506}$$

Where S_4 [mm] is the current storage in the base layer, refilled by drainage F_{34} from the sub-base zone and drained by evaporation E_4 [mm/d] and baseflow Y_5 [mm/d]. E_4 occurs at the potential rate E_p if water is available and the upper zones are empty. Baseflow Y_5 has a linear relationship with storage through time scale parameter D_1 [d^{-1}]. Total runoff:

$$Q_t = Y_1 + Y_2 + Y_3 + Y_4 + Y_5 \tag{B.507}$$

B.2.39 Midlands Catchment Runoff Model (model ID: 39)

The Midlands Catchment Runoff model (fig. B.40) is intended to be used in a flood-forecasting setting [222]. To reduce the number of free parameters, the original evaporation routines and routing are somewhat simplified here. The model has 5 stores and 16 parameters (S_{max}, c_{max}, c_0, c_1, c_e, D_{surp}, k_d, γ_d, $q_{p,max}$, k_g, τ, S_{bf}, k_{cr}, γ_{cr}, k_{or}, γ_{or}). The model aims to represent:

- Interception by vegetation;
- Direct runoff from a variable contributing area;
- A deficit-based approach to soil moisture accounting and interflow and percolation;
- Baseflow from groundwater;
- Uniform flood flood wave distribution in time;
- In-channel and out-of-channel flood routing.

B.2.39.1 File names

Model: m_39_mcrm_16p_5s

Parameter ranges: m_39_mcrm_16p_5s_parameter_ ranges

B.2.39.2 Model equations
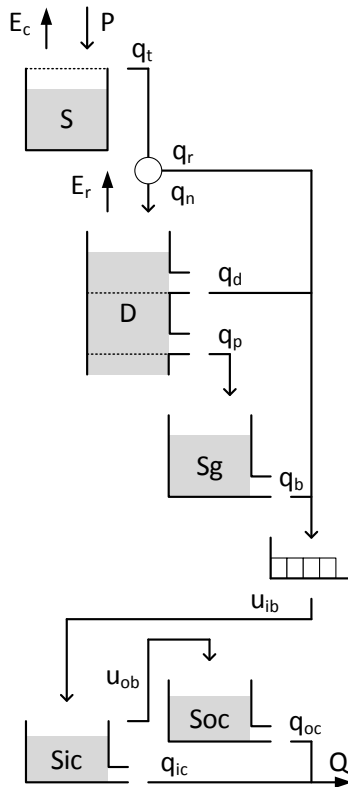
$$\frac{dS}{dt} = P - E_c - q_t \tag{B.508}$$

$$E_c = \begin{cases} E_p, & \text{if } S > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.509}$$

$$q_t = \begin{cases} P, & \text{if } S = S_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.510}$$

Where S [mm] is the current interception storage, refilled by precipitation P [mm/d] and drained by evaporation E_c [mm/d] and throughfall q_t [mm/d]. E_c occurs at the potential rate whenever possible. q_t occurs only when the store is at maximum capacity S_{max} [mm].

Figure B.40: Structure of the MCR model

248

$$\frac{dD}{dt} = q_n - E_r - q_d - q_p \tag{B.511}$$

$$q_n = q_t - q_r \tag{B.512}$$

$$q_r = min\left(c_{max}, c_0 + c_0 e^{c_1 D}\right) * q_t \tag{B.513}$$

$$E_r = \frac{1}{1 + e^{-c_e D}} * \left(E_p - E_c\right) \tag{B.514}$$

$$q_d = \begin{cases} k_d \left(D_{surp} - D\right)^{\gamma_d}, & \text{if } D > D_{surp} \\ 0, & \text{otherwise} \end{cases} \tag{B.515}$$

$$q_p = \begin{cases} q_{p,max}, & \text{if } D \geq D_{surp} \\ \frac{D}{D_{surp}} q_{p,max}, & \text{if } 0 < D < D_{surp} \\ 0, & \text{otherwise} \end{cases} \tag{B.516}$$

Where D [mm] is the current storage in soil moisture, refilled by net infiltration q_n [mm/d] and drained by evaporation E_r [mm/d], direct runoff q_d [mm/d] and percolation q_p [mm/d]. Negative D-values are possible and indicate a moisture deficit. Net inflow q_n is calculated as the difference between throughfall q_t and rapid runoff q_r [mm/d]. q_r varies depending on the current degree of saturation in the catchment, with a maximum fraction of the catchment area contributing to rapid runoff called c_{max} [-], a minimum contributing area of c_0 [-] and an exponential increase with increasing soil moisture storage, controlled through shape parameter c_1 [-], in between. E_r fulfils any remaining evaporation demand but decreases with increasing moisture deficit (negative D values). This relation is controlled through shape parameter c_2. q_d has a non-linear relation with storage above a threshold D_{surp} [mm] through time scale parameter k_d [d^{-1}] and non-linearity parameter γ_d [-]. Percolation q_p has a maximum rate of $q_{p,max}$ if D is above threshold D_{surp} and decreases linearly between $D = D_{surp}$ and $D = 0$.

$$\frac{dS_g}{dt} = q_p - q_b \tag{B.517}$$

$$q_b = k_g * S_g^{1.5} \tag{B.518}$$

Where S_g [mm] is the current groundwater storage, refilled by percolation q_p and drained by baseflow q_b [mm/d]. q_b uses time parameter k_b [d^{-1}] and a fixed non-linearity coefficient of 1.5. Next, q_r, q_d and q_b are summed together and distributed uniformly over timespan τ [d], giving delayed flow u_{ib} [mm/d].

$$\frac{dS_{ic}}{dt} = u_{ib} - u_{ob} - q_{ic} \tag{B.519}$$

$$u_{ob} = \begin{cases} u_{ib}, & \text{if } S_{ic} = S_{bf} \\ 0, & \text{otherwise} \end{cases} \tag{B.520}$$

$$q_{ic} = \begin{cases} k_{cr} * S_{ic}^{\gamma_{cr}}, & \text{if } q_{ic} < \frac{3}{4} S_{ic} \\ \frac{3}{4} S_{ic}, & \text{otherwise} \end{cases} \tag{B.521}$$

Where S_{ic} [mm] is the current in-channel storage, refilled by u_{ic} and drained by in-channel flow q_{ic} [mm/d] and out-of-bank flow u_{ob} [mm/d]. u_{ob} only occurs when the store is at maximum capacity S_{bf} [mm]. q_{ic} uses time parameter k_{cr} [d^{-1}] and non-linearity parameter γ_{cr} [-].

$$\frac{dS_{oc}}{dt} = u_{ob} - q_{oc} \tag{B.522}$$

$$q_{oc} = \begin{cases} k_{or} * S_{oc}^{\gamma_{or}}, & \text{if } q_{oc} < \frac{3}{4} S_{oc} \\ \frac{3}{4} S_{oc}, & \text{otherwise} \end{cases} \tag{B.523}$$

Where S_{oc} [mm] is the current out-of-channel storage, refilled by u_{ob} and drained by out-of-channel flow q_{oc} [mm/d]. q_{oc} uses time parameter k_{or} [d^{-1}] and non-linearity parameter γ_{or} [-]. Total flow:

$$Q_t = q_{oc} + q_{ic} \tag{B.524}$$

B.2.40 SMAR (model ID: 40)

The SMAR model (fig. B.41) is the result of a series of modifications to the original 'layers-model' [236] and summarized by Tan & O'Connor [312]. The model uses an arbitrary number of soil moistures stores connected in series, with each store having a depth of 25mm. The number of stores is an optimization parameter. The current storage in the upper 5 stores features in various equations. For consistency within this framework, the process is reversed: the model uses a fixed number of 5 soil moisture stores, but the depth of each store is variable and given as $S_{n,max} = S_{max}/5$. It has 6 stores and 8 parameters (H, Y, S_{max}, C, G, K_G, N, K). The model aims to represent:

- Saturation excess overland flow;
- Infiltration excess overland flow;
- Gradual infiltration into soil moisture and declining evaporation potential when water is sourced from further underground;
- Groundwater flow;
- Routing of non-groundwater flow.

B.2.40.1 File names

Model: m_40_smar_8p_6s
Parameter ranges: m_40_smar_8p_6s_parameter_ ranges

B.2.40.2 Model equations
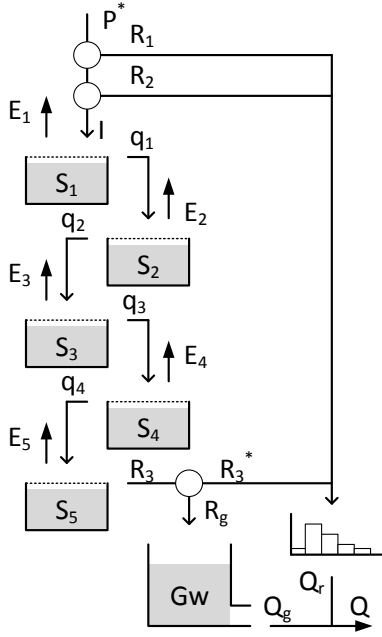
Figure B.41: Structure of the SMAR model

$$\frac{dS_1}{dt} = I - E_1 - q_1 \tag{B.525}$$

$$I = \begin{cases} Y, & \text{if } P^* - R_1 \geq Y \\ P^* - R_1, & \text{otherwise} \end{cases} \tag{B.526}$$

$$P^* = \begin{cases} P - E_p, & \text{if } P > E_p \\ 0, & \text{otherwise} \end{cases} \tag{B.527}$$

$$R_1 = P^* * H * \frac{\sum S_n}{S_{max}} \tag{B.528}$$

$$R_2 = \left(P^* - R_1\right) - I \tag{B.529}$$

$$E_1 = C^{(1-1)} * Ep^* \tag{B.530}$$

$$E_p^* = \begin{cases} E_p - P, & \text{if } E_p > P \\ 0, & \text{otherwise} \end{cases} \tag{B.531}$$

$$q_1 = \begin{cases} P^* - R_1 - R_2, & \text{if } S_1 \geq \frac{S_{max}}{5} \\ 0, & \text{otherwise} \end{cases} \tag{B.532}$$

Where S_1 [mm] is the current storage in the upper soil layer, I [mm/d] infiltration into the soil, P^* the effective precipitation [mm/d], R_1 [mm/d] is direct runoff, R_2 [mm/d] is infiltration excess runoff, E_1 [mm/d] evaporation and q_1 [mm/d] flow towards deeper soil layers. I uses a constant infiltration rate Y [mm/d]. Direct runoff R_1 relies on distribution parameter H [-] and is scaled by the current soil moisture storage in all layers compared to the maximum soil moisture storage S_{max} [mm] of all layers. Evaporation from this soil layer occurs at the effect potential rate E_p^*. Runoff to deeper layers q_1 only occurs when the current storage exceeds the store's maximum capacity.

$$S_2 = q_1 - E_2 - q_2 \tag{B.533}$$

$$E_2 = \begin{cases} C^{(2-1)} * Ep, & \text{if } S_1 = 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.534}$$

$$q_2 = \begin{cases} q1, & \text{if } S_2 \geq \frac{S_{max}}{5} \\ 0, & \text{otherwise} \end{cases} \tag{B.535}$$

Where S_2 [mm] is the current storage in the second soil layer, E_2 [mm/d] the evaporation scaled by parameter C [-], and q_2 [mm/d] overflow into the next layer. Evaporation is assumed to occur only when the storage in the upper layers has been exhausted.

$$S_3 = q_2 - E_3 - q_3 \tag{B.536}$$

$$E_3 = \begin{cases} C^{(3-1)} * Ep, & \text{if } S_2 = 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.537}$$

$$q_3 = \begin{cases} q2, & \text{if } S_3 \geq \frac{S_{max}}{5} \\ 0, & \text{otherwise} \end{cases} \tag{B.538}$$

Where S_3 [mm] is the current storage in the second soil layer, E_3 [mm/d] the evaporation scaled by parameter C^2 [-], and q_3 [mm/d] overflow into the next layer. Evaporation is assumed to occur only when the storage in the upper layers has been exhausted.

$$S_4 = q_3 - E_4 - q_4 \tag{B.539}$$

$$E_4 = \begin{cases} C^{(4-1)} * Ep, & \text{if } S_3 = 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.540}$$

$$q_4 = \begin{cases} q3, & \text{if } S_4 \geq \frac{S_{max}}{5} \\ 0, & \text{otherwise} \end{cases} \tag{B.541}$$

Where S_4 [mm] is the current storage in the second soil layer, E_4 [mm/d] the evaporation scaled by parameter C^3 [-], and q_4 [mm/d] overflow into the next layer. Evaporation is assumed to occur only when the storage in the upper layers has been exhausted.

$$S_5 = q_4 - E_5 - R_3 \tag{B.542}$$

$$E_5 = \begin{cases} C^{(5-1)} * Ep, & \text{if } S_4 = 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.543}$$

$$R_3 = \begin{cases} q4, & \text{if } S_5 \geq \frac{S_{max}}{5} \\ 0, & \text{otherwise} \end{cases} \tag{B.544}$$

Where S_5 [mm] is the current storage in the second soil layer, E_5 [mm/d] the evaporation scaled by parameter C^4 [-], and R_3 [mm/d] overflow towards groundwater. Evaporation is assumed to occur only when the storage in the upper layers has been exhausted.

$$\frac{dG_w}{dt} = R_g - Q_g \tag{B.545}$$

$$R_g = G * R_3 \tag{B.546}$$

$$Q_g = K_G * G_w \tag{B.547}$$

Where G_w [mm] is the current groundwater storage, refilled by fraction G [-] of R_3 [mm/d] and drained as a linear reservoir with time parameter K_G [d^{-1}]. This groundwater flow Q_g [mm/d] contributes directly to simulated streamflow Q. The fraction $R_3^* = (1-G) * R_3$ that does not reach the groundwater reservoir is combined with R_1 and R_2 and routed with a gamma function with parameters N and K. The routing function approximates a Nash-cascade consisting of N reservoirs with storage coefficient K:

$$h = \frac{1}{K\Gamma(N)} \left(\frac{t}{K} \right)^{N-1} e^{-t/K} \tag{B.548}$$

Integration over the time step length d provides the fraction of flow routed per time step Q_r [mm/d]. Total flow:

$$Q_t = Q_r + Q_g \tag{B.549}$$

B.2.41 NAM model (model ID: 41)

The NAM model (fig. B.42) is originally developed for use in Denmark [232]. Here a small modification is made by replacing runoff routing equations of the form $\frac{1}{k}e^{-t/k}$ with the linear reservoirs these equations represent. The model has 6 stores and 10 parameters (C_s, C_{if}, L^*, C_{L1}, U^*, C_{of}, C_{L2}, K_0, K_1, K_b). The model aims to represent:

- Snow accumulation and melt;
- Interflow when total soil moisture exceeds a threshold;
- Separation of saturation excess flow into overland flow and infiltration;
- Baseflow from groundwater.;

B.2.41.1 File names

Model: m_41_nam_10p_6s
Parameter ranges: m_41_nam_10p_6s_parameter_ ranges

B.2.41.2 Model equations
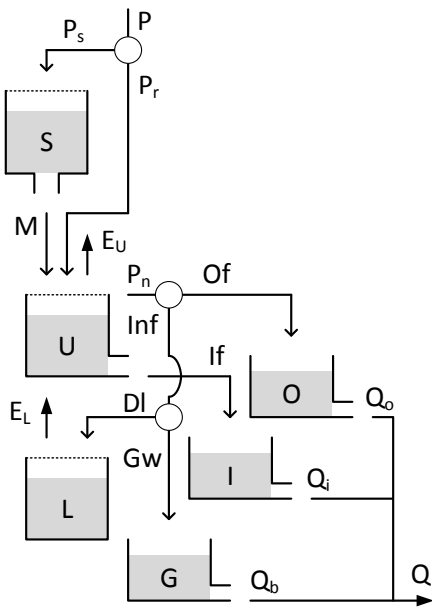
$$\frac{dS}{dt} = P_s - M \tag{B.550}$$

$$P_s = \begin{cases} P, & \text{if } T \le 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.551}$$

$$M = \begin{cases} C_s * T, & \text{if } T > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.552}$$

Where S is the current snow storage [mm], P_s [mm/d] the precipitation that falls as snow and M the snowmelt [mm/d] based on a degree-day factor (c_s, [mm/°C/d]). The freezing point of 0° [C] is used as a threshold for snowfall and melt.

Figure B.42: Structure of the NAM model

$$\frac{dU}{dt} = P_r + M - E_U - If - P_n \tag{B.553}$$

$$P_r = \begin{cases} P, & \text{if } T > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.554}$$

$$E_U = \begin{cases} E_p, & \text{if } U > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.555}$$

$$If = \begin{cases} C_{if} * \frac{L/L^* - C_{L1}}{1 - C_{L1}} U, & \text{if } L/L^* > C_{L1} \\ 0, & \text{otherwise} \end{cases} \tag{B.556}$$

$$P_n = \begin{cases} (P_r + M), & \text{if } U = U^* \\ 0, & \text{otherwise} \end{cases} \tag{B.557}$$

Where U [mm] is the current storage in the upper zone, refilled by precipitation as rain P_r [mm/d] and snowmelt M, and drained by evaporation E_U [mm/d], interflow If [mm/d] and net precipitation P_n [mm/d]. P_r occurs only when the current temperature exceeds the threshold of 0^oC. E_U occurs at the potential rate E_p whenever possible. If occurs only if the fractional storage in the lower zone L/L^* (L is current lower zone storage, L^* is lower zone maximum storage) exceeds a threshold C_{L1} [-]. If is further scaled current deficit in the lower zone and a second scaling coefficient C_{if} [-]. P_n occurs only when the upper zone exceeds its maximum storage capacity U^* [mm].

$$\frac{dL}{dt} = Dl - E_t \tag{B.558}$$

$$Dl = (P_n - Of)\left(1 - \frac{L}{L^*}\right) \tag{B.559}$$

$$Of = \begin{cases} C_{of} * \frac{L/L^* - C_{L2}}{1 - C_{L2}} * P_n, & \text{if } L/L^* > C_{L2} \\ 0, & \text{otherwise} \end{cases} \tag{B.560}$$

$$E_t = \begin{cases} \frac{L}{L^*} E_p, & \text{if } U = 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.561}$$

Where L [mm] is the current storage in the lower zone, refilled by a fraction of infiltration Dl [mm/d] and drained by evaporation E_t [mm/d]. Dl is calculated as a fraction of infiltration $P_n - Of$, dependent on the current deficit in the lower zone. Note that with the current formulation Dl might be larger than the lower zone deficit $L^* - L$ and a constraint of the form $Dl \leq L^* - L$ is needed. Overland flow Of [mm/d] is a fraction of P_n determined

using the relative storage in the lower zone L/L^* and two coefficients C_{of} [-] and C_{L2} [-]. E_t occurs only when the upper zone is empty, and at a reduced rate that uses the relative storage in the lower zone.

$$\frac{dO}{dt} = Of - Q_o \tag{B.562}$$

$$Q_o = K_0 * O \tag{B.563}$$

Where O [mm] is the current storage in the overland flow routing store. Q_o is the routed overland flow, using time coefficient K_0 $[d^{-1}]$.

$$\frac{dI}{dt} = If - Q_i \tag{B.564}$$

$$Q_i = K_1 * I \tag{B.565}$$

Where I [mm] is the current storage in the interflow routing store. Q_i is the routed interflow, using time coefficient K_1 $[d^{-1}]$.

$$\frac{dG}{dt} = Gw - Q_b \tag{B.566}$$

$$Gw = (P_n - Of)\left(\frac{L}{L^*}\right) \tag{B.567}$$

$$Q_b = K_b * O \tag{B.568}$$

Where G [mm] is the current storage in the overland flow routing store, refilled by groundwater flow Gw $[mm/d]$. Q_b is the routed baseflow, using time coefficient K_b $[d^{-1}]$. Total flow:

$$Q = Q_o + Q_i + Q_b \tag{B.569}$$

B.2.42 HYCYMODEL (model ID: 42)

The HYCYMODEL (fig. B.43) is originally developed for use in heavily forested catchments in Japan [115]. The original model specifies evaporation from the S_b store as $E_T = e_p(i) * Q_b/Q_{bc}$, if $S_u < 0$ & $S_b < S_{bc}$, with $Q_{bc} = f(S_{bc})$. However, no further details are given and S_{bc} is not listed as a parameter. We assume that S_{bc} [mm] is a threshold parameter and that evaporation potential declines linearly to zero when the store drops under this threshold. The model has 6 stores and 12 parameters (C, $I_{1,max}$, α, $I_{2,max}$, k_{in}, D_{50}, D_{16}, S_{bc}, k_b, p_b, k_h, k_c). The model aims to represent:

- Split between channel and ground precipitation;
- Interception by canopy and stems/trunks;
- Overland flow from a variable contributing area;
- Non-linear channel flow, hillslope flow and baseflow;
- Channel evaporation.

B.2.42.1 File names

Model: m_42_hycymodel_12p_6s

Parameter ranges: m_42_hycymodel_12p_6s_parameter_ ranges

B.2.42.2 Model equations

Figure B.43: Structure of the HYCY-MODEL

$$\frac{dI_c}{dt} = R_g - E_{ic} - q_{ie} \tag{B.570}$$

$$R_g = (1 - C)P \tag{B.571}$$

$$E_{ic} = \begin{cases} (1-C) * E_p, & \text{if } I_c > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.572}$$

$$q_{ie} = \begin{cases} R_g, & \text{if } I_c = I_{1,max} \\ 0, & \text{otherwise} \end{cases} \tag{B.573}$$

Where I_c [mm] is the current canopy storage, refilled by rainfall on ground R_g [mm/d] and drained by evaporation E_{ic} [mm/d] and canopy interception excess Q_{ie} [mm/d]. R_g is the fraction (1-C) [mm] of rainfall P [mm/d] that falls on ground (and not in the channel).

This fraction appears several times in the model to scale evaporation values according to surface area. E_{ic} occurs at the potential rate E_p [mm/d] when possible. q_{ie} only occurs when the canopy store is at maximum capacity $I_{1,max}$ [mm].

$$\frac{dI_s}{dt} = q_{is} - E_{is} - R_s \tag{B.574}$$

$$q_{is} = \alpha * q_{ie} \tag{B.575}$$

$$E_{is} = \begin{cases} (1-C)*E_p, & \text{if } I_s > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.576}$$

$$R_s = \begin{cases} q_{is}, & \text{if } I_s = I_{2,max} \\ 0, & \text{otherwise} \end{cases} \tag{B.577}$$

Where I_s [mm] is the current stem and trunk storage, refilled by a fraction of canopy excess q_{is} [mm/d] and drained by evaporation E_{is} [mm/d] and stem flow R_s [mm/d]. q_{is} is the fraction α [-] of canopy excess q_{ie}. The remainder $(1-\alpha)$ is throughfall R_t [mm/d]. E_{is} occurs at the potential rate E_p when possible. R_s occurs only when the store is at maximum capacity $I_{2,max}$.

$$\frac{dS_u}{dt} = R_n - Re - E_{su} - Q_{in} \tag{B.578}$$

$$R_n = R_t + R_s \tag{B.579}$$

$$Re = m * R_n \tag{B.580}$$

$$m = \int_{-\inf}^{\xi} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{\xi^2}{2}\right) d\xi \tag{B.581}$$

$$\xi = \frac{log(S_u/D_{50})}{log(D_{50}/D_{16})} \tag{B.582}$$

$$E_{su} = \begin{cases} (1-C)*E_p, & \text{if } E_{us} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.583}$$

$$Q_{in} = k_{in} * S_u \tag{B.584}$$

Where S_u [mm] is the current storage in the upper zone, refilled by net precipitation R_n [mm/d] and drained by effective rainfall R_e [mm/d], evaporation E_{su} [mm/d] and infiltration Q_{in} [mm/d]. R_n is the sum of throughfall R_t and stem flow R_s. R_e is a fraction

m [-] of R_e, determined from a variable contributing area concept. m is calculated is an integral from a regular normal distribution, scaled by the current storage S_u compared to two parameters D_{50} [mm] and D_{16} [mm]. These parameters represent the effective soil depths at which respectively 50% and 16% of the catchment area contribute to R_e. E_{su} occurs at the potential rate E_p when possible. Q_{in} has a linear relation with storage through time parameter k_{in} [d^{-1}].

$$\frac{dS_b}{dt} = Q_{in} - E_{sb} - Q_b \tag{B.585}$$

$$E_{sb} = \begin{cases} (1-C)*E_p, & \text{if } S_u = 0 \ \& \ S_b \geq S_{bc} \\ (1-C)*E_p \frac{S_b}{S_{bc}}, & \text{otherwise} \end{cases} \tag{B.586}$$

$$Q_b = k_b * S_b^{p_b} \tag{B.587}$$

Where S_b [mm] is the current storage in the lower zone, refilled by infiltration Q_{in} and drained by evaporation E_{sb} [mm/d] and baseflow Q_b [mm/d]. E_{sb} occurs at the potential rate when the store is above a threshold S_{bc} [mm], and declines linearly below that. Q_b has a potentially non-linear relation with storage through time parameter k_b [d^{-1}] and scale parameter p_b [-].

$$\frac{dS_h}{dt} = R_e - Q_h \tag{B.588}$$

$$Q_h = k_h * S_h^{p_h} \tag{B.589}$$

Where S_h [mm] is the current storage in the hillslope routing store, refilled by effective rainfall R_e and drained by hillslope runoff Q_h. Q_h has a potentially non-linear relation with storage through time parameter k_h [d^{-1}] and scale parameter p_h [-]. p_h is a fixed parameter in the original model with value 5/3.

$$\frac{dS_c}{dt} = R_c - Q_c \tag{B.590}$$

$$Q_c = k_c * S_c^{p_c} \tag{B.591}$$

Where S_c [mm] is the current storage in the channel routing store, refilled by rainfall on the channel R_c and drained by channel runoff Q_c. Q_c has a potentially non-linear relation with storage through time parameter k_c [d^{-1}] and scale parameter p_c [-]. p_c is a fixed parameter in the original model with value 5/3.

$$Q_t = Q_c + Q_h + Q_b - E_c \tag{B.592}$$

$$E_c = C * E_p \tag{B.593}$$

Where Q_t [mm/d] is the total flow as sum of the three individual flow fluxes minus channel evaporation E_c [mm/d].

B.2.43   GSM-SOCONT model (model ID: 43)

The Glacier and SnowMelt - SOil CONTribution model (GSM-SOCONT) model (fig. B.44) is a model developed for alpine, partly glaciated catchments [281]. For consistency with other models in this framework, several simplifications are used. The model does not use different elevation bands nor DEM data to estimate certain parameters, and does not calculate an annual glacier mass balance. The model has 6 stores and 12 parameters (f_{ice}, T_0, a_{snow}, T_m, k_s, a_{ice}, k_i, A, x, y, k_{sl}, β). The model aims to represent:

- Separate treatment of glacier and non-glacier catchment area;
- Snow accumulation and melt;
- Glacier melt;
- Soil moisture accounting in the non-glacier catchment area.

B.2.43.1   File names

Model: m_43_gsmsocont_12p_6s

Parameter ranges: m_43_gsmsocont_12p_6s_parameter_ ranges

B.2.43.2   Model equations
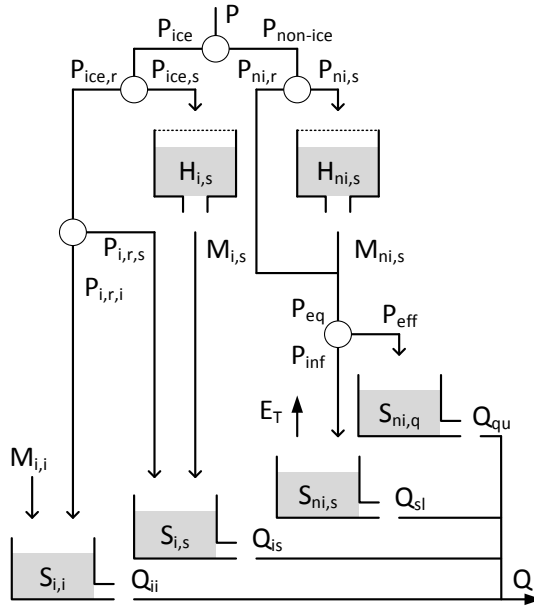
Figure B.44: Structure of the GSM-SOCONT model

$$\frac{dH_{i,s}}{dt} = P_{ice,s} - M_{i,s} \qquad (B.594)$$

$$P_{ice,s} = \begin{cases} P_{ice}, & \text{if } T \le T_0 \\ 0, & \text{otherwise} \end{cases} \qquad (B.595)$$

$$P_{ice} = f_{ice} * P \qquad (B.596)$$

$$M_{i,s} = \begin{cases} a_{snow}(T - T_m), & \text{if } T > T_m \\ 0, & \text{otherwise} \end{cases}$$

$$\qquad (B.597)$$

Where $H_{i,s}$ [mm] is the current storage in the snow pack, refilled by precipitation-as-snow $P_{ice,s}$ [mm/d] and depleted by melt $M_{i,s}$ [mm/d]. $P_{ice,s}$ occurs only when the temperature T [^oC] is below a threshold temperature for snowfall T_0 [^oC]. P_{ice} is the fraction f_{ice} [-] of precipitation P

$[mm/d]$ that falls on the ice-covered part of the catchment. $M_{i,s}$ uses a degree-day-factor a_{snow} $[mm/^oC/d]$ to estimate snow melt if temperature is above a threshold for snow melt T_s $[^oC]$.

$$\frac{dS_{i,s}}{dt} = M_{i,s} + P_{i,r,s,} - Q_{is} \tag{B.598}$$

$$P_{i,r,s} = P_{ice,r}, \quad \text{if } H_{i,s} > 0 \tag{B.599}$$

$$P_{ice,r} = \begin{cases} P_{ice}, & \text{if } T > T_0 \\ 0, & \text{otherwise} \end{cases} \tag{B.600}$$

$$Q_{is} = k_s * S_{i,s} \tag{B.601}$$

Where $S_{i,s}$ [mm] is the current storage in the snow-water routing reservoir, refilled by snow melt $M_{i,s}$ $[mm/d]$ and rain-on-snow $P_{ice,s}$ $[mm/d]$, and drained by runoff Q_{is}. $P_{i,r,s}$ occurs only if the current snow pack storage is above zero. $P_{ice,r}$ is precipitation-as-rain that occurs only if the temperature is above a snowfall threshold T_0. Q_{is} has a linear relation with storage through time parameter k_s $[d^{-1}]$.

$$\frac{dS_{i,i}}{dt} = M_{i,i} + P_{i,r,i,} - Q_{ii} \tag{B.602}$$

$$P_{i,r,i} = P_{ice,r}, \quad \text{if } H_{i,s} = 0 \tag{B.603}$$

$$M_{i,s} = \begin{cases} a_{ice}(T - T_m), & \text{if } T > T_m \text{ \& } H_{i,s} = 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.604}$$

$$Q_{ii} = k_i * S_{i,i} \tag{B.605}$$

Where $S_{i,i}$ [mm] is the current storage in the ice-water routing reservoir, refilled by glacier melt $M_{i,i}$ $[mm/d]$ and rain-on-ice $P_{ice,i}$ $[mm/d]$, and drained by runoff Q_{ii} $[mm/d]$. Both $M_{i,i}$ and $P_{ice,i}$ are assumed to only occur once the snow pack $H_{i,s}$ is depleted. $M_{i,i}$ uses a degree-day-factor a_{ice} $[mm/^oC/d]$ to estimate glacier melt. Ice storage in the glacier is assumed to be infinite. $P_{ice,r,i}$ is equal to $P_{ice,r}$ if $H_{i,s} = 0$. Q_{ii} has a linear relation with storage through time parameter k_i $[d^{-1}]$.

$$\frac{dH_{ni,s}}{dt} = P_{ni,s} - M_{ni,s} \tag{B.606}$$

$$P_{ni,s} = \begin{cases} P_{non-ice}, & \text{if } T \leq T_0 \\ 0, & \text{otherwise} \end{cases} \tag{B.607}$$

$$P_{non-ice} = (1 - f_{ice}) * P \tag{B.608}$$

$$M_{ni,s} = \begin{cases} a_{snow}(T - T_m), & \text{if } T > T_m \\ 0, & \text{otherwise} \end{cases} \tag{B.609}$$

Where $H_{ni,s}$ [mm] is the current snow pack storage on the non-ice covered fraction $1 - f_{ice}$ [-] of the catchment, which increases through snowfall $P_{ni,s}$ [mm/d] and decreases through snow melt $M_{ni,s}$ [mm/d]. Both fluxes are calculated in the same manner as those on the ice-covered part of the catchment (fluxes $P_{ice,s}$ and $M_{ice,s}$).

$$\frac{dS_{ni,s}}{dt} = P_{inf} - E_T - Q_{sl} \tag{B.610}$$

$$P_{inf} = P_{eq} - P_{eff} \tag{B.611}$$

$$P_{eff} = P_{eq} \left(\frac{S_{ni,s}}{A} \right)^y \tag{B.612}$$

$$P_{eq} = M_{ni,s} + P_{ni,r} \tag{B.613}$$

$$E_T = E_p \left(\frac{S_{ni,s}}{A} \right)^x \tag{B.614}$$

$$Q_{sl} = k_{sl} S_{ni,s} \tag{B.615}$$

Where $S_{ni,s}$ [mm] is the current storage in soil moisture, refilled by infiltrated precipitation P_{inf} [mm/d] and drained by evapotranspiration E_T [mm/d] and slow flow Q_{sl} [mm/d]. P_{inf} depends on the effective precipitation P_{eff}. P_{eq} is the total of snow melt $M_{ni,s}$ and precipitation-as-rain $P_{ni,r}$ [mm/d]. $P_{ni,r}$ is calculated in the same manner as $P_{i,r}$ (equation 7). E_T is a fraction potential evapotranspiration E_p [mm/d], calculated using A and non-linearity parameter y [-]. Q_{sl} has a linear relation with storage through time parameter k_{sl} [d^{-1}].

$$\frac{dS_{ni,q}}{dt} = P_{eff} - Q_{qu} \tag{B.616}$$

$$Q_{qu} = \beta S_{ni,q}^{5/3} \tag{B.617}$$

Where $S_{ni,q}$ [mm] is the current storage in the direct runoff reservoir, refilled by effective precipitation P_{eff} [mm/d] and by quick flow Q_{qu} [mm/d]. Q_{sl} has a non-linear relation with storage through time parameter β [$mm^{4/3}/d$] and the factor 5/3. Total flow:

$$Q = Q_{qu} + Q_{sl} + Q_{is} + Q_{ii} \tag{B.618}$$

265

B.2.44 ECHO model (model ID: 44)

The ECHO model (fig. B.45) is a single element from the Spatially Explicit Hydrologic Response (SEHR-ECHO) model [282]. Because the model is used as a lumped model here, the "SEHR" prefix was dropped intentionally. For consistency with other models, soil moisture storage S is given here in absolute terms [mm], rather than fractional terms that are used in the original reference. Rain- and snowfall equations are taken from Schaefli et al [281]. The model has 6 stores and 16 parameters (ρ, T_s, T_m, a_s, a_f, G_{max}, θ, ϕ, S_{max}, sw, sm, K_{sat}, c, L_{max}, k_f, k_s). The model aims to represent:

- Interception by vegetation;
- Snowfall, snowmelt, ground-heat flux and storage and refreezing of liquid snow;
- Infiltration, infiltration excess and saturation excess;
- Fast and slow runoff.

B.2.44.1 File names

Model: m_44_echo_16p_6s

Parameter ranges: m_44_echo_16p_6s_parameter_ ranges

B.2.44.2 Model equations
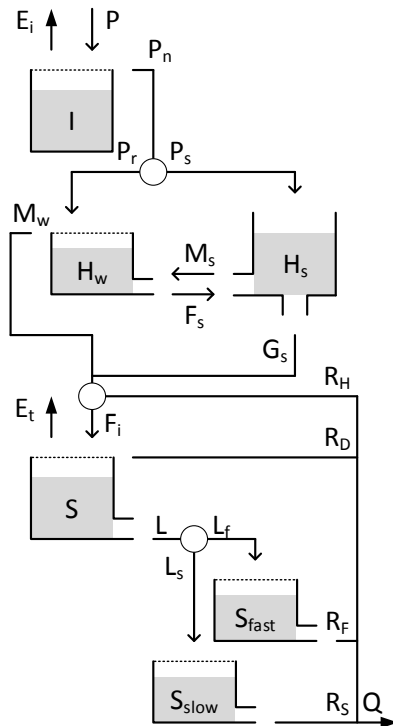
$$\frac{dI}{dt} = P - E_i - P_n \tag{B.619}$$

$$E_i = \begin{cases} E_p, & \text{if } I > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.620}$$

$$P_n = \begin{cases} P, & \text{if } I = \rho \\ 0, & \text{otherwise} \end{cases} \tag{B.621}$$

Where I [mm] is the current interception storage, refilled by precipitation P [mm/d] and drained by evaporation E_i [mm/d] and net precipitation P_n [mm/d]. E_i occurs at the potential rate E_p [mm/d] when possible. P_n only occurs when the store is at maximum capacity ρ [mm].

Figure B.45: Structure of the ECHO model

266

$$\frac{dH_s}{dt} = P_s + F_s - M_s - G_s \tag{B.622}$$

$$P_s = \begin{cases} P_n, & \text{if } T \le T_s \\ 0, & \text{otherwise} \end{cases} \tag{B.623}$$

$$M_s = \begin{cases} a_s(T - T_m), & \text{if } T > T_m, H_s > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.624}$$

$$F_s = \begin{cases} a_f a_s(T_m - T), & \text{if } T < T_m, H_w > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.625}$$

$$G_s = \begin{cases} G_{max}, & \text{if } H_s > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.626}$$

Where H_s [mm] is the current storage in the snow pack, refilled by precipitation-as-snow P_s [mm/d] and refreezing of melted snow F_s [mm/d], and drained by snowmelt M_s [mm/d] and the ground-heat flux G_s [mm/d]. P_s is calculated as all effective rainfall after interception, provided the temperature is below a threshold T_s [^oC]. M_s uses a degree-day factor a_s [$mm/^oC/d$] and threshold temperature for snowmelt T_m [^oC]. F_s occurs if the current temperature is below T_m and the degree-day rate reduced by factor a_f [-]. G_s occurs at a constant rate G_{max} [mm/d].

$$\frac{dH_w}{dt} = P_r + M_s - F_s - M_w \tag{B.627}$$

$$P_r = \begin{cases} P_n, & \text{if } T > T_s \\ 0, & \text{otherwise} \end{cases} \tag{B.628}$$

$$M_w = \begin{cases} P_r + M_s, & \text{if } H_w = \theta * H_s \\ 0, & \text{otherwise} \end{cases} \tag{B.629}$$

Where H_w [mm] is the current storage of liquid water in the snow pack, refilled by precipitation-as-rain P_r [mm/d] and snowmelt M_s [mm/d], and drained by refreezing F_s [mm/d] and outflow of melt water M_w [mm/d]. P_r is calculated as all effective rainfall after interception, provided the temperature is above a threshold T_s [^oC]. M_w occurs only if the store is at maximum capacity, which is a fraction θ [-] of the current snow pack height H_s [mm].

$$\frac{dS}{dt} = F_i - R_D - E_t - L \tag{B.630}$$

$$F_i = P_{eq} - R_H \tag{B.631}$$

$$P_{eq} = M_w + G_s \tag{B.632}$$

$$R_H = \begin{cases} max(P_{eq} - \phi, 0), & \text{if } S < S_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.633}$$

$$R_D = \begin{cases} P_{eq}, & \text{if } S = S_{max} \\ 0, & \text{otherwise} \end{cases} \tag{B.634}$$

$$E_t = min\left(max\left(0, E_{t,pot}\frac{S - sw}{sm - sw}\right), E_{t,pot}\right) \tag{B.635}$$

$$E_{t,pot} = E_p - E_i \tag{B.636}$$

$$L = K_{sat}S^c \tag{B.637}$$

Where S [mm] is the current storage in the soil moisture zone, refilled by infiltration F_i [mm/d] and drained by Dunne-type runoff R_D [mm/d], evapotranspiration E_t [mm/d] and leakage L [mm/d]. F_i is calculated as equivalent precipitation P_{eq} minus Horton-type runoff R_H. P_{eq} is the sum of melt water M_w and the ground-heat flux G_s. R_H occurs at fixed rate ϕ [mm/d] and only if the soil moisture is not saturated. R_D is equal to equivalent precipitation P_{eq} but occurs only when the store is at maximum capacity S_{max} [mm]. E_t fulfils any leftover evaporation demand after interception. E_t occurs at the potential rate until the plant stress point sm [mm], decreases linearly until the wilting point sw [mm] and is zero for any lower storage values. L has a non-linear relationship with storage through time parameter K_{sat} [d^{-1}] and coefficient c [-].

$$\frac{S_{fast}}{dt} = L_f - R_f \tag{B.638}$$

$$L_f = L - L_s \tag{B.639}$$

$$L_s = min(L, L_{max}) \tag{B.640}$$

$$R_f = k_f * S_{fast} \tag{B.641}$$

Where S_{fast} [mm] is the current storage in the fast runoff reservoir, refilled by leakage-to-fast-flow L_f [mm/d] and drained by fast runoff R_f [mm/d]. L_f depends on leakage L from soil moisture and the leakage-to-slow-flow L_s. L_s is calculated from a maximum leakage rate L_{max} [mm/d]. R_f has a linear relation with storage through time parameter k_f [mm/d].

$$\frac{dS_{slow}}{dt} = L_s - R_s \tag{B.642}$$

$$R_s = k_s * S_{slow} \tag{B.643}$$

Where S_{slow} [mm] is the current storage in the slow runoff reservoir, refilled by leakage-to-slow-flow L_s [mm/d] and drained by slow runoff R_s [mm/d]. R_s has a linear relation with storage through time parameter k_s [mm/d]. Total flow:

$$Q = R_H + R_D + R_F + R_S \tag{B.644}$$

B.2.45 Precipitation-Runoff Modelling System (PRMS) (model ID: 45)

The PRMS model (fig. B.46) is a modelling system that, in its most recent version, allows the user to specify a wide variety of catchment processes and flux equations [205]. The version presented here is a simplified version of the original PRMS model [192]. Simplifications involve the use of PET time series instead of within-model estimates based on temperature, and simpler interception and snow routines. The model has 7 stores and 18 parameters (TT, ddf, α, β, $STOR$, $RETIP$, SCN, SCX, $REMX$, $SMAX$, c_{gw}, $RESMAX$, k_1, k_2, k_3, k_4, k_5, k_6). The model aims to represent:

- Snow accumulation and melt;
- Interception by vegetation;
- Depression storage and impervious surface areas;
- Direct runoff based on catchment saturation;
- Infiltration into soil moisture and connection with deeper groundwater;
- Potentially non-linear interflow, baseflow and groundwater sink.

B.2.45.1 File names

Model: m_45_prms_18p_7s

Parameter ranges: m_45_prms_18p_7s_parameter_ ranges

B.2.45.2 Model equations
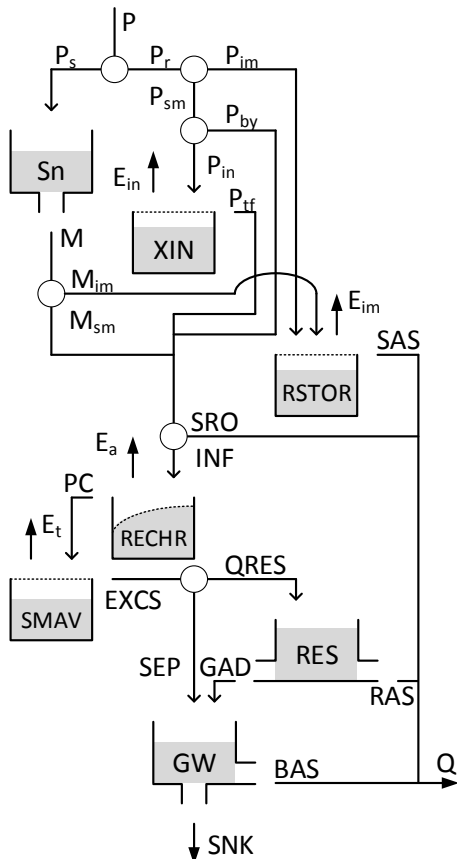
$$\frac{dSn}{dt} = P_s - M \tag{B.645}$$

$$P_s = \begin{cases} P, & \text{if } T \leq TT \\ 0, & \text{otherwise} \end{cases} \tag{B.646}$$

$$M = \begin{cases} ddf * (T - TT), & \text{if } T \geq TT \\ 0, & \text{otherwise} \end{cases} \tag{B.647}$$

Where S is the current snow storage [mm], P_s the rain that falls as snow [mm], M the snowmelt [mm] based on a degree-day factor (ddf, [mm/°C/d]) and threshold temperature for snowfall and snowmelt (TT, [°C]).

270

Figure B.46: Structure of the PRMS model

$$\frac{dXIN}{dt} = P_{in} - E_{in} - P_{tf} \tag{B.648}$$

$$P_{in} = \alpha * P_{sm} \tag{B.649}$$

$$P_{sm} = \beta * P_r \tag{B.650}$$

$$P_r = \begin{cases} P, & \text{if } T > TT \\ 0, & \text{otherwise} \end{cases} \tag{B.651}$$

$$E_{in} = \begin{cases} \beta * E_p, & \text{if } XIN > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.652}$$

$$P_{tf} = \begin{cases} P_{in}, & \text{if } XIN = STOR \\ 0, & \text{otherwise} \end{cases} \tag{B.653}$$

Where XIN [mm] is the current storage in the interception reservoir, recharged by intercepted rainfall P_{in} [mm/d] and drained by evaporation E_i [mm/d] and throughfall P_{tf} [mm/d]. P_{in} [mm/d] is the fraction α [-] of rainfall on non-impervious area P_{sm} [mm/d] that does not bypass the interception reservoir. P_{sm} [mm/d] is the fraction β [-] of rainfall P_r [mm/d] that does not fall on impervious area. Rainfall is given as all precipitation P [mm/d] that occurs when temperature T [°C] is above a threshold TT [°C]. E_i [mm/d] occurs at the potential rate E_p, corrected for the fraction of the catchment where interception can occur. Throughfall P_{tf} is all rainfall that reaches the interception reservoir when it is at maximum capacity $STOR$ [mm].

$$\frac{dRSTOR}{dt} = P_{im} + M_{im} - E_{im} - SAS \tag{B.654}$$

$$P_{im} = (1 - \beta) * P_r \tag{B.655}$$

$$M_{im} = (1 - \beta) * M \tag{B.656}$$

$$E_{im} = \begin{cases} (1 - \beta) * E_p, & \text{if } RSTOR > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.657}$$

$$SAS = \begin{cases} P_{im} + M_{im}, & \text{if } RSTOR = RETIP \\ 0, & \text{otherwise} \end{cases} \tag{B.658}$$

Where $RSTOR$ [mm] is current depression storage, refilled by rainfall and snowmelt on impervious area, P_{im} [mm/d] and M_{im} [mm/d] respectively, and drained by evaporation E_{im} [mm/d] and surface runoff SAS [mm/d]. P_{im} is given as the fraction $1 - \beta$ of rainfall

P_r. M_{im} is given as the fraction $1 - \beta$ of snowmelt M. E_{im} occurs at the potential rate E_p, corrected for the fraction of the catchment where impervious areas can occur. SAS occurs when the depression store is at maximum capacity $RETIP$ [mm].

$$\frac{dRECHR}{dt} = INF - E_a - PC \tag{B.659}$$

$$INF = M_{sm} + P_{tf} + P_{by} - SRO \tag{B.660}$$

$$M_{sm} = \beta * M \tag{B.661}$$

$$P_{by} = (1 - \alpha) * P_{sm} \tag{B.662}$$

$$SRO = \left[SCN + (SCX - SCN) * \frac{RECHR}{REMX} \right] * \left(M_{sm} + P_{tf} + P_{by} \right) \tag{B.663}$$

$$E_a = \frac{RECHR}{REMX} * \left(E_p - E_i - E_{im} \right) \tag{B.664}$$

$$PC = \begin{cases} INF, & \text{if } RECHR = REMX \\ 0, & \text{otherwise} \end{cases} \tag{B.665}$$

Where $RECHR$ [mm] is the current storage in the upper soil moisture zone, recharged by infiltration INF [mm/d] and drained by evaporation E_a [mm/d] and percolation PC [mm/d]. INF is the difference between incoming snowmelt M_{sm} [mm/d], throughfall P_{tf} [mm/d] and interception bypass P_{by} [mm/d], and surface runoff from saturated area SRO [mm/d]. S_{sm} is snowmelt from the fraction β [-] of the catchment that is not impervious. P_{by} is the fraction $1 - \alpha$ of rainfall over non-impervious area P_{sm} that bypasses the interception store. SRO has a linear relation between minimum contributing area SCN [-] and maximum contributing area SCX [-] based on current storage $RECHR$ and maximum storage $REMX$ [mm]. E_a uses a similar linear relationship and accounts for already fulfilled evaporation demand by interception and impervious areas. PC occurs when the store reaches maximum capacity.

$$\frac{dSMAV}{dt} = PC - E_t - EXCS \tag{B.666}$$

$$E_t = \begin{cases} \frac{SMAV}{SMAX} * \left(E_p - E_{in} - E_{im} - E_a \right), & \text{if } RECHR < \left(E_p - E_{in} - E_{im} \right) \\ 0, & \text{otherwise} \end{cases} \tag{B.667}$$

$$EXCS = \begin{cases} PC, & \text{if } SMAV = SMAX - REMX \\ 0, & \text{otherwise} \end{cases} \tag{B.668}$$

Where $SMAV$ [mm] is the current storage in the lower soil moisture zone, recharged by percolation from the upper zone PC [mm/d] and drained by transpiration E_t [mm/d]

and soil moisture excess $EXCS$ [mm/d]. E_t is corrected for already fulfilled evaporation demand and only occurs if the upper zone can not satisfy this demand. E_t uses a linear relationship between current storage and the maximum storage in the lower zone $SMAX - REMX$ [mm]. $EXCS$ only occurs when the store has reached maximum capacity $SMAX - REMX$.

$$\frac{dRES}{dt} = QRES - GAD - RAS \tag{B.669}$$

$$QRES = min(EXCS - SEP, 0) \tag{B.670}$$

$$GAD = k_1 \left(\frac{RES}{RESMAX} \right)^{k_2} \tag{B.671}$$

$$RAS = k_3 * RES + k_4 * RES^2 \tag{B.672}$$

$$\tag{B.673}$$

Where RES [mm] is the current storage in the runoff reservoir, filled by the difference between soil moisture excess $EXCS$ [mm/d] and constant groundwater recharge SEP [mm/d], and drained by groundwater drainage GAD [mm/d] and interflow component RAS [mm/d]. GAD is potentially non-linear using time coefficient k_1 [d^{-1}] and non-linearity coefficient k_2 [-], and is also scaled by the maximum reservoir capacity $RESMAX$ [mm]. RAS is non-linear interflow based on coefficients k_3 [d^{-1}] and k_4 [$mm^{-1}d^{-1}$].

$$\frac{dGW}{dt} = SEP + GAD - BAS - SNK \tag{B.674}$$

$$SEP = min(c_{gw}, EXCS) \tag{B.675}$$

$$BAS = k_5 * GW \tag{B.676}$$

$$SNK = k_6 * GW \tag{B.677}$$

Where GW [mm] is the current groundwater storage, refilled by groundwater recharge from soil moisture SEP and recharge from runoff reservoir GAD and drained by baseflow BAS [mm/d] and flow to deeper groundwater SNK [mm/d]. SEP occurs at the maximum rate c_{gw} [mm/d] if possible. BAS is a linear reservoir with time coefficient k_5 [d^{-1}]. SNK is a linear reservoir with time coefficient k_6 [d^{-1}]. Total flow Q_t [mm/d]:

$$Q_t = SAS + SRO + RAS + BAS \tag{B.678}$$

B.2.46 Climate and Land-use Scenario Simulation in Catchments model (model ID: 46)

The CLASSIC model (fig. B.47) is developed as a modular semi-distributed grid-based rainfall runoff model [85]. For comparability with other models the grid-based routing component is not included here, nor is the arable soil element because input data for this soil type is not supported. The model represents runoff from three different soil categories: permeable, semi-permeable and impermeable. It has 8 stores and 12 parameters (f_{ap}, f_{dp}, d_p, c_q, d_1, f_{as}, f_{ds}, d_s, d_2, c_{xq}, c_{xs}, c_u). The model aims to represent:

- Division into permeable, semi-permeable and impermeable areas;
- Infiltration into permeable soils and deficit-based soil moisture accounting;
- Infiltration into semi-permeable soils and direct runoff from semi-permeable soils (bypassing the moisture accounting);
- Fixed interception on impermeable soils;
- Linear flow routing from permeable soils;
- Fast and slow routing from semi-permeable soils;
- Linear flow routing from impermeable soils.

B.2.46.1 File names

Model: m_46_classic_12p_8s

Parameter ranges: m_46_classic_12p_8s_parameter_ ranges

B.2.46.2 Model equations
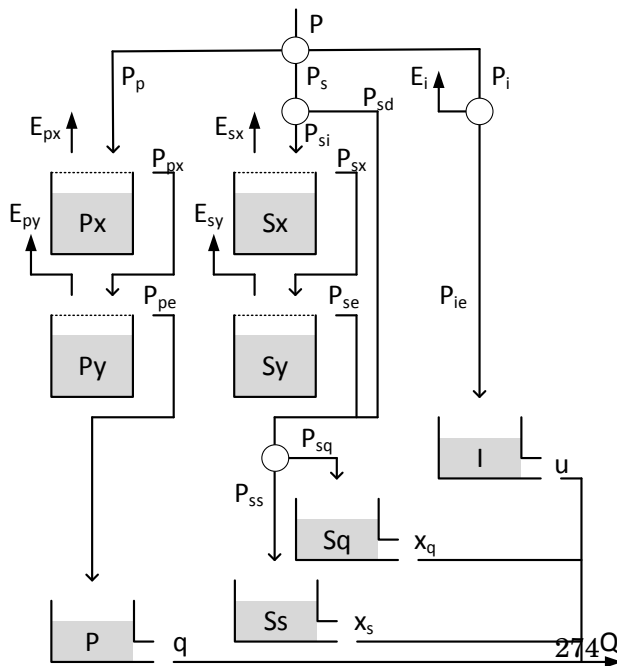
$$\frac{dP_x}{dt} = P_p - E_{px} - P_{px} \qquad \text{(B.679)}$$

$$P_p = f_{ap} * P \qquad \text{(B.680)}$$

$$E_{px} = \begin{cases} f_{ap} * E_p, & \text{if } P_x > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{(B.681)}$$

$$P_{px} = \begin{cases} P_p, & \text{if } P_x = f_{dp} * d_p \\ 0, & \text{otherwise} \end{cases}$$

$$\text{(B.682)}$$

Figure B.47: Structure of the CLASSIC model

Where P_x [mm] is the current storage in the upper permeable layer, refilled by precipitation P_p [mm/d] and drained by evaporation E_{px} [mm/d] and excess flow P_{px} [mm/d]. P_p is the fraction of precipitation P [mm/d] that falls on permeable area f_{ap} [-]. E_{px} occurs at the potential rate E_p [mm/d] whenever possible, adjusted for the fraction of area that is permeable soil. P_{px} only occurs when the store is at maximum capacity $f_{dp} * d_p$, where d_p is the total soil depth (sum of depths X and Y) in the permeable area and f_{dp} the fraction of this depth that is store X.

$$\frac{dP_y}{dt} = -P_{px} + E_{py} + P_{pe} \tag{B.683}$$

$$E_{py} = 1.9 * exp\left[\frac{-0.6523 * (P_y + f_{dp} * d_p)}{f_{dp} * d_p}\right] * \left(f_{ap} * E_p - E_{px}\right) \tag{B.684}$$

$$P_{pe} = \begin{cases} P_{px}, & \text{if } P_y = 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.685}$$

Where P_y [mm] is the current *deficit*, which is increased by evaporation E_{py} [mm/d] and decreased by inflow P_{px} [mm/d]. Effective precipitation P_{pe} [mm/d] is only generated when the deficit is 0. E_{py} decreases exponentially with increasing deficit.

$$\frac{dP}{dt} = P_{pe} - q \tag{B.686}$$

$$q = c_q * P \tag{B.687}$$

Where P [mm] is the current storage in the permeable soil routing store, refilled by effective rainfall on permeable soil P_{pe} [mm/d] and drained by baseflow q [mm/d]. q has a linear relation with storage through time scale parameter c_p [d^{-1}].

275

$$\frac{dS_x}{dt} = P_{si} - E_{sx} - P_{sx} \tag{B.688}$$

$$P_{si} = d_1 * P_s \tag{B.689}$$

$$P_s = f_{as} * P \tag{B.690}$$

$$E_{sx} = \begin{cases} f_{as} * E_p, & \text{if } S_x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.691}$$

$$P_{sx} = \begin{cases} P_s, & \text{if } S_x = f_{ds} * d_s \\ 0, & \text{otherwise} \end{cases} \tag{B.692}$$

Where S_x [mm] is the current storage in the upper semi-permeable layer, refilled by infiltration P_{si} [mm/d] and drained by evaporation E_{sx} [mm/d] and excess flow P_{sx} [mm/d]. P_{si} is the fraction d_1 [mm] of precipitation on semi-permeable area P_s that infiltrates into the soil. The complementary fraction $1 - d_1$ of P_s bypasses the soil and directly becomes effective rainfall as P_{sd}. P_s is the fraction of precipitation P [mm/d] that falls on semi-permeable area f_{as} [-]. E_{sx} occurs at the potential rate E_p [mm/d] whenever possible, adjusted for the fraction of area that is semi-permeable soil. P_{sx} only occurs when the store is at maximum capacity $f_{ds} * d_s$, where d_s is the total soil depth (sum of depths X and Y) in the semi-permeable area and f_{ds} the fraction of this depth that is store X.

$$\frac{dS_y}{dt} = -P_{sx} + E_{sy} + P_{se} \tag{B.693}$$

$$E_{sy} = 1.9 * exp \left[\frac{-0.6523 * (S_y + f_{ds} * d_s)}{f_{ds} * d_s} \right] * \left(f_{as} * E_p - E_{sx} \right) \tag{B.694}$$

$$P_{pe} = \begin{cases} P_{sx}, & \text{if } S_y = 0 \\ 0, & \text{otherwise} \end{cases} \tag{B.695}$$

Where S_y [mm] is the current *deficit*, which is increased by evaporation E_{sy} [mm/d] and decreased by inflow P_{sx} [mm/d]. Effective precipitation P_{se} [mm/d] is only generated when the deficit is 0. E_{sy} decreases exponentially with increasing deficit.

$$\frac{dS_q}{dt} = P_{sq} - x_q \tag{B.696}$$

$$P_{sq} = d_2 * (P_{se} + P_{sd}) \tag{B.697}$$

$$x_q = c_{xq} * S_q \tag{B.698}$$

Where S_q [mm] is the current storage in the semi-permeable quick soil routing store, refilled by a fraction of effective rainfall on semi-permeable soil P_{sq} [mm/d] and drained by quick flow x_q [mm/d]. P_{sq} is the fraction d_2 [-] of $(P_{se} + P_{Sd})$ that is quick flow. x_q has a linear relation with storage through time scale parameter c_{xq} [d^{-1}].

$$\frac{dS_s}{dt} = P_{ss} - x_s \tag{B.699}$$

$$P_{ss} = (1 - d_2) * (P_{se} + P_{sd}) \tag{B.700}$$

$$x_s = c_{xs} * S_s \tag{B.701}$$

Where S_s [mm] is the current storage in the semi-permeable quick soil routing store, refilled by a fraction of effective rainfall on semi-permeable soil P_{ss} [mm/d] and drained by slow flow x_s [mm/d]. P_{ss} is the fraction $1 - d_2$ [-] of $(P_{se} + P_{Sd})$ that is slow flow. x_s has a linear relation with storage through time scale parameter c_{xs} [d^{-1}].

$$\frac{dI}{dt} = P_{ie} - u \tag{B.702}$$

$$P_{ie} = P_i - E_i \tag{B.703}$$

$$P_i = P - P_p - P_s \tag{B.704}$$

$$u = c_u * I \tag{B.705}$$

Where I [mm] is the current storage in the impermeable soil routing store, refilled by effective rainfall on impermeable soil P_{ie} [mm/d] and drained by baseflow u [mm/d]. P_{ie} is the remained of precipitation on impermeable soils P_i [mm/d], after a constant evaporation E_i has been extracted. E_i is fixed at 0.5 [mm/d]. x_s has a linear relation with storage through time scale parameter c_{xs} [d^{-1}]. Total flow:

$$Q = q + x_s + x_q + u \tag{B.706}$$

B.3 Flux equations

Section B.2 gives descriptions of each model and provides both Ordinary Differential Equations and the constitutive functions that describe each model's fluxes. These constitutive functions and any relevant constraints are implemented in MARRMoT as individual *flux files*. Each *flux file* contains computer code that combines the consitituve function and constraints (if needed). *Flux files* are located in the folder "./MARRMoT/Models/Flux files/". The User Manual contains details on understanding, modifying and creating new *flux files*. Table B.1 shows a complete overview of fluxes currently implemented in MARRMoT.

Table B.1: Equations from model descriptions and their implementation in MARRMoT (Table starts on following page)

Process	Details	Function name	Constitutive function	Constraints	MARRMoT Code	Model
Abstraction	Groundwater abstraction at a constant rate	abstraction_1	$flux_{out} = \theta_1$	None, taken from a store with possible negative depth	$flux_{out} = \theta_1$	25
Baseflow	Linear reservoir	baseflow_1	$flux_{out} = \theta_1 * S$		$flux_{out} = \theta_1 * S$	2, 4, 6, 8, 9, 12, 13, 15, 16, 17, 18, 20, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 43, 44, 45, 46
	Non-linear outflow from a reservoir	baseflow_2	$flux_{out} = \left(\dfrac{1}{\theta_1}S\right)^{\frac{1}{\theta_2}}$	$flux_{out} \leq \dfrac{S}{\Delta t}$ To prevent complex numbers, S = [0,∞>	$flux_{out} = min\left(\dfrac{S}{\Delta t}, \left(\dfrac{1}{\theta_1}max(S,0)\right)^{\frac{1}{\theta_2}}\right)$	9, 11
	Empirical exponential outflow from a reservoir	baseflow_3	$flux_{out} = \dfrac{S_{max}^{-4}}{4}S^5$	Empirical equation, so interwoven with other equations that no constraints are needed. Also implicitly assumes time step $\Delta t = 1$	$flux_{out} = \dfrac{S_{max}^{-4}}{4}S^5$	7
	Exponential outflow from a deficit store	baseflow_4	$flux_{out} = \theta_1 e^{-\theta_2 S}$		$flux_{out} = \theta_1 e^{-\theta_2 S}$	14
	Non-linear outflow scaled by current relative storage	baseflow_5	$flux_{out} = \theta_1\left(\dfrac{S}{S_{max}}\right)^{\theta_2}$	$flux_{out} \leq \dfrac{S}{\Delta t}$ To prevent complex numbers, S = [0,∞>	$flux_{out} = min\left(\dfrac{S}{\Delta t}, \theta_1\left(\dfrac{max(0,S)}{S_{max}}\right)^{\theta_1}\right)$	22
	Quadratic outflow from reservoir if a storage threshold is exceeded	baseflow_6	$flux_{out} = \begin{cases} \theta_1 * S^2, & if\ S > \theta_2 \\ 0, & otherwise \end{cases}$	$flux_{out} \leq \dfrac{S}{\Delta t}$	$flux_{out} = min\left(\theta_1 * S^2, \dfrac{S}{\Delta t}\right) * [1 - logisticSmoother\_S(S, \theta_2)]$	25
	Non-linear outflow from a reservoir	baseflow_7	$flux_{out} = \theta_1 S^{\theta_2}$	$flux_{out} \leq \dfrac{S}{\Delta t}$ To prevent complex numbers, S = [0,∞>	$flux_{out} = min\left(\dfrac{S}{\Delta t}, \theta_1 max(0,S)^{\theta_2}\right)$	39, 42

Process	Details	Function name	Constitutive function	Constraints	MARRMoT Code	Model
	Exponential scaled outflow from a deficit store	baseflow_8	$flux_{out} = \theta_1\left(e^{\theta_2 S/S_{max}} - 1\right)$	$S \leq S_{max}$ $S \geq 0$	$flux_{out} = \theta_1\left(e^{\theta_2 * min\left(1, max\left(0, S/S_{max}\right)\right)} - 1\right)$	23
	Linear outflow from a reservoir if a storage threshold is exceeded	baseflow_9	$flux_{out} = \begin{cases} \theta_1(S - \theta_2), & if\ S > \theta_2 \\ 0, & otherwise \end{cases}$		$flux_{out} = \theta_1 * max(0, S - \theta_2)$	20
Capillary rise	Capillary rise scaled by relative deficit in receiving store	capillary_1	$flux_{out} = \theta_1\left[1 - \dfrac{S_1}{S_{1,max}}\right]$	$flux_{out} \leq \dfrac{S_2}{\Delta t}$	$flux_{out} = min\left(\theta_1\left[1 - \dfrac{S_1}{S_{1,max}}\right], \dfrac{S_2}{\Delta t}\right)$	37
	Capillary rise at a constant rate	capillary_2	$flux_{out} = \begin{cases} \theta_1, & if\ S \geq 0 \\ 0, & otherwise \end{cases}$	$flux_{out} \leq \dfrac{S}{\Delta t}$	$flux_{out} = min\left(\dfrac{S}{\Delta t}, \theta_1\right)$	13, 15
	Capillary rise if the receiving store is below a storage threshold	capillary_3	$flux_{out} = \begin{cases} \theta_1\left(1 - \dfrac{S_1}{\theta_2}\right), & if\ S_1 < \theta \\ 0, & otherwise \end{cases}$	$flux_{out} \leq \dfrac{S_2}{\Delta t}$	$flux_{out} = min\left(\dfrac{S_2}{\Delta t}, \theta_1\left(1 - \dfrac{S_1}{\theta_2}\right) * logisticSmoother\_S(S_1, \theta_2)\right)$	38
Depression storage	Exponential inflow rate into surface depressions	depression_1	$flux_{out} = \theta_1 * exp\left[-\theta_2 \dfrac{S}{S_{max} - S}\right] * flux_{in}$	$flux_{out} \leq \dfrac{S_{max} - S}{\Delta t}$ $S \leq S_{max}$	$flux_{out} = min\left(\theta_1 * exp\left[-\theta_2 \dfrac{S}{max(S_{max} - S, 0)}\right] * flux_{in}, \dfrac{S_{max} - S}{\Delta t}\right)$	36
Evaporation	Evaporation at the potential rate	evap_1	$E_a = \begin{cases} E_p, & if\ S \geq 0 \\ 0, & otherwise \end{cases}$	$E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left(E_p, \dfrac{S}{\Delta t}\right)$	2, 6, 12, 13, 16, 17, 18, 23, 25, 26, 27, 33, 34, 36, 38, 39, 41, 42, 44, 45, 46
	Evaporation at scaled plant-controlled rate	evap_2	$E_a = \theta_1 \dfrac{S}{S_{max}}$	$E_a \leq E_p$ $E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left(\theta_1 \dfrac{S}{S_{max}}, E_p, \dfrac{S}{\Delta t}\right)$	18, 36
	Evaporation scaled by relative storage below a wilting point and at the potential rate above wilting point	evap_3	$E_A = \begin{cases} E_p \dfrac{S}{\theta_1 S_{max}}, & if\ S < \theta_1 S_{max} \\ E_p, & otherwise \end{cases}$	$E_a \leq E_p$ $E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left(E_p \dfrac{S}{\theta_1 S_{max}}, E_p, \dfrac{S}{\Delta t}\right)$	3, 11, 14, 21, 26, 34, 37, 42
	Scaled evaporation if storage is above the wilting point, constrained by a limitation parameter	evap_4	$E_a = E_p * max\left(0, \theta_1 \dfrac{S - \theta_2 S_{max}}{S_{max} - \theta_2 S_{max}}\right)$	$E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left(E_p * max\left(0, \theta_1 \dfrac{S - \theta_2 S_{max}}{S_{max} - \theta_2 S_{max}}\right), \dfrac{S}{\Delta t}\right)$	15
	Evaporation from bare soil, scaled by relative storage	evap_5	$E_a = (1 - \theta_1)\dfrac{S}{S_{max}} E_p$	$E_a \leq E_p$ $E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left((1 - \theta_1)\dfrac{S}{S_{max}} E_p, \dfrac{S}{\Delta t}\right)$	4, 8, 9, 16

Process	Details	Function name	Constitutive function	Constraints	MARRMoT Code	Model
	Transpiration from vegetation at the potential rate if storage is above a wilting point and scaled by relative storage if not	evap_6	$E_A = \begin{cases} \theta_1 * E_p, & if\ S > \theta_2 * S_{max} \\ \theta_1 \dfrac{S}{\theta_2 S_{max}} E_p, & otherwise \end{cases}$	$E_a \leq \theta_1 E_p$ $E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left(\theta_1 E_p \dfrac{S}{\theta_2 S_{max}}, \theta_1 E_p, \dfrac{S}{\Delta t}\right)$	4, 9, 16
	Evaporation scaled by relative storage	evap_7	$E_a = \dfrac{S}{S_{max}} E_p$	$E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left(\dfrac{S}{S_{max}} E_p, \dfrac{S}{\Delta t}\right)$	1, 3, 10, 11, 19, 22, 24, 29, 30, 31, 32, 33, 35, 45
	Transpiration from vegetation, at potential rate if soil moisture is above the wilting point, and linearly decreasing if not. Also scaled by relative storage across all stores	evap_8	$E_A = \begin{cases} \dfrac{S_1}{S_1 + S_2} \theta_1 E_p, & if\ S_1 > \theta_2 \\ \dfrac{S_1}{\theta_2} * \dfrac{S_1}{S_1 + S_2} \theta_1 E_p, & otherwise \end{cases}$	$E_a \leq \dfrac{S_1}{\Delta t}$ $E_a \geq 0$	$E_a = max\left(min\left(\dfrac{S_1}{S_1 + S_2} \theta_1 E_p, \dfrac{S_1}{\theta_2} * \dfrac{S_1}{S_1 + S_2} \theta_1 E_p, \dfrac{S_1}{\Delta t}\right), 0\right)$	8
	Evaporation from bare soil scaled by relative storage and by relative water availability across all stores	evap_9	$E_a = \dfrac{S_1}{S_1 + S_2} * (1 - \theta_1) \dfrac{S_1}{S_{max} - S_2} E_p$	$E_a \leq \dfrac{S_1}{\Delta t}$ $E_a \geq 0$	$E_a = max\left(min\left(\dfrac{S_1}{S_1 + S_2} * (1 - \theta_1) \dfrac{S_1}{S_{max} - S_2} E_p, \dfrac{S_1}{\Delta t}\right), 0\right)$	8
	Evaporation from bare soil, scaled by relative storage	evap_10	$E_a = \theta_1 \dfrac{S}{S_{max}} E_p$	$E_a \leq E_p$ $E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left(\theta_1 \dfrac{S}{S_{max}} E_p, \dfrac{S}{\Delta t}\right)$	8
	Evaporation quadratically related to current soil moisture	evap_11	$E_a = \left(2 \dfrac{S}{S_{max}} - \left(\dfrac{S}{S_{max}}\right)^2\right) E_p$	$E_a \geq 0$	$E_a = max\left(0, \left(2 \dfrac{S}{S_{max}} - \left(\dfrac{S}{S_{max}}\right)^2\right) E_p\right)$	7
	Evaporation from deficit store, with exponential decline as deficit goes below a threshold	evap_12	$E_a = min\left(1, e^{2\left(1 - \frac{S}{\theta_1}\right)}\right) E_p$		$E_a = min\left(1, e^{2\left(1 - \frac{S}{\theta_1}\right)}\right) E_p$	5
	Exponentially scaled evaporation	evap_13	$E_a = \theta_1^{\theta_2} E_p$	$E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left(\theta_1^{\theta_2} E_p, \dfrac{S}{\Delta t}\right)$	40
	Exponentially scaled evaporation that only activates if another store goes below a certain threshold	evap_14	$E_A = \begin{cases} \theta_1^{\theta_2} E_p, & if\ S_2 \leq S_{2,min} \\ 0, & otherwise \end{cases}$	$E_a \leq \dfrac{S_1}{\Delta t}$	$E_a = min\left(\theta_1^{\theta_2} E_p, \dfrac{S_1}{\Delta t}\right) * logisticSmoother\_S(S_2, S_{2,min})$	40
	Scaled evaporation if another store is below a threshold	evap_15	$E_a = \begin{cases} \dfrac{S_1}{S_{max}} E_p, & if\ S_2 < \theta_1 \\ 0, & otherwise \end{cases}$	$E_a \leq \dfrac{S_1}{\Delta t}$	$E_a = min\left(\dfrac{S_1}{S_{1,max}} * E_p * logisticSmoother\_S(S_2, \theta_2), \dfrac{S_1}{\Delta t}\right)$	41, 45

Process	Details	Function name	Constitutive function	Constraints	MARRMoT Code	Model
	Scaled evaporation if another store is below a threshold	evap_16	$E_a = \begin{cases} \theta_1 E_p, & if\ S_2 < \theta_2 \\ 0, & otherwise \end{cases}$	$E_a \leq \dfrac{S_1}{\Delta t}$	$E_a = min\left(\theta_1 * E_p * logisticSmoother\_S(S_2, \theta_2), \dfrac{S_1}{\Delta t}\right)$	17, 25
	Scaled evaporation from a store that allows negative values	evap_17	$E_a = \dfrac{1}{1 + e^{-\theta_1 * S}} E_p$	None, because the store is allowed to go negative	$E_a = \dfrac{1}{1 + e^{-\theta_1 * S}} E_p$	39
	Exponentially declining evaporation from deficit store	evap_18	$E_a = \theta_1 e^{\frac{-\theta_2 S}{\theta_3}} E_p$		$E_a = \theta_1 e^{\frac{-\theta_2 S}{\theta_3}} E_p$	46
	Non-linear scaled evaporation	evap_19	$E_a = \theta_1 \left(\dfrac{S}{S_{max}}\right)^{\theta_2} E_p$	$E_a \leq E_p$ $E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left(\theta_1 * max\left(0, \dfrac{S}{S_{max}}\right)^{\theta_2} E_p, E_p, \dfrac{S}{\Delta t}\right)$	23, 43
	Evaporation limited by a maximum evaporation rate and scaled below a wilting point	evap_20	$E_A = \begin{cases} \theta_1 \dfrac{S}{\theta_2 S_{max}}, & if\ S < \theta_2 S_{max} \\ E_p, & otherwise \end{cases}$	$E_a \leq E_p$ $E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left(\theta_1 \dfrac{S}{\theta_2 S_{max}}, E_p, \dfrac{S}{\Delta t}\right)$	20
	Threshold-based evaporation with constant minimum rate	evap_21	$E_a = \begin{cases} E_p, & if\ S > \theta_1 \\ \dfrac{S}{\theta_1} E_p, & if\ \theta_2 \theta_1 \geq S \geq \theta_1 \\ \theta_2\, E_p & otherwise \end{cases}$	$E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left(max\left(\theta_2, min\left(\dfrac{S}{\theta_1}, 1\right)\right) * E_p, \dfrac{S}{\Delta t}\right)$	28
	Threshold-based evaporation rate	evap_22	$E_a = \begin{cases} E_p, & if\ S > \theta_1 \\ \dfrac{S - \theta_1}{\theta_1 - \theta_2} E_p, & if\ \theta_2 \theta_1 \geq S \geq \theta_1 \\ 0 & otherwise \end{cases}$	$E_a \leq \dfrac{S}{\Delta t}$	$E_a = min\left(\dfrac{S}{\Delta t}, min\left(E_p, max\left(0, \dfrac{S - \theta_1}{\theta_2 - \theta_1} E_p\right)\right)\right)$	44
Exchange	Water exchange between aquifer and channel	exchange_1	$flux_{out} = \begin{cases} \theta_1 * \left|\dfrac{S}{\Delta t}\right| + \theta_2\left(1 - exp\left[-\theta_3 * \left|\dfrac{S}{\Delta t}\right|\right]\right), & if\ S \geq 0 \\ -\left[\theta_1 * \left|\dfrac{S}{\Delta t}\right| + \theta_2\left(1 - exp\left[-\theta_3 * \left|\dfrac{S}{\Delta t}\right|\right]\right)\right], & if\ S < 0 \end{cases}$	$\begin{cases} No\ constraint \\ flux_{out} \leq flux_{in} \end{cases}$ The "channel" store in this model has 0 time delay, so the incoming flux to the channel is the maximum channel-to-groundwater flux size. Groundwater has infinite depth	$flux_{out} = max\left(\begin{array}{c}\left[\theta_1 * \left|\dfrac{S}{\Delta t}\right| + \theta_2 * \left(1 - exp\left[-\theta_3 * \left|\dfrac{S}{\Delta t}\right|\right]\right)\right] \\ * sign(S), -flux_{in}\end{array}\right)$	36
	Water exchange based on relative storages	exchange_2	$flux_{out} = \theta_1 \left(\dfrac{S_1}{S_{1,max}} - \dfrac{S_2}{S_{2,max}}\right)$		$flux_{out} = \theta_1 \left(\dfrac{S_1}{S_{1,max}} - \dfrac{S_2}{S_{2,max}}\right)$	38
	Water exchange with infinite size store based on threshold	exchange_3	$flux_{out} = \theta_1 * (S - \theta_2)$		$flux_{out} = \theta_1 * (S - \theta_2)$	36

282

Process	Details	Function name	Constitutive function	Constraints	MARRMoT Code	Model
Infiltration	Infiltration as exponentially declining based on relative storage (taken from a flux)	infiltration_1	$flux_{out} = \theta_1 * exp\left[-\theta_2 \dfrac{S}{S_{max}}\right]$	$flux_{out} \leq flux_{in}$	$flux_{out} = min\left(\theta_1 * exp\left[-\theta_2 \dfrac{S}{S_{max}}\right], flux_{in}\right)$	18, 36, 44
	Delayed infiltration as exponentially declining based on relative storage (taken from a store)	infiltration_2	$flux_{out} = \theta_1 * exp\left[-\theta_2 \dfrac{S_1}{S_{1,max}}\right] - flux_{used}$	$0 \leq flux_{out} \leq \dfrac{S_2}{\Delta t}$	$flux_{out} = max\left(min\left(\theta_1 * exp\left[-\theta_2 \dfrac{S_1}{S_{1,max}}\right] - flux_{used}, \dfrac{S_2}{\Delta t}\right), 0\right)$	36
	Infiltration to soil moisture of liquid water stored in snow pack	infiltration_3	$flux_{out} = \begin{cases} flux_{in}, & if\ S \geq S_{max} \\ 0, & otherwise \end{cases}$		$flux_{out} = flux_{in}[1 - logisticSmoother\_S(S, S_{max})]$	37
	Constant infiltration rate	infiltration_4	$flux_{out} = \theta_1$	$flux_{out} \leq flux_{in}$	$flux_{out} = min(flux_{in}, \theta_1)$	15, 23, 40, 44
	Maximum infiltration rate non-linearly based on relative deficit and storage	infiltration_5	$flux_{out} = \theta_1\left(1 - \dfrac{S_1}{S_{1,max}}\right)\left(\dfrac{S_2}{S_{2,max}}\right)^{-\theta_2}$	To prevent complex numbers, S = $[0,\infty>$ To prevent numerical issues with a theoretical infinite infiltration rate, flux$_{out}$ < 10^9	$flux_{out} = min\left(10^9, \theta_1\left(1 - \dfrac{S_1}{S_{1,max}}\right)max\left(0, \dfrac{S_2}{S_{2,max}}\right)^{-\theta_2}\right)$	23
	Infiltration rate non-linearly scaled by relative storage	infiltration_6	$flux_{out} = \theta_1\left(\dfrac{S}{S_{max}}\right)^{\theta_2} flux_{in}$	$flux_{out} \leq flux_{in}$	$flux_{out} = min\left(\theta_1 * max\left(0, \dfrac{S}{S_{max}}\right)^{\theta_2} flux_{in}, flux_{in}\right)$	43
Interception	Interception excess when maximum capacity is reached	interception_1	$flux_{out} = \begin{cases} flux_{in}, & if\ S \geq S_{max} \\ 0, & otherwise \end{cases}$		$flux_{out} = flux_{in}[1 - logisticSmoother\_S(S, S_{max})]$	16, 18, 22, 26, 34, 36, 39, 42, 44, 45
	Interception excess after a constant amount is intercepted	interception_2	$flux_{out} = \begin{cases} flux_{in} - \theta_1, & if\ flux_{in} \geq 0 \\ 0, & otherwise \end{cases}$	$flux_{out} \geq 0$	$flux_{out} = max(flux_{in} - \theta_1, 0)$	2, 13, 15
	Interception excess after a fraction is intercepted	interception_3	$flux_{out} = \theta_1$		$flux_{out} = \theta_1$	8
	Interception excess after a time-varying fraction is intercepted	interception_4	$flux_{out} = \left(\theta_1 + (1 - \theta_1) * cos\left(2\pi \dfrac{t * \Delta t - \theta_2}{t_{max}}\right)\right) * flux_{in}$	$flux_{out} \geq 0$	$flux_{out} = max\left(0, \theta_1 + (1 - \theta_1) * cos\left(2\pi \dfrac{t * \Delta t - \theta_2}{t_{max}}\right)\right) * flux_{in}$	32, 35
	Interception excess after a combined absolute amount and fraction are intercepted	interception_5	$flux_{out} = \begin{cases} \theta_1 * flux_{in} - \theta_2, & if\ flux_{in} \geq 0 \\ 0, & otherwise \end{cases}$	$flux_{out} \geq 0$	$flux_{out} = max(\theta_1 * flux_{in} - \theta_2, 0)$	23

Process	Details	Function name	Constitutive function	Constraints	MARRMoT Code	Model
Interflow	Interflow as a scaled fraction of an incoming flux	interflow_1	$flux_{out} = \theta_1 \dfrac{S}{S_{max}} * flux_{in}$		$flux_{out} = \theta_1 \dfrac{S}{S_{max}} * flux_{in}$	18, 36
	Non-linear interflow	interflow_2	$flux_{out} = \theta_1 S^{(1+\theta_2)}$	$flux_{out} \leq \dfrac{S}{\Delta t}$ To prevent complex numbers, S = $[0,\infty>$	$flux_{out} = min\left(\theta_1 max(S,0)^{(1+\theta_2)}, max\left(\dfrac{S}{\Delta t},0\right)\right)$	37
	Non-linear interflow (variant)	interflow_3	$flux_{out} = \theta_1 S^{\theta_2}$	$flux_{out} \leq \dfrac{S}{\Delta t}$ To prevent complex numbers, S = $[0,\infty>$	$flux_{out} = min\left(\theta_1 max(S,0)^{\theta_2}, max\left(\dfrac{S}{\Delta t},0\right)\right)$	10, 19, 42, 43
	Combined linear and scaled quadratic interflow	interflow_4	$flux_{out} = \theta_1 S + \theta_2 S^2$	$flux_{out} \leq \dfrac{S}{\Delta t}$ To prevent complex numbers, S = $[0,\infty>$	$flux_{out} = min\left(\theta_1 max(S,0) + \theta_2 max(S,0)^2, max\left(\dfrac{S}{\Delta t},0\right)\right)$	45
	Linear interflow	interflow_5	$flux_{out} = \theta_1 * S$		$flux_{out} = \theta_1 * S$	28, 33, 41
	Scaled linear interflow if a storage in the receiving store exceeds a threshold	interflow_6	$flux_{out} = \begin{cases} \theta_1 * S_1 * \dfrac{S_2/S_{2,max} - \theta_2}{1 - \theta_2}, & if \; S_2/S_{2,max} > \theta_2 \\ 0, & otherwise \end{cases}$	$\dfrac{S_2}{S_{2,max}} \leq 1$	$flux_{out} = \left(\theta_1 * S_1 * \dfrac{min\left(1, S_2/S_{2,max}\right) - \theta_2}{1 - \theta_2}\right)$ $* \left[1 - logisticSmoother\_S\left(\dfrac{S_2}{S_{2,max}}, \theta_2\right)\right]$	41
	Non-linear interflow if storage exceeds a threshold	interflow_7	$flux_{out} = \begin{cases} \left(\dfrac{S - \theta_1 S_{max}}{\theta_2}\right)^{\frac{1}{\theta_3}}, & if \; S > \theta_1 S_{max} \\ 0, & otherwise \end{cases}$	$flux_{out} \leq \dfrac{S - \theta_1 S_{max}}{\Delta t}$ To prevent complex numbers, S-θ_1Smax = $[0,\infty>$	$flux_{out} = min\left(max\left(0, \dfrac{S - \theta_1 S_{max}}{\Delta t}\right), \left(\dfrac{max(0, S - \theta_1 S_{max})}{\theta_2}\right)^{\frac{1}{\theta_3}}\right)$	9
	Linear interflow if storage exceeds a threshold	interflow_8	$flux_{out} = \begin{cases} \theta_1(S - \theta_2), & if \; S > \theta_2 \\ 0, & otherwise \end{cases}$		$flux_{out} = max\left(0, \theta_1(S - \theta_2)\right)$	3, 12, 27, 38
	Non-linear interflow if storage exceeds a threshold (variant)	interflow_9	$flux_{out} = \begin{cases} \left(\theta_1(S - \theta_2)\right)^{\theta_3}, & if \; S > \theta_2 \\ 0, & otherwise \end{cases}$	$flux_{out} \leq \dfrac{S - \theta_2}{\Delta t}$ To prevent complex numbers, S-θ_2 = $[0,\infty>$	$flux_{out} = min\left(\dfrac{S - \theta_2}{\Delta t}, \left(\theta_1 * max(0, S - \theta_2)\right)^{\theta_3}\right)$	4, 11, 16, 39
	Scaled linear interflow if storage exceeds a threshold	interflow_10	$flux_{out} = \begin{cases} \theta_1 \dfrac{(S - \theta_2)}{\theta_3}, & if \; S > \theta_2 \\ 0, & otherwise \end{cases}$		$flux_{out} = \theta_1 \dfrac{max(0, S - \theta_2)}{\theta_3}$	14
	Constant interflow if storage exceeds a threshold	interflow_11	$flux_{out} = \begin{cases} \theta_1, & if \; S > \theta_2 \\ 0, & otherwise \end{cases}$	$flux_{out} \leq \dfrac{S - \theta_2}{\Delta t}$	$flux_{out} = min\left(\theta_1, \dfrac{S - \theta_2}{\Delta t}\right) * [1 - logisticSmoother\_S(S, \theta_2)]$	20

Process	Details	Function name	Constitutive function	Constraints	MARRMoT Code	Model
Misc	Auxiliary function to find contributing area	area_1	$A = \begin{cases} \theta_1 \left[\dfrac{S - S_{min}}{S_{max} - S_{min}}\right]^{\theta_2}, & if\ S > S_{min} \\ 0, & otherwise \end{cases}$	$A \leq 1$	$A = min\left(1, \theta_1 \left[\dfrac{S - S_{min}}{S_{max} - S_{min}}\right]^{\theta_2}\right) * [1 - logisticSmoother\_S(S, S_{min})]$	23
	General effective flow (returns flux [mm/d])	effective_1	$flux_{out} = \begin{cases} flux_{in,1} - flux_{in,2}, & if\ flux_{in,1} > flux_{in,2} \\ 0, & otherwise \end{cases}$		$flux_{out} = max\left(0, flux_{in,1} - flux_{in,2}\right)$	22, 23, 25, 39, 40, 42, 43, 44, 45, 46
	Storage excess when store size changes (returns flux [mm/d])	excess_1	$flux_{out} = \dfrac{S - S_{max,new}}{\Delta t}$	$flux_{out} \geq 0$	$flux_{out} = max\left(\dfrac{S - S_{max,new}}{\Delta t}, 0\right)$	10, 19, 22, 37, 44
	Phenology-based correction factor for potential evapotranspiration (returns flux [mm/d])	phenology_1	$E_p^* = \begin{cases} 0, & if\ T(t) < \theta_1 \\ \dfrac{T(t) - \theta_1}{\theta_2 - \theta_1} * E_p, & if\ \theta_1 \leq T(t) < \theta_2 \\ E_p, & if\ T(t) \geq \theta_2 \end{cases}$		$E_p^* = min\left(1, max\left(0, \dfrac{T(t) - \theta_1}{\theta_2 - \theta_1}\right)\right) * E_p$	35
	Phenology-based maximum interception capacity (returns store size [mm])	phenology_2	$S_{max} = \theta_1\left(1 + \theta_2 sin\left(2\pi \dfrac{t * \Delta t - \theta_3}{t_{max}}\right)\right)$	Assumes $0 \leq \theta_2 \leq 1$ to guarantee $S_{max} \geq 0$	$S_{max} = \theta_1\left(1 + \theta_2 sin\left(2\pi \dfrac{t * \Delta t - \theta_3}{t_{max}}\right)\right)$	22
	Split flow (returns flux [mm/d])	split_1	$flux_{out} = \theta_1 * flux_{in}$		$flux_{out} = \theta_1 * flux_{in}$	5, 11, 13, 17, 21, 25, 26, 28, 29, 33, 34, 40, 41, 42, 43, 45, 46
Percolation	Percolation at a constant rate	percolation_1	$flux_{out} = \begin{cases} \theta_1, & if\ S \geq 0 \\ 0, & otherwise \end{cases}$	$flux_{out} \leq \dfrac{S}{\Delta t}$	$flux_{out} = min\left(\dfrac{S}{\Delta t}, \theta_1\right)$	37
	Percolation scaled by current relative storage	percolation_2	$flux_{out} = \theta_1 \dfrac{S}{S_{max}}$	$flux_{out} \leq \dfrac{S}{\Delta t}$	$flux_{out} = min\left(\dfrac{S}{\Delta t}, \theta_1 \dfrac{S}{S_{max}}\right)$	21, 26, 34
	Non-linear percolation (empirical)	percolation_3	$flux_{out} = \dfrac{S_{max}^{-4}}{4}\left(\dfrac{4}{9}\right)^{-4} S^5$		$flux_{out} = \dfrac{S_{max}^{-4}}{4}\left(\dfrac{4}{9}\right)^{-4} S^5$	7
	Demand-based percolation scaled by available moisture	percolation_4	$flux_{out} = \dfrac{S}{S_{max}}\left[\theta_1\left\{1 + \theta_2\left(\dfrac{\sum deficiencies}{\sum capacities}\right)^{\theta_3}\right\}\right]$	$flux_{out} \leq \dfrac{S}{\Delta t}$ To avoid erratic numerical behaviour, $flux_{out} \geq 0$	$flux_{out} = max\left(0, min\left(\dfrac{S}{\Delta t}, \dfrac{max(S,0)}{S_{max}} * \left[\theta_1\left\{1 + \theta_2\left(\dfrac{\sum deficiencies}{\sum capacities}\right)^{\theta_3}\right\}\right]\right)\right)$	33
	Non-linear percolation	percolation_5	$flux_{out} = \theta_1\left(\dfrac{S}{S_{max}}\right)^{\theta_2}$	$flux_{out} \leq \dfrac{S}{\Delta t}$ To prevent complex numbers, S = $[0, \infty>$	$flux_{out} = min\left(\dfrac{S}{\Delta t}, \theta_1\left(\dfrac{max(0,S)}{S_{max}}\right)^{\theta_2}\right)$	22

Process	Details	Function name	Constitutive function	Constraints	MARRMoT Code	Model
	Threshold-based percolation from a store that can reach negative values	percolation_6	$flux_{out} = \begin{cases} \theta_1, & if\ S \geq \theta_2 \\ \theta_1 \dfrac{S}{\theta_2}, & if\ 0 < S < \theta_2 \\ 0, & otherwise \end{cases}$	$flux_{out} \leq \dfrac{S}{\Delta t}$	$flux_{out} = min\left(\dfrac{S}{\Delta t}, \theta_1 min\left[1, \dfrac{max(0,S)}{\theta_2}\right]\right)$	39
Recharge	Recharge as scaled fraction of incoming flux	recharge_1	$flux_{out} = \theta_1 \dfrac{S}{S_{max}} * flux_{in}$		$flux_{out} = \theta_1 \dfrac{S}{S_{max}} * flux_{in}$	18, 36
	Recharge as non-linear scaling of incoming flux	recharge_2	$flux_{out} = \left(\dfrac{S}{S_{max}}\right)^{\theta_1} * flux_{in}$	To prevent complex numbers, S = $[0,\infty>$	$flux_{out} = \left(\dfrac{max(0,S)}{S_{max}}\right)^{\theta_1} * flux_{in}$	7, 37, 45
	Linear recharge	recharge_3	$flux_{out} = \theta_1 * S$		$flux_{out} = \theta_1 * S$	19, 23, 24, 27, 30, 31, 32, 35, 38, 42
	Constant recharge from a store	recharge_4	$flux_{out} = \begin{cases} \theta_1, & if\ S \geq 0 \\ 0, & otherwise \end{cases}$	$flux_{out} \leq \dfrac{S}{\Delta t}$	$flux_{out} = min\left(\dfrac{S}{\Delta t}, \theta_1\right)$	23, 44
	Recharge to fulfil evaporation demand if the receiving store is below a threshold	recharge_5	$flux_{out} = \begin{cases} \theta_1 S_1 \left(1 - \dfrac{S_2}{\theta_2}\right), & if\ S_2 < \theta_2 \\ 0, & otherwise \end{cases}$		$flux_{out} = \theta_1 S_1 \left[1 - min\left(1, \dfrac{S_2}{\theta_2}\right)\right]$	20
	Non-linear recharge	recharge_6	$flux_{out} = \theta_1 S^{\theta_2}$	$flux_{out} \leq \dfrac{S}{\Delta t}$ To prevent complex numbers, S = $[0,\infty>$	$flux_{out} = min\left(\theta_1 max(S,0)^{\theta_2}, max\left(\dfrac{S}{\Delta t}, 0\right)\right)$	44
	Constant recharge from a flux	recharge_7	$flux_{out} = \theta_1$	$flux_{out} \leq flux_{in}$	$flux_{out} = min(flux_{in}, \theta_1)$	45
Routing	Threshold-based non-linear routing	routing_1	$flux_{out} = \begin{cases} \theta_1 S^{\theta_2}, & if\ flux_{out} < \theta_3 S \\ \theta_3 S, & otherwise \end{cases}$	$flux_{out} \leq \dfrac{S}{\Delta t}$	$flux_{out} = min\left(\dfrac{S}{\Delta t}, \theta_1 max(S,0)^{\theta_2}, \theta_3 \dfrac{S}{\Delta t}\right)$	39
Saturation excess	Saturation excess from a store that has reached maximum capacity	saturation_1	$flux_{out} = \begin{cases} flux_{in}, & if\ S \geq S_{max} \\ 0, & otherwise \end{cases}$		$flux_{out} = flux_{in}\left[1 - logisticSmoother\_S(S, S_{max})\right]$	1, 3, 4, 6, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 22, 24, 25, 30, 31, 32, 33, 35, 36, 39, 40, 41, 44, 45, 46

continued …

Process	Details	Function name	Constitutive function	Constraints	MARRMoT Code	Model
	Saturation excess from a store with different degrees of saturation	saturation_2	$flux_{out} = \left(1 - \left(1 - \frac{S}{S_{max}}\right)^{\theta_1}\right) * flux_{in}$	To prevent complex numbers, S/Smax = [0,∞>	$flux_{out} = \left(1 - \left(min\left(1, max\left(0, \left(1 - \frac{S}{S_{max}}\right)\right)\right)\right)^{\theta_1}\right) * flux_{in}$	2, 13, 22, 28, 29
	Saturation excess from a store with different degrees of saturation (exponential variant)	saturation_3	$flux_{out} = \left(1 - \frac{1}{1 + exp\left(\frac{S/S_{max} + 0.5}{\theta_1}\right)}\right) * flux_{in}$		$flux_{out} = \left(1 - \frac{1}{1 + exp\left(\frac{S/S_{max} + 0.5}{\theta_1}\right)}\right) * flux_{in}$	21, 26, 34
	Saturation excess from a store with different degrees of saturation (quadratic variant)	saturation_4	$flux_{out} = \left(1 - \left(\frac{S}{S_{max}}\right)^2\right) * flux_{in}$	$0 \leq flux_{out}$	$flux_{out} = max\left(0, \left(1 - \left(\frac{S}{S_{max}}\right)^2\right) * flux_{in}\right)$	7
	Deficit store: exponential saturation excess based on current storage and a threshold parameter	saturation_5	$flux_{out} = \left(1 - min\left(1, \left(\frac{S}{\theta_1}\right)^{\theta_2}\right)\right) * flux_{in}$	To prevent complex numbers, S = [0,∞>	$flux_{out} = \left(1 - min\left(1, \left(\frac{max(S,0)}{\theta_1}\right)^{\theta_2}\right)\right) * flux_{in}$	5
	Saturation excess from a store with different degrees of saturation (linear variant)	saturation_6	$flux_{out} = \theta_1 \frac{S}{S_{max}} * flux_{in}$		$flux_{out} = \theta_1 \frac{S}{S_{max}} * flux_{in}$	40
	Saturation excess from a store with different degrees of saturation (gamma function variant)	saturation_7	$flux_{out} = flux_{in} \left\{ \int_{x=\theta_5*S+\theta_4}^{x=\infty} \frac{1}{\theta_1 \Gamma(\theta_2)} \left(\frac{x-\theta_3}{\theta_1}\right)^{\theta_2-1} e^{\left(-\frac{x-\theta_3}{\theta_1}\right)}, x > \theta_3 \atop 0 \right.$	To prevent numerical problems, S = [0,∞>	$flux_{out} = flux_{in} * integral\left(\frac{1}{\theta_1 \Gamma(\theta_2)} \left(\frac{max(x-\theta_3,0)}{\theta_1}\right)^{\theta_2-1} * e^{\left(-1*\frac{max(x-\theta_3,0)}{\theta_1}\right)}, \theta_5 * max(S,0) + \theta_4, \infty\right)$	14
	Saturation excess flow from a store with different degrees of saturation (min-max linear variant)	saturation_8	$flux_{out} = \left[\theta_1 + (\theta_2 - \theta_1)\frac{S}{S_{max}}\right] * flux_{in}$	$flux_{out} \leq flux_{in}$	$flux_{out} = \left[\theta_1 + (\theta_2 - \theta_1)\frac{S}{S_{max}}\right] * flux_{in}$	45
	Deficit store: saturation excess from a store that has reached maximum capacity	saturation_9	$flux_{out} = \begin{cases} flux_{in}, & if\ S = 0 \\ 0, & otherwise \end{cases}$		$flux_{out} = flux_{in} * logisticSmoother\_S(S,0)$	17, 25, 43, 46
	Saturation excess flow from a store with different degrees of saturation (min-max exponential variant)	saturation_10	$flux_{out} = min\left(\theta_1, \theta_2 + \theta_2 e^{\theta_3 S}\right) * flux_{in}$		$flux_{out} = min\left(\theta_1, \theta_2 + \theta_2 e^{\theta_3 S}\right) * flux_{in}$	39

continued …

Process	Details	Function name	Constitutive function	Constraints	MARRMoT Code	Model
	Saturation excess flow from a store with different degrees of saturation (min exponential variant)	saturation_11	$flux_{out} = \begin{cases} \left(\theta_1 \left[\frac{S - S_{min}}{S_{max} - S_{min}}\right]^{\theta_2}\right) flux_{in}, \ if \ S > S_{min} \\ 0, \ otherwise \end{cases}$	$flux_{out} \leq flux_{in}$	$flux_{out} = min\left(1, \theta_1\left[\frac{S - S_{min}}{S_{max} - S_{min}}\right]^{\theta_2}\right) flux_{in} * [1 - logisticSmoother\_S(S, S_{min})]$	23
	Saturation excess flow from a store with different degrees of saturation (min-max linear variant)	saturation_12	$flux_{out} = \frac{\theta_1 - \theta_2}{1 - \theta_2} flux_{in}$	$flux_{out} \geq 0$	$flux_{out} = max\left(0, \frac{\theta_1 - \theta_2}{1 - \theta_2}\right) flux_{in}$	23
	Saturation excess flow from a store with different degrees of saturation (normal distribution variant)	saturation_13	$flux_{out} = flux_{in} * \int_{-\infty}^{\xi} \frac{1}{\sqrt{2\pi}} exp\left[-\frac{\xi^2}{2}\right] d\xi, \ with \ \xi$ $= \frac{log(S/\theta_1)}{log(\theta_1/\theta_2)}$		$flux_{out} = flux_{in} * normcdf\left(\frac{log(max(0,S)/\theta_1)}{log(\theta_1/\theta_2)}\right)$	42
	Saturation excess flow from a store with different degrees of saturation (two-part exponential variant)	saturation_14	$flux_{out}$ $= flux_{in} \begin{cases} (0.5 - \theta_1)^{1-\theta_2}\left(\frac{S}{S_{max}}\right)^{\theta_3}, if \frac{S}{S_{max}} \leq 0.5 - \theta_1 \\ 1 - (0.5 - \theta_1)^{1-\theta_2}\left(1 - \frac{S}{S_{max}}\right)^{\theta_3}, otherwise \end{cases}$		$flux_{out} = \left(\begin{array}{c} \left((0.5 - \theta_1)^{1-\theta_2} max\left(0, \frac{S}{S_{max}}\right)^{\theta_3}\right) * \\ \left(\frac{S}{S_{max}} \leq 0.5 - \theta_1\right) + \\ \left(1 - (0.5 + \theta_1)^{1-\theta_2} max\left(0, 1 - \frac{S}{S_{max}}\right)^{\theta_3}\right) * \\ \frac{S}{S_{max}} > 0.5 - \theta_1 \end{array}\right) * flux_{in}$	28
Snow	Snowfall based on temperature threshold	snowfall_1	$flux_{out} = \begin{cases} flux_{in}, if \ T \leq T_{threshold} \\ 0, \ otherwise \end{cases}$		$flux_{out} = flux_{in} * [logisticSmoother\_T(T, T_{threshold})]$	6, 12, 30, 31, 32, 34, 35, 41, 43, 44, 45
	Snowfall based on a temperature threshold interval	snowfall_2	$flux_{out}$ $= \begin{cases} flux_{in}, \ if \ T \leq \theta_1 - \frac{1}{2}\theta_2 \\ flux_{in} * \frac{\theta_1 + \frac{1}{2}\theta_2 - T}{\theta_2}, \ if \ \theta_1 - \frac{1}{2}\theta_2 < T < \theta_1 + \frac{1}{2}\theta_2 \\ 0, \ if \ T \geq \theta_1 + \frac{1}{2}\theta_2 \end{cases}$		$flux_{out} = min\left(flux_{in}, max\left(0, flux_{in} * \frac{\theta_1 + \frac{1}{2}\theta_2 - T}{\theta_2}\right)\right)$	37
	Rainfall based on temperature threshold	rainfall_1	$flux_{out} = \begin{cases} flux_{in}, if \ T > T_{threshold} \\ 0, \ otherwise \end{cases}$		$flux_{out} = flux_{in} * [1 - logisticSmoother\_T(T, T_{threshold})]$	6, 12, 30, 31, 32, 34, 35, 41, 43, 44, 45
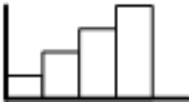
Process	Details	Function name	Constitutive function	Constraints	MARRMoT Code	Model
	Snowfall based on a temperature threshold interval	rainfall_2	$flux_{out}$ $= \begin{cases} 0, \ if \ T \leq \theta_1 - \frac{1}{2}\theta_2 \\ flux_{in} * \dfrac{\theta_1 + \frac{1}{2}\theta_2 - T}{\theta_2}, \ if \ \theta_1 - \frac{1}{2}\theta_2 < T < \theta_1 + \frac{1}{2}\theta_2 \\ flux_{in}, \ if \ T \geq \theta_1 + \frac{1}{2}\theta_2 \end{cases}$		$flux_{out} = min\left(flux_{in}, max\left(0, flux_{in} * \dfrac{T - \left(\theta_1 - \frac{1}{2}\theta_2\right)}{\theta_2}\right)\right)$	37
	Refreezing of stored melted snow	refreeze_1	$flux_{out} = \begin{cases} \theta_1 * \theta_2 * (T_{threshold} - T), \ if \ T \leq T_{threshold} \\ 0, \ otherwise \end{cases}$	$flux_{out} \leq \dfrac{S}{\Delta t}$	$flux_{out} = min\left(\dfrac{S}{\Delta t}, max\left(0, \theta_1 * \theta_2 * (T_{threshold} - T)\right)\right)$	37, 44
	Snowmelt from degree-day-factor	melt_1	$flux_{out} = \begin{cases} \theta_1 * (T - T_{threshold}), \ if \ T \geq T_{threshold} \\ 0, \ otherwise \end{cases}$	$flux_{out} \leq \dfrac{S}{\Delta t}$	$flux_{out} = min\left(\dfrac{S}{\Delta t}, max\left(0, \theta_1 * (T - T_{threshold})\right)\right)$	6, 12, 30, 31, 32, 34, 35, 37, 43, 44, 45
	Snowmelt at a constant rate	melt_2	$flux_{out} = \begin{cases} \theta_1, \ if \ S \geq 0 \\ 0, \ otherwise \end{cases}$	$flux_{out} \leq \dfrac{S}{\Delta t}$	$flux_{out} = min\left(\dfrac{S}{\Delta t}, \theta_1\right)$	44
	Glacier melt provided no snow is stored on the ice layer	melt_3	$flux_{out} = \begin{cases} \theta_1 * (T - T_{threshold}), \ if \ T \geq T_{threshold} , \ S_2 = 0 \\ 0, \ otherwise \end{cases}$	$flux_{out} \leq \dfrac{S_1}{\Delta t}$	$flux_{out} = min\left(\dfrac{S_1}{\Delta t}, max\left(0, \theta_1 * \theta_2 * (T_{threshold} - T)\right)\right)$ $* logisticSmoother\_S(S_2, 0)$	43
Soil moisture	Water rebalance to equal relative storage (2 stores)	soilmoisture_1	$flux_{out} = \begin{cases} \dfrac{S_2 S_{1,max} - S_1 S_{2,max}}{S_{1,max} + S_{2,max}}, \ if \ \dfrac{S_1}{S_{1,max}} < \dfrac{S_2}{S_{2,max}} \\ 0, \ otherwise \end{cases}$		$flux_{out} = \left(\dfrac{S_2 S_{1,max} - S_1 S_{2,max}}{S_{1,max} + S_{2,max}}\right) * logisticSmoother\_S\left(\dfrac{S_1}{S_{1,max}}, \dfrac{S_2}{S_{2,max}}\right)$	33
	Water rebalance to equal relative storage (3 stores)	soilmoisture_2	$flux_{out} = \begin{cases} S_2 \dfrac{S_1(S_{2,max} + S_{3,max}) + S_{1,max}(S_2 + S_3)}{(S_{2,max} + S_{3,max})(S_{1,max} + S_{2,max} + S_{3,max})}, \\ \quad if \ \dfrac{S_1}{S_{1,max}} < \dfrac{S_2 + S_3}{S_{2,max} + S_{3,max}} \\ 0, \quad otherwise \end{cases}$		$flux_{out} = \left(S_2 \dfrac{S_1(S_{2,max} + S_{3,max}) + S_{1,max}(S_2 + S_3)}{(S_{2,max} + S_{3,max})(S_{1,max} + S_{2,max} + S_{3,max})}\right)$ $* logisticSmoother\_S\left(\dfrac{S_1}{S_{1,max}}, \dfrac{S_2 + S_3}{S_{2,max} + S_{3,max}}\right)$	33

B.4   Unit Hydrographs

This section provides details on the implementation of various Unit Hydrographs. An overview of the 7 UHs is given in Table B.2. Computational implementation of each UH is given in sections B.4.1 to B.4.7. Unit Hydrograph files can be found in "./MARRMoT/Models/Unit Hydrograph files/".

Table B.2: Overview of Unit Hydrograph schemes implemented in MARRMoT

File name	Inputs	Diagram	Description	In model ...
uh_1_half	1: amount to be routed 2: time base 3: Δt		Exponentially increasing scheme	7
uh_2_full	1: amount to be routed 2: time base (time is doubled inside the function) 3: Δt		Exponential triangular scheme	7
uh_3_half	1: amount to be routed 2: time base 3: Δt		Triangular scheme: linearly increasing	13, 15, 21, 26 34
uh_4_full	1: amount to be routed 2: time base 3: Δt		Triangular scheme: linearly increasing and decreasing	0 (template), 16, 37, nn (example)
uh_5_half	1: amount to be routed 2: time base 3: Δt		Exponentially decreasing scheme	5
uh_6_gamma	1: amount to be routed 2: gamma parameter [-] 3: time for flow to reduce by factor e [d] 4: length of time series		Gamma function-based	40
uh_7_uniform	1: amount to be routed 2: time base 3: Δt		Uniform distribution	39

B.4.1   Code: uh_1_half

This section provides the computational implementation of a unit hydrograph with an increasing exponential distribution of flows.

File location   ./MARRMoT/Models/Unit Hydrograph files/uh_1_half

References   E.g. GR4J [255]

```matlab
1  function [ out,UH ] = uh_1_half( in, d_base, delta_t )
2  %uh_1_half Unit Hydrograph [days] with half a bell curve.
      GR4J-based
3  %
4  % Copyright (C) 2018 W. Knoben
5  % This program is free software (GNU GPL v3) and distributed
      WITHOUT ANY
6  % WARRANTY. See <https://www.gnu.org/licenses/> for details.
7  %
8  %    Inputs
9  %    in      - volume to be routed
10 %    d_base  - time base of routing delay [d]
11 %    delta_t - time step size [d]
12 %
13 %    Unit hydrograph spreads the input volume over a time
      period x4.
14 %    Percentage of input returned only increases.
15 %    I.e. d_base = 3.8 [days], delta_t = 1:
16 %    UH(1) = 0.04  [% of inflow]
17 %    UH(2) == 0.17
18 %    UH(3) = 0.35
19 %    UH(4) = 0.45
20
21 %%INPUTS
22 if any(size(in)) > 1; error('UH input should be a single
      value.'); end
23
24 %%TIME STEP SIZE
25 delay = d_base/delta_t;
26 if delay == 0; delay = 1; end        % any value below t = 1
      means no delay,
27                                        % but zero leads to
                                            problems
```

```matlab
28  tt = 1:ceil(delay);                    % Time series for which
       we need UH
29                                          % ordinates [days]
30
31  %%EMPTIES
32  SH = zeros(1,length(tt)+1); SH(1) = 0;
33  UH = zeros(1,length(tt));
34
35  %%UNIT HYDROGRAPH
36  for t = tt
37      if     t <  delay; SH(t+1) = (t./delay).^(5./2);
38      elseif t >= delay; SH(t+1) = 1;
39      end
40
41      UH(t) = SH(t+1)-SH(t);
42  end
43
44  %%DISPERSE VOLUME
45  out = in.*UH;
46
47  end
```

B.4.2 Code: uh_2_full

This section provides the computational implementation of a unit hydrograph with an exponential triangular distribution of flows.

File location ./MARRMoT/Models/Unit Hydrograph files/uh_2_full
References   E.g. GR4J [255]

```matlab
1  function [ out, UH ] = uh_2_full( in,d_base,delta_t )
2  %uh_2_full Unit Hydrograph [days] with a full bell curve.
       GR4J-based
3  %
4  % Copyright (C) 2018 W. Knoben
5  % This program is free software (GNU GPL v3) and distributed
       WITHOUT ANY
```

```matlab
 6 | % WARRANTY. See <https://www.gnu.org/licenses/> for details.
 7 | %
 8 | %   Inputs
 9 | %   in      - volume to be routed
10 | %   d_base  - time base of routing delay [d]
11 | %   delta_t - time step size [d]
12 | %
13 | %   Unit hydrograph spreads the input volume over a time
   |     period 2*x4.
14 | %   Percentage of input returned goes up (till x4), then down
   |     again.
15 | %   I.e. d_base = 3.8 [days], delta_t = 1:
16 | %   UH(1) = 0.02  [% of inflow]
17 | %   UH(2) = 0.08
18 | %   UH(3) = 0.18
19 | %   UH(4) = 0.29
20 | %   UH(5) = 0.24
21 | %   UH(6) = 0.14
22 | %   UH(7) = 0.05
23 | %   UH(8) = 0.00
24 |
25 | %%INPUTS
26 | if any(size(in)) > 1; error('UH input should be a single
   |     value.'); end
27 |
28 | %%TIME STEP SIZE
29 | delay = d_base/delta_t;
30 | tt = 1:2*ceil(delay); % time series for which we need UH
   |     ordinates [days]
31 |
32 | %%EMPTIES
33 | SH = zeros(1,length(tt)+1); SH(1) = 0;
34 | UH = zeros(1,length(tt));
35 |
36 | %%UNIT HYDROGRAPH
37 | for t = tt
```

```
38      if      (t <= delay)
39          SH(t+1) = 0.5*(t./delay).^(5./2);
40      elseif (t > delay) && (t < 2*delay);
41          SH(t+1) = 1-0.5*(2-t./delay).^(5./2);
42      elseif (t >= 2*delay);
43          SH(t+1) = 1;
44      end
45
46      UH(t) = SH(t+1)-SH(t);
47 end
48
49 %%DISPERSE VOLUME
50 out = in.*UH;
51
52 end
```

B.4.3 Code: uh_3_half

This section provides the computational implementation of a unit hydrograph with an linearly increasing distribution of flows.

> File location   ./MARRMoT/Models/Unit Hydrograph files/uh_3_half
> References   E.g. FLEX-Topo [274]

```
1 function [ out,UH ] = uh_3_half( in, d_base, delta_t )
2 %uh_3_half Unit Hydrograph [days] with half a triangle (
     linear)
3 %
4 % Copyright (C) 2018 W. Knoben
5 % This program is free software (GNU GPL v3) and distributed
     WITHOUT ANY
6 % WARRANTY. See <https://www.gnu.org/licenses/> for details.
7 %
8 %    Inputs
9 %    in      - volume to be routed
10 %   d_base  - time base of routing delay [d]
11 %   delta_t - time step size [d]
```

```matlab
12  %
13  %    Unit hydrograph spreads the input volume over a time
        period delay.
14  %    Percentage of input returned only increases.
15  %    I.e. d_base = 3.8 [days], delta_t = 1:
16  %    UH(1) = 0.04   [% of inflow]
17  %    UH(2) = 0.17
18  %    UH(3) = 0.35
19  %    UH(4) = 0.45
20
21  %%INPUTS
22  if any(size(in)) > 1; error('UH input should be a single
        value.'); end
23
24  %%TIME STEP SIZE
25  delay = d_base/delta_t;
26  if delay == 0; delay = 1; end   % any value below t = 1 means
         no delay,
27                                  % but zero leads to problems
28  tt = 1:ceil(delay);             % time series for which we
        need UH
29                                  % ordinates [days]
30
31  %%UNIT HYDROGRAPH
32  % The area under the unit hydrograph by definition sums to 1.
         Thus the area
33  % is S(t=0 to t = delay) t*[ff: fraction of flow to move per
        time step] dt
34  % Analytical solution is [1/2 * t^2 + c]*ff, with c = 0. Thus
         the fraction
35  % of flow step size is:
36  ff = 1/(0.5*delay^2);
37
38  %%EMPTIES
39  UH = zeros(1,length(tt));
40
```

```matlab
41  %%UNIT HYDROGRAPH
42  for t = 1:length(tt)
43      if t <= delay
44          UH(t) = ff.*(0.5*t^2 - 0.5*(t-1)^2);
45      else
46          UH(t) = ff.*(0.5*delay^2 - 0.5*(t-1)^2);
47      end
48  end
49
50  %%DISPERSE VOLUME
51  out = in.*UH;
52
53  end
```

B.4.4 Code: uh_4_full

This section provides the computational implementation of a unit hydrograph with an
linear triangular distribution of flows.

File location	./MARRMoT/Models/Unit Hydrograph files/uh_4_full
References	E.g. HBV-96
citeLindstrom1997	

```matlab
1  function [ out,UH ] = uh_4_full( in, d_base, delta_t )
2  %uh_4_half Unit Hydrograph [days] with a triangle (linear)
3  %
4  % Copyright (C) 2018 W. Knoben
5  % This program is free software (GNU GPL v3) and distributed
       WITHOUT ANY
6  % WARRANTY. See <https://www.gnu.org/licenses/> for details.
7  %
8  %   Inputs
9  %   in     - volume to be routed
10 %   d_base  - time base of routing delay [d]
11 %   delta_t - time step size [d]
12 %
13 %   Unit hydrograph spreads the input volume over a time
       period delay.
```

```matlab
14  %    Percentage runoff goes up, peaks, and goes down again.
15  %    I.e. d_base = 3.8 [days], delta_t = 1:
16  %    UH(1) = 0.14   [% of inflow]
17  %    UH(2) = 0.41
18  %    UH(3) = 0.36
19  %    UH(4) = 0.09
20
21  %%INPUTS
22  if any(size(in)) > 1; error('UH input should be a single
        value.'); end
23
24  %%TIME STEP SIZE
25  delay = d_base/delta_t;
26  if delay == 0; delay = 1; end        % any value below t = 1
        means no delay,
27                                       % but zero leads to
                                            problems
28  tt = 1:ceil(delay);                  % time series for which
        we need UH
29                                       % ordinates [days]
30
31  %%UNIT HYDROGRAPH
32  % The area under the unit hydrograph by definition sums to 1.
        Thus the area
33  % is S(t=0 to t = delay) t*[ff: fraction of flow to move per
        time step] dt
34  % Analytical solution is [1/2 * t^2 + c]*ff, with c = 0.
35  % Here, we use two half triangles t make one big one, so the
        area of half a
36  % triangle is 0.5. Thus the fraction of flow step size is:
37  ff  = 0.5/(0.5*(0.5*delay)^2);
38  d50 = 0.5*delay;
39
40  %%TRIANGLE FUNCTION
41  tri = @(t) max(ff.*(t-d50).*sign(d50-t)+ff.*d50,0);
42
```

```
43  %%EMPTIES
44  UH = zeros (1,length(tt));
45
46  %%UNIT HYDROGRAPH
47  for t = 1:length(tt)
48      UH(t) = integral(tri,t-1,t);
49  end
50
51  %%ENSURE UH SUMS TO 1
52  tmp_diff    = 1-sum(UH);
53  tmp_weight = UH./sum(UH);
54  UH          = UH + tmp_weight.*tmp_diff;
55
56  %%DISPERSE VOLUME
57  out = in.*UH;
58
59  end
```

B.4.5 Code: uh_5_half

This section provides the computational implementation of a unit hydrograph with an decreasing exponential distribution of flows.

File location	./MARRMoT/Models/Unit Hydrograph files/uh_5_half
References	E.g. IHACRES [84, 199]

```
1  function [ out,UH ] = uh_1_half( in, d_base, delta_t )
2  %uh_1_half Unit Hydrograph [days] with half a bell curve.
       GR4J-based
3  %
4  % Copyright (C) 2018 W. Knoben
5  % This program is free software (GNU GPL v3) and distributed
       WITHOUT ANY
6  % WARRANTY. See <https://www.gnu.org/licenses/> for details.
7  %
8  %   Inputs
9  %   in      - volume to be routed
```

```matlab
10  %    d_base  - time base of routing delay [d]
11  %    delta_t - time step size [d]
12  %
13  %    Unit hydrograph spreads the input volume over a time
       period x4.
14  %    Percentage of input returned only increases.
15  %    I.e. d_base = 3.8 [days], delta_t = 1:
16  %    UH(1) = 0.04   [% of inflow]
17  %    UH(2) = 0.17
18  %    UH(3) = 0.35
19  %    UH(4) = 0.45
20
21  %%INPUTS
22  if any(size(in)) > 1; error('UH input should be a single
       value.'); end
23
24  %%TIME STEP SIZE
25  delay = d_base/delta_t;
26  if delay == 0; delay = 1; end        % any value below t = 1
       means no delay,
27                                       % but zero leads to
                                            problems
28  tt = 1:ceil(delay);                  % Time series for which
       we need UH
29                                       % ordinates [days]
30
31  %%EMPTIES
32  SH = zeros(1,length(tt)+1); SH(1) = 0;
33  UH = zeros(1,length(tt));
34
35  %%UNIT HYDROGRAPH
36  for t = tt
37      if      t <  delay; SH(t+1) = (t./delay).^(5./2);
38      elseif t >= delay; SH(t+1) = 1;
39      end
40
```

```
41        UH(t) = SH(t+1)-SH(t);
42   end
43
44   %%DISPERSE VOLUME
45   out = in.*UH;
46
47   end
```

B.4.6   Code: uh_6_gamma

This section provides the computational implementation of a unit hydrograph with a gamma distribution of flows.

File location	./MARRMoT/Models/Unit Hydrograph files/uh_6_gamma
References	E.g. SMAR [236, 312]

```
1   function [ out,UH,frac_routing_beyond_time_series ] = ...
2                                   uh_6_gamma( in,n,k,
                                      t_end,delta_t )
3   %uh_6_gamma Unit Hydrograph [days] from gamma function.
4   %
5   % Copyright (C) 2018 W. Knoben
6   % This program is free software (GNU GPL v3) and distributed
         WITHOUT ANY
7   % WARRANTY. See <https://www.gnu.org/licenses/> for details.
8   %
9   %   Inputs
10  %   n       = shape parameter [-]
11  %   k       = time delay for flow reduction by a factor e [d]
12  %   t_end   = length of time series [d]
13  %   delta_t = time step size [d]
14  %
15  %   Unit hydrograph spreads the input volume over a time
         period delay.
16  %   Percentage of input returned only decreases.
17  %   I.e. n = 1, k = 3.8 [days], delta_t = 1:
18  %   UH(1) = 0.928  [% of inflow]
```

```matlab
19  %      UH(2) = 0.067
20  %      UH(3) = 0.005
21  %      UH(4) = 0.000
22
23  %%INPUTS
24  if any(size(in)) > 1; error('UH input should be a single
        value.'); end
25
26  %%TIME STEP SIZE
27  tmax = t_end/delta_t;
28  tt   = 1:tmax;            % time series for which we need UH
        ordinates [days]
29
30  %%EMPTIES
31  UH_full  = zeros(1,length(tt));
32  frac_routing_beyond_time_series = 0;
33
34  %%UNIT HYDROGRAPH
35  % The Unit Hydrograph follows a gamma distribution. For a
        given
36  % delay time, the fraction of flow per time step is thus the
        integral of
37  % t-1 to t of the gamma distrubtion. The curve has range [0,
        Inf>.
38  % We need to choose a point at which to cap the integration,
        but this
39  % depends on the parameters n & k, and the total time step.
        We choose the
40  % cutoff point at the time step where less than 0.1% of the
        peak flow
41  % is still on route.
42
43  %%Unit hydrograph
44  for t = 1:length(tt)
45      UH_full(t) = integral(@(x) 1./(k.*gamma(n)).*(x./k).^(n
            -1).* ...
```

```matlab
46                                             exp(-1.*x./k),(t-1)*
                                                delta_t,t*delta_t);
47  end
48
49  %%Find cutoff point where less than 0.1% of the peak flow is
        being routed
50  [max_val,max_here] = max(UH_full);
51  end_here = find(UH_full(max_here:end)./max_val<0.001,1) +
        max_here;
52
53  %%Take action depending on whether the distribution function
        exceeds the
54  %%time limit or not
55  if ~isempty(end_here)
56      %%Construct the Unit Hydrograph
57      UH = UH_full(1:end_here);
58
59      %%Account for the truncated part of the full UH.
60      % find probability mass to the right of the cut-off point
61      tmp_excess = 1-sum(UH);
62
63      % find relative size of each time step
64      tmp_weight = UH_full(1:end_here)./sum(UH_full(1:end_here)
            );
65
66      % distribute truncated probability mass proportionally to
             all elements
67      % of the routing vector
68      UH = UH+tmp_weight.*tmp_excess;
69
70  else
71      %%Construct the Unit Hydrograph
72      UH = UH_full;
73
74      %%The UH is longer than the provided time series length.
            Track the
```

```
75        %%percentage of flow that is routed beyond the simulation
              duration
76        frac_routing_beyond_time_series = 1-sum(UH);
77
78  end
79
80  %%DISPERSE VOLUME
81  out = in.*UH;
82
83  end
```

B.4.7 Code: uh_7_uniform

This section provides the computational implementation of a unit hydrograph with a uniform distribution of flows.

File location	./MARRMoT/Models/Unit Hydrograph files/uh_7_uniform
References	E.g. MCRM [28, 222]

```
1   function [ out,UH ] = uh_7_uniform( in, d_base, delta_t )
2   %uh_7_uniform Unit Hydrograph [days] with uniform spread
3   %
4   % Copyright (C) 2018 W. Knoben
5   % This program is free software (GNU GPL v3) and distributed
        WITHOUT ANY
6   % WARRANTY. See <https://www.gnu.org/licenses/> for details.
7   %
8   %   Inputs
9   %   in      - volume to be routed
10  %   d_base  - time base of routing delay [d]
11  %   delta_t - time step size [d]
12  %
13  %   Unit hydrograph spreads the input volume over a time
        period delay.
14  %   I.e. d_base = 3.8 [days], delta_t = 1:
15  %   UH(1) = 0.26   [% of inflow]
16  %   UH(2) = 0.26
```

```matlab
17 %    UH(3) = 0.26
18 %    UH(4) = 0.22
19
20 %%INPUTS
21 if any(size(in)) > 1; error('UH input should be a single
       value.'); end
22
23 %%TIME STEP SIZE
24 delay = d_base/delta_t;
25 tt = 1:ceil(delay); % time series for which we need UH
       ordinates [days]
26
27 %%EMPTIES
28 UH = NaN.*zeros(1,length(tt));
29
30 %%FRACTION FLOW
31 ff = 1/delay; % fraction of flow per time step
32
33 %%UNIT HYDROGRAPH
34 for t=1:ceil(delay)
35     if t < delay
36         UH(t) = ff;
37     else
38         UH(t) = mod(delay,t-1)*ff;
39     end
40 end
41
42 %%DISPERSE VOLUME
43 out = in.*UH;
44
45 end
```

B.5   Parameter ranges

Each model function in MARRMoT is accompanied by a file that specifies suitable sampling ranges for each parameter used in the model, that could be applied if the user chooses to pair MARRMoT with a calibration or parameter sampling procedure. This section gives the reasoning behind our choices of parameter ranges used within MARRMoT.

B.5.1   Model-specific ranges versus generalised process-specific ranges

There are two different approaches to determining parameter ranges for model calibration or parameter sampling studies: (1) make a choice for appropriate parameter ranges per model, based on previous applications of the model, or (2) try to make consistent choices for all models based on literature (e.g. ensure that all 'slow' linear reservoirs, regardless of which model they are part of, have the same limits for the drainage time scale parameter). Generalization of parameter ranges across models is difficult because models use different flux formulations and thus different parameter values might be appropriate, even if the fluxes are intended to represent the same hydrologic process. On the other hand, using model-specific parameter ranges based on earlier studies might limit a model's potential. Especially if the model has only been applied to a small number of places, published 'appropriate' parameter ranges might also reflect the climate or catchment characteristics of the few study catchments the model has been applied to. MARRMoT is intended as a model comparison framework. We thus attempt to generalize parameter ranges across all models in the framework, to facilitate fair comparison of different models. We try to err on the side of caution and intentionally set these ranges wide. Table B.3 shows the parameter ranges used in MARRMoT and specifies in which model(s) each parameter range is used.

Table B.3: Parameter ranges used in MARRMoT

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Snow							
Threshold temperature for snowfall (and melt, if not specified otherwise) [^oC]	Table B.4	Table B.4	-3	5	[171, 181]		6, 12, 30, 31, 32, 34, 35, 37, 43, 44, 45
Threshold interval width for snowfall [^oC]	0	7	0	17	[171]	0 is a physical limit	37
Threshold temperature for melt [^oC]			-3	3		Not easy to find any interval. Temperature for melt tends be treated as constant at 0	37, 43, 44
Degree-day-factor for snow or ice melt [$mm/^oC/d$]	0	Table B.5	0	20		0 is a physical limit	6, 12, 30, 31, 32, 34, 35, 37, 41, 43, 44, 45
Water holding content of snow pack [-]	0	0.8	0	1	[181]	[0,1] are physical limits	37, 44
Refreezing factor of retained liquid water [-]	0	1	0	1		[0,1] are physical limits	37, 44 (given as fraction [0,1] of degree-day-factor)

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Maximum melt rate due to ground-heat flux [mm/d]	0	2	0	2	[282]		44
Interception							
Maximum store depth [mm]	0	Table B.6	0	5	[69, 118]	0 is a physical limit. [118] (table 1.1) reports 3.8mm as maximum value used out of 15 studies. [69] (table 3) report 5.6mm as a maximum value for 28 catchments	2, 13, 15, 16, 18, 22, 23, 26, 34, 36, 39, 42, 44, 45
Maximum intercepted fraction of precipitation [-]	0	0.42	0	1	[118]	[0,1] are physical limits. [118] (table 1.1) reports 42% as maximum intercepted fraction out of 15 studies	8, 23, 32, 35, 45
Seasonal variation in LAI as fraction of mean [-]			0	1		0 is a physical limit	22
Timing of maximum Leaf Area Index [d]			1	365		Refers to days in a normal calendar year	22, 32, 35

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Surface depression							
Maximum surface area contributing to store [-]	0	1	0	1		[0,1] are physical limits	36, 45
Maximum store depth [mm]	0	Table B.7	0	50	[69]	0 is physical limit. 50 is recommended in [69]	36, 45
Filling parameter [-]	1	1	0.99	1	[71, 258]	Controls the shape of the depression store inflow flux but is usually set at 1 because no studies are (were?) available about how a depression store fills	36
Infiltration							
Maximum loss [mm]	0	400	0	600	[70]	Fig 11.11a shows calibrated parameter values for 339 catchments. Pattern indicates that limit was set at 400	18, 36

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Loss exponent [-]	0	12	0	15	[70]	Fig 11.11a shows calibrated parameter values for 339 catchments. Pattern indicates that limit was set at 10	18, 36
Maximum infiltration rate [mm/d]	Table B.8	Table B.8	0	200		Infiltration rates can be very high. However, to have a practical effect on modelling, (i.e. generate infiltration excess flow), Inf_rate < P(t). In the context of a follow-up study, Inf_rate is capped at 200mm/d because the maximum daily P in the study area is 200mm/d.	15, 20, 23, 40, 44
Infiltration decline nonlinearity parameter [-]			0	5	[302]	Very difficult to find information for (original paper mentions nothing)	23, 43

310

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Evaporation							
Plant-controlled maximum rate [mm/d]	5	24.5	0	20	[69]	Although the study reports an upper value of 24.5, the recommended range is capped at 20 (paper appendix)	20, 36
Wilting point as fraction of Soil moisture capacity [-]	0.1	0.25	0.05	0.95	[309]	0 is a physical limit but can break model equations through "divide-by-zero" errors. 1 is a physical limit	3, 4, 8, 9, 10, 12, 14, 15, 16, 19, 20, 21, 26, 31, 32, 34, 35, 37, 44
Moisture constrained rate parameter [-]			0	1		[0,1] are physical limits	15
Forest fraction for separate soil/vegetation evap [-]	0	1	0.05	0.95		[0,1] are physical limits, but using these limits can result in divide-by-zero-errors in certain fluxes	3, 4, 8, 16
Phenology: minimum temperature where transpiration stops [^oC]	-5	-5	0	-10	[349]		35

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Phenology: maximum temperature above which transpiration fully utilizes Ep [^oC]	10	10	1	20	[349]	The setup of minimum and maximum temperature used in Ye et al. (2012) is here changed to a minimum temperature + temperature range (Tmax = Tmin + Trange) to avoid overlap in parameter values	35
Evaporation reduction with depth coefficient [-]	0.083	1	0	1	[251, 312]	[0,1] are physical limits	17, 23, 25, 40
Shape parameter for evaporation reduction in a deficit store [-]			0	1	[222]	This uses a sigmoid function to determine a fraction of Ep to evaporate. Values >1 make the transition very steep	39
Evaporation non-linearity coefficient [-]			0	10	[302]	Very difficult to find information for. Assumption made to be in line with other non-linearity coefficients.	23, 43

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Soil moisture							
Maximum store depth [mm]	1	Table B.9	1	2000		0 is a physical limit	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46
Capillary rise [mm/d]	0	Table B.10	0	4		0 is a physical limit	13, 15, 37, 38
						SMHI gives a default value of 1 mm/d for use with HBV. We use a wider range here	
Percolation rate [mm/d]	0	Table B.11	0	20	[31]	Some modelling studies report very large percolation rates (100 mm/d). Field observations give approx. 11mm/d [31].	21, 26, 34, 37, 39, 44, 45
Percolation fraction [-]	0.013	0.533	0	1	[349] (Table 1)	[0,1] are physical limits	14, 22, 23, 24, 27, 30, 31, 32, 35, 45

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Recharge nonlinearity [-]	0	7	0	10	[181]	Also seen as a soil depth distribution	5, 22, 33, 37
Soil depth distribution [-]	0	Table B.12	0	10		For cases where the soil depth is not considered constant. Most studies limit this to 0-2.5 but this seems based on a single source [335] which is UK only. Thus we use a wider range here	2, 13, 15, 21, 22, 26, 28, 29, 34
Porosity [-]	0.35	0.5	0.05	0.95	[309]	[0,1] are theoretical physical limits, but no (0) porosity and full (1) porosity are not sensible: there would be no soil moisture or soil respectively	10, 19
Gamma distribution for topographic indices - phi [-]	0.4	3.5	0.1	5	[78]		14
Gamma distribution for topographic indices - chi [-]	2	5	1	7.5	[78]		14

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Fraction area with permeable soils [-]			0	1	[85]	[0,1] are physical limits	46
Fraction area with semipermeable soils [-]			0	1	[85]	[0,1] are physical limits	46
Fraction area with impermeable soils [-]			0	1	[85]	[0,1] are physical limits	46
Variable contributing area scaling [-]			0	5	[302]	Very difficult to find information about this. Assumption made	23
Variable contributing area non-linearity [-]					[302]	See: **Soil depth distribution** above	23
Fraction of D50 that is D16 [-]			0.01	0.99		Note: re-writing of D16 parameter in [115]	42
Variable contributing area equation inflection point [-]	-0.5	0.5	-0.5	0.5	[153]		28
Groundwater							
Leakage coefficient [-]	0.07	0.13	0	0.5	[69]	0 is physical limit. 0.5 is recommended in the paper's appendix	36

316

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Leakage rate [mm/d]						See: **Percolation rate** above	
Level compared to channel level [mm]	-2.8	3.9	-10	10	[69]	Range recommended in appendix of the paper	36
Base flow rate at no deficit [mm/d]	0	201.6	0.1	200	[33]	Based on Table 2 [33]	14, 23
Baseflow deficit scaling parameter [-]			0	1		[0,1] are physical limits	14, 23
Flow distribution							
Interflow and saturation excess [-]	0	1	0	1		[0,1] are physical limits	18, 36
Preferential recharge [-]	0	2	0	1	[69]	0 is a physical limit. Later paper sets max limit to 1	18, 25, 36, 46
Surface/groundwater division [-]			0	1		[0,1] are physical limits	13, 17, 33
Fast and slow flow [-]	0	1	0	1		[0,1] are physical limits	21, 26, 29, 34, 46
Groundwater recharge and interflow [-]	0.05	0.3	0	1	[309]	[0,1] are physical limits	10, 11, 20, 40
Infiltration and direct runoff [-]	0.161	0.422	0	1	[312]	[0,1] are physical limits	40

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Impervious and infiltration area [-]			0	1		[0,1] are physical limits	28, 33, 45
Contributing area to overland flow [-]			0	1		[0,1] are physical limits	39, 45
Tension water and free water [-]			0	1		[0,1] are physical limits	33
Threshold for overland flow generation [-]	0	<1	0	0.99	[232]	[0,1] are physical limits	41
Threshold for overland flow generation [-]	0	<1	0	0.99	[232]	[0,1] are physical limits	41
Channel and land division [-]			0	1		[0,1] are physical limits	42
Throughfall/stem flow division [-]			0	1		[0,1] are physical limits	42
Glacier/non-glacier precipitation [-]			0	1		[0,1] are physical limits	43

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Flow time scale and shape							
Fast reservoir time scale [d-1]	0.05	Table B.13	0	1		0 is a physical limit	12, 21, 24, 26, 28, 29, 30, 31, 32, 33, 34, 35, 37, 39, 41, 42, 43, 44, 46
Slow reservoir time scale [d-1]	0.01	Table B.14	0	1		0 is a physical limit	2, 3, 4, 6, 8, 10, 13, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 28, 29, 30, 31, 32, 33, 34, 35, 37, 39, 40, 41, 42, 43, 44, 46
Flow non-linearity S^x [-]	0	Table B.15	1	5			4, 9, 10, 11, 16, 19, 22, 23, 37, 39, 42, 44, 45
Flow reduction (S/X) [mm]	5	40	1	50	[309]		9
Exponential shape parameter [mm-1]			0	2	[222]	Very difficult to find documentation for	39
Routing							
Routing delay to fast flow [d]	0	1	1	5	[106]		5, 21, 26, 34

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Routing delay to slow flow [d]	0	8	1	15			5, 7, 21, 26, 34
Routing delay [d]	1	Table B.16	1	120	[181]	1 is the limit (water shouldn't speed up). 120 because it seems very high	13, 15, 16, 21, 37, 39, 40
Routing store depth [mm]	1	300	1	300	[255]		7, 20, 39, 45
Gamma function, number of Nash cascade reservoirs [-]	0.75	9.79	1	10	[312]	0 would mean no routing, so slightly above that	40
Water exchange parameters							
Coefficient 1 [-]	0.005	0.54	0	1	[69]	Although the study only reports values up to 0.54, an upper range of 1 is recommended in the study's appendix	36

319

...continued

Description	Min(lit)	Max(lit)	Min(used)	Max(used)	Reference(s)	Notes	Model
Coefficient 2 [-]	0.01	0.29	0	1	[69]	Although the study only reports values up to 0.29, an upper range of 1 is recommended in the study's appendix	36
Coefficient 3 [-]	0	13	0	100	[69]	Although the study only reports values up to 13, an upper range of 100 is recommended in the study's appendix	36
Water exchange coefficient [mm/d]	-10	14	-10	15	[255, 271]	Parameter x_2 in GR4J model	7

Table B.4: Literature-based ranges for snow melt parameter "threshold temperature for snowfall"

Threshold temperature for snowfall [^{o}C]	Min	Max
Table 2 in [287]	-2.5	2.5
Table 1 in [181]	-3	3
Table 2 in [171] Note: always coupled with a snow interval [10,17]	1.1	4.5
Table A3 in [291]	-1.5	2.5

Table B.5: Literature-based ranges for snow melt parameter "degree-day-factor"

Degree-day factor for snowmelt [$mm/^{o}C/d$]	Min	Max
Table 2 in [287]	1	10
Table 1 in [181]	0	20
Table A3 in [291]	1	10

Table B.6: Literature-based ranges for interception parameter "maximum interception capacity"

Interception bucket [mm]	Min	Max
Figure 11.11a in [70]	0	5
Table 3 in [69]	0.5	5.6
Table 1.1 in [118]	0	3.8
Table 2 in [309]		0.4

Table B.7: Literature-based ranges for depression parameter "maximum depression capacity"

Depression bucket [mm]	Min	Max
Table 3 in [69]	1	100
Table 1 in [9]	5	110

Table B.8: Literature-based ranges for infiltration parameter "maximum infiltration rate"

Infiltration rate	Min	Max
Figure 2 in [17] [mm/d]	40	100
Table 3.3 in [157] [mm/h]	6	76
Table 3 in [66] [mm/h]	50	770

Table B.9: Literature-based ranges for soil moisture parameter "maximum soil moisture capacity"

Soil moisture bucket [mm]	Min	Max
Figure 11.11b in [70]	0	500
Table 3 in [69]	65	400
Table 2 in [287]	50	500
Table 1 in [269]	100	800
Table 1 in [181]	0	2000
Table A3 in [291]	50	500
Table 3 in [311]	1	500

Table B.10: Literature-based ranges for capillary rise parameter "maximum capillary rise rate"

Capillary rise [mm/d]	Min	Max
Table 1 in [269]	0.1	1
Default value in [306]	1	1
Figure 3 in [31]	0	0.06

Table B.11: Literature-based ranges for percolation parameter "maximum percolation rate"

Percolation rate [mm/d]	Min	Max
Table 2 in [287]	0	6
Table 1 in [269]	0.1	5
Table 1 in [181]	0	100
Figure 3 in [31]	0	10.4
Table A3 in [291]	0	3

Table B.12: Literature-based ranges for soil moisture parameter "soil depth distribution non-linearity"

Soil depth distribution [-]	Min	Max
Table 3 in [311]	0	2
Figure 9 in [189]	0	2.5
Table 4 in [59]	0	2.5
Figure 4.12 in [335]	0	2
Page 700 in [304]		4.03
Figure 4 in [147] Note: estimated values, $97\% < 6$	0	11.5

Table B.13: Literature-based ranges for flow parameter "fast flow time scale"

Fast flow time scale [d^{-1}]	Min	Max
Table 2 in [287]	0.05	0.5
Table 1 in [269]	0.05	0.8
Table 1 in [181]	0.01	1
Table A3 in [291]	0.01	0.4
Table 3 in [311]	0.5	1.2

Table B.14: Literature-based ranges for flow parameter "slow flow time scale"

Slow flow time scale [d^{-1}]	Min	Max
Figure 11.11b in [70]	0	0.3
Table 2 in [309]	2.40E-05	0.1
Table 2 in [287]	0.001	0.1
Table 1 in [269]	0.0005	0.1
Table 1 in [181]	0.00005	0.05
Table A3 in [291]	0.001	0.15
Table 3 in [311]	0.001	0.5

Table B.15: Literature-based ranges for flow parameter "flow non-linearity"

Flow non-linearity	Min	Max
Table 3 in [195] – non-linearity shape $= S^{1+var}$	0	3
Table 1 in [309] – non-linearity shape $= S^{1/var}$	0.45	0.5
Table 3 in [158]	0.5	0.5

Table B.16: Literature-based ranges for routing parameter "routing delay"

Routing delay [d]	Min	Max
Table 2 in [287]	1	5
Table 1 in [181]	24	120
Table 3 in [195]	1	4
Table 1 in [255]	0.5	4
Table A3 in [291]	1	7
Table 2 in [18] Note: converted from a flow speed of 0.5m/s and catchment area of $47km^2$		<1
Table 3 in [121]	12	36
Table 2 in [328] Note: approximated from flow velocities and catchment sizes	0.01	4

SUPPORTING INFORMATION FOR CHAPTER 5

This appendix is based on a draft manuscript that will be submitted as Supporting Information for a research article in Water Resources Research.

Citation: Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A. & Woods, R. A. (in prep). Exploring conceptual model structure uncertainty through a large-sample approach: Comparative analysis of 36 models across 559 catchments. To be submitted to Water Resources Reseach.
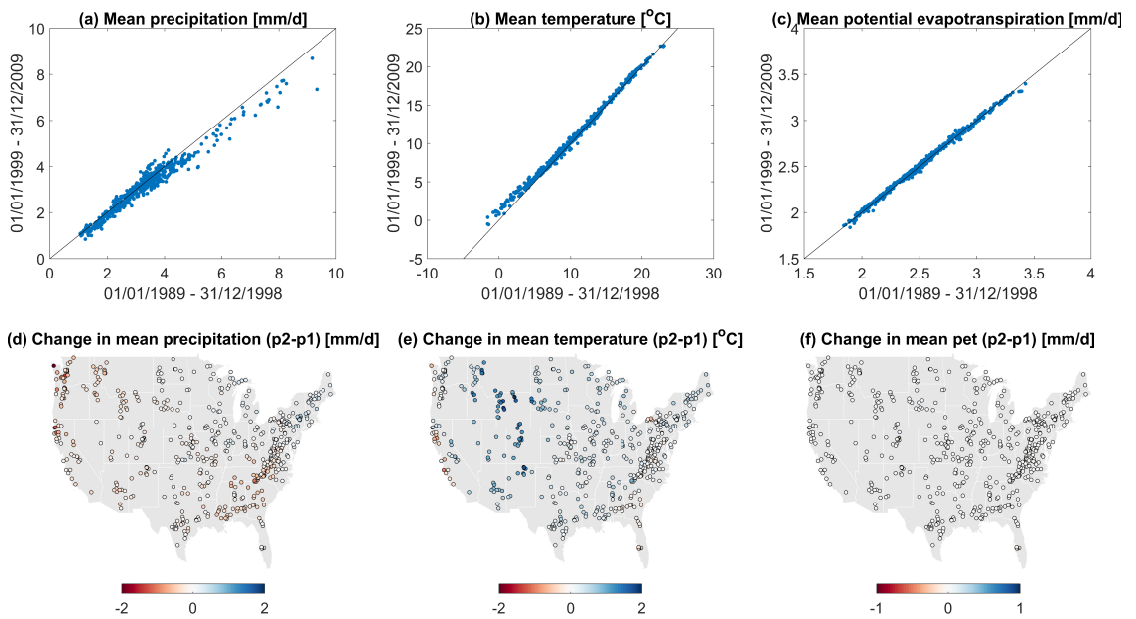
Figure C.1: Changes in mean precipitation, temperature and potential evapotranspiration between the two 10-year periods used in this study. Period 1 (p1) covers 01-Jan 1989 to 31-Dec 1999. Period 2 (p2) covers 01-Jan 2000 to 31-Dec 2009.
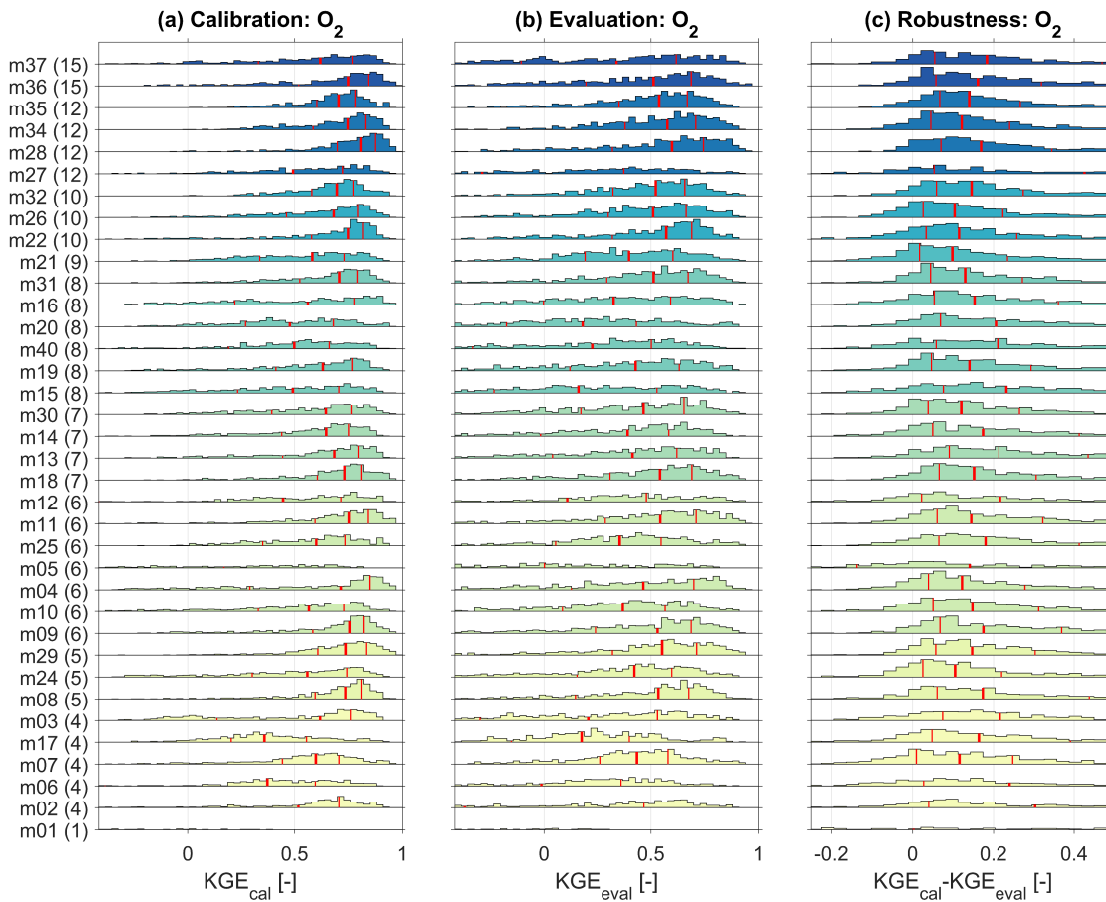
Figure C.2: Summary of model performance during low flow simulation (KGE(1/Q)) in all 559 catchments. Models are sorted by number of parameters, with the model with the most parameters at the top (shown in brackets after each model ID). Coloring and model IDs match those used in Figure 2 in the main manuscript. Each histogram is based on 559 values. Red lines show the 25th, 50th (thick line) and 75th percentiles. Bin widths have been set at 0.03 KGE [-] through trial-and-error, settling on this value because it shows the distributions in a sufficient, but not overly high, level of detail. (a) Calibration efficiency. (b) Evaluation efficiency. (c) Difference in efficiency between calibration and evaluation period.
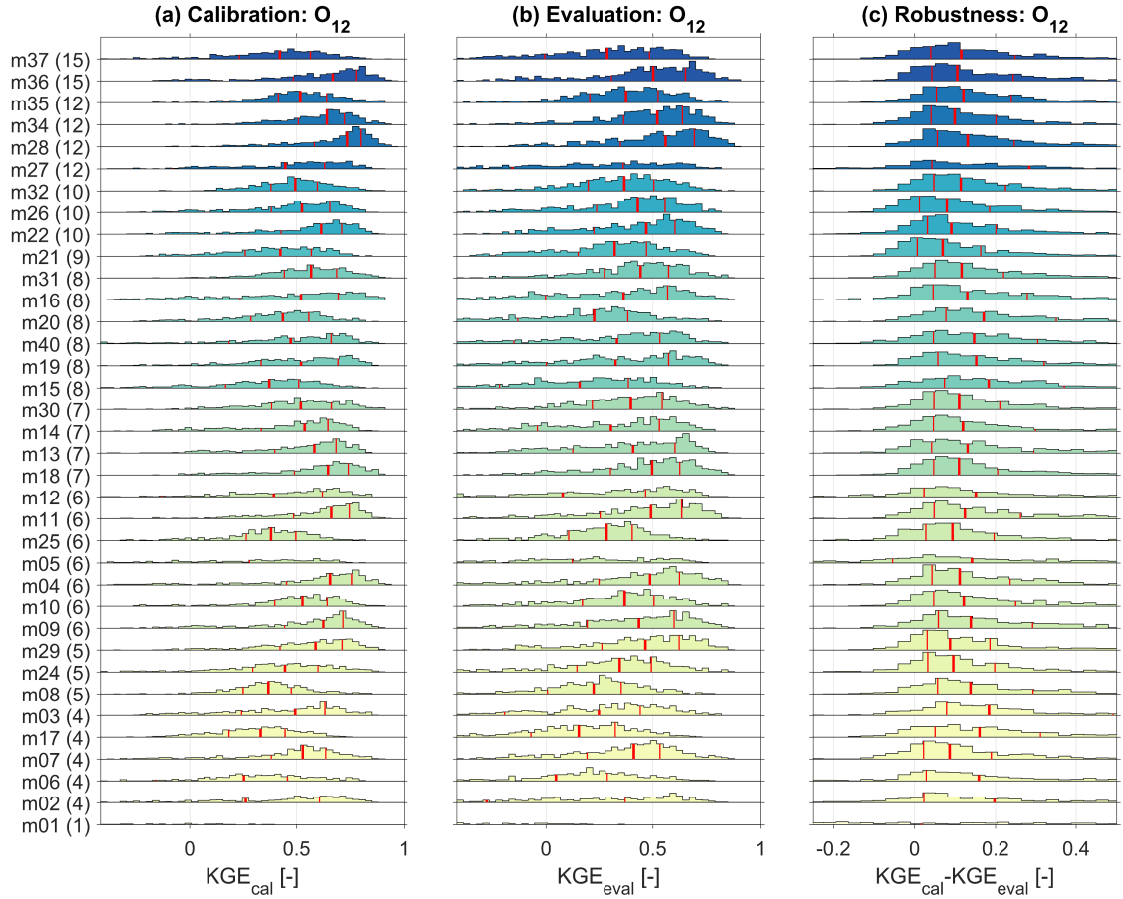
Figure C.3: Summary of model performance during combined flow simulation ($\frac{1}{2}[KGE(Q) + KGE(1/Q)]$) in all 559 catchments. Models are sorted by number of parameters, with the model with the most parameters at the top (shown in brackets after each model ID). Coloring and model IDs match those used in Figure 2 in the main manuscript. Each histogram is based on 559 values. Red lines show the 25th, 50th (thick line) and 75th percentiles. Bin widths have been set at 0.03 KGE [-] through trial-and-error, settling on this value because it shows the distributions in a sufficient, but not overly high, level of detail. (a) Calibration efficiency. (b) Evaluation efficiency. (c) Difference in efficiency between calibration and evaluation period.
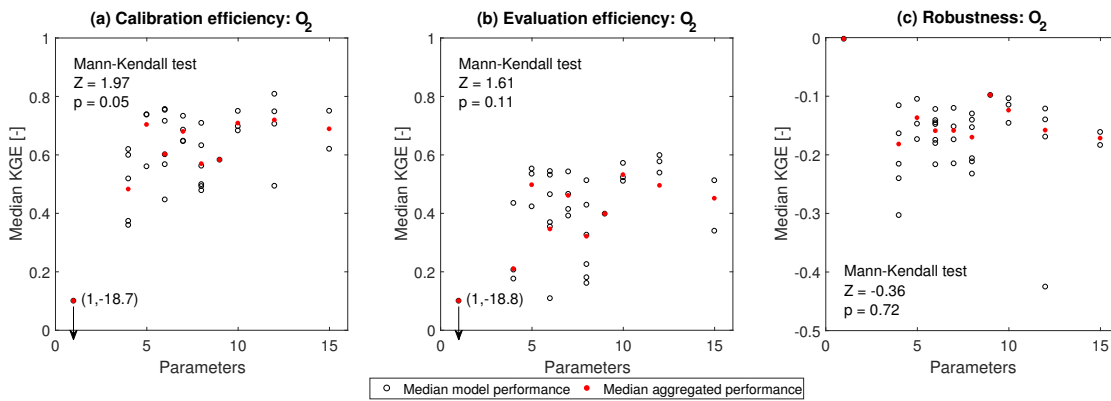
Figure C.4: Overview of median model performance across 559 catchments (open circles) and median aggregated model performance per number of free parameters (red dots) for the low flow ($KGE(1/Q)$) objective function. Statistics are shown for (a) the calibration period, (b) the evaluation period, and (c) the robustness of model performance given as $KGE_{val} - KGE_{cal}$. Mann-Kendall statistical tests using aggregated model performance show the relative strength and direction of trends (Z-value, negative values indicate downward trend whereas positive values indicate an upward trend) and the associated probability level (p).
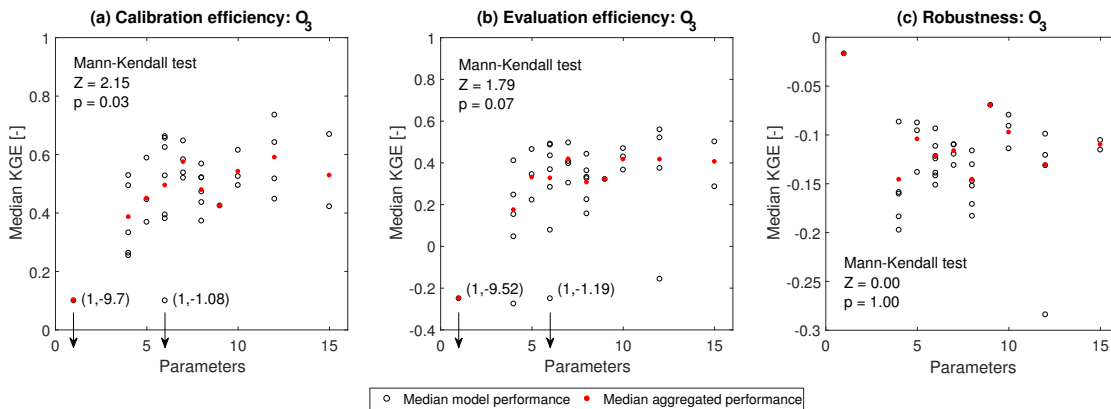

Figure C.5: Overview of median model performance across 559 catchments (open circles) and median aggregated model performance per number of free parameters (red dots) for the combined flow ($\frac{1}{2}[KGE(Q) + KGE(1/Q)]$) objective function. Statistics are shown for (a) the calibration period, (b) the evaluation period, and (c) the robustness of model performance given as $KGE_{val} - KGE_{cal}$. Mann-Kendall statistical tests using aggregated model performance show the relative strength and direction of trends (Z-value, negative values indicate downward trend whereas positive values indicate an upward trend) and the associated probability level (p).
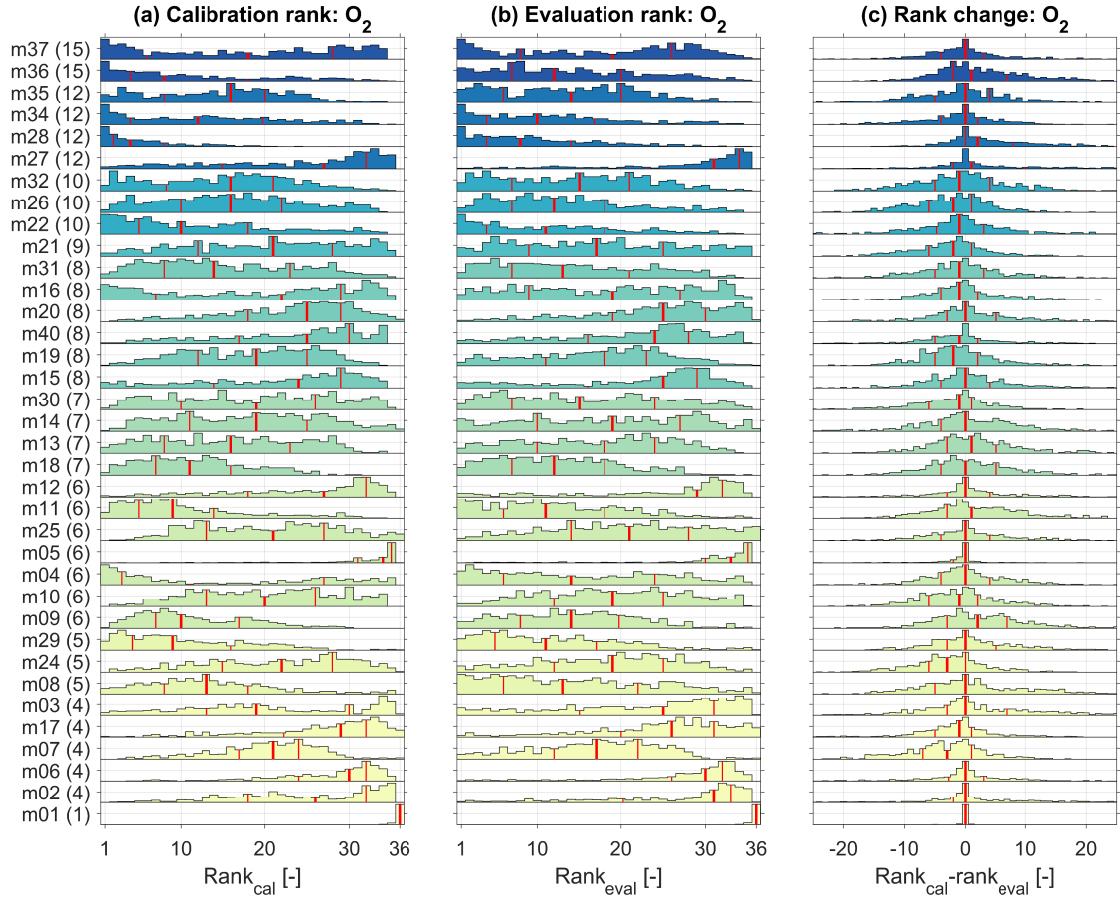
Figure C.6: Summary of model ranking during low flow simulation ($KGE(1/Q)$) in all 559 catchments. Rank 1 is assigned to the model with the highest efficiency value in a given catchment. Models are sorted by number of parameters, with the model with the most parameters at the top (shown in brackets after each model ID). Colouring and model IDs match those in Figure 2 in the main manuscript. Each histogram is based on 559 values. Note that each histogram is scaled relative to its own maximum, meaning that absolute heights should not be compared between different rows. Red lines show the 25th, 50th (thick line) and 75th percentiles. (a) Calibration ranking. (b) Evaluation ranking. (c) Difference in ranking between calibration and evaluation period.
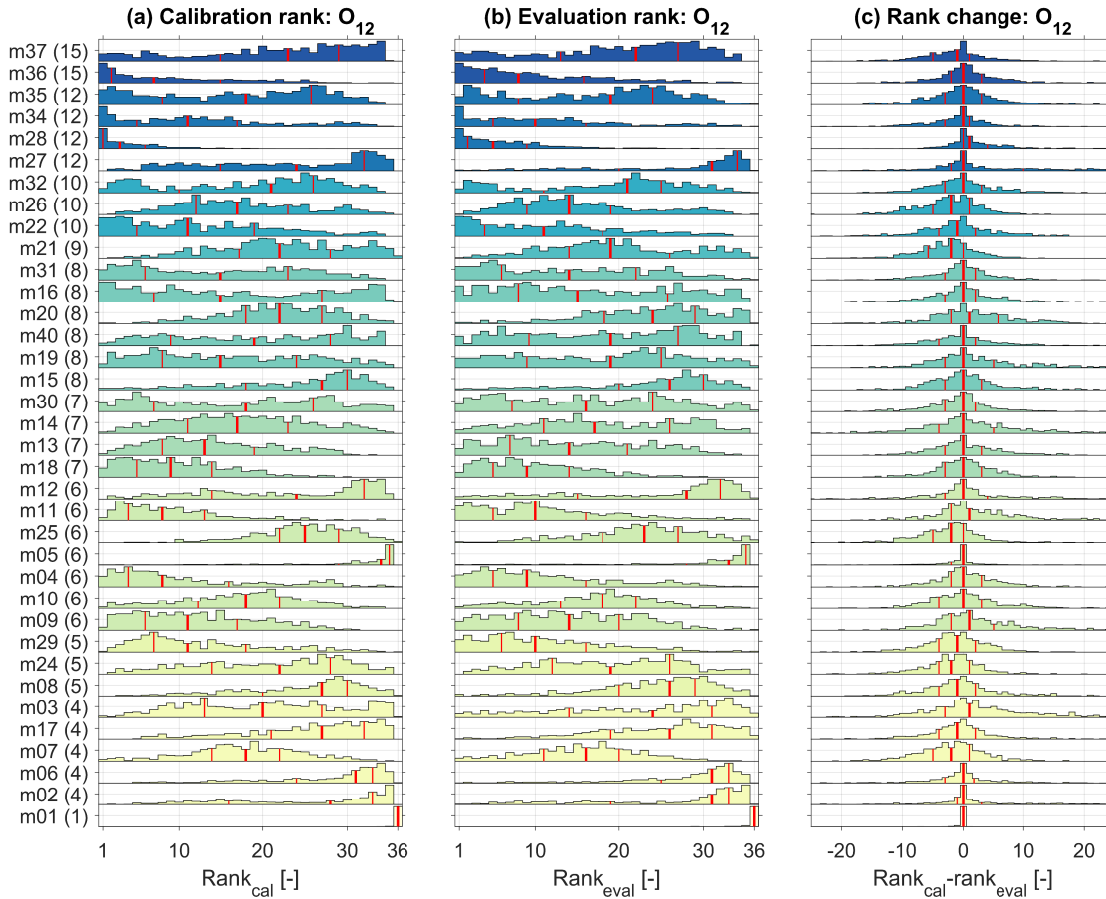
Figure C.7: Summary of model ranking during combined flow simulation ($\frac{1}{2}(KGE(Q) + KGE(1/Q))$) in all 559 catchments. Rank 1 is assigned to the model with the highest efficiency value in a given catchment. Models are sorted by number of parameters, with the model with the most parameters at the top (shown in brackets after each model ID). Colouring and model IDs match those in Figure 2 in the main manuscript. Each histogram is based on 559 values. Note that each histogram is scaled relative to its own maximum, meaning that absolute heights should not be compared between different rows. Red lines show the 25th, 50th (thick line) and 75th percentiles. (a) Calibration ranking. (b) Evaluation ranking. (c) Difference in ranking between calibration and evaluation period.

Figure C.8: Model efficiency trade-off during evaluation of all three objective functions at various percentiles of model performance. Performance for the combined flow objective is shown through the colour scheme. 5 models are identified on each plot through their model ID.
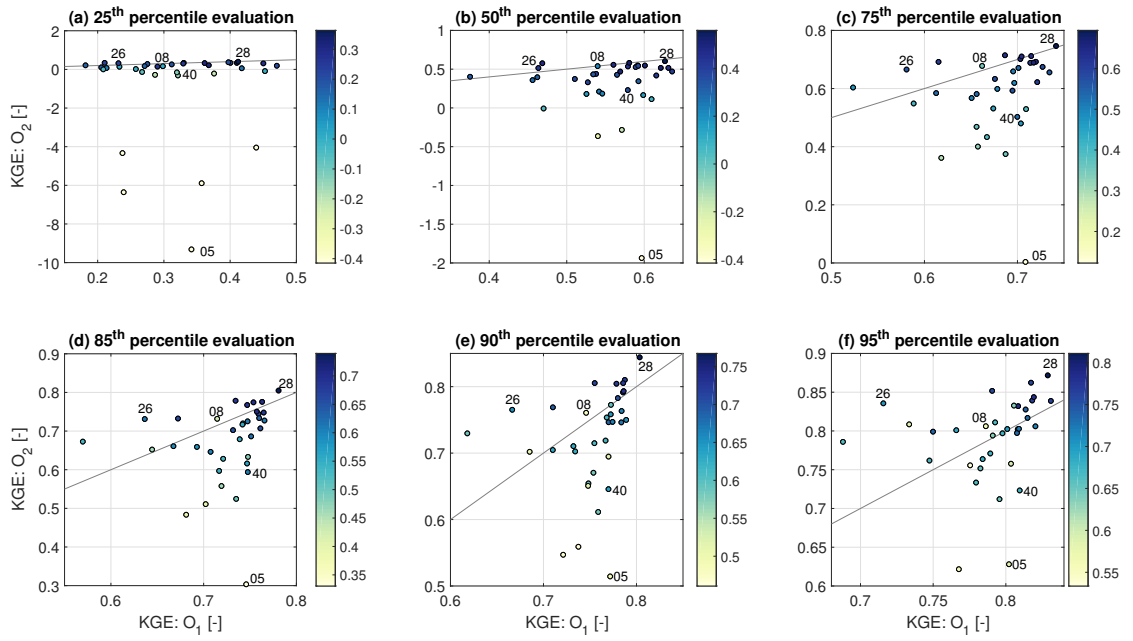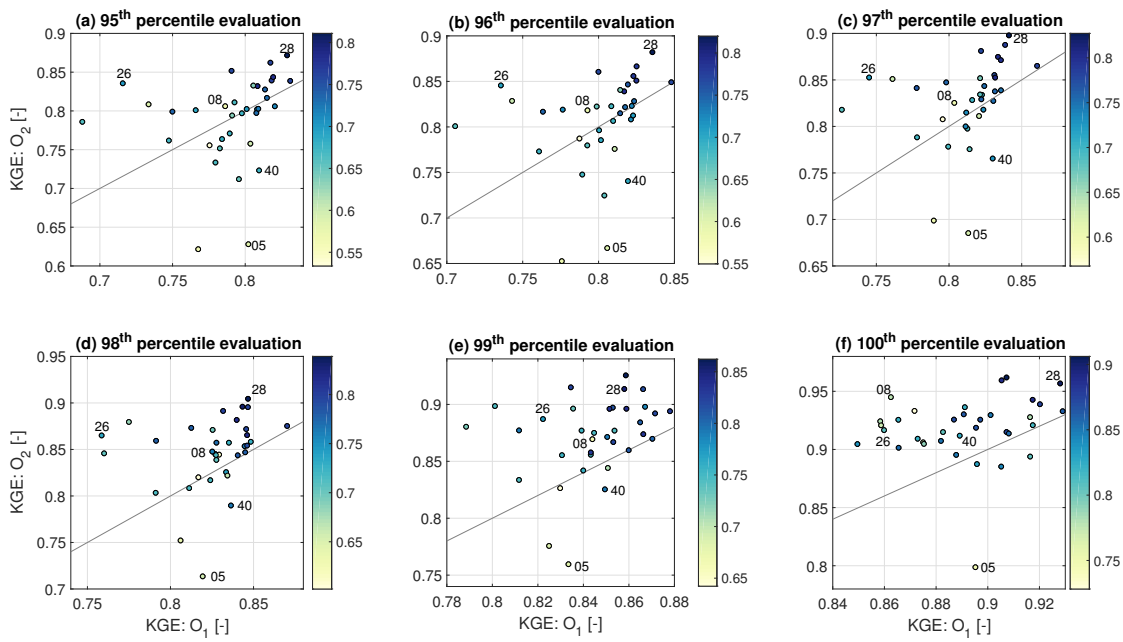
Figure C.9: Model efficiency trade-off during evaluation of all three objective functions at various percentiles of model performance. Performance for the combined flow objective is shown through the colour scheme. 5 models are identified on each plot through their model ID.

Rank correlation between high flow evaluation rank and catchment attributes (559 basins)

Figure C.10: Spearman rank correlation between model ranks (obtained during high flow (KGE(Q)) objective function evaluation) and CAMELS catchment attributes. Models are ranked from best to worst, with rank 1 indicating the best model. Only correlations with p-value < 0.05 are shown. Marker size corresponds to the strength of the correlation, as does the intensity of the colours. Models are sorted manually in an attempt to place models with similar correlation patterns close together. The sorting order is the same as used in Figure 9. This makes the models with snow components stand out slightly. See m06, m12, m30 and m37 which would here be better placed near models m31, m32, m35, m21 and m26.

Figure C.11: Spearman rank correlation between model ranks (obtained during low flow (KGE(1/Q)) objective function evaluation) and CAMELS catchment attributes. Models are ranked from best to worst, with rank 1 indicating the best model. Only correlations with p-value < 0.05 are shown. Marker size corresponds to the strength of the correlation, as does the intensity of the colours. Models are sorted manually in an attempt to place models with similar correlation patterns close together. The sorting order is the same as used in Figure C.12.

Figure C.12: Spearman rank correlation between model ranks (obtained during low flow (KGE(1/Q)) objective function evaluation) and CAMELS catchment attributes. Only catchments where less than 10% of annual precipitation occurs as snowfall are included in this figure. Models are ranked from best to worst, with rank 1 indicating the best model. Only correlations with p-value < 0.05 are shown. Marker size corresponds to the strength of the correlation, as does the intensity of the colours. Models are sorted manually in an attempt to place models with similar correlation patterns close together. The sorting order is the same as used in Figure C.11.

Figure C.13: Spearman rank correlation between model ranks (obtained during combined flow ($\frac{1}{2}[KGE(Q) + KGE(1/Q)]$) objective function evaluation) and CAMELS catchment attributes. Models are ranked from best to worst, with rank 1 indicating the best model. Only correlations with p-value < 0.05 are shown. Marker size corresponds to the strength of the correlation, as does the intensity of the colours. Models are sorted manually in an attempt to place models with similar correlation patterns close together. The sorting order is the same as used in Figure C.14.

Figure C.14: Spearman rank correlation between model ranks (obtained during combined flow ($\frac{1}{2}[KGE(Q) + KGE(1/Q)]$) objective function evaluation) and CAMELS catchment attributes. Only catchments where less than 10% of annual precipitation occurs as snowfall are included in this figure. Models are ranked from best to worst, with rank 1 indicating the best model. Only correlations with p-value < 0.05 are shown. Marker size corresponds to the strength of the correlation, as does the intensity of the colours. Models are sorted manually in an attempt to place models with similar correlation patterns close together. The sorting order is the same as used in Figure C.13.

This appendix has been published as a blog post on the website of the Young Hydrologic Society, in co-operation with Wouter Berghuijs, Shaun Harrigan and David Wright. This publication does not relate directly to the main topic of the thesis and is therefore not included in the main manuscript.

Citation: Knoben, W. J. M., Wright, D., Harrigan, S, & Berghuijs, W. R. (2017), Is research at different spatial scales connected, Streams of Thought (Young Hydrologic Society), Published May 2017. URL: https://younghs.com/2017/05/03/meet-the-expert-in-hydrology-is-research-at-different-spatial-scales-connected/

D.1 Introduction

Scaling (i.e. the transfer of knowledge across scales) and scale issues (i.e. the associated problems) are at the heart of most hydrologic puzzles. In the most recent "Meet the Expert in Hydrology" session, organized at the EGU General Assembly 2017 in Vienna, YHS invited three speakers to identify to what degree their research is connected, influenced by, and influencing research at other spatial scales. By evaluating the current state of research and discussing future directions we tried to shed some new light on the question "Is hydrological research at different spatial scales connected?". This is what we learned. . .

Hydrological research takes place across extremely different spatial scales (from pore-scale to global). Experimentalists have an understanding of detailed hydrological processes of hillslopes and headwater catchments, model developers generate larger-scale hydrological descriptions to simulate streamflow, while global hydrologists use such models to

simulate the global water cycle. In theory, research at these different spatial scales could be strongly connected. For example, field experiments lead to new understanding and models, which lead to better hydrologic predictions at larger scales. In practice, hydrological research across different spatial scales can, at times, be disconnected. Barriers, such as limited data availability and consistency and the enormous heterogeneity across diverse landscapes, may limit the interaction between scales.

In the "Meet the Expert in Hydrology: Is research at different spatial scales connected?" we invited three experts who in their work focus on different spatial scales: Prof. Jeffrey McDonnell (University of Saskatchewan) focuses on novel experimental work at the hillslope and catchment scale, Dr. Markus Hrachowitz (Delft University of Technology) focuses on catchment scale models, and Prof. Reed Maxwell (Colorado School of Mines) develops models for hydrologic simulations across continental and global scales. The three experts were invited to share their experiences on bridging the connections between different research scales, the difficulties, and opportunities of transferring knowledge between these scales and their views on future developments in hydrological knowledge from the field to the global scale.

D.2   The role of fieldwork at small spatial scales

Jeffrey McDonnell first talked about the tremendous amount of heterogeneity that is present in small hillslopes. An example he highlighted was the study that led to the "fill-and-spill" hypothesis [321]. This analysis of a single hillslope over various storm events shows that subsurface depressions first have to fill up before the hillslope starts to produce runoff. This is potentially very similar to the processes that generate runoff on the land surface.

Jeffrey McDonnell argued similar threshold processes occur in many different parts of the hydrological cycle and that the heterogeneities are essentially unknowable. Emergent behaviour might be the best option to deduce hydrological functioning for scales larger than experimental plots. He then proceeded with an example of comparing two relatively close catchments that show wildly different dominant features. In such cases, tracer hydrology can be valuable to confirm (or to challenge) hypotheses of catchment functioning; tracer experiments can be used to better understand how catchments work. He showed a further example of how experimental-scale research can be expanded to a global scale using a meta-analysis of experimental studies across the planet [99].
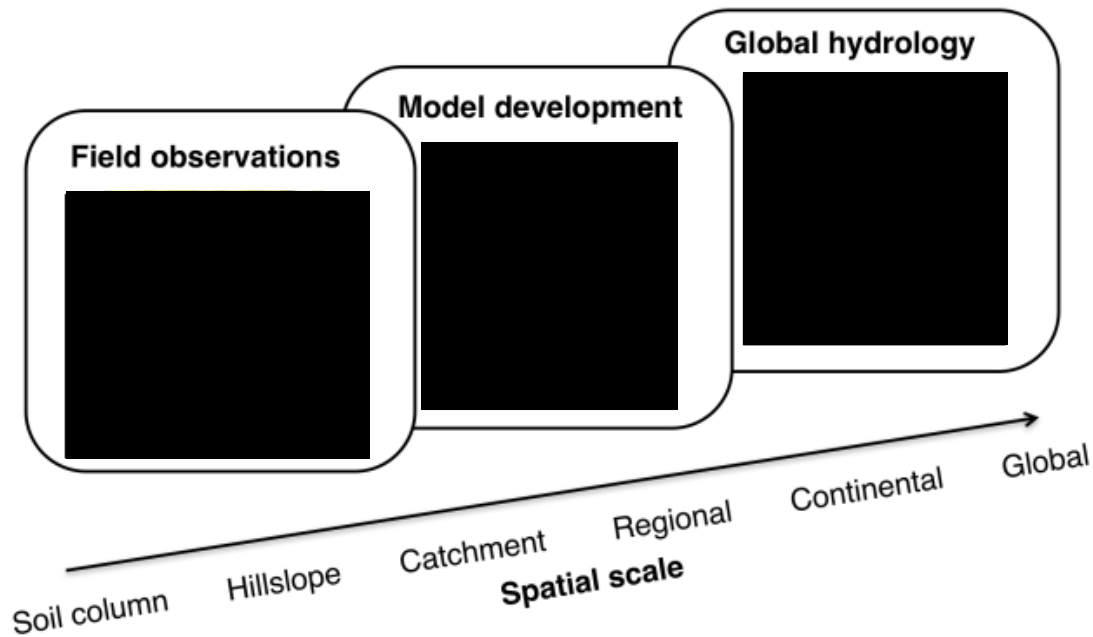
FIGURE D.1. Three experts were invited to talk about their respective research scale, and the connection to other spatial scales: Prof. Jeffrey McDonnell (University of Saskatchewan) focuses on novel experimental work at the hillslope and catchment scale, Dr. Markus Hrachowitz (Delft University of Technology) focuses on catchment scale models, and Prof. Reed Maxwell (Colorado School of Mines) develops models for hydrologic simulations across continental and global scales.

D.3   The role of catchment-scale model development

Markus Hrachowitz argued that hydrology has an observation problem, not a science problem. He gave examples of how during his Ph.D. research he found that it was extremely challenging to find hillslopes that were representative for other parts of the (very small) catchment he studied. He gave another example of measurement challenges by emphasizing that snow measurements may not be representative for snowfall only just a few meters further in the catchment.

He subsequently argued that we can still make effective predictions at the catchment scale by looking at emergent relationships. While we may not be able to measure and model the full heterogeneity present at the catchment scale simple emergent relationships at larger scales can help to develop useful models that we can parametrize and that can represent the catchment's overall functioning. He discussed the example of root-zone

storage, which while highly heterogeneous at small scales, is predictable at the catchment scale [116].

D.4   The role of large-scale simulations

Reed Maxwell argued that computational models are a key tool in bridging the gap between research scales. In his diverse research group (working both on field studies and large-scale modelling), they use models as a place to integrate the knowledge and observations across scales and test whether that matches with our understanding described in hydrological models.

He gave examples where modelling and field observations have been integrated; a recent publication tested the role of increased radiation versus the role of rain to snow transition for total runoff generation [200]. A second example he discussed is how including lateral flow seems to strongly improve the match between modelled transpiration and observed transpiration [208].

However, he carefully stated that, while models are useful we need to realize that all models are predicting the water cycle in a comparable manner, and they may all be wrong.

D.5   Discussion on increasing the synergy between research scales

During the final 40 minutes of the session, there was a panel discussion on how we can increase the synergy between research scales.

- The speakers agreed that **focussing at emergent behaviour on larger scales can inform hypotheses, that then should be tested by modelling and experimental work**.

- **Hydrologists have focused too long on reproducing the hydrograph.** There are plenty of models that can accurately reproduce the hydrograph. While this is very useful for operational purposes, it does not allow much progress in better understanding the water cycle.

- **Hydrologists have been "farmers" for too long** (i.e. collecting data and running models because we can) **and we should become "hunters" instead** (i.e. actively pursuing the relevant and interesting hypothesis that are testable). In this way, we can become much more efficient in making progress.

- As the resolution of models is improved, the hope is that the heterogeneity will disappear and the governing equation that we use can represent the processes that we will model. **In reality, heterogeneity will not disappear, at any scale.** We need to critically think how we better deal with sub-grid heterogeneity as currently we often apply the equations that are representative for other scales and hope that the heterogeneity can be adequately described by parameters that we tune. This premise may be wrong for particular processes. For example, is Darcy-Richards equation really the best equation for gridded models covering large areas?

- The large watershed is not a linear superposition of soils. **We should go after the characteristic forms of nonlinearity that can be used in models.** For example, it is currently unclear if large scale models can represent the filling and spilling of the landscape.

- **Observations indicate that water in catchments can be thousands and sometimes even billions of years old.** This is a part of reality that is not often represented by current-day hydrological models. However, Reed Maxwell emphasized that large-scale models can generate very old water ages.

D.6   About the authors

Wouter Knoben is a PhD student at the University of Bristol. Shaun Harrigan (@shaunharrigan) is a research associate at the Centre for Ecology & Hydrology (CEH) in the UK and Early Career Scientist representative for the EGU Hydrological Sciences Division. Wouter Berghuijs (@wberghuijs) is a postdoctoral researcher at ETH Zurich. David Wright is a PhD Candidate at the University of Adelaide.
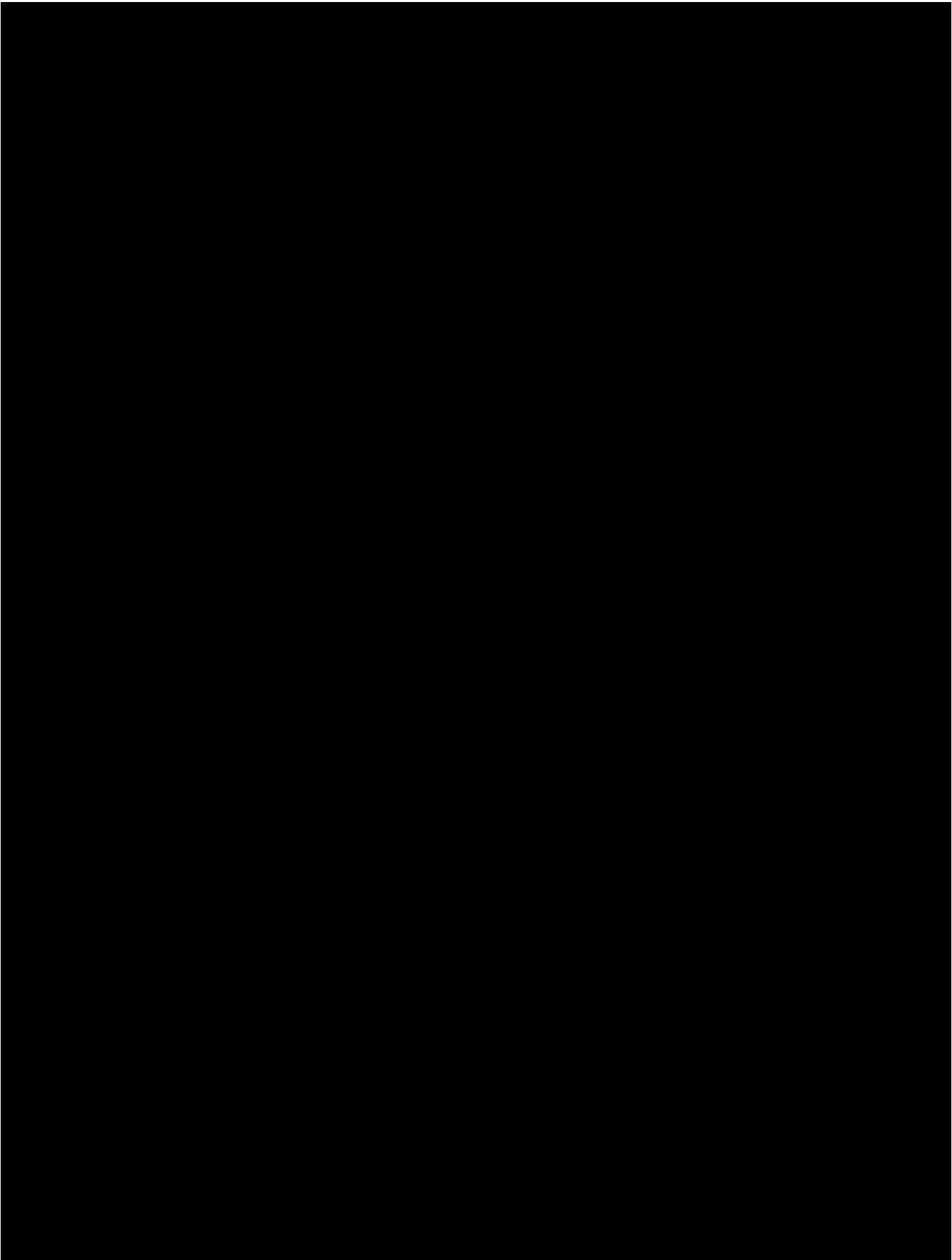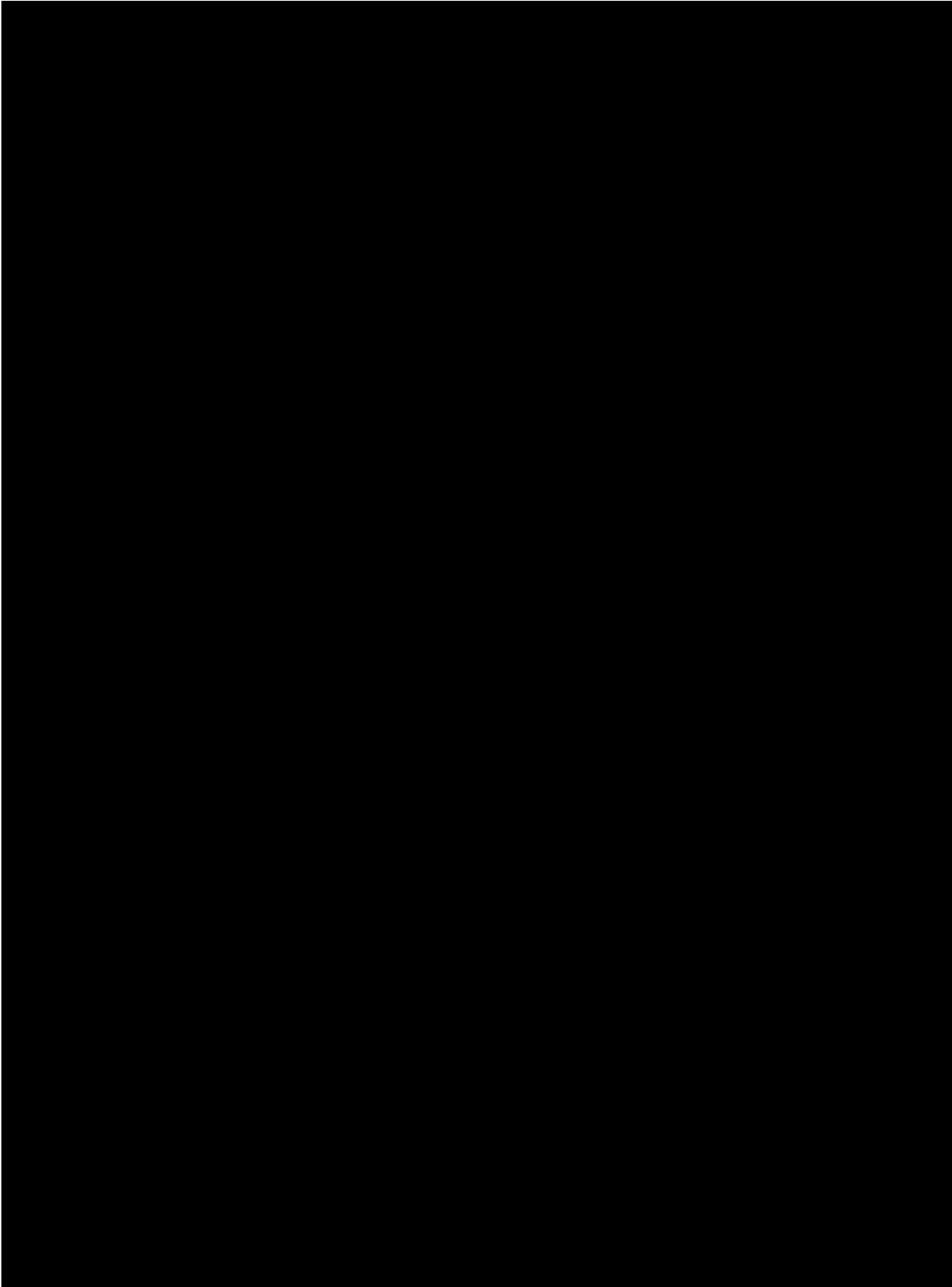
D.7   Acknowledgements
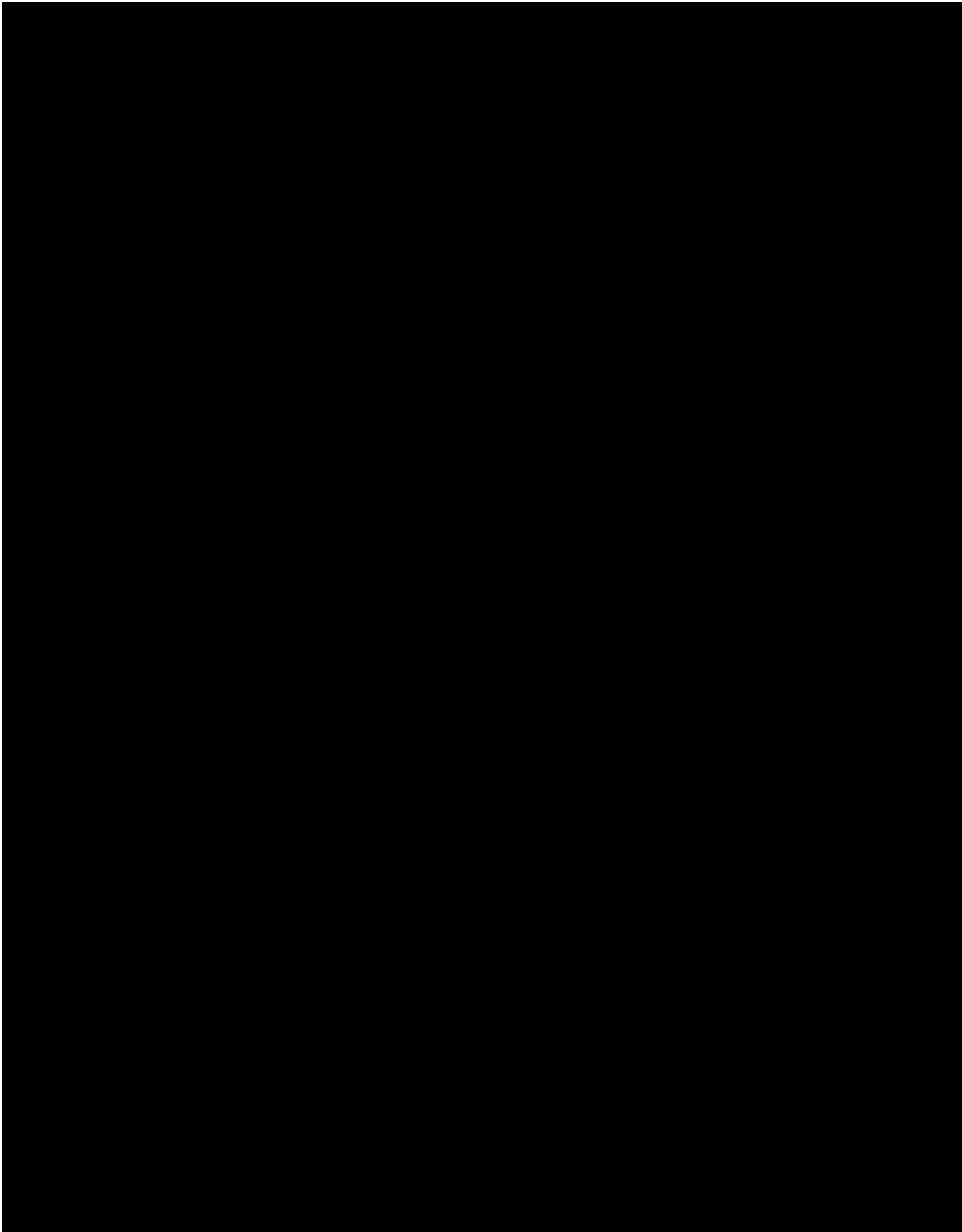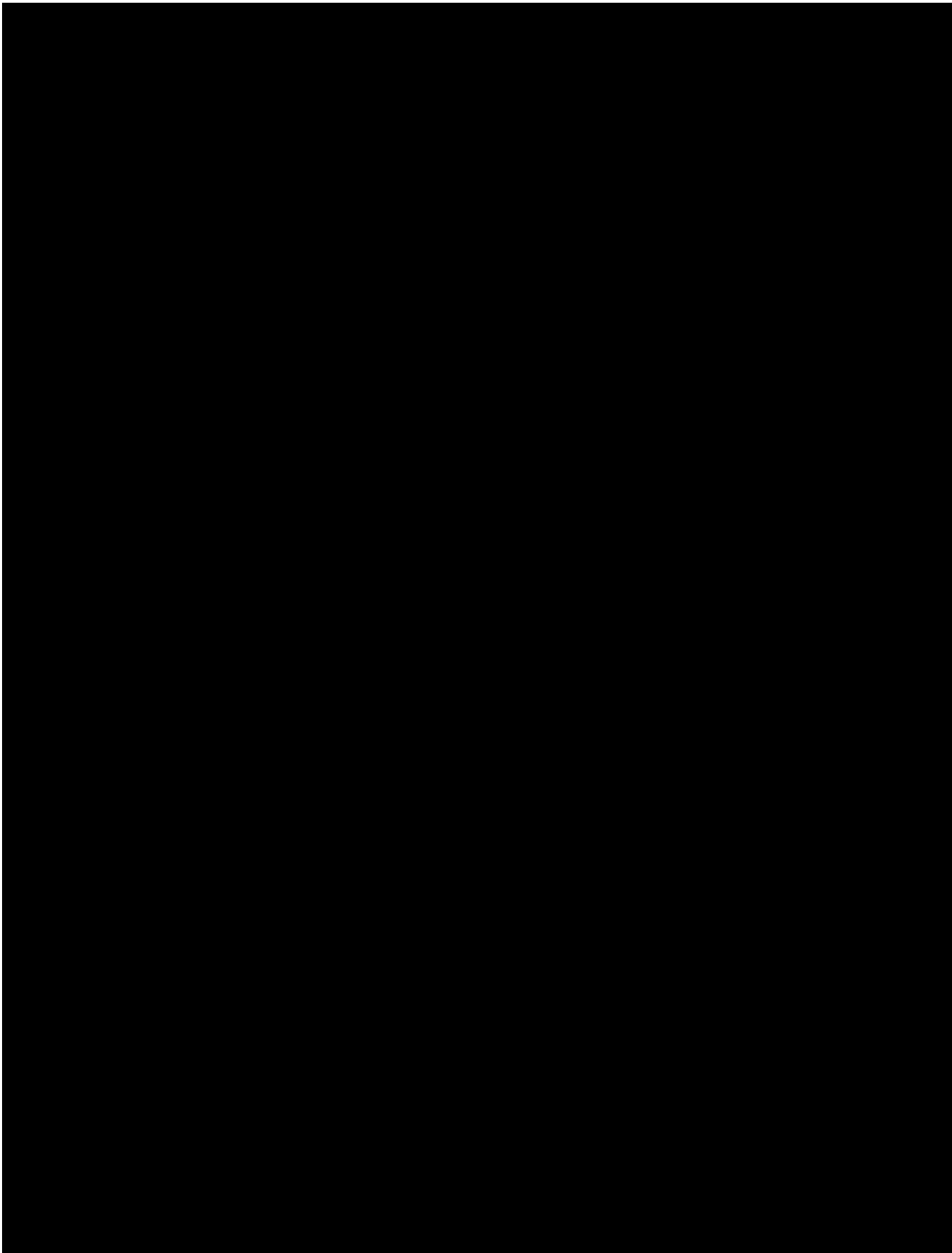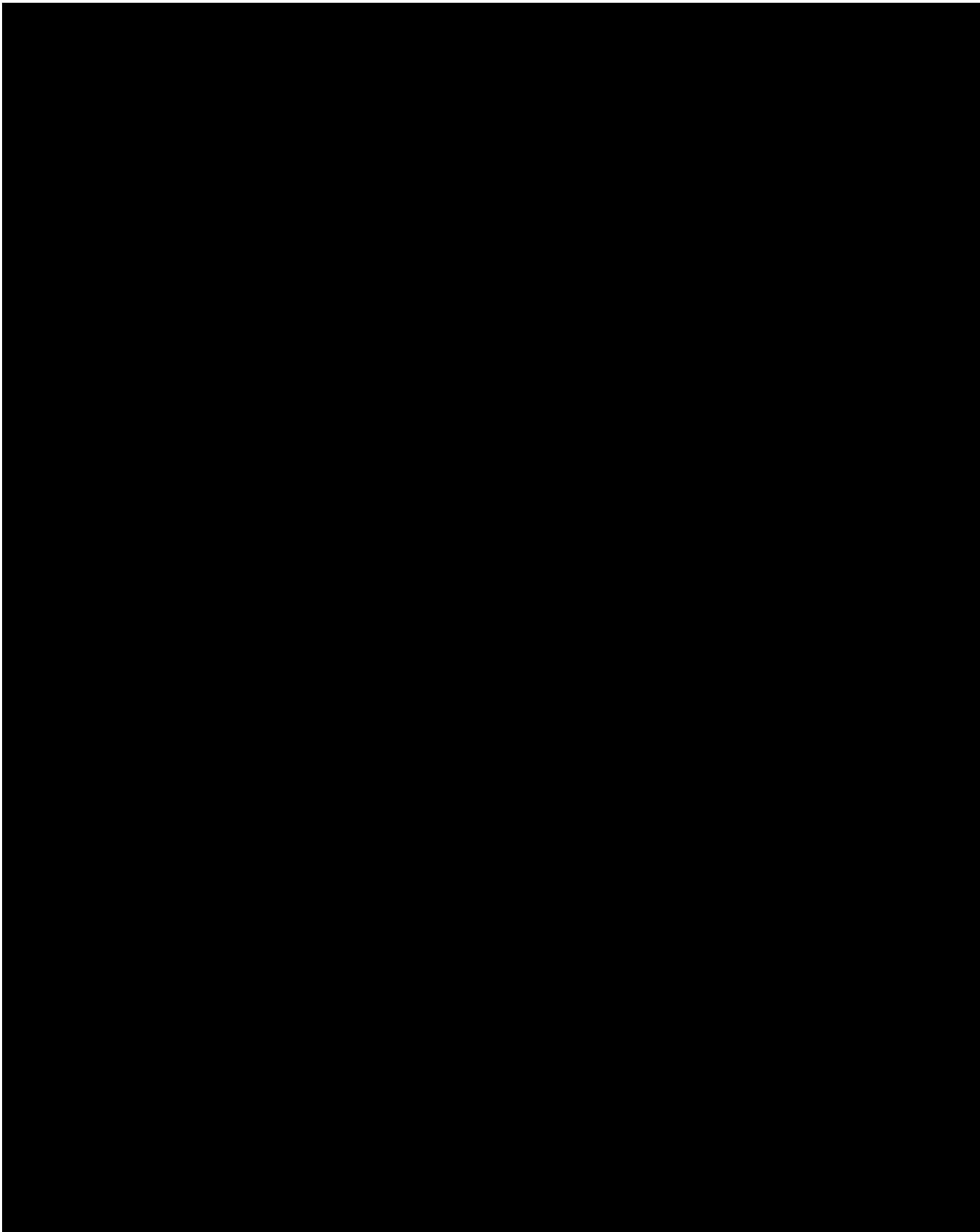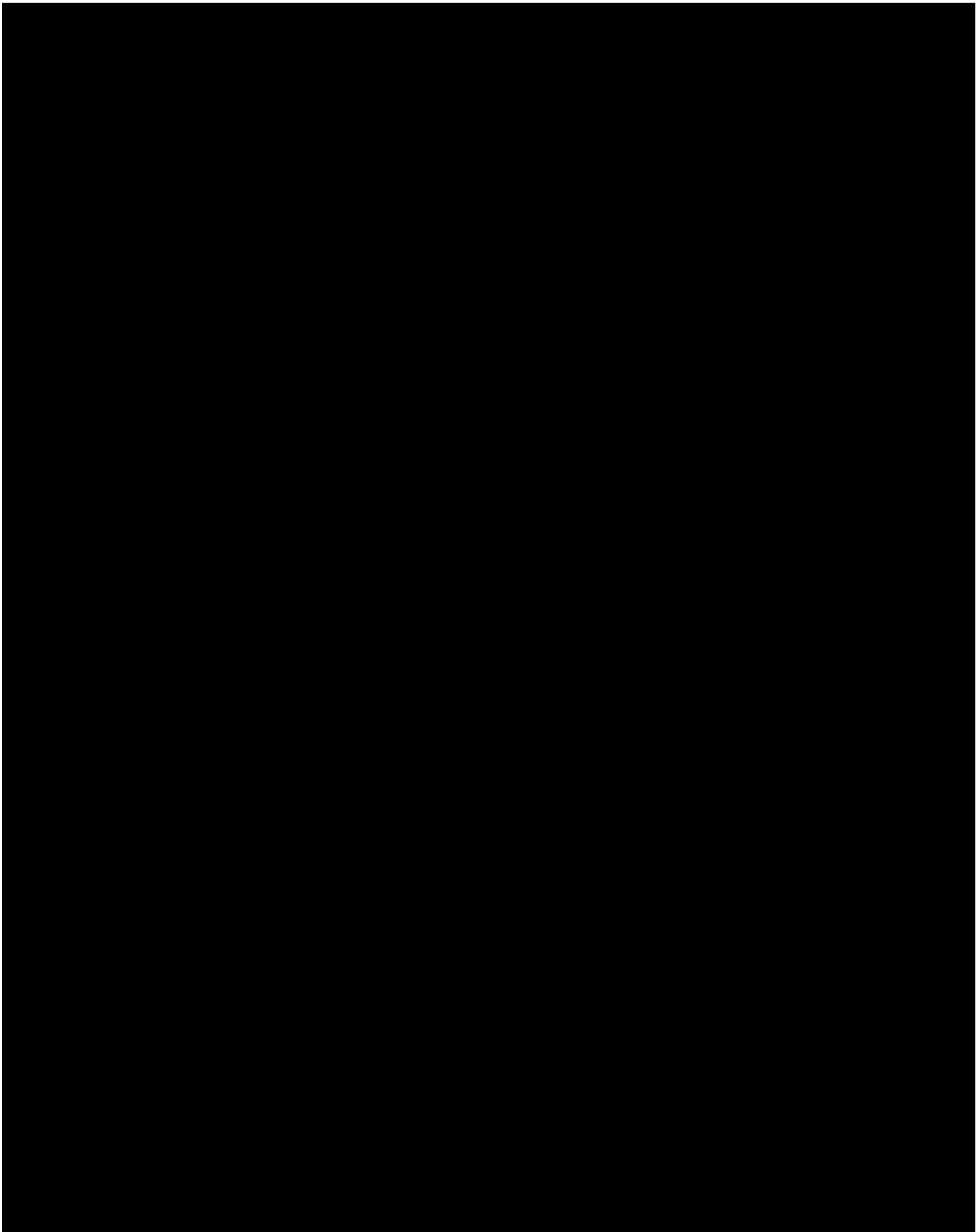
[1] G. ABRAMOWITZ AND H. GUPTA, *Toward a model space and model independence metric*, Geophysical Research Letters, 35 (2008), p. L05705.

[2] N. ADDOR AND L. A. MELSEN, *Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models*, Water Resources Research, 55 (2019), pp. 378–390.

[3] N. ADDOR, G. NEARING, C. PRIETO, A. J. NEWMAN, N. LE VINE, AND M. P. CLARK, *A Ranking of Hydrological Signatures Based on Their Predictability in Space*, Water Resources Research, 54 (2018), pp. 8792–8812.

[4] N. ADDOR, A. J. NEWMAN, N. MIZUKAMI, AND M. P. CLARK, *The CAMELS data set: catchment attributes and meteorology for large-sample studies*, Hydrology and Earth System Sciences, 21 (2017), pp. 5293–5313.

[5] R. F. ADLER, G. J. HUFFMAN, A. CHANG, R. FERRARO, P.-P. XIE, J. JANOWIAK, B. RUDOLF, U. SCHNEIDER, S. CURTIS, D. BOLVIN, A. GRUBER, J. SUSSKIND, P. ARKIN, AND E. NELKIN, *The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present)*, Journal of Hydrometeorology, 4 (2003), pp. 1147–1167.

[6] G. ALI, C. J. OSWALD, C. SPENCE, E. L. H. CAMMERAAT, K. J. MCGUIRE, T. MEIXNER, AND S. M. REANEY, *Towards a unified threshold-based hydrological theory: necessary components and recurring challenges*, Hydrological Processes, 27 (2013), pp. 313–318.

[7] R. ALLEN, L. PEREIRA, D. RAES, AND M. SMITH, *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56*, Rome, 1998.

[8] C. ALVAREZ-GARRETON, P. A. MENDOZA, J. P. BOISIER, N. ADDOR, M. GALLEGUILLOS, M. ZAMBRANO-BIGIARINI, A. LARA, C. PUELMA, G. CORTES, R. GARREAUD, J. MCPHEE, AND A. AYALA, *The CAMELS-CL dataset: catchment*

attributes and meteorology for large sample studies – Chile dataset, Hydrology and Earth System Sciences, 22 (2018), pp. 5817–5846.

[9]   J. K. O. AMOAH, D. M. AMATYA, AND S. NNAJI, *Quantifying watershed surface depression storage: determination and application in a hydrologic model*, Hydrological Processes, 27 (2013), pp. 2401–2413.

[10] V. ANDRÉASSIAN, N. LE MOINE, C. PERRIN, M.-H. RAMOS, L. OUDIN, T. MATHEVET, J. LERAT, AND L. BERTHET, *All that glitters is not gold: the case of calibrating hydrological models*, Hydrological Processes, 26 (2012), pp. 2206–2210.

[11] V. ANDRÉASSIAN, C. PERRIN, L. BERTHET, N. LE MOINE, J. LERAT, C. LOUMAGNE, L. OUDIN, T. MATHEVET, M. H. RAMOS, AND A. VALÉRY, *Crash tests for a standardized evaluation of hydrological models*, Hydrology and Earth System Sciences, 13 (2009), pp. 1757–1764.

[12] V. ANDRÉASSIAN, C. PERRIN, AND C. MICHEL, *Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models*, Journal of Hydrology, 286 (2004), pp. 19–35.

[13] V. ANDRÉASSIAN, C. PERRIN, E. PARENT, AND A. BÁRDOSSY, *The Court of Miracles of Hydrology: can failure stories contribute to hydrological science?*, Hydrological Sciences Journal, 55 (2010), pp. 849–856.

[14] M. E. ANGELES, J. E. GONZÁLEZ, N. D. RAMÍREZ-BELTRÁN, C. A. TEPLEY, AND D. E. COMARAZAMY, *Origins of the Caribbean Rainfall Bimodal Behavior*, Journal of Geophysical Research, 115 (2010), p. D11106.

[15] S. ARCHFIELD, J. KENNEN, D. CARLISLE, AND D. WOLOCK, *An Objective and Parsimonious Approach for Classifying Natural Flow Regimes at a Continental Scale*, River research and applications, 30 (2014), pp. 1085–1095.

[16] R. ARSENAULT, A. POULIN, P. CÔTÉ, AND F. BRISSETTE, *Comparison of Stochastic Optimization Algorithms in Hydrological Model Calibration*, Journal of Hydrologic Engineering, 19 (2014), pp. 1374–1384.

[17] S. ASSOULINE, *Infiltration into soils: Conceptual approaches and solutions*, Water Resources Research, 49 (2013), pp. 1755–1772.

[18] S. ATKINSON, M. SIVAPALAN, R. WOODS, AND N. VINEY, *Dominant physical controls on hourly flow predictions and the role of spatial variability: Mahurangi catchment, New Zealand*, Advances in Water Resources, 26 (2003), pp. 219–235.

[19] S. E. ATKINSON, R. A. WOODS, AND M. SIVAPALAN, *Climate and landscape controls on water balance model complexity over changing timescales*, Water Resources Research, 38 (2002), pp. 50–1–50–17.

[20] Y. BAI, T. WAGENER, AND P. REED, *A top-down framework for watershed model evaluation and selection under uncertainty*, Environmental Modelling & Software, 24 (2009), pp. 901–916.

[21] D. BAIZE AND M.-C. GIRARD, *Référentiel pédologique 2008*, Editions Quae, 2009.

[22] A. BÁRDOSSY AND S. K. SINGH, *Robust estimation of hydrological model parameters*, Hydrology and Earth System Sciences, 12 (2008), pp. 1273–1283.

[23] J. C. BATHURST, J. EWEN, G. PARKIN, P. E. O'CONNELL, AND J. D. COOPER, *Validation of catchment models for predicting land-use and climate change impacts. 3. Blind validation for internal and outlet responses*, Journal of Hydrology, 287 (2004), pp. 74–94.

[24] H. E. BECK, A. DE ROO, AND A. I. J. M. VAN DIJK, *Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments**, Journal of Hydrometeorology, 16 (2014), pp. 1478–1501.

[25] H. E. BECK, A. I. J. M. V. DIJK, V. LEVIZZANI, J. SCHELLEKENS, D. G. MIRALLES, B. MARTENS, AND A. DE ROO, *MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data*, Hydrology and Earth System Sciences, 21 (2017), pp. 589–615.

[26] R. BECKINSALE, *River regimes*, in Water, earth, and man, R. Chorley, ed., London: Methuen, 1969, pp. 455–471.

[27] M. BELDA, E. HOLTANOVÁ, T. HALENKA, AND J. KALVOVÁ, *Climate classification revisited: from Köppen to Trewartha*, Climate Research, 59 (2014), pp. 1–13.

[28] V. BELL, D. CARRINGTON, AND R. MOORE, *Comparison of rainfall-runoff models for flood forecasting - part 2*, tech. rep., Environment Agency, 2001.

[29] W. R. BERGHUIJS, M. SIVAPALAN, R. A. WOODS, AND H. H. G. SAVENIJE, *Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales*, Water Resources Research, 50 (2014), pp. 5638–5661.

[30] W. R. BERGHUIJS AND R. A. WOODS, *A simple framework to quantitatively describe monthly precipitation and temperature climatology*, International Journal of Climatology, (2016), pp. 3161–3174.

[31] M. G. BETHUNE, B. SELLE, AND Q. J. WANG, *Understanding and predicting deep percolation under surface irrigation*, Water Resources Research, 44 (2008), pp. 1–16.

[32] K. BEVEN, *Changing ideas in hydrology — The case of physically-based models*, Journal of Hydrology, 105 (1989), pp. 157–172.

[33] ——, *TOPMODEL: a critique*, Hydrological Processes, 11 (1997), pp. 1069–1085.

[34] ——, *Uniqueness of place and process representations in hydrological modelling*, 2000.

[35] ——, *On hypothesis testing in hydrology*, Hydrological Processes, 15 (2001), pp. 1655–1657.

[36] ——, *On landscape space to model space mapping*, Hydrological Processes, 15 (2001), pp. 323–324.

[37] ——, *Towards a coherent philosophy for modelling the environment*, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 458 (2002), pp. 2465–2484.

[38] ——, *Environmental modelling: an uncertain future?*, Routledge, London, 2009.

[39] ——, *Preferential flows and travel time distributions: defining adequate hypothesis tests for hydrological process models*, Hydrological Processes, 24 (2010), pp. 1537–1547.

[40] ——, *Rainfall-Runoff Modelling: The Primer*, John Wiley and Sons Ltd, 2nd ed., 2012.

[41] ——, *Facets of uncertainty: Epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication*, Hydrological Sciences Journal, 61 (2016), pp. 1652–1665.

[42] ——, *On hypothesis testing in hydrology : Why falsification of models is still a really good idea*, WIRE's Water, 5 (2018).

[43] K. BEVEN AND A. BINLEY, *The future of distributed models: model calibration and uncertainty prediction*, Hydrological processes, 6 (1992), pp. 279–298.

[44] ——, *GLUE: 20 years on*, Hydrological Processes, 28 (2014), pp. 5897–5918.

[45] K. BEVEN AND H. L. CLOKE, *Comment on "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water" by Eric F. Wood et al.*, Water Resources Research, 48 (2012), p. W01801.

[46] K. BEVEN AND J. FREER, *A dynamic topmodel*, Hydrological Processes, 15 (2001), pp. 1993–2011.

[47] ——, *Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology*, Journal of hydrology, 249 (2001), pp. 11–29.

[48] K. BEVEN AND M. KIRKBY, *A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant*, Hydrological Sciences Bulletin, 24 (1979), pp. 43–69.

[49] K. BEVEN, M. KIRKBY, N. SCHOFIELD, AND A. TAGG, *Testing a physically-based flood forecasting model (TOPMODEL) for three U.K. catchments*, Journal of Hydrology, 69 (1984), pp. 119–143.

[50] K. BEVEN, R. LAMB, P. QUINN, R. ROMANOWICZ, AND J. FREER, *TOPMODEL*, in Computer Models of Watershed Hydrology, V. P. Singh, ed., Water Resources Publications, USA, Baton Rouge, 1995, ch. 18, pp. 627–668.

[51] K. BEVEN AND I. WESTERBERG, *On red herrings and real herrings: disinformation and information in hydrological inference*, Hydrological Processes, 25 (2011), pp. 1676–1680.

[52] J. C. BEZDEK, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Springer US, Boston, MA, 1981.

[53] G. BLÖSCHL AND A. MONTANARI, *Climate change impacts - Throwing the dice?*, Hydrological Processes, 24 (2010), pp. 374–381.

[54] G. BLÖSCHL, S. MURUGESU, T. WAGENER, A. VIGLIONE, AND H. SAVENIJE, eds., *Runoff prediction in ungauged basins - Synthesis across processes, places and scales*, Cambridge University Press, 2013.

355

[55] G. BLÖSCHL AND M. SIVAPALAN, *Scale issues in hydrological modelling: A review*, Hydrological Processes, 9 (1995), pp. 251–290.

[56] C. I. BOVOLO, R. PEREIRA, G. PARKIN, C. KILSBY, AND T. WAGNER, *Fine-scale regional climate patterns in the Guianas, tropical South America, based on observations and reanalysis data*, International Journal of Climatology, 32 (2012), pp. 1665–1689.

[57] D. P. BOYLE, *Multicriteria calibration of hydrologic models*, phd thesis, University of Arizona, 2001.

[58] M. I. BUDYKO, *Climate and life*, New York: Academic Press, 1974.

[59] N. BULYGINA, N. MCINTYRE, AND H. WHEATER, *Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis*, Hydrol. Earth Syst. Sci, 13 (2009), pp. 893–904.

[60] R. BURNASH, *The NWS River Forecast System - catchment modeling*, in Computer Models of Watershed Hydrology, V. Singh, ed., 1995, ch. 10, pp. 311–366.

[61] D. E. CANFIELD JR, K. A. LANGELAND, M. J. MACEINA, W. T. HALLER, J. V. SHIREMAN, AND J. R. JONES, *Trophic state classification of lakes with aquatic macrophytes*, Canadian Journal of Fisheries and Aquatic Sciences, 40 (1983), pp. 1713–1718.

[62] G. CARRILLO, P. A. TROCH, M. SIVAPALAN, T. WAGENER, C. HARMAN, AND K. SAWICZ, *Catchment classification: Hydrological analysis of catchment behavior through process-based modeling along a climate gradient*, Hydrology and Earth System Sciences, 15 (2011), pp. 3411–3430.

[63] S. CASTIGLIONI, A. CASTELLARIN, A. MONTANARI, J. SKØIEN, G. LAAHA, AND G. BLÖSCHL, *Smooth regional estimation of low-flow indices: physiographical space based interpolation and top-kriging*, Hydrology and Earth System Sciences, 15 (2011), pp. 715–727.

[64] CENTRE FOR ECOLOGY & HYDROLOGY, *Plynlimon Experimental Catchments*. https://www.ceh.ac.uk/our-science/projects/plynlimon-experimental-catchments, 2014, Accessed on 2019-03-04.

[65] S. CEOLA, B. ARHEIMER, E. BARATTI, G. BLÖSCHL, R. CAPELL, A. CASTELLARIN, J. FREER, D. HAN, M. HRACHOWITZ, Y. HUNDECHA, C. HUTTON, G. LINDSTRÖM, A. MONTANARI, R. NIJZINK, J. PARAJKA, E. TOTH, A. VIGLIONE,

AND T. WAGENER, *Virtual laboratories: new opportunities for collaborative water science*, Hydrology and Earth System Sciences, 19 (2015), pp. 2101–2117.

[66] A. CERDÀ, *Seasonal variability of infiltration rates under contrasting slope conditions in southeast Spain*, Geoderma, 69 (1996), pp. 217–232.

[67] C. P. CHANG, Z. WANG, J. MCBRIDE, AND C. H. LIU, *Annual cycle of Southeast Asia - Maritime continent rainfall and the asymmetric monsoon transition*, Journal of Climate, 18 (2005), pp. 287–301.

[68] S. CHIDAMBARAM, M. V. PRASANNA, A. L. RAMANATHAN, K. VASU, S. HAMEED, U. K. WARRIER, K. SRINIVASAMOORTHY, R. MANIVANNAN, K. TIRUMALESH, P. ANANDHAN, AND G. JOHNSONBABU, *A study on the factors affecting the stable isotopic composition in precipitation of Tamil Nadu, India*, Hydrological Processes, 23 (2009), pp. 1792–1800.

[69] F. CHIEW AND T. MCMAHON, *Application of the daily rainfall-runoff model MOD-HYDROLOG to 28 Australian catchments*, Journal of Hydrology, 153 (1994), pp. 383–416.

[70] F. CHIEW, M. PEEL, AND A. WESTERN, *Application and testing of the simple rainfall-runoff model SIMHYD*, in Mathematical Models of Small Watershed Hydrology, V. Singh and D. Frevert, eds., Water Resources Publications LLC, USA, Chelsea, Michigan, USA, 2002, ch. 11, pp. 335–367.

[71] F. H. S. CHIEW, *Estimating groundwater recharge using an integrated surface and groundwater model*, PhD thesis, University of Melbourne, 1990.

[72] M. P. CLARK AND D. KAVETSKI, *Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes*, Water Resources Research, 46 (2010).

[73] M. P. CLARK, D. KAVETSKI, AND F. FENICIA, *Pursuing the method of multiple working hypotheses for hydrological modeling*, Water Resources Research, 47 (2011).

[74] M. P. CLARK, H. K. MCMILLAN, D. B. G. COLLINS, D. KAVETSKI, AND R. A. WOODS, *Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses*, Hydrological Processes, 25 (2011), pp. 523–543.

357

[75] M. P. CLARK, B. NIJSSEN, J. D. LUNDQUIST, D. KAVETSKI, D. E. RUPP, R. A. WOODS, J. E. FREER, E. D. GUTMANN, A. W. WOOD, L. D. BREKKE, J. R. ARNOLD, D. J. GOCHIS, AND R. M. RASMUSSEN, *A unified approach for process-based hydrologic modeling: 1. Modeling concept*, Water Resources Research, 51 (2015), pp. 2498–2514.

[76] M. P. CLARK, B. NIJSSEN, J. D. LUNDQUIST, D. KAVETSKI, D. E. RUPP, R. A. WOODS, J. E. FREER, E. D. GUTMANN, A. W. WOOD, D. J. GOCHIS, R. M. RASMUSSEN, D. G. TARBOTON, V. MAHAT, G. N. FLERCHINGER, AND D. G. MARKS, *A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies*, Water Resources Research, 51 (2015), pp. 2515–2542.

[77] M. P. CLARK, B. SCHAEFLI, S. J. SCHYMANSKI, L. SAMANIEGO, C. H. LUCE, B. M. JACKSON, J. E. FREER, J. R. ARNOLD, R. D. MOORE, E. ISTANBULLUOGLU, AND S. CEOLA, *Improving the theoretical underpinnings of process-based hydrologic models*, Water Resources Research, 52 (2016), pp. 2350–2365.

[78] M. P. CLARK, A. G. SLATER, D. E. RUPP, R. A. WOODS, J. A. VRUGT, H. V. GUPTA, T. WAGENER, AND L. E. HAY, *Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models*, Water Resources Research, 44 (2008).

[79] E. COOPERSMITH, M. A. YAEGER, S. YE, L. CHENG, AND M. SIVAPALAN, *Exploring the physical controls of regional patterns of flow duration curves – Part 3: A catchment classification system based on regime curve indicators*, Hydrology and Earth System Sciences, 16 (2012), pp. 4467–4482.

[80] L. CORON, V. ANDRÉASSIAN, C. PERRIN, J. LERAT, J. VAZE, M. BOURQUI, AND F. HENDRICKX, *Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments*, Water Resources Research, 48 (2012), p. W05552.

[81] A. COURT, *Measures of streamflow timing*, Journal of Geophysical Research, 67 (1962), pp. 4335–4339.

[82] G. COXON, J. FREER, R. LANE, T. DUNNE, N. J. K. HOWDEN, T. WAGENER, AND R. WOODS, *DECIPHeR v1 : Dynamic fluxEs and ConnectIvity for Predictions of HydRology*, Geoscientific Model Development Discussions, (2018).

[83] G. COXON, J. FREER, T. WAGENER, N. A. ODONI, AND M. CLARK, *Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments*, Hydrological Processes, 28 (2014), pp. 6135–6150.

[84] B. CROKE AND A. JAKEMAN, *A catchment moisture deficit module for the IHACRES rainfall-runoff model*, Environmental Modelling & Software, 19 (2004), pp. 1–5.

[85] S. M. CROOKS AND P. S. NADEN, *CLASSIC: a semi-distributed rainfall-runoff modelling system*, Hydrology and Earth System Sciences, 11 (2007), pp. 516–531.

[86] H. DAKHLAOUI, D. RUELLAND, Y. TRAMBLAY, AND Z. BARGAOUI, *Evaluating the robustness of conceptual rainfall-runoff models under climate variability in northern Tunisia*, Journal of Hydrology, 550 (2017), pp. 201–217.

[87] A. DAVISON AND D. HINKLEY, *Bootstrap methods and their application*, Cambridge University Press, Cambridge, 1997.

[88] T. DE BOER-EUSER, L. BOUAZIZ, J. DE NIEL, C. BRAUER, B. DEWALS, G. DROGUE, F. FENICIA, B. GRELIER, J. NOSSENT, F. PEREIRA, H. SAVENIJE, G. THIREL, AND P. WILLEMS, *Looking beyond general metrics for model comparison - lessons from an international model intercomparison study*, Hydrology and Earth System Sciences, 21 (2017), pp. 423–440.

[89] K. DE QUEIROZ AND J. GAUTHIER, *Phylogenetic Taxonomy*, Annual Review of Ecology and Systematics, 23 (1992), pp. 449–480.

[90] N. DE VOS, T. RIENTJES, AND H. GUPTA, *Diagnostic evaluation of conceptual rainfall-runoff models using temporal clustering*, Hydrological Processes, 24 (2010), pp. 2840–2850.

[91] G. DI BALDASSARRE, L. BRANDIMARTE, AND K. BEVEN, *The seventh facet of uncertainty: wrong assumptions, unknowns and surprises in the dynamics of human-water systems*, Hydrological Sciences Journal, 61 (2016), pp. 1748–1758.

[92] G. DI BALDASSARRE AND A. MONTANARI, *Uncertainty in river discharge observations: A quantitative analysis*, Hydrology and Earth System Sciences, 13 (2009), pp. 913–921.

[93] H. X. DO, L. GUDMUNDSSON, M. LEONARD, AND S. WESTRA, *The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily*

streamflow archive and metadata, Earth System Science Data, 10 (2018), pp. 765–785.

[94] B. A. EBEL AND K. LOAGUE, *Physics-based hydrologic-response simulation: Seeing through the fog of equifinality*, Hydrological Processes, 20 (2006), pp. 2887–2900.

[95] G. EDER, M. SIVAPALAN, AND H. P. NACHTNEBEL, *Modelling water balances in an Alpine catchment through exploitation of emergent properties over changing time scales*, Hydrological Processes, 17 (2003), pp. 2125–2149.

[96] A. EFSTRATIADIS AND D. KOUTSOYIANNIS, *One decade of multi-objective calibration approaches in hydrological modelling: a review*, Hydrological Sciences Journal, 55 (2010), pp. 58–78.

[97] R. ESSERY, S. MORIN, Y. LEJEUNE, AND C. B. MÉNARD, *A comparison of 1701 snow models using observations from an alpine site*, Advances in Water Resources, 55 (2013), pp. 131–148.

[98] T. EUSER, H. C. WINSEMIUS, M. HRACHOWITZ, F. FENICIA, S. UHLENBROOK, H. H. G. SAVENIJE, W. EDUCATION, P. O. BOX, AND D. A. DELFT, *A framework to assess the realism of model structures using hydrological signatures*, Hydrology and Earth System Sciences, 17 (2013), pp. 1893–1912.

[99] J. EVARISTO AND J. J. MCDONNELL, *Prevalence and magnitude of groundwater use by vegetation: a global stable isotope meta-analysis*, Scientific Reports, 7 (2017), p. 44110.

[100] J. EWEN AND G. PARKIN, *Validation of catchment models for predicting land-use and climate change impacts. 1. Method*, Journal of Hydrology, 175 (1996), pp. 583–594.

[101] D. FARMER, M. SIVAPALAN, AND C. JOTHITYANGKOON, *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 (2003).

[102] J. J. FEDDEMA, *A Revised Thornthwaite-Type Global Climate Classification*, Physical Geography, (2005).

[103] F. FENICIA, D. KAVETSKI, AND H. H. G. SAVENIJE, *Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development*, Water Resources Research, 47 (2011).

[104] F. FENICIA, D. KAVETSKI, H. H. G. SAVENIJE, M. P. CLARK, G. SCHOUPS, L. PFISTER, AND J. FREER, *Catchment properties, function, and conceptual model representation: is there a correspondence?*, Hydrological Processes, 28 (2014), pp. 2451–2467.

[105] F. FENICIA, D. KAVETSKI, H. H. G. SAVENIJE, AND L. PFISTER, *From spatially variable streamflow to distributed hydrological models: Analysis of keymodeling decisions*, Water Resources Research, 52 (2016), pp. 954–989.

[106] F. FENICIA, H. H. G. SAVENIJE, P. MATGEN, AND L. PFISTER, *Understanding catchment behavior through stepwise model concept improvement*, Water Resources Research, 44 (2008).

[107] P. L. FINKELSTEIN AND L. E. TRUPPI, *Spatial-Distribution of Precipitation Seasonality in the United-States*, Journal of Climate, 4 (1991), pp. 373–385.

[108] F. A. FOREL, *Températures lacustres: Recherches sur la température du Lac Léman et d'autres lacs d'eau-douce*, Archives des sciences physiques et naturelles (ser 3), 35 (1880), pp. 505–575.

[109] K. FOWLER, G. COXON, J. FREER, M. PEEL, T. WAGENER, A. WESTERN, R. WOODS, AND L. ZHANG, *Simulating Runoff Under Changing Climatic Conditions: A Framework for Model Improvement*, Water Resources Research, 54 (2018).

[110] K. J. FOWLER, M. C. PEEL, A. W. WESTERN, L. ZHANG, AND T. J. PETERSON, *Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models*, Water Resources Research, 52 (2016), pp. 1820–1846.

[111] M. FRANCHINI AND M. PACCIANI, *Comparative analysis of several conceptual rainfall-runoff models*, Journal of Hydrology, 122 (1991), pp. 161–219.

[112] J. E. FREER, K. BEVEN, AND B. AMBROISE, *Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the GLUE Approach*, Water Resources Research, 32 (1996), pp. 2161–2173.

[113] J. E. FREER, K. J. BEVEN, AND N. E. PETERS, *Multivariate seasonal period model rejection within the Generalised Likelihood Uncertainty Estimation procedure*, in Calibration of Watershed Models, Q. Duan, H. Gupta, S. Sorooshian, A. N. Rousseau, and R. Turcotte, eds., AGU Books, Washington, USA, 2002, pp. 69–87.

[114] J. E. FREER, H. MCMILLAN, J. J. MCDONNELL, AND K. J. BEVEN, *Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures*, Journal of Hydrology, 291 (2004), pp. 254–277.

[115] Y. FUKUSHIMA, *A model of river flow forecasting for a small forested mountain catchment*, Hydrological Processes, 2 (1988), pp. 167–185.

[116] H. GAO, M. HRACHOWITZ, S. J. SCHYMANSKI, F. FENICIA, N. SRIWONGSITANON, AND H. H. G. SAVENIJE, *Climate controls how ecosystems size the root zone storage capacity at catchment scale*, Geophysical Research Letters, 41 (2014), pp. 7916–7923.

[117] R. GEIGER, *Klassifikation der Klimate nach W. Köppen" [Classification of climates after W. Köppen]*, in Landolt-Börnstein – Zahlenwerte und Funktionen aus Physik, Chemie, Astronomie, Geophysik und Technik, alte Serie, Springer, Berlin, 1954, pp. 603–607.

[118] M. GERRITS, *The role of interception in the hydrological cycle*, PhD thesis, Technische Universiteit Delft, Netherlands, 2010.

[119] A. GERSHUNOV AND J. MICHAELSEN, *Climatic-scale space-time variability of tropical precipitation*, Journal of Geophysical Research-Atmospheres, 101 (1996), pp. 26297–26307.

[120] R. O. GILBERT, *Statistical Methods for Environmental Pollution Monitoring*, Wiley, New York, 1987.

[121] M. GOSWAMI AND K. M. O'CONNOR, *A 'monster' that made the SMAR conceptual model "right for the wrong reasons"*, Hydrological Sciences Journal, 55 (2010), pp. 913–927.

[122] L. GOTTSCHALK, E. LEBLOIS, AND J. O. SKØIEN, *Distance measures for hydrological data having a support*, Journal of Hydrology, 402 (2011), pp. 415–421.

[123] R. B. GRAYSON AND G. BLÖSCHL, eds., *Spatial Patterns in Catchment Hydrology: Observations and Modelling*, Cambridge University Press, 2001.

[124] R. B. GRAYSON, I. D. MOORE, AND T. A. MCMAHON, *Physically based hydrologic modeling: 1. A terrain based model for investigative purposes*, Water Resources Research, 28 (1992), pp. 2639–2658.

[125] ———, *Physically Based hydrologic Modeling 2. Is the Concept Realistic?*, Water Resources Research, 26 (1992), pp. 2659–2666.

[126] J. P. GRIME, *Vegetation classification by reference to strategies*, Nature, 250 (1974), pp. 26–31.

[127] L. GUDMUNDSSON, H. X. DO, M. LEONARD, AND S. WESTRA, *The Global Streamflow Indices and Metadata Archive (GSIM)-Part 2: Quality control, time-series indices and homogeneity assessment*, Earth System Science Data, 10 (2018), pp. 787–804.

[128] H. V. GUPTA, M. P. CLARK, J. A. VRUGT, G. ABRAMOWITZ, AND M. YE, *Towards a comprehensive assessment of model structural adequacy*, Water Resources Research, 48 (2012).

[129] H. V. GUPTA, H. KLING, K. K. YILMAZ, AND G. F. MARTINEZ, *Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling*, Journal of Hydrology, 377 (2009), pp. 80–91.

[130] H. V. GUPTA, C. PERRIN, G. BLÖSCHL, A. MONTANARI, R. KUMAR, M. CLARK, AND V. ANDRÉASSIAN, *Large-sample hydrology: A need to balance depth with breadth*, Hydrology and Earth System Sciences, 18 (2014), pp. 463–477.

[131] H. V. GUPTA, T. WAGENER, AND Y. LIU, *Reconciling theory with observations : elements of a diagnostic approach to model evaluation*, Hydrological Processes, 3813 (2008), pp. 3802–3813.

[132] A. GUSTARD, A. BULLOCK, AND J. DIXON, *Low flow estimation in the United Kingdom*, tech. rep., Institute of Hydrology, 1992.

[133] A. T. HAINES, B. L. FINLAYSON, AND T. A. MCMAHON, *A global classification of river regimes*, Applied Geography, 8 (1988), pp. 255–272.

[134] N. HANSEN, *References to CMA-ES Applications*.
http://www.cmap.polytechnique.fr/~nikolaus.hansen/cmaapplications.pdf, 2009, Accessed on 2019-01-02.

[135] N. HANSEN, A. AUGER, R. ROS, S. FINCK, AND P. POŠÍK, *Comparing Results of 31 Algorithms from the Black-Box Optimization Benchmarking BBOB-2009*, in Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference 2010, 2010, pp. 1689–1696.

[136] N. Hansen, S. D. Müller, and P. Koumoutsakos, *Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES)*, Evolutionary Computation, 11 (2003), pp. 1–18.

[137] I. Harris, P. D. Jones, T. J. Osborn, and D. H. Lister, *Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset*, International Journal of Climatology, 34 (2014), pp. 623–642.

[138] R. Hereford, R. H. Webb, and S. Graham, *Precipitation History of the Colorado Plateau Region*, in U.S. Geological Survey Fact Sheet 119-02, 2002, pp. 1–4.

[139] S. M. Herrmann and K. I. Mohr, *A continental-scale classification of rainfall seasonality regimes in Africa based on gridded precipitation and land surface temperature products*, Journal of Applied Meteorology and Climatology, 50 (2011), pp. 2504–2513.

[140] A. Hewitt, *Soil classification in New Zealand - Legacy and lessons*, Australian Journal of Soil Research, 30 (1992), pp. 843–854.

[141] T. S. Hogue, L. A. Bastidas, H. V. Gupta, and S. Sorooshian, *Evaluating model performance and parameter behavior for varying levels of land surface model complexity*, Water Resources Research, 42 (2006), p. W08430.

[142] L. Holdridge, *Life zone ecology*, Tropical Science Center, San Jose, Costa Rica, 1967.

[143] H. M. Holländer, T. Blume, H. Bormann, W. Buytaert, G. B. Chirico, J. F. Exbrayat, D. Gustafsson, H. Hölzel, P. Kraft, C. Stamm, S. Stoll, G. Blöschl, and H. Flühler, *Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data*, Hydrology and Earth System Sciences, 13 (2009), pp. 2069–2094.

[144] M. Hrachowitz and M. Clark, *HESS Opinions: The complementary merits of top-down and bottom-up modelling philosophies in hydrology*, Hydrology and Earth System Sciences Discussions, (2017), pp. 1–22.

[145] M. Hrachowitz, H. Savenije, G. Blöschl, J. McDonnell, M. Sivapalan, J. Pomeroy, B. Arheimer, T. Blume, M. Clark, U. Ehret, F. Fenicia, J. Freer, A. Gelfan, H. Gupta, D. Hughes, R. Hut, A. Montanari, S. Pande, D. Tetzlaff, P. Troch, S. Uhlenbrook, T. Wagener, H. Winsemius, R. Woods, E. Zehe, and C. Cudennec, *A decade of Predictions in*

Ungauged Basins (PUB) - a review, Hydrological Sciences Journal, 58 (2013), pp. 1198–1255.

[146] C.-P. F. HSU AND J. M. WALLACE, *The global Distribution of the Annual and Semi-annual Cycles in Precipitation*, Monthly Weather Review, 104 (1976), pp. 1093–1101.

[147] M. HUANG, X. LIANG, AND Y. LIANG, *A transferability study of model parameters for the variable infiltration capacity land surface scheme*, J. Geophys. Res., 108 (2003), p. 8864.

[148] D. A. HUGHES, *Hydrological models: mathematics or science?*, Hydrological Processes, 2201 (2010).

[149] G. HUTCHINSON AND H. LÖFFLER, *The Thermal Classification of Lakes*, Proceedings of the National Academy of Sciences, 42 (1956), pp. 84–86.

[150] R. P. IBBIT AND T. O'DONNELL, *Designing conceptual catchment models for automatic fitting methods*, in IAHS Publication, vol. 101, 1974, pp. 461–475.

[151] IUSS WORKING GROUP WRB, *World Reference Base for Soil Resources 2014, update 2015 - International soil classification system for naming soils and creating legends for soil maps.*, FAO, Rome, world soil ed., 2015.

[152] A. J. JAKEMAN AND G. M. HORNBERGER, *How much complexity is warranted in a rainfall-runoff model?*, Water Resources Research, 29 (1993), pp. 2637–2649.

[153] A. W. JAYAWARDENA AND M. C. ZHOU, *A modified spatial soil moisture storage capacity distribution curve for the Xinanjiang model*, Journal of Hydrology, 227 (2000), pp. 93–113.

[154] F. U. JEHN, L. BREUER, T. HOUSKA, K. BESTIAN, AND P. KRAFT, *Incremental model breakdown to assess the multi-hypotheses problem*, Hydrology and Earth System Sciences, 22 (2018), pp. 4565–4581.

[155] P. JIMENO-SÁEZ, J. SENENT-APARICIO, J. PÉREZ-SÁNCHEZ, AND D. PULIDO-VELAZQUEZ, *A comparison of SWAT and ANN models for daily runoff simulation in different climatic zones of peninsular Spain*, Water (Switzerland), 10 (2018).

[156] P. JOHNES, B. MOSS, AND G. PHILLIPS, *Lakes - Classification & Monitoring: A strategy for the classification of lakes*, tech. rep., National Rivers Authority, 1994.

[157] J. A. A. JONES, *Global Hydrology: Processes, Resources and Environmental Management*, Routledge, 1997.

[158] C. JOTHITYANGKOON, M. SIVAPALAN, AND D. FARMER, *Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development*, Journal of Hydrology, 254 (2001), pp. 174–198.

[159] J. M. JUSTON, A. KAUFFELDT, B. Q. MONTANO, J. SEIBERT, K. J. BEVEN, AND I. K. WESTERBERG, *Smiling in the rain: Seven reasons to be positive about uncertainty in hydrological modelling*, Hydrological Processes, 27 (2013), pp. 1117–1122.

[160] D. KAVETSKI AND M. P. CLARK, *Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction*, Water Resources Research, 46 (2010), pp. 1–27.

[161] ——, *Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing*, Hydrological Processes, 25 (2011), pp. 661–670.

[162] D. KAVETSKI AND F. FENICIA, *Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights*, Water Resources Research, 47 (2011), pp. n/a–n/a.

[163] D. KAVETSKI AND G. KUCZERA, *Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration*, Water Resources Research, 43 (2007), pp. n/a–n/a.

[164] D. KAVETSKI, G. KUCZERA, AND S. W. FRANKS, *Semidistributed hydrological modeling: A "saturation path" perspective on TOPMODEL and VIC*, Water Resources Research, 39 (2003), pp. n/a–n/a.

[165] ——, *Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory*, Water Resources Research, 42 (2006), p. W03407.

[166] ——, *Bayesian analysis of input uncertainty in hydrological modeling: 2. Application*, Water Resources Research, 42 (2006), p. W03408.

[167] ——, *Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts*, Journal of Hydrology, 320 (2006), pp. 173–186.

[168] M. J. KEABLES, *A Synoptic Climatology of the Bimodal Precipitation Distribution in the Upper Midwest*, Journal of Climate, 2 (1989), pp. 1289–1294.

[169] M. G. KENDALL, *Rank Correlation Methods*, Charles Griffin, London, 4th ed., 1975.

[170] M. KENNARD, B. PUSEY, J. OLDEN, S. MACKAY, J. STEIN, AND N. MARSH, *Classification of natural flow regimes in Australia to support environmental flow management*, Freshwater Biology, 55 (2010), pp. 171–193.

[171] S. W. KIENZLE, *A new temperature based method to separate rain and snow*, Hydrological Processes, 22 (2008), pp. 5067–5085.

[172] K. B. KIM, H. H. KWON, AND D. HAN, *Exploration of warm-up period in conceptual hydrological modelling*, Journal of Hydrology, 556 (2018), pp. 194–210.

[173] J. W. KIRCHNER, *Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology*, Water Resources Research, 42 (2006).

[174] ——, *Aggregation in environmental systems: catchment mean transit times and young water fractions under hydrologic nonstationarity*, Hydrology and Earth System Sciences Discussions, 12 (2016), pp. 3105–3167.

[175] J. W. KIRCHNER, R. P. HOOPER, C. KENDALL, C. NEAL, AND G. LEAVESLEY, *Testing and validating environmental models*, The Science of the Total Environment, 183 (1996), pp. 33–47.

[176] V. KLEMEŠ, *Operational testing of hydrological simulation models*, Hydrological Sciences Journal, 31 (1986), pp. 13–24.

[177] W. J. M. KNOBEN, J. E. FREER, K. J. A. FOWLER, M. C. PEEL, AND R. A. WOODS, *Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT): User Manual Version 1.0*, tech. rep., University of Bristol, 2018.

[178] ——, *Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT) v1.0: an open- source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous space-state formulations*, Geoscientific Model Development Discussions, (2019), pp. 1–26.

[179] W. J. M. KNOBEN, R. A. WOODS, AND J. E. FREER, *A Quantitative Hydrological Climate Classification Evaluated with Independent Streamflow Data*, Water Resources Research, 54 (2018), pp. 5088–5109.

[180] ——, *Climate data from paper "A Quantitative Hydrological Climate Classification Evaluated with Independent Streamflow Data"*, 2018.

[181] J. B. KOLLAT, P. M. REED, AND T. WAGENER, *When are multiobjective calibration trade-offs in hydrologic models meaningful?*, Water Resources Research, 48 (2012).

[182] V. I. KOREN, M. SMITH, D. WANG, AND Z. ZHANG, *Use of soil property data in the derivation of conceptual rainfall-runoff model parameters*, Proceedings of the 15th Conference on Hydrology, AMS, Long Beach, CA, (2000), pp. 103–106.

[183] M. KOTTEK, J. GRIESER, C. BECK, B. RUDOLF, AND F. RUBEL, *World map of the Köppen-Geiger climate classification updated*, Meteorologische Zeitschrift, 15 (2006), pp. 259–263.

[184] P. KRAFT, K. B. VACHÉ, H.-G. FREDE, AND L. BREUER, *CMF: A Hydrological Programming Language Extension For Integrated Catchment Models*, Environmental Modelling & Software, 26 (2011), pp. 828–830.

[185] T. KRUEGER, J. FREER, J. N. QUINTON, C. J. A. MACLEOD, G. S. BILOTTA, R. E. BRAZIER, P. BUTLER, AND P. M. HAYGARTH, *Ensemble evaluation of hydrological model hypotheses*, Water Resources Research, 46 (2010).

[186] G. KUCZERA AND M. MROCZKOWSKI, *Assessment of hydrologic parameter uncertainty and the worth of multiresponse data*, Water Resources Research, 34 (1998), pp. 1481–1489.

[187] A. KUENTZ, B. ARHEIMER, Y. HUNDECHA, AND T. WAGENER, *Understanding hydrologic variability across Europe through catchment classification*, Hydrology and Earth System Sciences, 21 (2017), pp. 2863–2879.

[188] J. LAGARIAS, J. REEDS, M. WRIGHT, AND P. WRIGHT, *Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions*, SIAM Journal on Optimization, 9 (1998), pp. 112–147.

[189] R. LAMB, *Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation*, Water Resources Research, 35 (1999), pp. 3103–3114.

[190] R. A. LANE, G. COXON, J. E. FREER, T. WAGENER, P. J. JOHNES, P. JOHN, S. GREENE, C. J. A. MACLEOD, AND S. M. REANEY, *Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across a large-sample of catchments in Great Britain*, Hydrology and Earth System Sciences Discussions, (2019).

[191] G. LEAVESLEY, P. RESTREPO, S. MARKSTROM, M. DIXON, AND L. STANNARD, *The Modular Modeling System - MMS, User's Manual*, tech. rep., U.S. Geol. Surv. Open File Rep, Denver, Col., 1996.

[192] G. H. LEAVESLEY, R. LICHTY, B. TROUTMAN, AND L. SAINDON, *Precipitation-Runoff Modeling System: User's Manual*, U.S. Geological Survey, Water-Resources Investigations Report 83-4238, (1983), p. 207.

[193] W. M. LEWIS JR, *A revised classification of lakes based on mixing*, Canadian Journal of Fisheries and Aquatic Sciences, 40 (1983), pp. 1779–1787.

[194] X. LIANG, D. P. LETTENMAIER, E. F. WOOD, AND S. J. BURGES, *A simple hydrologically based model of land surface water and energy fluxes for general circulation models*, Journal of Geophysical Research, 99 (1994), pp. 14415–14428.

[195] R. LIDÉN AND J. HARLIN, *Analysis of conceptual rainfall-runoff modelling performance in different climates*, Journal of Hydrology, 238 (2000), pp. 231–247.

[196] A. LILLY, *A hydrological classification of UK soils based on soil morphological data*, in Proceedings of the 19th World Congress of Soil Science: Soil solutions for a changing world, no. August, 2010, pp. 1–4.

[197] G. LINDSTRÖM, B. JOHANSSON, M. PERSSON, M. GARDELIN, AND S. BERGSTRÖM, *Development and test of the distributed HBV-96 hydrological model*, Journal of hydrology, 201 (1997), pp. 272–288.

[198] R. K. LINSLEY, *Rainfall-Runoff models - an overview*, in Rainfall-Runoff Relationship, V. P. Singh, ed., Water Resources Publications, USA, 1982, ch. 1, p. 582.

[199] I. G. LITTLEWOOD, K. DOWN, J. PARKER, AND D. A. POST, *IHACRES v1.0 User Guide*, tech. rep., Centre for Ecology and Hydrology, Wallingford, UK & Integrated Catchment Assessment and Mangament Centre, Australian National University, 1997.

[200] L. M FOSTER, L. A BEARUP, N. P MOLOTCH, P. D BROOKS, AND R. M MAXWELL, *Energy budget increases reduce mean streamflow more than snow–rain transitions: using integrated modeling to isolate climate change impacts on Rocky Mountain hydrology*, Environmental Research Letters, 11 (2016), p. 044015.

[201] V. MAGAÑA, J. A. AMADOR, AND S. MEDINA, *The midsummer drought over Mexico and Central America*, Journal of Climate, 12 (1999), pp. 1577–1588.

[202] M. M. MALESU, *Rainwater harvesting innovations in response to water scarcity: The late experience*, 2006.

[203] B. A. MALMGREN, R. HULUGALLA, Y. HAYASHI, AND T. MIKAMI, *Precipitation trends in Sri Lanka since the 1870s and relationships to El Niño-southern oscillation*, International Journal of Climatology, 23 (2003), pp. 1235–1252.

[204] H. B. MANN, *Non-parametric tests against trend*, Econometrica, 13 (1945), pp. 163–171.

[205] S. L. MARKSTROM, S. REGAN, L. E. HAY, R. J. VIGER, R. M. T. WEBB, R. A. PAYN, AND J. H. LAFONTAINE, *PRMS-IV, the Precipitation-Runoff Modeling System, Version 4*, in U.S. Geological Survey Techniques and Methods, book 6, chap. B7, 2015, p. 158.

[206] G. F. MARTINEZ AND H. V. GUPTA, *Hydrologic consistency as a basis for assessing complexity of monthly water balance models for the continental United States*, Water Resources Research, 47 (2011), pp. 1–18.

[207] K. MATSUURA AND C. J. WILLMOTT, *Terrestrial Precipitation: 1900-2014 Gridded Monthly Time Series (Version 4.01)*, 2015.

[208] R. M. MAXWELL AND L. E. CONDON, *Connections between groundwater flow and transpiration partitioning*, Science, 353 (2016), pp. 377–380.

[209] J. MCDONNELL AND R. WOODS, *On the need for catchment classification*, Journal of Hydrology, 299 (2004), pp. 2–3.

[210] B. L. MCGLYNN, J. J. MCDONNEL, AND D. D. BRAMMER, *A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchments , New Zealand*, Journal of Hydrology, 257 (2002).

[211] T. A. MCMAHON, M. C. PEEL, L. LOWE, R. SRIKANTHAN, AND T. R. MCVICAR, *Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: A pragmatic synthesis*, Hydrology and Earth System Sciences, 17 (2013), pp. 1331–1363.

[212] ——, *Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: A pragmatic synthesis - supplement*, Hydrology and Earth System Sciences, 17 (2013), pp. 1331–1363.

[213] H. MCMILLAN, J. FREER, F. PAPPENBERGER, T. KRUEGER, AND M. CLARK, *Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions*, Hydrological Processes, 1284 (2010).

[214] H. MCMILLAN, T. KRUEGER, AND J. FREER, *Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality*, Hydrological Processes, 26 (2012), pp. 4078–4111.

[215] H. MCMILLAN, I. WESTERBERG, AND F. BRANGER, *Five guidelines for selecting hydrological signatures*, Hydrological Processes, 31 (2017), pp. 4757–4761.

[216] H. K. MCMILLAN, I. K. WESTERBERG, AND T. KRUEGER, *Hydrological data uncertainty and its implications*, Wiley Interdisciplinary Reviews: Water, 5 (2018).

[217] L. MELSEN, N. ADDOR, N. MIZUKAMI, A. NEWMAN, P. TORFS, M. CLARK, R. UIJLENHOET, AND A. TEULING, *Mapping (dis)agreement in hydrologic projections*, Hydrology and Earth System Sciences Discussions, 22 (2018), pp. 1775–1791.

[218] R. MERZ, J. PARAJKA, AND G. BLÖSCHL, *Time stability of catchment model parameters: Implications for climate impact analyses*, Water Resources Research, 47 (2011).

[219] P. C. D. MILLY, *Climate, soil water storage, and the average annual water balance*, Water Resources Research, 30 (1994), pp. 2143–2156.

[220] P. C. D. MILLY, J. BETANCOURT, M. FALKENMARK, R. M. HIRSCH, Z. W. KUNDZEWICZ, D. P. LETTENMAIER, AND R. J. STOUFFER, *Stationarity Is Dead : Whither Water Management?*, Science, 319 (2008), pp. 573–574.

[221] R. J. MOORE, *The PDM rainfall-runoff model*, Hydrology and Earth System Sciences, 11 (2007), pp. 483–499.

[222] R. J. MOORE AND V. A. BELL, *Comparison of rainfall-runoff models for flood forecasting. Part 1: Literature review of models*, tech. rep., Environment Agency, Bristol, 2001.

[223] D. N. MORIASI, J. G. ARNOLD, M. W. VAN LIEW, R. L. BINGNER, R. D. HARMEL, AND T. L. VEITH, *Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations*, Transactions of the ASABE, 50 (2007), pp. 885–900.

371

[224] R. NATHAN AND T. MCMAHON, *SFB model part l . Validation of fixed model parameters*, in Civil Eng. Trans., no. CE32, 1990, pp. 157–161.

[225] NATIONAL WEATHER SERVICE, *II.3-SAC-SMA: Conceptualization of the Sacramento Soil Moisture Accounting model*, in National Weather Service River Forecast System (NWSRFS) User Manual, 2005, pp. 1–13.

[226] NCAR/UCAR, *Climate Data Guide*. https://climatedataguide.ucar.edu/, 2019, Accessed on 2019-01-27.

[227] G. S. NEARING, Y. TIAN, H. V. GUPTA, M. P. CLARK, K. W. HARRISON, AND S. V. WEIJS, *A philosophical basis for hydrological uncertainty*, Hydrological Sciences Journal, 61 (2016), pp. 1666–1678.

[228] J. NELDER AND R. MEAD, *A Simplex Method for Function Minimization*, The Computer Journal, 7 (1965), pp. 308–313.

[229] A. J. NEWMAN, M. P. CLARK, K. SAMPSON, A. WOOD, L. E. HAY, A. BOCK, R. J. VIGER, D. BLODGETT, L. BREKKE, J. R. ARNOLD, T. HOPSON, AND Q. DUAN, *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 (2015), pp. 209–223.

[230] A. J. NEWMAN, N. MIZUKAMI, M. P. CLARK, A. W. WOOD, B. NIJSSEN, AND G. NEARING, *Benchmarking of a Physically Based Hydrologic Model*, Journal of Hydrometeorology, 18 (2017), pp. 2215–2225.

[231] P. NICOLLE, R. PUSHPALATHA, C. PERRIN, D. FRANÇOIS, D. THIÉRY, T. MATHEVET, M. LE LAY, F. BESSON, J. M. SOUBEYROUX, C. VIEL, F. REGIMBEAU, V. ANDRÉASSIAN, P. MAUGIS, B. AUGEARD, AND E. MORICE, *Benchmarking hydrological models for low-flow simulation and forecasting on French catchments*, Hydrology and Earth System Sciences, 18 (2014), pp. 2829–2857.

[232] S. A. NIELSEN AND E. HANSEN, *Numerical simulation of he rainfall-runoff process on a daily basis*, Nordic Hydrology, (1973), pp. 171–190.

[233] R. NIJZINK, C. HUTTON, I. PECHLIVANIDIS, R. CAPELL, B. ARHEIMER, J. FREER, D. HAN, T. WAGENER, K. MCGUIRE, H. SAVENIJE, AND M. HRACHOWITZ, *The evolution of root zone moisture capacities after land use change: a step towards*

predictions under change?, Hydrology and Earth System Sciences Discussions, 20 (2016), pp. 4775–4799.

[234] P. NOBRE AND J. SHUKLA, *Variations of Sea Surface Temperature, Wind Stress, and Rainfall over the Tropical Atlantic and South America*, 1996.

[235] B. V. NORTH, D. CURTIS, AND P. C. SHAM, *A note on the calculation of empirical P values from Monte Carlo procedures.*, American journal of human genetics, 71 (2002), pp. 439–41.

[236] P. O'CONNELL, J. NASH, AND J. FARRELL, *River flow forecasting through conceptual models part II - the Brosna catchment at Ferbane*, Journal of Hydrology, 10 (1970), pp. 317–329.

[237] J. D. OLDEN AND N. L. POFF, *Redundancy and the choice of hydrologic indices for characterizing streamflow regimes*, River Research and Applications, 19 (2003), pp. 101–121.

[238] J. E. OLIVER, *Monthly precipitation distribution: A comparative index*, Professional Geographer, 32 (1980), pp. 300–309.

[239] N. ORESKES, K. SHRADER-FRECHETTE, AND K. BELITZ, *Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences*, Science, 263 (1994), pp. 641–646.

[240] R. ORTH, M. STAUDINGER, S. I. SENEVIRATNE, J. SEIBERT, AND M. ZAPPA, *Does model performance improve with complexity? A case study with three hydrological models*, Journal of Hydrology, 523 (2015), pp. 147–159.

[241] L. OUDIN, V. ANDRÉASSIAN, J. LERAT, AND C. MICHEL, *Has land cover a significant impact on mean annual streamflow? An international assessment using 1508 catchments*, Journal of Hydrology, 357 (2008), pp. 303–316.

[242] L. OUDIN, F. HERVIEU, C. MICHEL, C. PERRIN, V. ANDRÉASSIAN, F. ANCTIL, AND C. LOUMAGNE, *Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling*, Journal of Hydrology, 303 (2005), pp. 290–306.

[243] L. OUDIN, A. KAY, V. ANDRÁASSIAN, AND C. PERRIN, *Are seemingly physically similar catchments truly hydrologically similar?*, Water Resources Research, 46 (2010).

373

[244] L. OUDIN, C. PERRIN, T. MATHEVET, V. ANDRÉASSIAN, AND C. MICHEL, *Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models*, Journal of Hydrology, 320 (2006), pp. 62–83.

[245] L. OUDIN, B. SALAVATI, C. FURUSHO-PERCOT, P. RIBSTEIN, AND M. SAADI, *Hydrological impacts of urbanization at the catchment scale*, Journal of Hydrology, 559 (2018), pp. 774–786.

[246] Z. OWITI AND W. ZHU, *Spatial distribution of rainfall seasonality over East Africa*, Journal of Geography and Regional Planning, 5 (2012), pp. 409–421.

[247] B. PARTHASARATHY, A. A. MUNOT, AND D. R. KOTHAWALE, *All-India monthly and seasonal rainfall series: 1871-1993*, Theoretical and Applied Climatology, 49 (1994), pp. 217–224.

[248] I. PECHLIVANIDIS, B. JACKSON, N. MCINTYRE, AND H. WHEATER, *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 (2011), pp. 193–214.

[249] M. C. PEEL AND G. BLÖSCHL, *Hydrological modelling in a changing world*, Progress in Physical Geography, 35 (2011), pp. 249–261.

[250] M. C. PEEL, B. L. FINLAYSON, AND T. A. MCMAHON, *Updated world map of the Köppen-Geiger climate classification*, Hydrology and Earth System Sciences, 11 (2007), pp. 1633–1644.

[251] H. L. PENMAN, *The Dependence of Transpiration on Weather and Soil Conditions*, Journal of Soil Science, 1 (1950), pp. 74–89.

[252] ———, *Weather, Plant and Soil Factors in Hydrology*, Weather, 16 (1961), pp. 207–219.

[253] C. PERRIN, V. ANDRÉASSIAN, AND C. MICHEL, *Simple benchmark models as a basis for model efficiency criteria*, River Systems, 17 (2006), pp. 221–244.

[254] C. PERRIN, C. MICHEL, AND V. ANDRÉASSIAN, *Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments*, Journal of Hydrology, 242 (2001), pp. 275–301.

[255] C. PERRIN, C. MICHEL, AND V. ANDRÉASSIAN, *Improvement of a parsimonious model for streamflow simulation*, Journal of Hydrology, 279 (2003), pp. 275–289.

[256] L. PFISTER AND J. W. KIRCHNER, *Debates—Hypothesis testing in hydrology: Theory and practice*, Water Resources Research, 53 (2017), pp. 1792–1798.

[257] F. PIANOSI, F. SARRAZIN, AND T. WAGENER, *SAFE Toolbox v1.0.* https://www.safetoolbox.info/, 2015.

[258] J. PORTER AND T. MCMAHON, *Application of a catchment model in southeastern Australia*, Journal of Hydrology, 24 (1971), pp. 121–134.

[259] N. J. POTTER, L. ZHANG, P. C. D. MILLY, T. A. MCMAHON, AND A. J. JAKEMAN, *Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments*, Water Resources Research, 41 (2005).

[260] G. POVEDA, P. R. WAYLEN, AND R. S. PULWARTY, *Annual and inter-annual variability of the present climate in northern South America and southern Mesoamerica*, Palaeogeography, Palaeoclimatology, Palaeoecology, 234 (2006), pp. 3–27.

[261] C. H. B. PRIESTLEY AND R. J. TAYLOR, *On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters*, Monthly Weather Review, 100 (1972), pp. 81–92.

[262] R. PUSHPALATHA, C. PERRIN, N. L. MOINE, AND V. ANDRÉASSIAN, *A review of efficiency criteria suitable for evaluating low-flow simulations*, Journal of Hydrology, 420-421 (2012), pp. 171–182.

[263] M. M. RAHMAN, M. LU, AND K. H. KYI, *Seasonality of hydrological model spin-up time: a case study using the Xinanjiang model*, Hydrology and Earth System Sciences Discussions, (2016), pp. 1–22.

[264] G. RASUL, A. MAHMOOD, A. SADIQ, AND S. I. KHAN, *Vulnerability of the Indus Delta to Climate Change in Pakistan*, Pakistan Journal of Meteorology, 8 (2012), pp. 89–107.

[265] J. REFSGAARD AND J. KNUDSEN, *Operational Validation and Intercomparison of Different Types of Hydrological Models*, Water Resources Research, 32 (1996), pp. 2189–2202.

[266] D. REGÜÉS AND F. GALLART, *Seasonal patterns of runoff and erosion responses to simulated rainfall in a badland area in Mediterranean mountain conditions (Vallcebre, Southeastern Pyrenees)*, Earth Surface Processes and Landforms, 29 (2004), pp. 755–767.

[267] B. Renard, D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks, *Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors*, Water Resources Research, 46 (2010).

[268] J. Reumert, *Vahls climatic divisions. An explanation.*, Geografisk Tidsskrift, 48 (1946).

[269] S. R. Rusli, D. Yudianto, and J. tao Liu, *Effects of temporal variability on HBV model calibration*, Water Science and Engineering, 8 (2015), pp. 291–300.

[270] A. Sankarasubramanian and R. M. Vogel, *Annual hydroclimatology of the continental United States*, Geophysical Research Letters, 38 (2003), pp. 1–4.

[271] L. Santos, G. Thirel, and C. Perrin, *Continuous state-space representation of a bucket-type rainfall-runoff model: a case study with the GR4 model using state-space GR4 (version 1.0)*, Geoscientific Model Development, 11 (2018), pp. 1591–1605.

[272] ⸺, *Technical note: Pitfalls in using log-transformed flows within the KGE criterion*, Hydrology and Earth System Sciences, 22 (2018), pp. 4583–4591.

[273] H. H. G. Savenije, *HESS Opinions "The art of hydrology"*, Hydrology and Earth System Sciences, 13 (2009), pp. 157–161.

[274] H. H. G. Savenije, *"Topography driven conceptual modelling (FLEX-Topo)"*, Hydrology and Earth System Sciences, 14 (2010), pp. 2681–2692.

[275] K. A. Sawicz, C. Kelleher, T. Wagener, P. Troch, M. Sivapalan, and G. Carrillo, *Characterizing hydrologic change through catchment classification*, Hydrology and Earth System Sciences, 18 (2014), pp. 273–285.

[276] K. A. Sawicz, T. Wagener, M. Sivapalan, P. A. Troch, and G. Carrillo, *Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA*, Hydrology and Earth System Sciences, 15 (2011), pp. 2895–2911.

[277] M. Sayemuzzaman and M. K. Jha, *Seasonal and annual precipitation time series trend analysis in North Carolina, United States*, Atmospheric Research, 137 (2014), pp. 183–194.

[278] E. Scerri, *The periodic table: its story and its significance*, Oxford University Press, New York, 2007.

[279] B. SCHAEFLI AND H. V. GUPTA, *Do Nash values have value?*, Hydrological Processes, 21 (2007), pp. 2075–2080.

[280] B. SCHAEFLI, C. J. HARMAN, M. SIVAPALAN, AND S. J. SCHYMANSKI, *Hydrologic predictions in a changing environment : behavioral modeling*, Hydrology and Earth System Sciences, 15 (2011), pp. 635–646.

[281] B. SCHAEFLI, B. HINGRAY, M. NIGGLI, AND A. MUSY, *A conceptual glacio-hydrological model for high mountainous catchments*, Hydrology and Earth System Sciences, 9 (2005), pp. 95–109.

[282] B. SCHAEFLI, L. NICOTINA, C. IMFELD, P. DA RONCO, E. BERTUZZO, AND A. RINALDO, *SEHR-ECHO v1.0: A spatially explicit hydrologic response model for ecohydrologic applications*, Geoscientific Model Development, 7 (2014), pp. 2733–2746.

[283] U. SCHNEIDER, A. BECKER, P. FINGER, A. MEYER-CHRISTOFFER, B. RUDOLF, AND M. ZIESE, *GPCC Full Data Reanalysis Version 7.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data*, 2015.

[284] G. SCHOUPS, N. C. VAN DE GIESEN, AND H. H. SAVENIJE, *Model complexity control for hydrologic prediction*, Water Resources Research, 44 (2008).

[285] G. SCHOUPS, J. A. VRUGT, F. FENICIA, AND N. C. VAN DE GIESEN, *Corruption of accuracy and efficiency of Markov chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models*, Water Resources Research, 46 (2010).

[286] V. SCHWÄMMLE AND O. N. JENSEN, *A simple and fast method to determine the parameters for fuzzy c-means cluster analysis*, Bioinformatics, 26 (2010), pp. 2841–2848.

[287] J. SEIBERT, *Estimation of Parameter Uncertainty in the HBV Model*, Nordic Hydrology, 28 (1997), pp. 247–262.

[288] ——, *On the need for benchmarks in hydrological modelling*, Hydrological Processes, 15 (2001), pp. 1063–1064.

[289] J. SEIBERT AND J. J. MCDONNELL, *On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration*, Water Resources Research, 38 (2002), pp. 23–1–23–14.

[290] J. SEIBERT AND H. J. VAN MEERVELD, *Hydrological change modeling: Challenges and opportunities*, Hydrological Processes, 30 (2016), pp. 4966–4971.

[291] J. SEIBERT AND M. J. VIS, *Teaching hydrological modeling with a user-friendly catchment-runoff-model software package*, Hydrology and Earth System Sciences, 16 (2012), pp. 3315–3325.

[292] J. SEIBERT, M. J. P. VIS, E. LEWIS, AND H. VAN MEERVELD, *Upper and lower benchmarks in hydrological modelling*, Hydrological Processes, (2018), pp. 1120–1125.

[293] G. SEILLER, F. ANCTIL, AND C. PERRIN, *Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions*, Hydrology and Earth System Sciences, 16 (2012), pp. 1171–1189.

[294] M. SHAHIN, *Hydrology and Water Resources of Africa*, Springer Science & Business Media, Dordrecht, illustrate ed., 2006.

[295] E. SHAW, K. BEVEN, N. CHAPPELL, AND R. LAMB, *Hydrology in practice*, Spon Press, 4th ed., 2011.

[296] V. P. SINGH, *Computer Models of Watershed Hydrology*, Water Resources Publications, USA, 1995.

[297] V. P. SINGH AND D. K. FREVERT, eds., *Mathematical Models of Small Watershed Hydrology and Applications*, Water Resources Publications LLC, USA, Chelsea, Michigan, USA, 2002.

[298] V. P. SINGH AND D. A. WOOLHISER, *Mathematical Modeling of Watershed Hydrology*, Journal of Hydrologic Engineering, 7 (2002), pp. 270–292.

[299] W. T. SITTNER, *WMO project on intercomparison of conceptual models used in hydrological forecasting*, Hydrological Sciences Bulletin, 21 (1976), pp. 203–213.

[300] M. SIVAPALAN, K. BEVEN, AND E. F. WOOD, *On hydrologic similarity: 2. A scaled model of storm runoff production*, Water Resources Research, 23 (1987), pp. 2266–2278.

[301] M. SIVAPALAN, G. BLÖSCHL, L. ZHANG, AND R. VERTESSY, *Downward approach to hydrological prediction*, Hydrological Processes, 17 (2003), pp. 2101–2111.

[302] M. SIVAPALAN, J. K. RUPRECHT, AND N. R. VINEY, *Water and salt balance modelling to predict the effects of land-use changes in forested catchments. 1. Small catchment water balance model*, Hydrological Processes, 10 (1996).

[303] M. SIVAPALAN, N. R. VINEY, AND C. G. JEEVARAJ, *Water and salt balance modelling to predict the effects of land-use changes in forested catchments. 3. The large catchment model*, Hydrological Processes, 10 (1996), pp. 429–446.

[304] M. SIVAPALAN AND R. A. WOODS, *Evaluation of the effects of general circulation models' subgrid variability and patchiness of rainfall and soil moisture on land surface water balance fluxes*, Hydrological Processes, 9 (1995), pp. 697–717.

[305] J. O. SKØIEN, G. BLÖSCHL, AND A. W. WESTERN, *Characteristic space scales and timescales in hydrology*, Water Resources Research, 39 (2003).

[306] SMHI, *Integrated Hydrological Modelling System (IHMS) Manual Version 4.5*, tech. rep., Norrköping, 2004.

[307] SOIL CLASSIFICATION WORKING GROUP, *Soil classification: a taxonomic system for South Africa*, Department of Agricultural Development, Pretoria, 1991.

[308] K. H. SOLIMAN, *Rainfall over Egypt*, Quarterly Journal of the Royal Meteorological Society, 79 (1953), pp. 389–397.

[309] K. SON AND M. SIVAPALAN, *Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data*, Water Resources Research, 43 (2007).

[310] M. SUGAWARA, *Tank model*, in Computer models of watershed hydrology, V. P. Singh, ed., Water Resources Publications, USA, 1995, ch. 6, pp. 165–214.

[311] W. SUN, H. ISHIDAIRA, S. BASTOLA, AND J. YU, *Estimating daily time series of streamflow using hydrological model calibrated based on satellite observations of river water surface width: Toward real world applications*, Environmental Research, 139 (2015), pp. 36–45.

[312] B. Q. TAN AND K. M. O'CONNOR, *Application of an empirical infiltration equation in the SMAR conceptual model*, Journal of Hydrology, 185 (1996), pp. 275–295.

[313] M. TANG AND E. REITER, *Plateau Monsoons of the Northern Hemisphere: A Comparison between North America and Tibet*, Monthly Weather Review, 112 (1984), pp. 617–637.

[314] THE GLOBAL RUNOFF DATA CENTRE, *Watershed Boundaries of GRDC Stations*. https://www.bafg.de/GRDC/EN/02_srvcs/22_gslrs/222_WSB/watershedBoundaries_node.html, 2011, Accessed on 2017-07-19.

[315] ——, *Climate Sensitive Stations Dataset (Pristine River Basins)*. http://www.bafg.de/GRDC/EN/04_spcldtbss/46_CSS/css.html?nn=201574, 2017, Accessed on 2017-07-19.

[316] ——, *GRDC pristine catchments data set, 1984-2014*, 2017.

[317] ——, *Global Runoff Data Base - Status, Development, Use*. https://www.bafg.de/GRDC/EN/01_GRDC/13_dtbse/database_node.html, 2019, Accessed on 2019-01-27.

[318] C. W. THORNTHWAITE, *Problems in the Classification of Climates*, American Geographical Society, 33 (1943), pp. 233–255.

[319] ——, *An Approach toward a Rational Classification of Climate*, Geographical Review, 38 (1948), pp. 55–94.

[320] G. TREWARTHA AND L. HORN, *An introduction to climate 5th edition*, McGraw-Hill, New York, NY, USA, 1968.

[321] H. J. TROMP-VAN MEERVELD AND J. J. MCDONNELL, *Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope*, Water Resources Research, 42 (2006).

[322] U.S. GEOLOGICAL SURVEY, *What is a Water Year?* https://water.usgs.gov/nwc/explain_data.html, 2016, Accessed on 2017-09-15.

[323] US GEOLOGICAL SURVEY, *National Water Information System: Web Interface*. https://nwis.waterdata.usgs.gov/nwis/dv, 2018, Accessed on 2018-01-29.

[324] UTAH STATE UNIVERSITY, *Water year*. https://climate.usurf.usu.edu/reports/waterYear.php, 2017, Accessed on 2017-05-24.

[325] W. R. VAN ESSE, C. PERRIN, M. J. BOOIJ, D. C. M. AUGUSTIJN, F. FENICIA, D. KAVETSKI, AND F. LOBLIGEOIS, *The influence of conceptual model structure on model performance: A comparative study for 237 French catchments*, Hydrology and Earth System Sciences, 17 (2013), pp. 4227–4239.

[326] K. VAN WERKHOVEN, T. WAGENER, P. REED, AND Y. TANG, *Characterization of watershed model behavior across a hydroclimatic gradient*, Water Resources Research, 44 (2008).

[327] L. A. VIERECK, C. T. DYRNESS, A. R. BATTEN, AND K. WENZLICK, *The Alaska Vegetation Classification*, tech. rep., United States Department of Agriculutre, 1992.

[328] Y. B. VINOGRADOV, O. M. SEMENOVA, AND T. A. VINOGRADOVA, *An approach to the scaling problem in hydrological modelling: The deterministic modelling hydrological system*, Hydrological Processes, 25 (2011), pp. 1055–1073.

[329] R. M. VOGEL AND A. SANKARASUBRAMANIAN, *Validation of a watershed model without calibration*, Water Resources Research, 39 (2003).

[330] T. WAGENER, D. P. BOYLE, M. J. LEES, H. S. WHEATER, V. GUPTA, HOSHIN, AND S. SOROOSHIAN, *A framework for development and application of hydrological models*, Hydrology and Earth System Sciences, 5 (2001), pp. 13–26.

[331] T. WAGENER, M. J. LEES, AND H. S. WHEATER, *A toolkit for the development and application of parsimonious hydrological models*, in Mathematical Models of Small Watershed Hydrology - Volume 2, Singh, Frevert, and Meyer, eds., Water Resources Publications LLC, USA, 2002, pp. 91–139.

[332] T. WAGENER, N. MCINTYRE, M. J. LEES, H. S. WHEATER, AND H. V. GUPTA, *Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis*, Hydrological Processes, 17 (2003), pp. 455–476.

[333] T. WAGENER, M. SIVAPALAN, P. TROCH, B. MCGLYNN, C. HARMAN, H. GUPTA, KUMAR, P. RAO, N. BASU, AND J. WILSON, *The future of hydrology: An evolving science for a changing world*, Water Resources Research, 46 (2010).

[334] T. WAGENER, M. SIVAPALAN, P. TROCH, AND R. WOODS, *Catchment Classification and Hydrologic Similarity*, Geography Compass, 1 (2007), pp. 901–931.

[335] T. WAGENER, H. S. WHEATER, AND H. V. GUPTA, *Rainfall-Runoff Modelling in Gauged and Ungauged Catchments*, Imperial College Press, 2004.

[336] R. E. WALPOLE, *Introduction to Statistics*, The MacMillan Company, New York, NY, USA, 1968.

[337] M. WEILER AND K. BEVEN, *Do we need a community hydrological model?*, Water Resources Research, 51 (2015).

[338] I. WESTERBERG, A. WALTHER, J. GUERRERO, Z. COELLO, S. HALLDIN, C. XU, D. CHEN, AND L. LUNDIN, *Precipitation data in a mountainous catchment in Honduras: Quality assessment and spatiotemporal characteristics*, Theoretical and Applied Climatology, 101 (2010), pp. 381–396.

[339] I. K. WESTERBERG AND H. K. MCMILLAN, *Uncertainty in hydrological signatures*, Hydrology and Earth System Sciences, 19 (2015), pp. 3951–3968.

[340] I. K. WESTERBERG, T. WAGENER, G. COXON, H. K. MCMILLAN, A. CASTELLARIN, A. MONTANARI, AND J. FREER, *Uncertainty in hydrological signatures for gauged and ungauged catchments*, Water Resources Research, 52 (2016), pp. 1847–1865.

[341] WIKIPEDIA, *Category:Hydrology models*.
https://en.wikipedia.org/wiki/Category:Hydrology_models, 2019, Accessed on 2019-01-27.

[342] F. WILCOXON, *Individual Comparisons by Ranking Methods*, Biometrics Bulletin, 1 (1945), p. 80.

[343] C. J. WILLMOTT AND J. J. FEDDEMA, *A more Rational Climatic Moisture Index*, The Professional Geographer, 44 (1992), pp. 84–87.

[344] C. L. WINTER AND D. NYCHKA, *Forecasting skill of model averages*, Stochastic Environmental Research and Risk Assessment, 24 (2010), pp. 633–638.

[345] E. F. WOOD, J. A. N. BOLL, P. BOGAART, AND P. TROCH, *The Need for a Virtual Hydrologic Laboratory for PUB*, in Predictions in Ungauged Basins: International Perspectives on State-of-the-Art and Pathways Forward, Proceedings of the Australia-Japan Workshop on Pub Working Groups, F. S. W., S. M., T. K., and T. Y., eds., IAHS Press: Wallingford, 2005.

[346] R. A. WOODS, *The relative roles of climate, soil, vegetation and topography in determining seasonal and long-term catchment dynamics*, Advances in Water Resources, 26 (2003), pp. 295–309.

[347] ———, *Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks*, Advances in Water Resources, 32 (2009), pp. 1465–1481.

[348] M. YADAV, T. WAGENER, AND H. GUPTA, *Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins*, Advances in Water Resources, 30 (2007), pp. 1756–1774.

[349] S. YE, M. YAEGER, E. COOPERSMITH, L. CHENG, AND M. SIVAPALAN, *Exploring the physical controls of regional patterns of flow duration curves - Part 2: Role of seasonality, the regime curve, and associated process controls*, Hydrology and Earth System Sciences, 16 (2012), pp. 4447–4465.

[350] W. YE, B. C. BATES, N. R. VINEY, AND M. SIVAPALAN, *Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments*, Water Resources Research, 33 (1997), pp. 153–166.

[351] R.-J. ZHAO, *The Xinanjiang model applied in China*, Journal of Hydrology, 135 (1992), pp. 371–381.