From the Department of Landscape, Water and Biogeochemical Cycles, Faculty of Agricultural Sciences, Nutritional Sciences, and Environmental Management of the Justus-Liebig University Gießen

A dissertation to obtain the degree of Doctor of Natural Sciences (Dr. rer. nat.)

Inferring hydrological process understanding using models and large-sample data sets

Florian Ulrich Jehn

Referees from the Justus-Liebig University Gießen:

Prof. Dr. Lutz Breuer (1st Supervisor)

Prof. Dr. Jan Siemens (2nd Supervisor)

Submitted: 10th February 2021

Abstract

Catchments are complex systems, which have evolved under the influence of environmental processes over long periods of time. Due to this inherent complexity, it is often difficult to understand what forces the hydrological behavior of a given catchment. The two most common approaches to better understand catchments are creating hydrological models to test which hypothesis of catchment functioning works best or to look at the catchment characteristics and try to infer the most important forcing directly from this. This dissertation uses both approaches to reach a more holistic understanding of catchment functioning.

The starting point of this dissertation was an earlier publication of mine, which used an innovative way for model development, the so called "incremental model breakdown". This new approach starts with a complex model, incrementally deactivates processes and checks which process deactivation causes the model to fail. All processes that lead to a failure when deactivated show that they are important for the model. This enables a more thorough exploration of the space of possible model structures than traditional approaches, which start from predefined structures. However, during the development of this approach it became apparent that model parameters are able to compensate extensively for the omission of processes. Therefore, larger and well-understood data sets are needed to form hypotheses of catchment functioning that could be tested by incremental model breakdown. Based on this prior knowledge and to lay the foundation for future research, this dissertation builds on the incremental model breakdown approach and examines two large sample data sets with different methods.

As the main problem of the new model building approach was the way it handles the model parameters, the first part of this dissertation focusses on the intricate interaction between model complexity and parameter uncertainty. This is done by exploring the trade-offs between tightly constrained parameters and the ability of the hydrological models to predict hydrological signatures that capture the behavior of a river. The results show that there is a clear trade-off along the axis of complexity for those models. The simpler a model is, the better its ability to constrain parameters, but the worse are the results of an independent validation of its realism using hydrological signatures.

Those results highlighted again that hydrological models can only be as good as the hypothesis forming their basis and those hypotheses can only be found and improved by looking at real catchments' data. These datasets need to contain the hydrological behavior and characteristics of catchments to facilitate deriving hydrological process understanding – and develop appropriate models that reflect this catchment's understanding. Therefore, in the second part of my thesis, is about the exploration of two large hydrological datasets. The first step was an analysis of the CAMELS dataset, a large-scale open access dataset that contains catchments from all over the continental United States. This allows to determine the most important factor for the overall hydrological behavior, namely the climate, and more specifically, the aridity and the frequency of large precipitation events. However, the results also show that

this climatic forcing can be found more directly in some catchments than in others. This was likely a problem of scale, given the continental domain of this study.

To better understand why this is the case, we established a second dataset which only contained catchments from Hesse, Germany, for the third part of my work. This allowed looking at how catchments with different characteristics behave under the same climatic forcing. The focus here was the complexity of the storage-discharge relationship. The results showed that the hydrological signal of the climatic forcing is mainly influenced by the catchment's permeability, conductivity, geology, soil and, to a lesser extent, its topography. It also showed that the complexity of the hydrological response differs strongly between catchments. While some catchments show a storage-discharge relationship that is almost exactly an exponential function, others show a more erratic behavior. The properties of the simple catchments all facilitate a higher interconnectedness of the storage system of the catchment; this indicates that the complexity of a catchment's behavior is strongly linked to its overall connectivity of water pathways.

To finally use this improved hydrological understanding and connect it to model structures, preliminary tests link the catchment complexity with modelling ease. For this, I used the simple HYMOD model and evaluated its performance for catchments of differing complexity. Those results showed that the simplicity in behavior is connected to the model performance. The simpler the storage-discharge relationship, the simpler the model for the catchment can be.

The findings of this dissertation highlight that even though it is possible to change model structures and calibrate parameters to get results with high values for the objective function, those models can still have difficulties in independent verification. This indicates that the models are often "right for the wrong reasons". Only if we thoroughly understand datasets that capture a wide variety of hydrological behavior and catchments characteristics will we be able to construct more realistic hydrological models that are "right for the right reasons".

Table of Contents

Extended Summary

Introduction

The challenges of hydrological model construction

Models have been used in hydrology longer than in most other scientific fields. Since the early beginnings in the 19th century, their concepts have undergone some drastic changes (Todini, 2007). The first widely used hydrological model was developed by Thomas James Mulvaney and only consisted of one single equation (Mulvaney, 1851). Over time, many different approaches to forecast the runoff were tested. For example splitting the catchment into different zones, divided by the time the water needed to get from the zone to the outlet of the catchment (Turner and Burdoin, 1941) or the well-known unit hydrograph (Sherman, 1932). Building on insights of those early experiments the focus in the 1960s went to the idea of representing the hydrological cycle as combination of connected, conceptual elements (Todini, 2007). An example for one of those early conceptual models is the Tank model (Sugawara, 1967). As those models have relatively large parameter sets, the concept of model calibration was introduced to find optimal values for every parameter (Dawdy and O'Donnel, 1965). Since the introduction of conceptual models in combination with efficient calibration routines, vast amounts of new models have been developed. However, there has not been a great improvement in model predictive performance since the introduction of the TOPMODEL (Buytaert et al., 2008).

As is nowadays often agreed, simply building more models does not necessarily result in finding better ones. This is partly because we still have to face many, and sometimes large sources of predictive uncertainty in our models. Four main sources of uncertainty exist, i. e. structural, stochastic and input uncertainty and finally the subjectivity of the modellers themselves (Renard et al., 2010). While the first one touches the design of a model, its philosophy and processes represented, the stochastic uncertainty comprises many different aspects including the values for the model parameters, uncertainties in spatial inputs (e. g. soil and land use maps) and validation data (e. g. errors in discharge measurements).

During the historical development of hydrological models the issues of parameter and input data uncertainty were addressed first, while structural uncertainty was only approached in the past decade (Breuer et al., 2009; Son and Sivapalan, 2007) and gained more momentum in the last few years (Clark et al., 2015; Fenicia et al., 2011; Hublart et al., 2015). Till today, hydrologists still debate whether a hydrological model should contain all known hydrological processes (and can be used as a universal model), or whether it should only contain the prevailing processes to generate a parsimonious model. However, it was also noted that problems arise when trying to build one model that is meant to work equally well for all catchments (Fenicia et al., 2011). Therefore, models were introduced that have interchangeable elements. First approaches used existing model structures like TOPMODEL, and varied the representation of processes such as hydraulic conductivity, channel routing (Wang et al., 2005) or evapotranspiration (Andréassian et al., 2004b). This way of model testing yielded some insight to why models work or not, but was restrained by problems with comparability and the lack of a clear theoretical background. Those insights resulted in the application of the theory of multiple hypotheses in hydrology (Clark et al., 2011). This theory enabled a more structured approach to model building, as it identified a given model not as a single hypothesis, but as an assemblage of coupled hypotheses. Clark et al. (2011) proposed that a model should be constructed in a way that allows the testing of every single hypothesis of every sub-process separately.

One additional problem in the exploration of uncertainty is that often models are compared without an underlying framework. This hinders comparisons, as it is not easy to determine if the models themselves or the way they are implemented and calibrated cause the change of predictive capability (Breuer et al., 2009). This was also noted by other authors like Andréassian et al. (2004b). They state the many comparisons of lumped and semi distributed approaches are hindered by differences in the models. This is something which can be avoided when a fixed modelling framework is used, which standardizes all steps of model structure development. Therefore, this study uses the Catchment Modelling Framework (CMF) (Kraft et al., 2011) to build and deploy all models. CMF was used as it is more flexible than other modelling frameworks, which mainly use large building blocks (Clark et al., 2008; Craig and Raven Development Team, 2018). In such a framework, all setscrews like data handling or the solving of the differential equations are handled the same way and so the differences in performance between models can only be caused by the models themselves. Modelling frameworks made of easy interchangeable parts for each process allow to split the underlying hypotheses network of models into testable units. Examples of such frameworks are hydrological model kits like the Framework for Understanding Structural Errors (FUSE) (Clark et al., 2008), SUPERFLEX (Fenicia et al., 2011), RAVEN (Craig and Raven Development Team, 2018), and the Catchment Modelling Framework (CMF) (Kraft et al., 2011). Model frameworks enable a stepwise modification of the model structure. Nevertheless, even with those model kits, the process of model building is time consuming and subjective. Therefore, most studies still consider only a few different model structures (van Esse et al., 2013; Fenicia et al., 2014; Gharari et al., 2014). Despite having the potential to create a wide range of models using model frameworks, we still explore only a tiny quantity in the vast space of possible model structures.

Two pieces in the puzzle of uncertainty, which have been identified as important are model structure and the approach to estimate the potential evapotranspiration (PET) (Clark et al., 2011; Orth et al., 2015). In both cases, it seems that it is not finally settled which approaches yield the best results. In the case of model structure it has become common to compare a range of models of different complexity (Fenicia et al., 2008; Gao et al., 2014; Lobligeois et al., 2014; Orth et al., 2015). Studies looking at the influence of model structure come to vastly different results. Some find that there is no difference between lumped and semi distributed models (Lobligeois et al., 2014), the lumped ones (Andréassian et al., 2004b) perform better or the (semi) distributed perform better (Patil et al., 2014). Nevertheless, all agree that the performance of the model type is linked to the quality of the forcing data and the internal structure of the catchment itself. For the PET only few studies focus on the influence of the PET on model performance and uncertainty. Most studies only compare different calculation of PET to measured values and do not test them within a model structure, e.g. (Lu et al., 2005; Xu and Singh, 2002). Seiller and Anctil (2016) tested 24 PET methods and their influence on hydrological response, found all lacking and recommend avoiding temperature based ones, while other studies state that simple temperature based ones are sufficient for hydrological modelling (Kannan et al., 2007; Oudin et al., 2005). This question of how model structure, model parameters and the method of the potential evapotranspiration interact is addressed in chapter 1 (Jehn et al., 2019).

Preliminary work

The work presented here is based on my master thesis that focussed on a new way of building hydrological model structures (Jehn et al., 2018). This preliminary work was the starting point of my PhD. As stated above the multi-hypotheses approach in hydrology (Clark et al., 2011) called for a new way to construct and evaluate models within modelling frameworks. In the last years, this lead to some interesting studies, which built their models incrementally to find out if small modifications of the model structure allow for a better simulation (Bai et al., 2009; Westerberg and Birkel, 2015). In contrast, other studies compared predefined model structures within a single framework (van Esse et al., 2013; Kavetski and Fenicia, 2011). In all cases, researchers stopped improving the models once a sufficient performance was reached. However, there is a chance that they have missed an even better model performance by including further modifications. To avoid this, my master thesis tackled the problem the other way around. I started with one complex model structure (Fig. 1, left), which included all processes deemed to be important for the catchment.

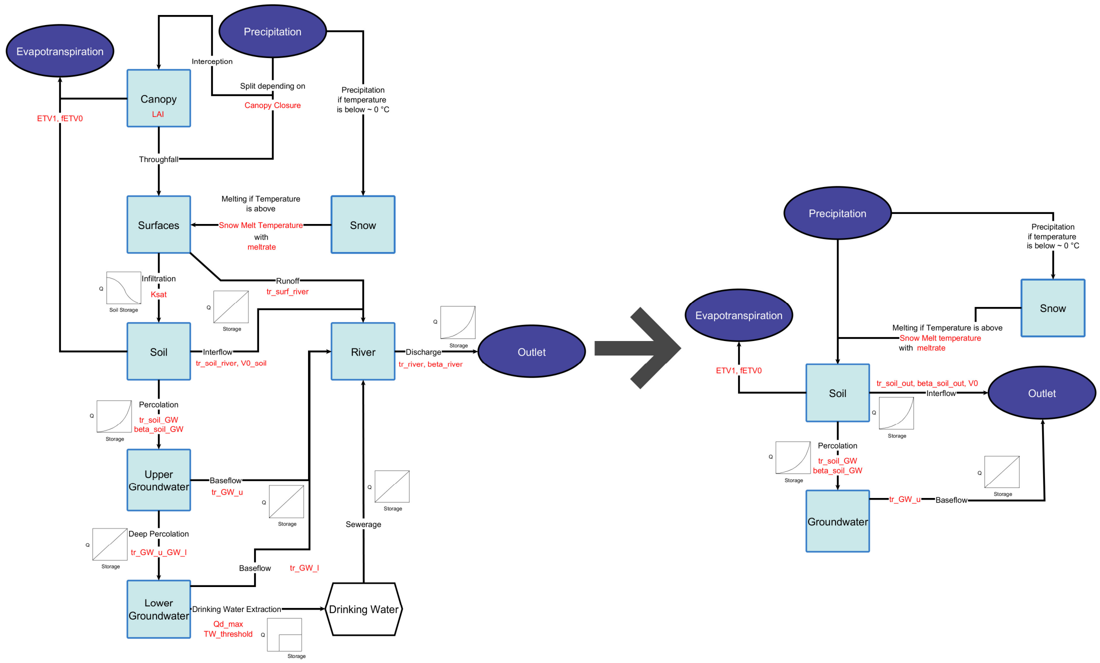

Figure 1*: First complex model (left) and final (15th) model (right). The first model included all processes which were deemed important for the catchment, while the final model only contains those processes which have proven to be important for the catchment.*

The next step was to create 13 additional simplified models, where some of the processes from the starting model structure were disabled. The performances of the models were evaluated using four objective functions (logarithmic Nash-Sutcliffe, Nash-Sutcliffe, percentage bias and the ratio between root mean square error to the standard deviation of the measured data). This allowed to identify the most important processes model, as the idea was that models lacking important processes fail. This allowed constructing a more streamlined subsequent 15th model (Fig. 1, right) with improved model performance and reduced uncertainty. Benchmarking the original Model 1 with the final Model 15 reveals that the *incremental model breakdown* led to a structure with good performance, while having fewer processes and parameters. Overall, this method of incremental model breakdown enables researchers to scrutinize existing models and to improve their structure to capture all relevant environmental processes (for schematic description see Fig. 2). This new approach also led to the development of the so called Fluxogram - a dynamic visualisation of the flow processes in hydrological models (https://youtu.be/mvwUz3pRlgA). For a more detailed description of the preliminary work, see Jehn et al. (2018).

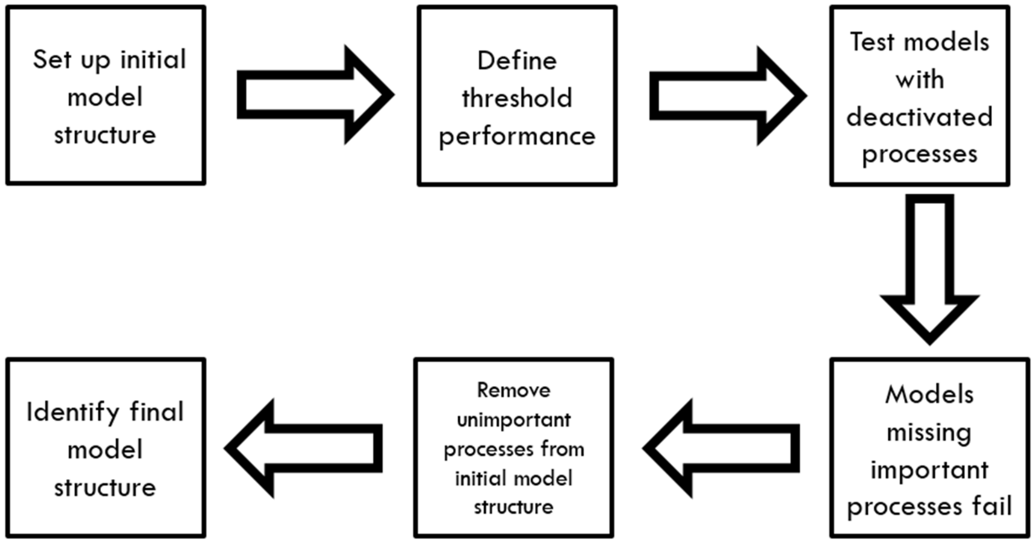

***Figure 2**: Schematic description of incremental model breakdown.*

Hydrology on different scales

Every catchment has a unique behavior shaped by its climate, topography and geology. However, the hydrological community still needs to find ways to generalize and understand the behavior of catchments in order to be able to create appropriate hydrological models (Beven, 2000). On way to approach this problem is catchment classification, with which shared behavior and characteristics can be identified (Sivapalan, 2003). First iterations of this approach used geographic, administrative or physiographic considerations, but this proved to be too simplistic (Burn, 1997). Another natural choice for hydrological classification are seasonality measures, but those are hard to obtain (Burn, 1997). This changed in recent years as a series of large sample and large scale datasets became available (Addor et al., 2017; Alvarez-Garreton et al., 2018; Newman et al., 2014; Schaake et al., 2006), which allowed to gain new insights that were out of reach in small sample studies (Gupta et al., 2014). Those new datasets were used to find similarities in flow duration curves (Coopersmith et al., 2012), hydro climate (Potter et al., 2005) or hydrological signatures (Kuentz et al., 2017a). Large scale studies attribute hydrological differences mostly to the climate (Kuentz et al., 2017a; Sawicz et al., 2014) and this connection can also be found when we use climate to predict hydrological behavior (Knoben et al., 2018). However, the studies conducted on a smaller scale usually find a higher influence of catchment attributes besides climate (Ali et al., 2012; Trancoso et al., 2017). This is especially the case when we look at the hillslope scale (Loritz et al., 2018; Tromp-van Meerveld and McDonnell, 2006). On this scale, hydrological behavior is strongly influenced by the geology and soils. Those insights led to the formulation of the fill and spill hypotheses, which sees a catchment's behavior mainly in the light of interconnectedness and the amount of water stored in it (Tromp-van Meerveld and McDonnell, 2006). Overall, the description of the storage dynamics have received more attention in the last few years that range from thermodynamic approaches (Loritz et al., 2019) and mathematical approaches

(Kirchner, 2009) to the comparative study of several catchments (Buttle, 2016; Spence, 2010). However, all these studies on storage only used few catchments (Cheng et al., 2017; Creutzfeldt et al., 2014; Geris et al., 2015; Staudinger et al., 2017) thus making it hard to generalize from them. Therefore, further research is needed to explore the connection between storage-discharge dynamics and weather conditions (Loritz et al., 2019) in order to determine how the climatic forcing is shaped by catchment attributes and how this changes on different scales. The connection between hydrological behavior and catchments characteristics is explored on the continental scale in chapter 2 (Jehn et al., 2020) and on the mesoscale in chapter 3 (Jehn et al., 2021).

Objectives

As the literature review and the preliminary work showed, there is still a lack in deep understanding of hydrological processes and how they relate to the structure of hydrological models. Therefore, the main goal of this study is to infer a deeper process understanding by using hydrological models and large datasets.

Building on the preliminary work in incremental model breakdown, parameter uncertainty showed itself as one of the main problem in building hydrological model structures. Therefore, the first aim is to understand the interactions and trade-offs between model parameters and model structures.

As models are only as good as their underlying data, the next aim of this dissertation is to use large datasets to find the underlying processes of hydrological behavior. As the importance of different processes changes depending on the scale, two separate datasets were used: the CAMELS dataset to explore large scale effects and the Hesse dataset for small scale effects.

Finally, this study aims to relate those findings of catchment processes and complexity back to hydrological model structures by exploring who catchment complexity relates to model complexity.

Tools and Datasets

Catchment Modelling Framework (CMF)

All models in this dissertation were constructed with the open source modular Catchment Modelling Framework (CMF) (Kraft et al., 2011). CMF consists of a large selection of possible model parts like different routing routines, storages and models for the canopy and the evapotranspiration. These building blocks can be put together in flexible ways, which allows to construct new model structures that are suited for the needs at hand. The spatial resolution reaches from 1d soil profiles (Djabelkhir et al., 2017) to fully distributed hydrological models (Maier et al., 2017). It is also possible to link CMF with other existing models and combine it with calibration tools like SPOTPY (Houska et al., 2017). Additional information can be found at the framework's website (CMF, 2018). To avoid numerical problems (Kavetski and Clark, 2011), the CVode Integrator (Hindmarsh et al., 2005) was used as the numerical solver. Following the findings of Singh (Singh, 2002), all connections in the model (Fig. 2) are described as kinematic waves (Eq. 1):

$$Q = Q_0 \left(\frac{V - V_{residual}}{V_0}\right)^{\beta} \qquad \text{(Eq. 1)}$$

where Q *[m³]* is the amount of water transferred from one storage to the other, $V_{residual}$ [m³] is the volume of water remaining in the storage at each time step, V_0 [m³] is the reference volume (calibrated) to scale the exponent, V is the current volume of water in the storage [m³] at each time step, and β is a parameter to shape the response curve [-]. Q_0 is the flux in [m³ d^{-1}], when $\frac{V - V_{residual}}{V_0} = 1$.

In the preliminary work, CMF was used to build a single complex model, which was then scrutinized with the incremental model breakdown method. It was also used in chapter 1 to construct three lumped models and two semi-lumped models. The code for those models can be found in the repositories for the papers (Jehn, 2017, 2018a).

HYdrological MODel (HYMOD)

While the models created with CMF can be adapted well to the characteristics of a single catchment, this also means that the resulting model structure is often uniquely tailored to a specific catchment and only used a single time. This can make comparing the results of CMF models with more commonly used models with a fixed structure difficult. Therefore, the already established HYMOD model (Quan et al., 2015) was used for the evaluation of the connection between the catchment complexity and modelling ease. HYMOD is a simple lumped, conceptual model with five parameters. It simulates a soil moisture routine, which drains in either a single slow release reservoir or three quick release reservoirs connected in series. For an detailed description, see Quan et al. (2015).

Hydrological signatures

There are many ways to assess the behavior of catchments. Most commonly used are objective functions like the Nash-Sutcliffe Efficiency (Nash and Sutcliffe, 1970) or the Kling Gupta Efficiency (Gupta et al., 2009) that simply compare the observed hydrograph with the simulated model values. However, in many cases it is also important to get a better understanding of the overall behavior of the river, like how often it floods or the seasonality of its discharge. Those so called hydrological signatures give a quick overview of the overall behavior of a catchment. Over the years, a wide variety of hydrological signatures has been developed, but recent research has shown that some of them have a large uncertainty and only have dubious hydrological value. Therefore only those hydrological signatures were used that have shown to be reliable descriptors of catchment behavior (Addor et al., 2018; Westerberg and McMillan, 2015). For a list of the hydrological signatures used, see chapter 1 and 2.

CAMELS dataset

Part of this work uses the CAMELS dataset (Addor et al., 2017; Newman et al., 2014). This dataset contains 671 catchments from a wide range of environments in the continental United States with data about atmospheric forcing, discharge and a selection of catchment attributes that are likely to have an influence on hydrological behavior. This includes attributes from vegetation, climate, geology, soil and topography. In addition, 15 hydrological signatures were added to characterize hydrograph features. As the dataset contains catchments from all over the continental United States, it allows to compare very different climates (Addor et al., 2017) and covers most of the ecoregions found on earth (Omernik and Griffith, 2014). In addition, the topography of the United States has a distinct boundary around the 100th meridian. This border marks the difference between the flat, humid east and the mountainous, arid west and allows comparing those quite different hydrological regimes. The CAMELS dataset has been used in a wide set of different studies, which enables building on existing research and opens many opportunities to discuss the results (e.g. Addor et al., 2018, Gauch et al., 2020; Tyralis et al., 2019).

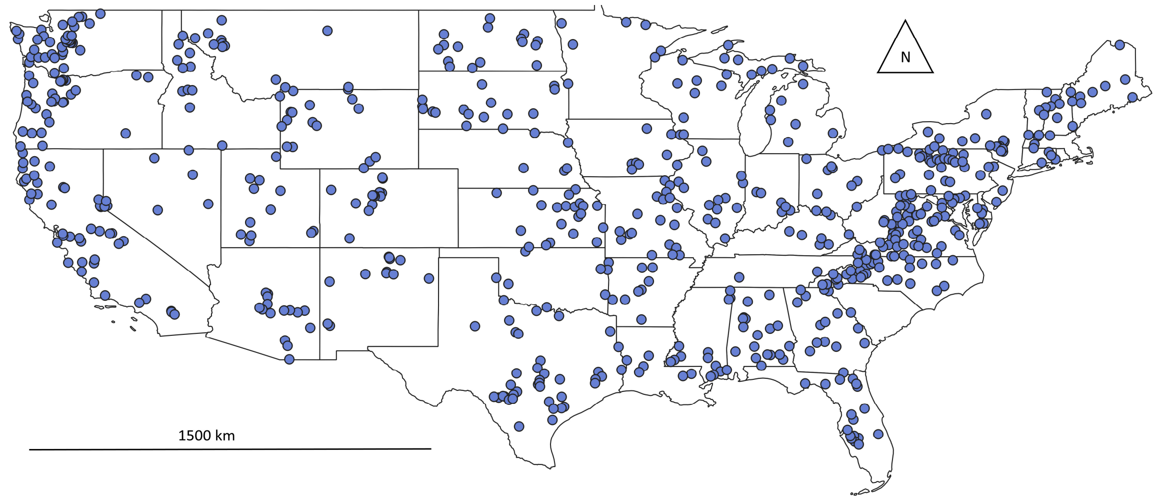

Figure 3: *Location of the CAMELS catchments in the continental United States.*

Hesse dataset

The CAMELS dataset allowed to identify the influence of different catchment attributes on the hydrological behavior on a large scale. However, to get a deeper understanding of hydrological processes it is also important to know how the climatic signal is shaped by the catchment itself. Studying this is only possible in a diverse and large set of catchments with a uniform climate. Therefore, data for 88 catchments in Hesse (Fig. 3) was collected. Through its mesoscale, the Hesse dataset is a good companion dataset to CAMELS. Both datasets contain similar catchment characteristics, but allow a different look at the hydrological processes through their different scale. Hesse has a very diverse geology and land use, but a relatively uniform climate. This dataset included discharge, precipitation, evapotranspiration, soils, geology, topography and climate. Catchment sizes range from small scale (6 km²) to mesoscale (2,793 km²). All hydrological variables show a wide range. For example the runoff-ratio ranges from 0.16 to 0.61 and the discharge from 101 mm to 670 mm per year. As part of the Hesse dataset, the Fulda catchment in the Northeast of Hesse has been used in the first part of my thesis as a test case.

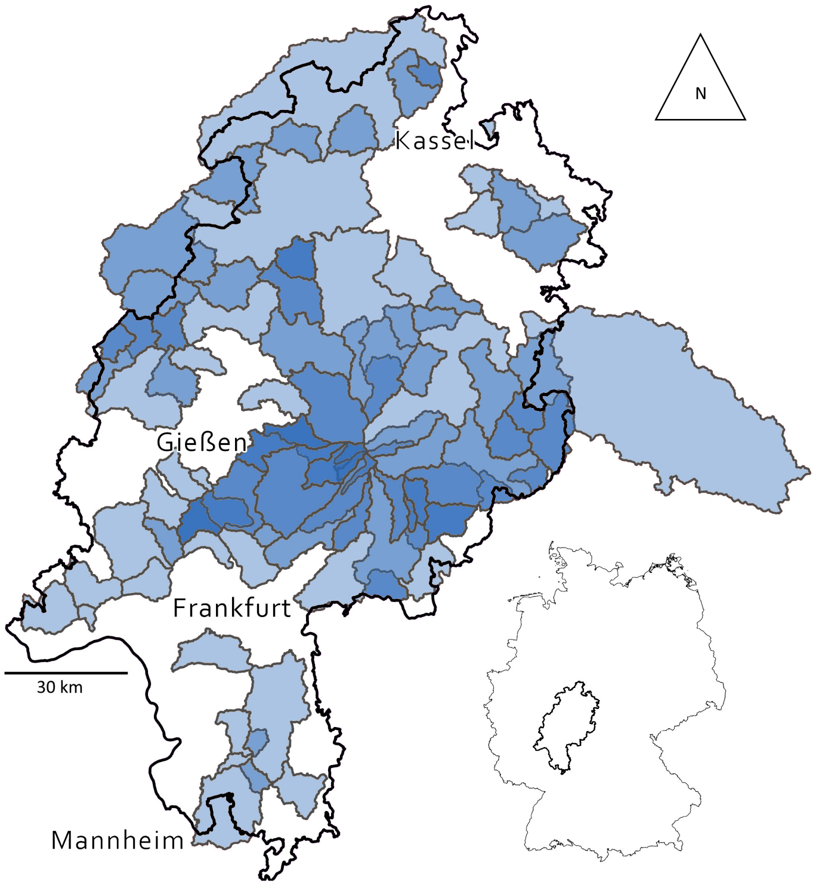

***Figure 4**: Location of the catchments in Hesse. Darker blue indicates nested catchments.*

Results and Discussion

Understanding the trade-offs between stable parameters and realistic model outputs

As preliminary work by Jehn et al. (2018) has shown, the complexity of model structures can be reduced significantly by removing all unnecessary hydro-meteorological processes to sufficiently simulate discharge. The next step was to address the interaction between model structure complexity and model parameters. This is described in detail in chapter I:

Jehn, F. U., Chamorro, A., Houska, T. and Breuer, L.: Trade-offs between parameter constraints and model realism: a case study, Sci Rep, 9(1), 10729, doi:10.1038/s41598-019-46963-6, 2019.

To create a hydrological model that works well, many things have to be done right. Apart from the model structure itself the model parameters are very important, as they allow the model to be calibrated to the data. Between model structure and model parameters exists a trade-off. The more complex the model becomes, the more parameters it needs, but it will also be able to give more realistic results as hydrological processes are depicted more accurately (Boyle et al., 2001; Reed et al., 2004). However, more parameters mean more degrees of freedom and thus a higher difficulty in constraining the parameters. This leads to new problems, as more constrained parameters are also seen as a sign of model realism (Beven, 2008a). In order to explore this conundrum of model parameters and model structure the Catchment Modelling Framework was used to build a set of five model structures with increasing complexity (lumped to semi-lumped). As the correct simulation of evapotranspiration is a further source of uncertainty, all models were run in two versions: one with the Hargreaves and one with Penman-Monteith approach to estimate potential evapotranspiration. Getting the evapotranspiration right is very important to be able to close the water balance of a catchment (Beven, 2006b). While it is already difficult to calculate the potential evapotranspiration, it is even more complex to conclude the actual evapotranspiration from it, because this requires knowing how much water is available in the catchment. Due to this complexity there exist many ways to calculate the evapotranspiration (Seiller and Anctil, 2016). Hargreaves and Penman-Monteith are used here as they are well established and highlight two different approaches to tackle the problem of evapotranspiration. The 10 models were run with data from the Fulda catchment (Hesse, Germany), an upland catchment with diverse land use. The models were calibrated with the ROPE algorithm (implemented in SPOTPY (Houska et al., 2015)) and the Kling-Gupta Efficiency as the objective function. To test the realism of the models, the best 1000 parameter sets for each model were used to simulate a set of hydrological signatures on which the models were not calibrated. This approach allows to test how good the model parameters can be constrained, how realistic their results are and the interaction between those two aspects. The results show that the simpler models had easier constrainable parameters and an overall more consistent behavior, while the more complex models show a wider plausible range for their parameter values. It is the other way around for the hydrological signatures (Fig. 5). The simpler models have difficulties in getting the hydrological signatures right, especially concerning flood frequency and duration. That indicates that the simple models are "right for the wrong reasons" (Kirchner, 2006): While they are able to get the overall shape of the hydrograph right, they fail to reproduce the hydrological signatures. Interestingly, the selected approach to simulate potential evapotranspiration did not influence the results by much (Fig. 5), even though evapotranspiration is an important factor for hydrology (Orth et al., 2015). This shows that the influence of model structure and model parameters has a larger effect on the results than the potential evapotranspiration. One possible explanation is that the larger number of parameters in the more complex models allow the models to find more realistic parameter combinations. This hints that the simpler models (like Lumped 1 in this study) do not show reality, but merely hide the uncertainties inherent in the data. Hence, models should include additional data like landscape related process heterogeneity - land cover if possible, as it allows for a more realistic prediction without hiding uncertainties. Overall, the results show that the simpler models constrain their parameters better, but their simple structure does not allow them to create realistic simulations.

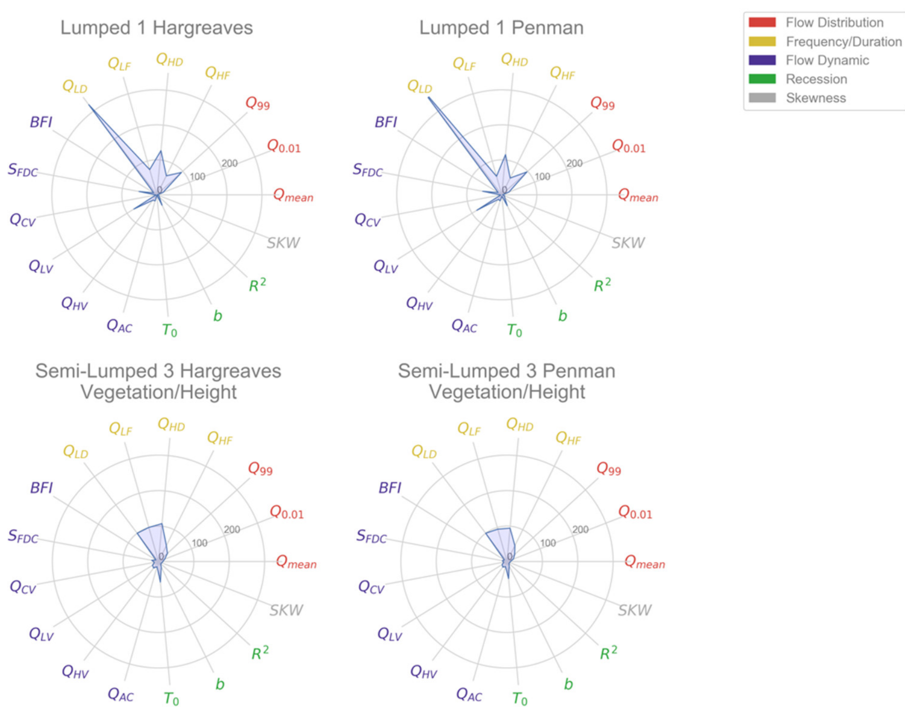

***Figure 5**: Radar plots for the simplest and most complex model structure and how well they are able to simulate hydrological signatures on which they are not calibrated (in depth description of the signatures in chapter 1). Larger blue areas indicate a larger error in comparison with the real hydrological signatures of the Fulda river.*

It also became apparent that an adequate model structure can only be built when the catchments at hand were understood well. Therefore, the next step was to gain a deep understanding of a large group of catchments and their characteristics.

Influence of climate and catchment characteristics on different scales

The first contribution to this dissertation has shown the importance of model parameters but could be improved upon by using a larger sample of catchments. However, it is difficult to find large samples of catchments that are clustered based on their behavior. Therefore, the second contribution of this dissertation tackles this problem by finding clusters of similar hydrological behavior in the CAMELS dataset, as described in chapter II:

Jehn, F. U., Bestian, K., Breuer, L., Kraft, P. and Houska, T.: Using hydrological and climatic catchment clusters to explore drivers of catchment behavior, Hydrol. Earth Syst. Sci., 24(3), 1081–1100, doi:10.5194/hess-24-1081-2020, 2020.

Hydrological signatures allow to concentrate the behavior of a river into a few key indicators. However, not all hydrological signatures have the same validity (Addor et al., 2018; Westerberg and McMillan, 2015). Therefore, I only used the six hydrological signatures with the highest predictability in space (Addor et al., 2018) to cluster the CAMELS dataset. As the hydrological signatures contained redundant information, I used principal component analysis to extract the main differences in the catchments. When combining the significant principal components with the catchment attributes like aridity, we can see clear differences in how the catchment attributes distribute themselves in the principal component space (Fig. 6).

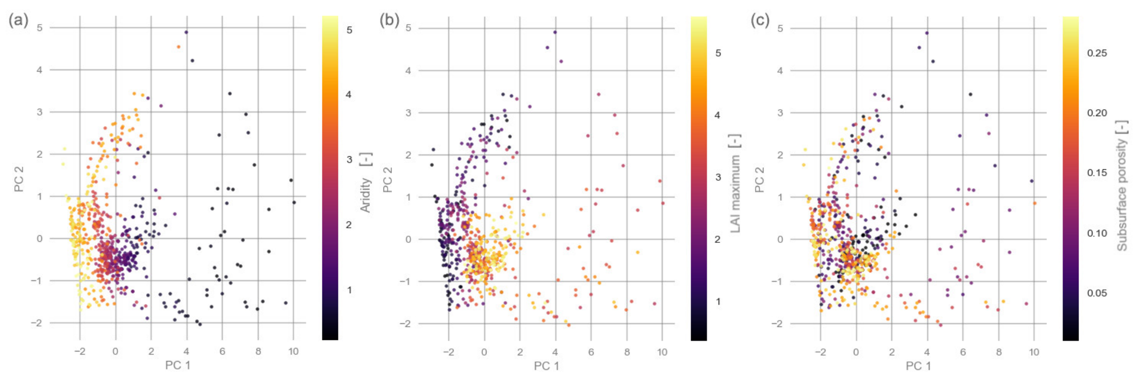

Figure 6: *Patterns of catchment attributes in the PCA space of the hydrological signatures, with decreasing strength of the observed pattern from left (aridity) to right (subsurface porosity).*

Clustering the principal component space using agglomerative hierarchical clustering with ward linkage (Ward, 1963) results in 10 distinct clusters of catchments. Those clusters roughly follow the ecoregions of the United States and had distinct differences in their catchment attributes and hydrological behavior (for an interactive map of the catchment see: https://zutn.github.io/Catchment-Classification/map.html).

Aridity in particular and climate in general are the most important drivers for catchment behavior for the whole of the United States. However, the results also indicate that the catchment clusters themselves show a more diverse behavior. Climate is more important in the clusters of the eastern United States. In the western United States, the influence of other catchment attributes increases, as it has a more diverse topography, which leads to a more diverse set of soils and geology. In addition, the results also showed that rivers with very different catchment attributes could have a surprisingly similar behavior. We can see that the climate is the major forcing of hydrological behavior, but its signal gets shaped more or less strongly by the catchment attributes. However, it remains unclear if the less clear climatic signal that can mainly be found in the western United States is caused by intra-catchment variability of the climate or by a larger influence from other catchment attributes. To explore this, a dataset is needed that contains a set of catchments with similar climatic forcing, but diverse catchment attributes.

Catchment complexity as a proxy for catchment interconnectedness

As was shown in the last chapter, climate is the main driving force for catchment behavior on a continental scale, but on smaller scales the importance of other factors increases. Therefore, I collected a dataset of 88 catchments from Hesse, Germany. Those catchments show a wide range of catchment attributes, while having a similar climate (see also chapter III).

Jehn, F. U., Breuer, L., Kraft, P., Bestian, K. and Houska, T.: Simple Catchments and Where to Find Them: The Storage-Discharge Relationship as a Proxy for Catchment Complexity, Front. Water, 3, 631651, https://doi.org/10.3389/frwa.2021.631651, 2021.

This research was also partly inspired by the extreme drought of 2018. After several months with almost no rain, some catchments in Hesse showed a surprisingly quick reaction after the first onset of rainfall (Fig. 7). According to common textbook knowledge, catchment storages should fill up first before substantial increases of discharge can be measured. However, field reconnaissance showed the opposite. When looking at rainfall reactions of a large set of Hessian streams, a number of catchments showed up that had been drained considerably by the drought, but flooded shortly after precipitation events, even though the catchments were missing several months' worth of precipitation. To explore this unexpected drought behavior and to address the impact of catchment attributes on a large sample of catchments with similar behavior, the focus of the following analysis is on the storage-discharge relationship, as it is a good proxy of overall catchment behavior (Tetzlaff et al., 2011). Discharge, precipitation, and evapotranspiration data, was used to calculate the storage change for all catchments and all years (1992-2018). This resulted in a storage-discharge relationship for every catchment and every year (a total of 2492), which we compared with an exponential function by using the Kling-Gupta Efficiency. The more the storage-discharge relationship fits an exponential function, the simpler the behavior of that catchment. The results showed quite a stark difference between the simplest and most complex catchments, but also that there are both catchments that behave very simple consistently for most of the years. The same is true for the complex catchments. However, even for those very simple and very complex catchments there exist years in which they show a behavior that is atypical for them (complex behavior for simple catchments and vice versa). This shows that even though the catchment's characteristics are quite important for its behavior, the hydrological signal is overwritten by the impact of strong climatic signals like long droughts or extreme precipitation.

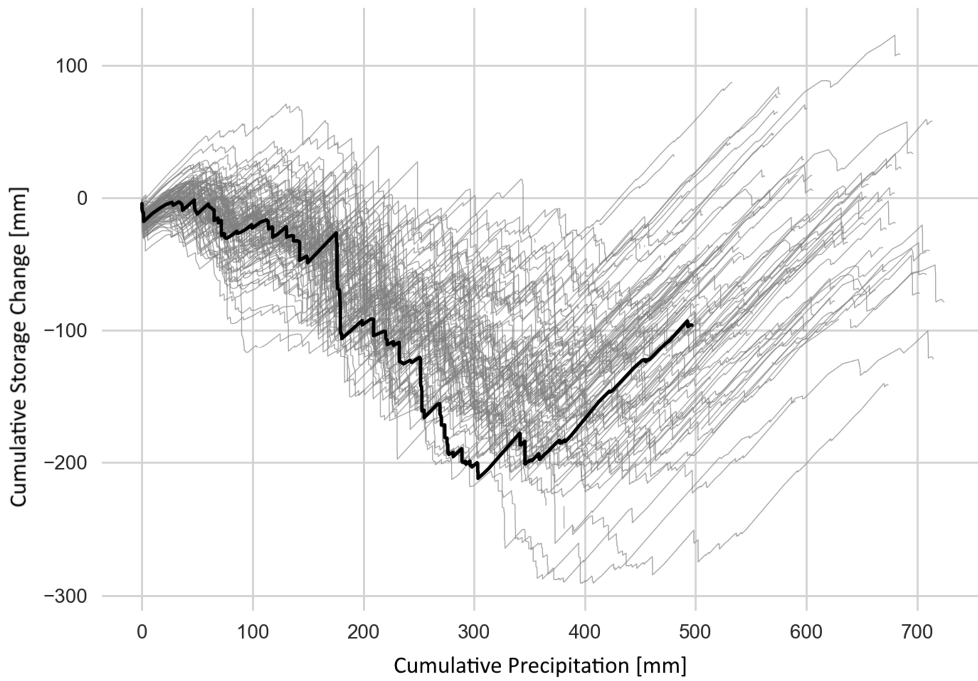

Figure 7: *Cumulative precipitation plotted versus cumulative storage change for all catchments in the Hesse dataset for the year 2018. Grey lines are the single catchments. Black line is the median of all catchments.*

While the simple catchments have a good permeability, igneous geology, a clay-silt soil texture and tend to be steep, the complex catchments have low permeability, sedimentary geology, loamy sand and are less steep. All those properties lead to the conclusion that catchments with a higher permeability and interconnectedness behave in a simpler way. In a simple catchment there is a more direct connection between the overall storage and the discharge, while in complex catchments the amount of water stored in the catchment matters less because it exists in less connected substorages. Those less connected substorages drain more erratically in the river than the more connected storages of the simpler catchments. The results helped to clarify the impact of catchment characteristics and climate. Therefore, the next step was to test how this newfound knowledge relates back to hydrological models by comparing how easy a catchment can be modelled depending on the complexity of its behavior.

Linking catchment and model complexity

The final contribution of this dissertation is a preliminary exploration of the connection between the idea of model complexity and the idea of catchment complexity. Is the complexity of a catchment's hydrological response correlated with the ease of modelling this catchment? This was tested by running the HYMOD model (Quan et al.,

2015) for all catchments and years from the Hesse dataset and evaluate the performance with the Kling-Gupta Efficiency. This resulted in having a catchment complexity value and a model performance value for every catchment year. The correlation between those two variables has a significant, linear trend (Fig. 8). This shows that simpler catchments are easier to simulate than complex catchments. The HYMOD model is a very simple lumped model, but it also seems that this simple structure is sufficient when the catchment is simple as well. However, the more complex catchments show a too complicated behavior for the simple model structure and would probably need more sophisticated models to be able to better represent the more complex behavior. Further research is needed to determine the degree to which complex model structures are needed to simulate complex catchments.

Due to this and the other contributions of this dissertation this research is now possible, as the dissertation provides a deep understanding of catchments on different scales and new tools to test how model structures and hydrological processes are connected.

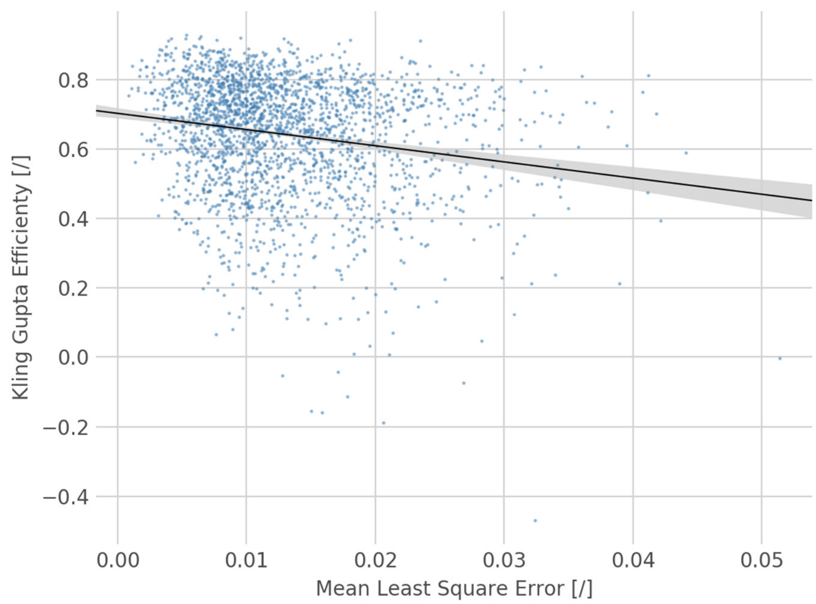

Figure 8: *Relationship between the model efficiency of the HYMOD Model (measured with the Kling-Gupta Efficiency) and the catchment complexity (measured with the Mean Least Square Error). The trend of the regression is significant with p-value < 0.001*

Conclusions and Outlook

This dissertation shows how hydrological models and large datasets can be used to improve our understanding of hydrological processes. In the preliminary work and the first paper, hydrological models were used to determine both the influence of model structure and model parameters on the model performance and realism. The findings of this section highlight that even though it is possible to change model structures and calibrate parameters to get good results, those results are often right for the wrong reasons. This is especially the case for simple model structures. Even though they are able to constrain parameters quite well, this focus on tightly constrained parameters can be deceiving, as the simple models create less realistic results. A more reliable approach seems to be to include spatial variability (in this dissertation in the form of semi-lumped model structures), as it only slightly increases the number of parameters while improving the overall realism. The evapotranspiration method on the other hand seems to be of lesser importance. However, this finding might only hold up for humid catchments, as different evapotranspiration methods deliver very similar results in this climate.

The second part of this dissertation disentangled the influence of climate and other catchment attributes on different scales. The larger the scale, the larger the influence of the climate. This becomes especially apparent when we look at the CAMELS dataset that covers the whole United States. While the eastern United States shows large homogenous regions concerning climate, hydrological clusters and ecoregions, the western United States shows a patchier pattern. This indicates that the climatic forcing has a more direct influence on the hydrological behavior in the eastern part. To explore this scale dependent influence of the climate the Hesse dataset was used, as it contains a set of catchments that have a similar climate but a wide variety of catchment attributes. This shows that on smaller scales we can see a higher influence of the geology and soils. Those smaller scale catchment attributes all link to the interconnectedness of the catchment, which in turn defines how complex a catchment behaves. The more connected, the simpler the catchment. This relates to the fill and spill hypothesis. The more connected catchments have lower thresholds for their spill behavior and thus a more direct connection between storage and discharge. Finally, I show that this simple behavior also has a connection to the model structure. The simpler the catchments, the easier it can be simulated.

Overall, this dissertation increased the understanding for interaction of model structure and model parameters and how this is linked to catchment complexity. From this point several promising new directions of research emerge. First, the clusters of the CAMELS dataset can be used for an in-depth study of model parameter transfer. This has been partly done in the first chapter of this dissertation but can be improved upon with the large sample size of the CAMELS dataset. Second, while the results achieved by analyzing the Hesse dataset are promising, it would be interesting to see if they also hold true for other regions (both in humid and other climates). Third, while a simple model structure is sufficient to simulate a simple catchment, it is unclear if complex catchments just need a complex model structure or if they are simply too complex to be modelled with current hydrological models.

What remains true for this dissertation and the hydrological science in general is that only if we thoroughly understand datasets that capture a wide variety of hydrological behavior and catchments characteristics will we be able to improve hydrological models.

I. Trade-offs between parameter constraints and model realism: a case study

This chapter is published in *Scientific Reports*:

Jehn, F. U., Chamorro, A., Houska, T. and Breuer, L.: Trade-offs between parameter constraints and model realism: a case study, Sci Rep, 9(1), 10729, doi:10.1038/s41598-019-46963-6, 2019.

Introduction

How complex should a hydrological model be? This question is still unsolved in hydrology. Recent advancements in experimental hydrology provide more data and lead to a better understanding of hydrology. This additional data and knowledge could be used to build more complex hydrological models. However, it is questioned if this will lead to better models. A higher complexity means more parameters and more parameters lead to a more difficult calibration (Beven, 2008a) and are sometimes seen as the main source of uncertainty (Perrin et al., 2001). Besides the parametric uncertainty, different kinds of uncertainty sources exist. Namely, model structure, evaluation, and forcing data. All of which are treated differently depending on the choice of the objective function and the calibration scheme applied. As uncertainty causes so many problems, like potentially undermining the trustworthiness or decreasing the forecasting ability of models (Beven, 2006a), it is sometimes referred to as the biggest problem in hydrology (Li et al., 2010).

However, hydrological models need to incorporate at least those hydrological features of a landscape that are needed to reflect dominant hydrological processes (Beven, 2008b; Boyle et al., 2001; Patil et al., 2014; Reed et al., 2004). To address this, it has become common practice to compare a range of models of different complexity. Complexity of the model structure in this context refers to the amount of processes in a model structure and its spatial subdivision (lumped vs semi-distributed vs distributed). Thus a model structure is more complex when it includes more processes and/or has a finer spatial resolution (Fenicia et al., 2008; Gao et al., 2014; Lobligeois et al., 2014; Orth et al., 2015). Those studies examine the influence of model structure and spatial layout and come to contrasting results. Some find that there is no difference between lumped and semi distributed models (Lobligeois et al., 2014), the lumped ones perform better (Andréassian et al., 2004a) or the (semi) distributed perform better (Patil et al., 2014).

Those problems of model structural uncertainty get aggravated as models are often compared without an underlying framework. This hinders comparisons of model structures and implemented processes (Breuer et al., 2009). This was also noted by other authors (Andréassian et al., 2004a), who state that many comparisons of lumped and semi-distributed models are hindered by different selections of included processes. This can be avoided by using a fixed modelling framework, which standardizes all steps of model structure development. In such a framework, all models are treated the same way, so that the differences in performance between models are only caused by the model structure itself.

Looking at the statements above it becomes clear that hydrological modellers are in a dilemma. Their models should avoid over-parametrization, but their models should also include all relevant hydrological processes. This implies a trade-off between the realism of the model and its ability to constrain its parameters. To explore this dilemma, this study will look at five different model structures, ranging from a simple lumped model to a semi-lumped model that takes vegetation and topography into account. All of those five models are run with two different methods to calculate the potential evapotranspiration (PET). PET has been identified as one very important process in models of this complexity concerning the simulation of discharge (Orth et al., 2015), and it is still not clear if simple temperature-based calculations can better help constraining the parameters of a model or not (Kannan et al., 2007; Oudin et al., 2005; Seiller and Anctil, 2016). To ensure comparability (Andréassian et al., 2004a), all models are built with the Catchment Modelling Framework (CMF) (CMF, 2018; Kraft et al., 2011). CMF is one of the few existing modelling frameworks that allows the isolation of the effects of the model structures and processes like the PET. The ROPE algorithm (Bárdossy and Singh, 2008) is used to calibrate the models, as it is capable of generating parameter sets with a small range of potential parameter values (Bárdossy and Singh, 2008). Using those tools, the aim of this study is to explore the trade-offs between the ability of a model structure to constrain its parameters, and the realism of the model structure. Realism is expressed as the performance of a model to simulate a variety of hydrologic signatures (Euser et al., 2013; Westerberg and McMillan, 2015) for which the model has not been calibrated.

Materials and Methods

Study area

The study area is the upper part of the Fulda catchment (Catchment area 562 km², gauging station Kämmerzell). The catchment has Mid-European temperate climatic conditions. To the east and west, the river receives water from two ridges: the Wasserkuppe and the Vogelsberg. Elevation ranges from 237 m a.s.l. to 950 m a.s.l. Land use is dominated by agriculture (~50%) and forests (~40%) (Fig. I.1). For more details see Jehn et al. (2018). Meteorological data for model forcing and discharge data for model calibration and validation are obtained from the Hessisches Landesamt für Naturschutz, Umwelt und Geologie (HLNUG, https://www.hlnug.de/messwerte.html) for the period 1979–1989. The discharge is measured at the Kämmerzell gauging station. Windspeed, relative humidity, sunshine duration, and temperature are taken from nine weather stations located in close vicinity to the catchment (Eschwege, Wasserkuppe, Grebenhain, Melsungen, Wartenberg, Neukirchen, Kassel, Bad Hersfeld and Fulda). Both the model time step and the temporal resolution of the input data are daily. This is in line with recommended temporal resolution based on results obtained for mesoscale model applications (Sikorska and Seibert, 2018).

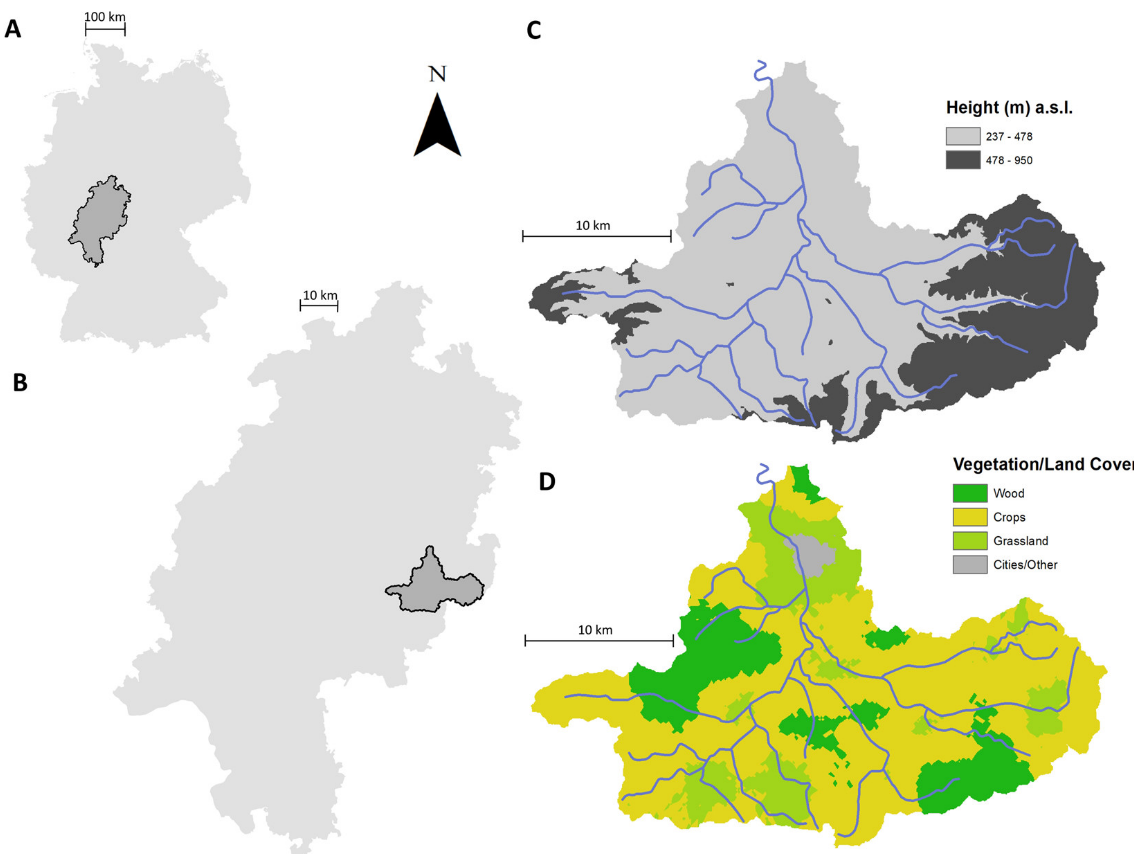

Figure I.1: *Location of Hesse in Germany (**A**), Location of the Fulda catchment in Hesse (**B**) (gauging station Kämmerzell) and separation of the catchment by height (**C**) and vegetation/land cover (**D**).*

Model framework

All models were constructed using the open source, modular Catchment Modelling Framework (CMF) (Kraft et al., 2011). Additional information can be found at the framework's website (CMF, 2018). To avoid numerical problems (Kavetski and Clark, 2011), we selected the CVode Integrator (Hindmarsh et al., 2005) as the numerical solver. The CMF version used for this study was 1.1.1.

The base model structure consists of a one storage set up with a simple snow storage and actual evapotranspiration (Fig. I.2). The storage receives precipitation when it is warmer than 0 °C. Otherwise, the precipitation is stored as snow. Water in the storage gets either evapotranspirated or is transferred to the outlet. Following the findings of Singh (Singh, 2002), all connections in the model (Fig. I.2) are described as kinematic waves (Eq. I.1):

$$Q = Q_0 \left(\frac{V - V_{residual}}{V_0}\right)^{\beta} \qquad \text{(Eq. I.1)}$$

where Q is the amount of water transferred from one storage to the other, $V_{residual}$ [m³] is the volume of water remaining in the storage at each time step, V_0 [m³] is the reference volume (calibrated) to scale the exponent, V is the current volume of water in the storage [m³] at each time step, and β is a parameter to shape the response curve [−]. Q_0 is the flux in [m³ d⁻¹], when $\frac{V - V_{residual}}{V_0} = 1$.

The code for all models is freely available on GitHub and is stored in a citable repository (Jehn, 2018a). In the following it will be explained how this base structure is built upon to create the more complex models.

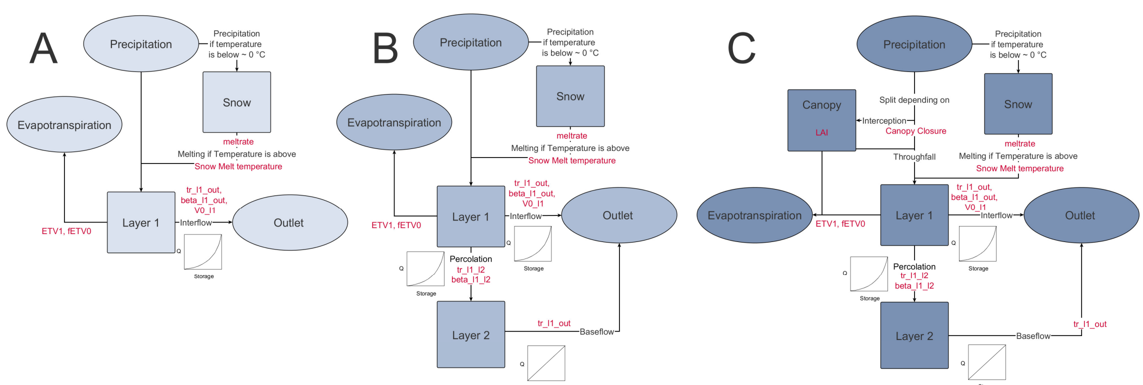

***Figure I.2**: Model structure for Lumped 1 (**A**), Lumped 2 (**B**) and Lumped 3 and Semi-Lumped 3 (**C**). Calibration parameters shown in red.*

Model structures

A total of five model structures were constructed, three lumped and two semi-lumped models. Semi-lumped is used here in line with in Andréassian et al. (2004b), meaning a lumped model with a spatial subdivision, but with the same parameters for each spatial subdivision. The models differ in three complexities (1–3). While the most simple lumped model Lumped 1 consists of only one storage *Layer 1* (Fig. I.2), evapotranspiration and a snow storage (7 parameters), the moderate complex lumped model Lumped 2 uses an second storage *Layer 2* (10 parameters). In addition to this, the most complex lumped model Lumped 3 features a simulation of the canopy storage *Canopy* (12 parameters). A detailed description of the parameters is given in Table I.1. The number of parameters is similar to other studies that compared models of differing complexity (Orth et al., 2015).

For the two semi-lumped models we used the model structure of the most complex lumped model Lumped 3. The spatial subdivision for the first semi-lumped model Semi-Lumped 3-Vegetation is based on vegetation (forest, arable land, grassland and settlements/other) (Fig. I.1). For the second semi-lumped model Semi-Lumped 3-Vegetation/Height an additional split between high (above 478 m a.s.l.; 25% of the catchment) and low (equal or below 478 m a.s.l.; 75% of the catchment) elevation was considered, resulting in eight spatial subdivisions. For those spatial subdivisions, the point measurements for the forcing data were interpolated, using external drift kriging with the height as external drift. For the lumped models, the interpolated data was arithmetically averaged for the whole catchment. In case of the semi-lumped models, the interpolated data were split into the separate spatial subdivisions, and the averages were calculated separately. This was necessary to bring the data in an appropriate format for the semi-lumped models.

Table I.1: *Parameter for all models with their intended meaning and ranges considered during calibration. Parameter related processes are shown in Fig. I.1.*

Name	Unit	Intended meaning	Model Structure	Min	Max
tr_l1_l2	day	Residence time from layer 1 to layer 2	B, C	1	400
tr_l1_out	day	Residence time from layer 1 to outlet	A, B, C	1	200
tr_l2_out	day	Residence time from layer 2 to outlet	B, C	1	650
V0_l1	mm	Field capacity of the soil	A, B, C	1	300
beta_l1_l2	—	Exponent the changes the shape of the flow curve	B, C	0.5	6
beta_l1_out	—	Exponent the changes the shape of the flow curve	A, B, C	0.3	8
ETV1	mm	Volume under which the evapotranspiration is lowered	A, B, C	1	300
fETV0	%	Factor by what the evapotranspiration is lowered	A, B, C	0	0.9
meltrate	mm $°C^{-1}$ day^{-1}	Melt rate of the snow	A, B, C	0	12
snow_melt_temp	°C	Temperature of snow melt	A, B, C	−3	3
LAI	—	Leaf area index	C	1	12
CanopyClosure	%	Canopy closure	C	0.1	0.9

Potential evapotranspiration

In addition, every model exists in two versions, depending on the methodology used for the calculation of the PET. For this, we considered the methods according to Hargreaves (Samani, 2000) and Penman-Monteith (Allen et al., 1998) (also referred to as Penman). A detailed description of the calculation of the PET methods can be found in the Supplementary Information.

Calibration and validation

The models were calibrated using the ROPE algorithm (Bárdossy and Singh, 2008), as implemented in the SPOTPY package (Houska et al., 2015). The algorithm itself was run 100,000 times. For further analysis the 1,000 best runs of the last set were used, as proposed Bardossy and Singh (2008). The performance of all models was evaluated using the Kling-Gupta Efficiency (KGE) (Gupta et al., 2009). The time series was split into a warm up period (1979), the calibration (1980–1984), and validation period (1985–1989).

All parameters (Tab. I.1) were sampled from a uniform distribution. The ranges for V0 and ETV1 were in agreement with typical field capacity values for German soils (Blume et al., 2016), while canopy parameters were taken from Breuer et al. (2003) All other parameters were subjectively set, as their conceptual nature does not allow to link them directly to physical processes. However, their ranges were in line with other studies that explored the Fulda catchment using models (Fink and Koch, 2010; Jehn et al., 2018) and field experimental approaches like tritium (Wittmann, 2002).

Model evaluation

The realism of all models was subsequently evaluated by how much it was possible to constrain their parameters and their ability to correctly simulate a selection of hydrological signatures, which they were not calibrated for (Table I.2). This way of assessing the models realism allows to evaluate both, their ability to constrain parameters and the realism of their simulations.

The parameter distribution is evaluated by comparing the parameters before and after calibration. A range reduction factor is determined to indicate how much those differ in their range [in %]. We choose the constraint of the parameters as one criteria in this study, as unconstrained parameters are often stated as a core problem in hydrology (Beven, 2008a; Perrin et al., 2001).

For the hydrological signatures, we selected a number of those signatures presented by Westerberg and McMillan (2015) (Table I.2). Those signatures capture the behaviour of a river concerning its flow distribution (high, mean and low flows), the frequency and duration of high and low flow events and the dynamics of the flow. They are widely used for catchment classification, and model calibration (Westerberg and McMillan, 2015). The signatures were calculated for the whole time period on daily data. We choose hydrological signatures to assess the realism of the simulation, as in recent years hydrological signatures are used more and more often to detect weaknesses in hydrological models (Euser et al., 2013; Gupta et al., 2008).

Table I. 2: *Hydrological signatures used in this study were taken from Westerberg and McMillan (2015). All signatures are calculated on daily data and for the whole time period.*

	Signature	**Name**	**Description**	**Unit**
Flow distribution	Q_{mean}	Mean flow	Mean flow for the analysis period	mm d^{-1}
	$Q_{0.01}$, Q_{99}	Flow percentiles	Low- and high-flow exceedance percentiles from the flow duration curve (FDC)	mm d^{-1}
Event frequency and duration	Q_{HF}	High-flow event frequency	Average number of daily high-flow events per year with a threshold of 9 times the median daily flow (Clausen and Biggs, 2000)	yr^{-1}
	Q_{HD}	High-flow event duration	Average duration of daily flow events higher days than 9 times the median daily flow (Clausen and Biggs, 2000)	days
	Q_{LF}	Low-flow event frequency	Average number of daily low-flow events per year with a threshold of 0.2 times the mean daily flow (Olden and Poff, 2003)	yr^{-1}
	Q_{LD}	Low-flow event duration	Average duration of daily flow events lower days than 0.2 times the mean daily flow (Olden and Poff, 2003)	days
Flow dynamics	BFI	Base-flow index	Contribution of base flow to total streamflow calculated from daily flows using the Flood Estimation Handbook method (Gustard et al., 1992)	—
	S_{FDC}	Slope of normalized FDC	Slope of the FDC between the 33 and 66% exceedance values of streamflow normalized by its mean (Yadav et al., 2007)	—
	Q_{CV}	Overall flow variability	Coefficient of variation in streamflow, i.e. standard deviation divided by mean flow (Clausen and Biggs, 2000; Jowett and Duncan, 1990)	—
	Q_{LV}	Low-flow variability	Mean of annual minimum flow divided by the median flow (Jowett and Duncan, 1990)	—
	Q_{HV}	High-flow variability	Mean of annual maximum flow divided by the median flow (Jowett and Duncan, 1990)	—
	Q_{AC}	Flow autocorrelation	Autocorrelation for 1 day (24 h) (Euser et al., 2013; Singh and Xu, 1997; Winsemius et al., 2009)	—

Results

Model performance

All models were able to produce runs that have KGEs above 0.8. In addition, all models performed better in the validation than in the calibration period (Fig. I.3), with the exception of the model Lumped 3 Hargreaves. The semi-lumped models reach slightly higher maximal KGE values than the lumped models. However, the semi-lumped models in combination with the Hargreaves PET method also show the overall largest spread and the lowest KGEs values. This tendency of a comparatively large KGE spread is also found for the more complex Lumped 3 models. For the more simple models Lumped 1 and particular for Lumped 2 it is the other way around. Here the models with the Penman PET method have a marginally larger spread of the objective function.

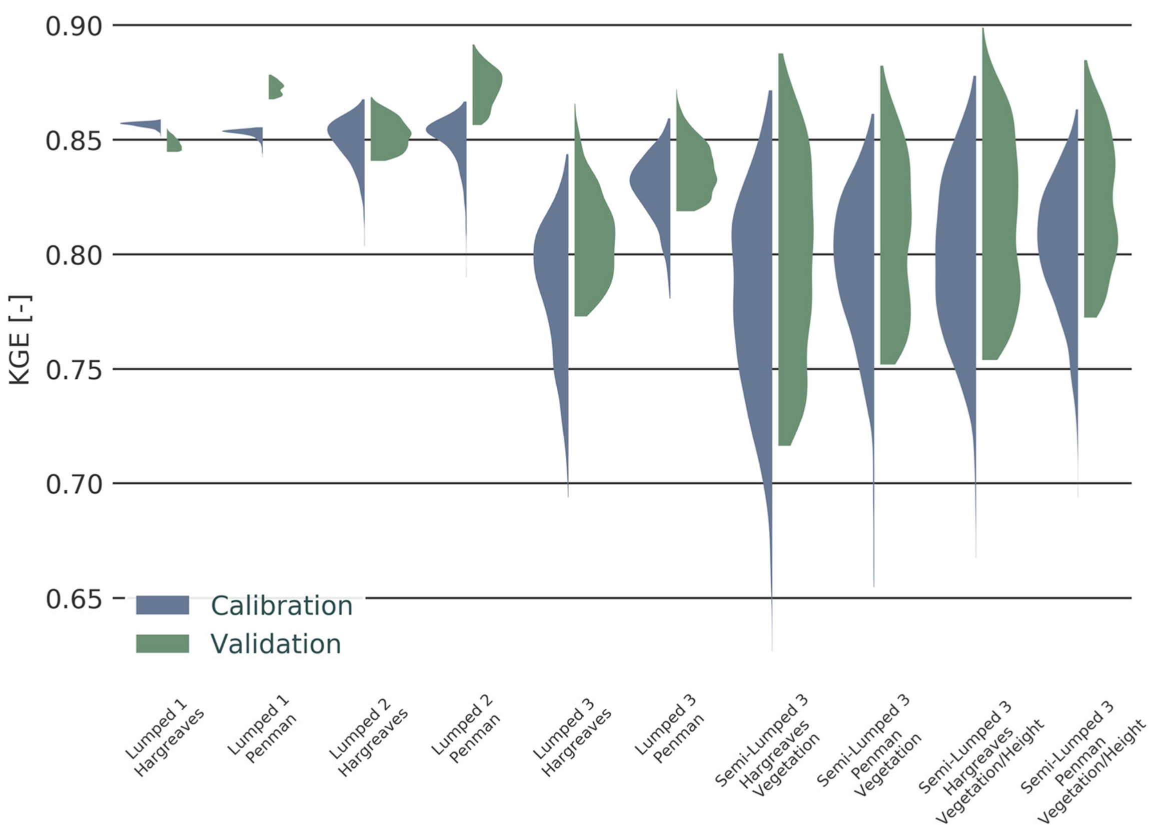

Figure I.3: *Model performance according to the Kling-Gupta-Efficiency (KGE) for all models, seperated by the calibration and validation period.*

Parameter constraints

When looking at the parameter distribution for all single model structures, the simpler models show a smaller range in the parameter distribution (Fig. I.4). Lumped 1 is the model structure that is most able to constrain its parameters. This is true for both PET version, with a median parameter constraint of 95% (Fig. I.5). All other model structures are less able to constrain their parameters (Figs I.4 and I.5). Especially the model structures Lumped 3 and Semi-Lumped 3 both have a median parameter constraint below 50% and contain parameters like *tr_l2_out* (Residence time from layer 2 to outlet), which can only be constrained by 25%.

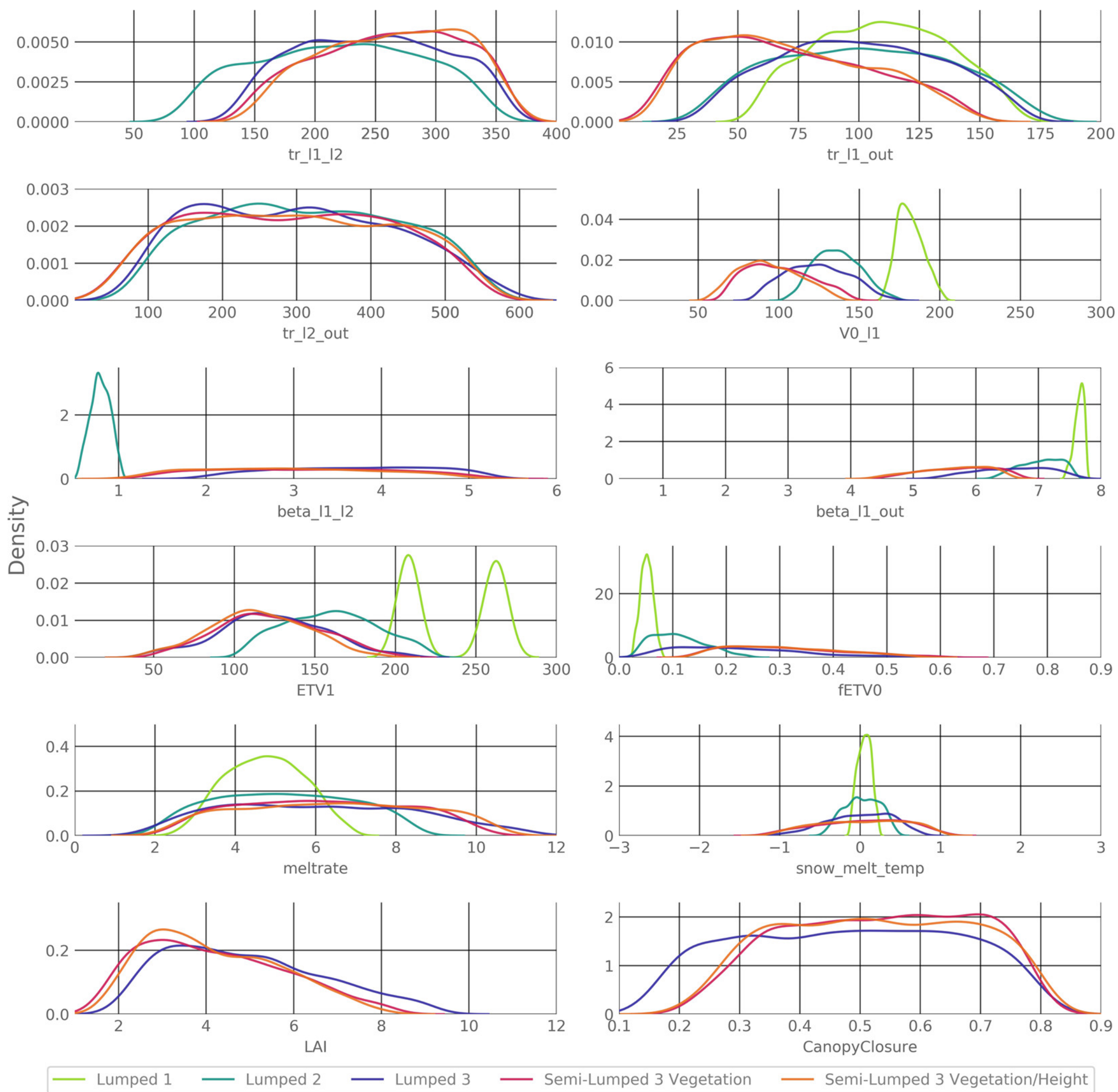

Figure I.4: *Posterior parameter distribution separated by model structures shown in different coloured lines. Different PET calculations for a model structure are pooled. X-axes scales equal the a priori distribution of the parameters before calibration. Lines are fitted with a Gaussian kernel density function.*

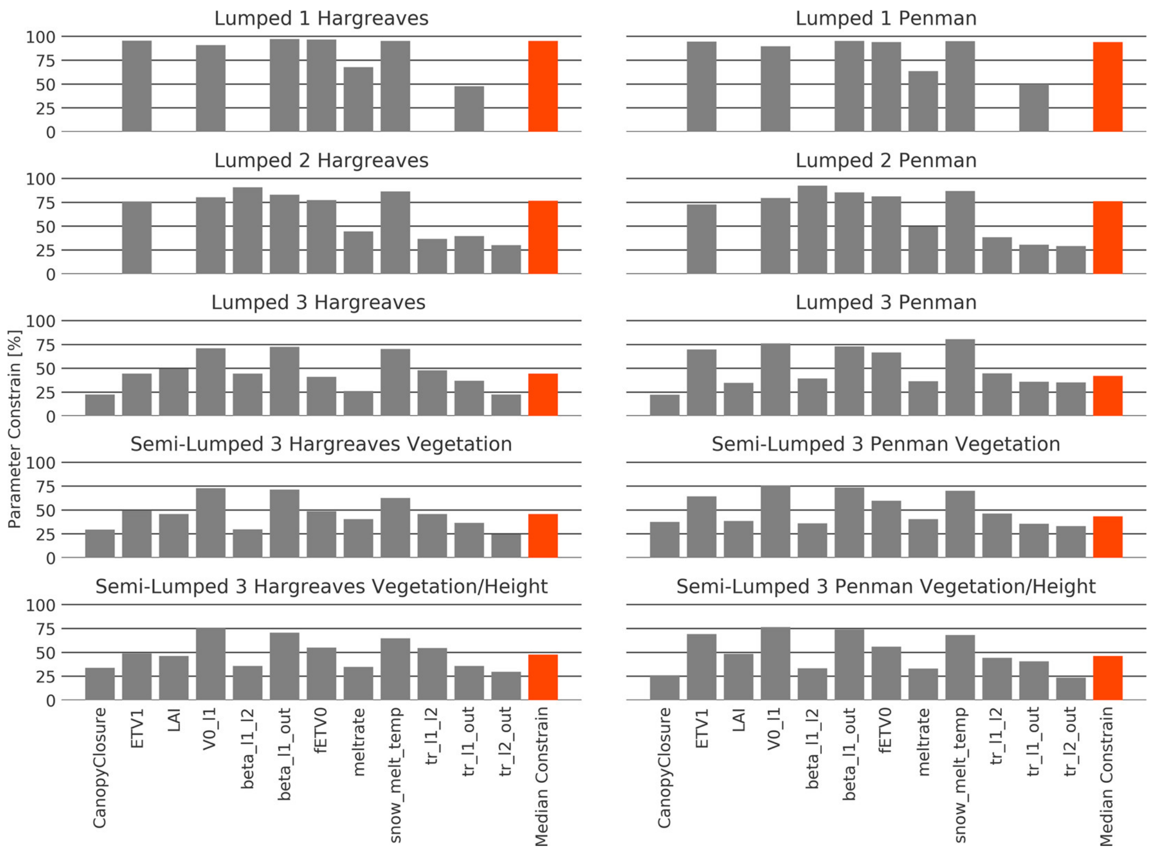

Figure I.5: *Parameter constrainability for all model structures separated by parameters. Red bar marks the median parameter constrainability for each model. Larger bars indicate larger constrained parameters. Parameter constrainability is defined as the difference*

The ability of the different model structures to constrain a parameter is also highly dependent on the parameter itself. We find three classes of parameters. Parameters like *V0_l1* (field capacity of the soil) or *snow_melt_temp* (temperature of the snow melt) have a very clear peak in the distribution after the calibration and are constrainable. Other parameters such as *tr_l2_out* (transition time from lower layer to outlet) or other residence time parameters are difficult to be constrained at all. A third class of parameters like *fETV0* (reduction of the PET under dry conditions) and *beta_l1_l2* (shapes the flow curve) show an ambiguous behaviour with better constrainability for the lumped model structures. Overall parameters, which can be constrained best by the models, are related to the evapotranspiration, the snow melt, and the water flux from the first layer to the outlet. Parameters related to the second layer and the canopy structure cannot be constrained well by the different model structures.

The distributions of the parameters are influenced more by the spatial subdivision than by the PET (Fig. I.6A). When all model structures are pooled and only the difference between Hargreaves and Penman is considered (Fig. I.6A), the only parameter where larger differences can be found is *ETV1* (Volume below which the PET is lowered by *fETV0*). For *ETV1* the models with Penman have a peak in the distribution of the parameter at around 270 [mm], while the Hargreaves models peak at 210 [mm]. The second parameter that is influenced by the PET is the *LAI* parameter. The peak in the distribution of *LAI* is slightly shifted to the left for the Penman models in comparison with the Hargreaves models.

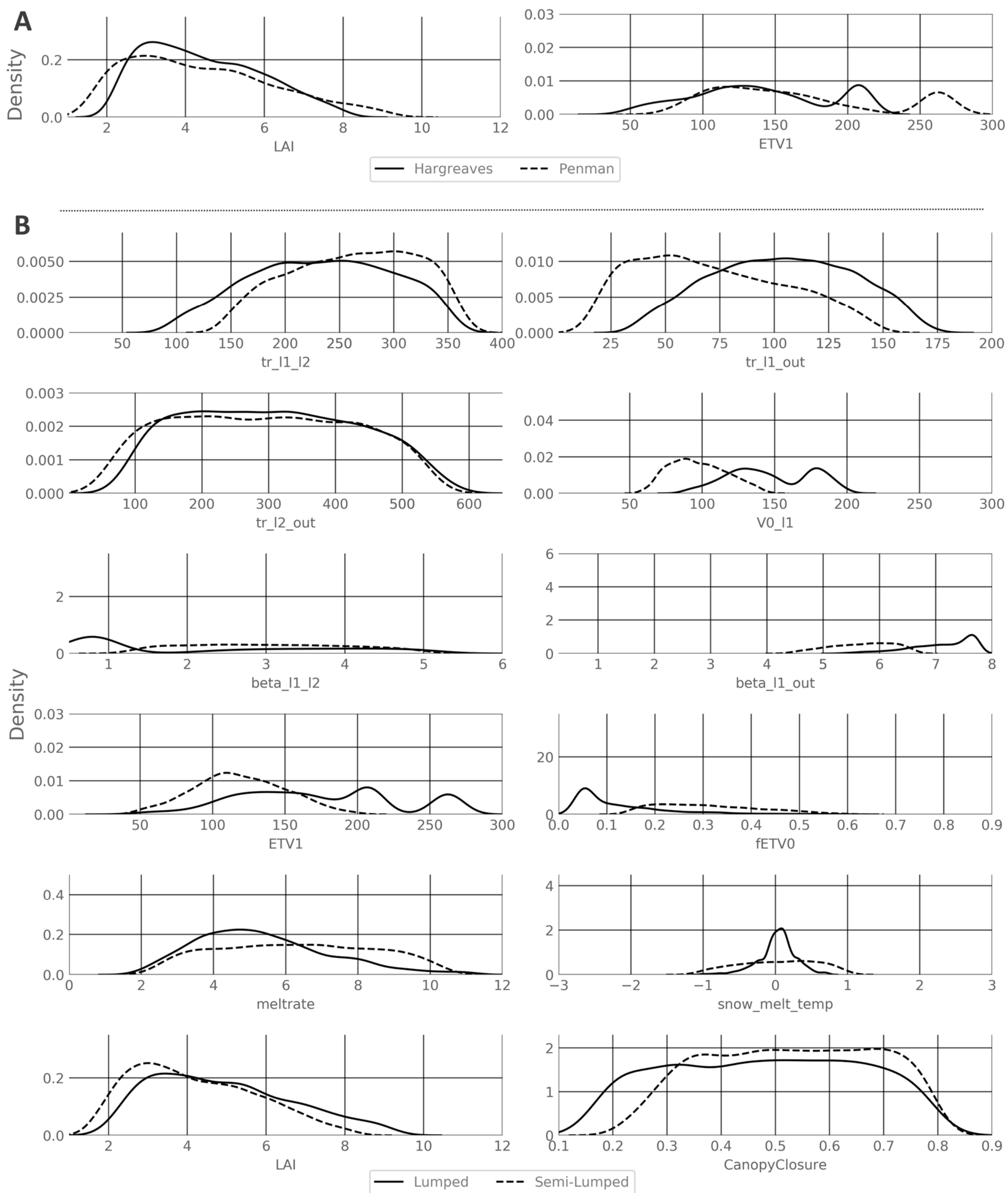

Figure I.6: *Posterior parameter distributions separated by PET method for the parameters influenced by PET method. Different model complexities are pooled (A). And distributions separated by spatial subdivision for the parameters influenced by spatial subdivision.*

The differences become clearer when all lumped and semi-lumped models are pooled (Fig. I.6B). Here most parameters show at least some deviations. Parameters like *V0_l1* or *ETV1* even depicting very different distributions. The only parameter that experiences a shift in both comparisons (lumped vs semi-lumped and Hargreaves vs Penman) is *ETV1* and to some extent the *LAI*. While the shape of the distribution for the *LAI* has its peak at around 3.5 for both the PET method and the comparison between lumped and semi-lumped, *EVT1* shows a different shape of the distribution. The distinction is clearer in the comparison of the lumped and semi-lumped models. The unimodal distribution for the semi-lumped models is very different to the bimodal distribution of the lumped models.

Model realism in regard to hydrological signatures

In the next step, we challenge the various model structures to simulate a large set of hydrological signatures, and relate their performance to the information on parameter distribution and KGEs. The simulated hydrological signatures (explanation of signatures in Table I.2) shown in Fig. I.7 depict different model performances compared to the previously described parameter distributions and KGEs. The simpler models, especially Lumped 1 and Lumped 2, are able to achieve consistently high KGEs and can constrain their parameters quite good. In spite of this, they show a larger deviation from the measured signatures than the more complex models Lumped 3 and Semi-Lumped 3. This is most apparent for the signatures regarding the frequency and duration (marked yellow in Fig. I.7). In this case, the model Lumped 1 completely fails to get the low flow event duration right (Q_{LD}). The model also reveals a large error in the prediction of high flow event duration (Q_{HD}) and the low flow exceedance percentiles (Q_{99}). To a lesser degree the slope of the flow duration curve (S_{FDC}), the low flow variability (Q_{LV}), and the high and low flow frequency (Q_{LF}, Q_{HF}), are also challenging for the model Lumped 1. Contrary, Lumped 2 does have a smaller error in its simulation of its hydrological signatures. This model only has problems in predicting the low and high flow durations (Q_{LD}, Q_{HD}) and the characteristic recession time at median flow (T_0). Similarly, Lumped 3 has the same problems as Lumped 2, but is able to get the low flow duration (QLD) more correct. Although, this comes at the cost that it has a larger error in the characteristic recession time at the median flow (T_0) and the low flow duration (Q_{LF}). The Semi-Lumped 3 models with both spatial set ups of vegetation and vegetation/height have overall smaller errors than the lumped models. Nevertheless, they also have problems in getting the low and high flow durations right (Q_{LD}, Q_{HD}), but to a lesser extent than the lumped models. At the same time, they have smaller errors in the characteristic recession time at the median flow (T_0), while Lumped 3 fails at that.

All models behave very similar for both PET methods in regard to the hydrological signatures. Only the low flow duration error (Q_{LD}) in Lumped 2 is considerably higher for the Penman version, while the low flow frequency error is lower (Q_{LF}).

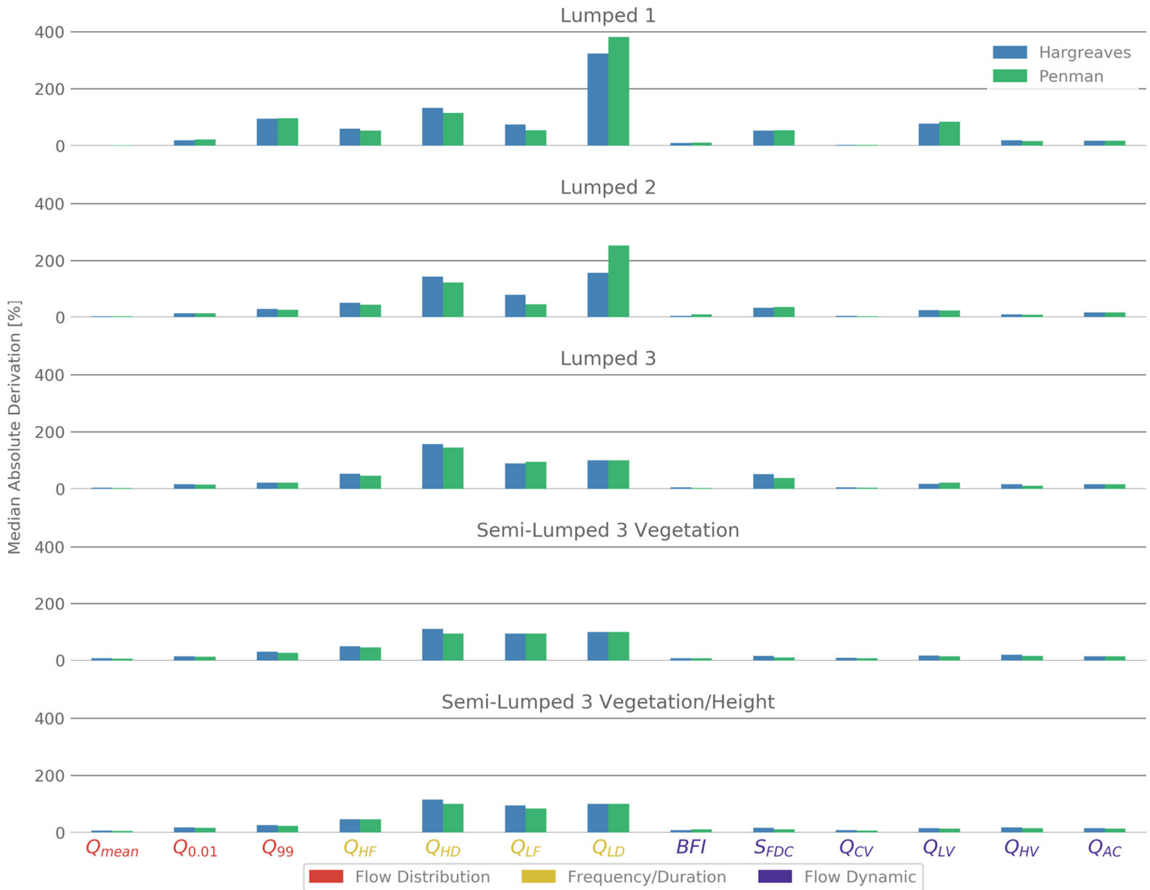

Figure I.7: *Median absolute deviations (%) of simulated versus observed hydrological signatures. Smaller values indicate smaller error in the simulation.*

Discussion

When we look at the model performances as indicated by the KGE (Fig. I.3) the two most simple model structures Lumped 1 and Lumped 2 seem to perform fairly well, showing only a very small range of the KGE at a high level, both during the calibration and validation. All other models have much larger spread for their KGE, even though the ROPE algorithm is intended to avoid that (Bárdossy and Singh, 2008). When we compare the KGE values for calibration and validation all models except Lumped 1 perform better in the validation period. A better performance during validation is usually considered as a sign for models of an appropriate complexity, which have an adequate number of parameters (Fenicia et al., 2008; Her and Chaubey, 2015). However, this might also be caused by less extreme rainfall events or reduced discharge variability in the validation period in comparison to the calibration period (Breuer et al., 2009). This drop in performance from calibration to validation of Lumped 1 hints that the model is not able to predict well, which often is the case when a model is too simple (Wilby, 2005).

The models, which have a small range for the KGE, also have tightly constrained parameters (Fig. I.5). Again, the parameters of the two most simple models (Lumped 1 and Lumped 2) can be constrained most. Lumped 1 has a median parameter constrain of 95%. This is quite high, since other studies with a comparable number of parameters could not constrain their parameters this much (Teweldebrhan et al., 2018; Yaduvanshi et al., 2018). However, studies with fewer parameters found similar constraints (Zhang et al., 2016). This shows that hydrological models with fewer parameters can usually be constrained more easily. Nevertheless, this relationship is not linear and difficult to be generalized. For example, Shen et al. (2012) used the SWAT model with twenty parameters and could constrain around half of them while (Seibert, 1997) was only able to constrain one out of 12 parameters in HBV.

When all models are pooled by the PET method, we could only find large differences in the distributions *ETV1* (volume under which the evapotranspiration is lowered). Therefore, we conclude that the PET method only affects those parameters that are directly related to it. In addition, when the parameter constraint is quantified (Fig.I.7) Hargreaves is slightly better for all models. However, the effect is small compared to the strong effects on the parameter values by the PET calculation as also found by other studies (Vázquez, 2003).

The main shift in the distribution of the parameters is caused by the switch from the lumped to the semi-lumped model structure (Figs I.4 and I.6). Here, several parameters experience a shift or reshape of their distribution. This is especially the case for *V0_l1* (field capacity of the soil) and *ETV1*. Further, the parameters of the semi-lumped models are less constrained than the parameters in the lumped models (Fig. I.5). Nevertheless, they are similar constrained in comparison with models of similar complexity (Samadi et al., 2017; Yaduvanshi et al., 2018). We conclude that the lumped models, especially the more simple ones, are markedly better in constraining the parameters than the more complex models and this can be mainly attributed to the switch from a lumped to a semi-lumped structure.

The patterns found in the hydrological signatures are different to the ones concerning parameter constrainability. Here, the lumped models struggle more than the semi-lumped ones to correctly simulate the hydrological signatures. Especially their ability to simulate the low flows shows larger errors. This is in line with other studies (Gharari et al., 2014; Orth et al., 2015) who found that models that do not get the groundwater behavior right or miss a groundwater component fail to simulate discharge minima. Generally, it is stated that models must incorporate as much of the catchments landscape characteristics as possible to come up with reasonable explanatory power (Clark et al., 2016) and many studies find a performance increase when switching from a lumped to a semi-distributed model layout (Fenicia et al., 2008; Gharari et al., 2014). Usually, this is attributed to the accounting of rainfall variability (Andréassian et al., 2004b) and topography (Gao et al., 2014). This might also be the case for the semi-lumped models, as the spatial subdivision might contain a more accurate representation of rainfall. However, there seems to be an upper limit on how much spatial subdivisions make sense for a given amount of data (Boyle et al., 2001; Rouhier et al., 2017), which also seems to be the case for this study. Not much improvement can be found when going from four to eight spatial subdivisions.

Concerning the PET method there seems to be almost no influence on the hydrological signatures (Fig. I.7). This is in contrast to other studies (Kannan et al., 2007; Krysanova et al., 1999), who state that getting the PET right is essential to model the discharge successful. The PET method is often attributed to cause large differences between hydrological models (Breuer et al., 2009). In spite of that, the calculation of the PET might mainly influence the overall water balance, while not having a large effect on the daily discharge. In our study, the Hargreaves and Penman methods were similar enough not to cause any differences between the simulation of the hydrological signatures. The only exception from this is Lumped 2, where the Penman version depicts a larger error in the low flow duration (QLD) and a smaller error in the low flow frequency (QLF). This is caused by the shift in the parameter ETV1 and LAI, which both control the evapotranspiration. The simpler model Lumped 1 has such a large error in its signatures that it overlays the differences between the different PET methods. On the other hand, the more complex models are able to correctly simulate the low flow characteristics due to their more realistic structure.

Overall, the models used in this study show two patterns along their axis of complexity. While the simple models (Lumped 1 and Lumped 2) are quite good at constraining their parameter and not so good at getting hydrological signatures right, it is the other way around for the more complex models (Semi-Lumped 3, both spatial versions). They have problems with constraining their parameters, but manage to have a lower error at their hydrological signatures. This seems counterintuitive, as tightly constrained parameters are seen as a property of good models, but it highlights that is important to use several criteria to evaluated models to avoid one sided results (Melsen et al., 2016). A better model performance in the calibration than in the validation period is often seen as a sign of an overfitting of the more complex models (Das et al., 2008; Her and Chaubey, 2015; Orth et al., 2015; Perrin et al., 2001). This does not apply here as all models perform better in validation. One possible explanation for the good performance of the more complex models concerning the hydrological signatures, can be found in the study of Shen et al. (2012). They used a semi-distributed model (SWAT) with twenty parameters and found that they could not constrain most of their parameters. However, they stated that unconstrained parameters do not imply that those parameters are not important for the model, but simply that they interact with other parameters in the model. Similar results were also stated by Zhao et al. (2018, 2019). They also used the SWAT model and found that in such a more complex model set up, the parameters seem more disperse. Still, the added complexity of the model allows SWAT to more accurately reflect the real conditions, but this complexity must be constrained with additional data (Zhao et al., 2019), like it was done in this study by using information about the land use and topography of the catchment.

This interaction of parameters could be caused by an increase in uncertainty due to the introduction of additional data to the semi-lumped models. Therefore, simple models will not show the reality but merely hide the uncertainties inherent in the data (Houska et al., 2017). Hence, models should include additional data like landscape related process heterogeneity (Gharari et al., 2014), land cover (Oudin et al., 2005) if possible, as it allows for a more realistic prediction without hiding uncertainties.

Overall, the results in this study show that it is easier to constrain parameters of simple models. However, their simple structure does not allow them to provide realistic simulations. We analysed this behaviour with the ability to simulate hydrological characteristics. It turned out that the simply structured models have strong weaknesses here. For the more complex models, the story is different. Their parameters are harder to constrain, but they outperform the simple models regarding the hydrological characteristics. This indicates a clear trade-off between the ability to constrain the parameters of these models and the ability to realistically simulate the discharge.

Conclusion

This study explored five hydrological models of differing complexity implemented with two PET methods concerning the trade-offs between parameter constrainability and their ability to simulate hydrological signatures. We used the same model building framework, numerical solver, calibration algorithm, and forcing data to ensure that the results are only influenced by the model structure itself. The results show that parameters of the more complex models are less constrained, still the models have a smaller error in simulating hydrological signatures in comparison with the simpler models. The selection of the PET method only affected canopy parameters, but had hardly any influence on parameters of the flow generating processes. We note that the results depend on the investigated site and period and may not be generalizable. However, the catchment used has typical properties for a Central German Upland catchment and thus the findings should at least be applicable in this region. This study also shows the benefits of comparing model in a modelling framework, as it ensures that all models are handled equal. Finally, this study highlights the importance of not focusing too narrowly on the parameter uncertainty, as models that incorporate more relevant hydrological processes are able to simulate a river more realistically concerning hydrological signatures, even though their parameters are less constrained.

II. Using hydrological and climatic catchment clusters to explore drivers of catchment behavior

This chapter is published in *Hydrology and Earth System Sciences*:

Jehn, F. U., Bestian, K., Breuer, L., Kraft, P. and Houska, T.: Using hydrological and climatic catchment clusters to explore drivers of catchment behavior, Hydrol. Earth Syst. Sci., 24(3), 1081–1100, doi:10.5194/hess-24-1081-2020, 2020.

Introduction

Every hydrological catchment is composed of a unique combination of topography and climate, which makes their discharge heterogeneous. This, in turn, makes it hard to generalize behavior beyond individual catchments (Beven, 2000). Catchment classification is used to find patterns and laws in the heterogeneity of landscapes and climatic inputs (Sivapalan, 2003). Historically, this classification was often done by simply using geographic, administrative or physiographic considerations. However, those regions proved to be not sufficiently homogenous (Burn, 1997). Therefore, it was proposed to use seasonality measures with physiographic and meteorological characteristics, but it was deemed difficult to obtain this information for a large number of catchments (Burn, 1997), even if only simple catchment attributes (e.g., aridity) are used (Wagener et al., 2007). Nonetheless, in the last decade datasets with hydrologic and geological data were made available, comprising information on hundreds of catchments around the world (Addor et al., 2017; Alvarez-Garreton et al., 2018; Newman et al., 2014; Schaake et al., 2006). This is a significant step forward as those large-sample datasets can generate new insights, which are impossible to obtain when only a few catchments are considered (Gupta et al., 2014). Different attributes have been used to classify groups of catchments in those kind of datasets: flow duration curve (Coopersmith et al., 2012; Yaeger et al., 2012), catchment structure (McGlynn and Seibert, 2003), hydro-climatic regions (Potter et al., 2005), function response (Sivapalan, 2005a) and, more recently, a variety of hydrological signatures (Kuentz et al., 2017a; Sawicz et al., 2014; Toth, 2013). Quite often, climate has been identified as the most important driving factor for different hydrological behavior (Berghuijs et al., 2014; Kuentz et al., 2017a; Sawicz et al., 2014). Still, it is also noted that this does not hold true for all regions and scales (Ali et al., 2012; Singh et al., 2014; Trancoso et al., 2017). In addition, a recent large study of Addor et al. (2018) has shown that many of the hydrological signatures often used for classification are easily affected by data uncertainties and cannot be predicted using catchment attributes. Another recent study by Kuentz et al. (2017) used an extremely large datasets of 35 000 catchments in Europe and classified them using hydrological signatures. For their classification, they used hierarchical clustering and evaluated the result of the clustering by comparing variance between different numbers of clusters. They were able to find 10 distinct classes of catchments. However, Kuentz et al. (2017) used some of the signatures identified to have a low spatial predictability by Addor et al. (2018). In addition, one-third of their catchments was aggregated in one large class with no distinguishable attributes. Overall, we conclude that no large-sample study exists that uses only hydrological signatures with a good spatial predictability. In addition, if the climate is the dominant driver of catchment behavior, clustering catchments based on their hydrological behavior should result in clusters with a similar climate.

Therefore, we selected the best six hydrological signatures with spatial predictability to classify catchments of the CAMELS (Catchment Attributes and MEteorology for Large-Sample Studies) dataset (Addor et al., 2017). Those six hydrological signatures are evaluated together with the 16 catchment attributes that were shown to have a large influence on hydrological signatures (Addor et al., 2018). The connection between the hydrological signatures and the catchment attributes is determined by using quadratic regression of the principal components (of the hydrological signatures) and the catchment attributes. This will help to explore whether a clustering with hydrological signatures that have a high predictability in space provides hydrologically meaningful clusters and how those are related to catchment attributes. In addition, we compare the hydrologically derived clusters with climatic clusters and determine the spatial distance between the most hydrologically similar catchments. This will determine whether grouping catchments by climate or by hydrologic behavior will yield the same results and whether the signatures identified by Addor et al. (2018) as having the highest spatial predictability can be used to delineate hydrologically meaningful clusters, even though they do not consider low flows.

Material and methods

Database

This work is based on a detailed analysis of catchment attributes and information contained in hydrological signatures. The CAMELS dataset contains 671 catchment in the continental United States (Addor et al., 2017) with additional meta information such as slope and vegetation parameters. For our study, we used a selection of the available metadata. We excluded all catchments that had missing data, which left us with 643 catchments. Those catchments come from a wide spectrum of characteristics like different climatic regions, elevations ranging from 10 to almost 3600 m a.s.l. and catchment areas ranging from 4 to almost 26 000 km^2. We used the following attributes per class:

- climate: aridity, frequency of high-precipitation events, fraction of precipitation falling as snow, precipitation seasonality;
- vegetation: forest fraction, green vegetation fraction maximum, leaf area index (LAI) maximum;
- topography: mean slope, mean elevation, catchment area;
- soil: clay fraction, depth to bedrock, sand fraction;
- geology: dominant geological class, subsurface porosity, subsurface permeability.

Those catchment attributes were chosen due to their ability to improve the prediction of hydrological signatures (Addor et al., 2018) and because they are relatively easy to obtain, which will allow a transfer of this method to other groups of catchments worldwide.

Table II. 1: *Applied hydrological signatures on the discharge data of the CAMELS data set (Addor et al., 2018).*

Signature	Unit
Mean annual daily discharge	mm d^{-1}
Mean winter daily discharge (Nov. – Apr.)	mm d^{-1}
Mean half-flow date; Date on which the cumulative discharge since October first reaches half of the annual discharge	day of year
95 % Flow quantile (high flow)	mm d^{-1}
Runoff ratio	-
Mean summer daily discharge (May – Oct.)	mm d^{-1}

Hydrological signatures cover different behaviors of catchments. However, many of the published signatures have large uncertainties (Westerberg and McMillan, 2015) and lack in predictive power (Addor et al., 2018). Therefore, we used the six hydrological signatures with the best predictability in space (Table II.1) (Addor et al., 2018). Those signatures were calculated for all catchments. Due to this selection, no signatures that capture low flow behavior were used, as those signatures have a very low spatial predictability.

Data analysis

The workflow of the data analysis considers a data reduction approach with a principal component analysis and a subsequent clustering of the principal components, similar to Kuentz et al. (2017) and McManamay et al. (2014). For the principal component analysis and the clustering, we used the Python package sklearn (0.19.1). The code is available at GitHub (Jehn, 2018b). Validity was checked by also clustering a random selection of 50 % and 75 % of all catchments. This showed that the clustering stayed the same, independently of the number of catchments used (not shown). In all further analysis, we used all catchments to get a sample as large as possible to be able to make statements that are more general.

Calculation of the principal component analysis

The principal components were calculated from the six hydrological signatures described above (Table II.1). We used a principal component analysis on the hydrological signatures to remove correlations between the single hydrological signatures. We only used principal components that together account for at least 80 % of the total variance of the hydrological signatures, which resulted in two principal components. Those two principal components contain the uncorrelated information of all hydrological signatures used and thus can be seen as describers of the hydrological behavior in regard to the overall amount of discharge, its distribution throughout the year, high flows and runoff ratio. Therefore, catchments with similar principal components have similar hydrological behavior along those signatures.

Evaluating the connection between the principal components and the catchment attributes

First, we calculated quadratic regressions between the two principal components and the catchment attributes (with the principal component as the dependent variable). This resulted in one coefficient of determination (R^2) for each pair of principal component and catchment attribute (e.g., PC 1 and aridity).

We then weighted the R^2 by the explained variance of the principal components. This addresses the differences in the explained variance of the principal components (e.g., PC 1 explained 75 % of the variance, PC 2 explained 19 % of the variance).

The weighted coefficients of determination of the two principal components were subsequently added to obtain one coefficient of determination for every catchment attribute.

Quadratic regression was selected as interactions in natural hydrological systems are known to have unclear patterns and can therefore often not be fitted with a simple straight line (Addor et al., 2017; Costanza et al., 1993). This was done first for the whole dataset and then for all clusters separately. This procedure captures the pattern on the catchment attributes in the PCA space of the hydrological signatures (for examples of this pattern see Appendix Fig. II.A1).

Clustering the principal components

The principal components of the hydrological signatures were clustered following agglomerative hierarchical clustering with ward linkage (Ward, 1963), similar to previous studies (Kuentz et al., 2017a; Li et al., 2018; Yeung and Ruzzo, 2001). Therefore, the clusters are based on the hydrological signatures of the catchments. From the previous studies, Kuentz et al. (2017) provides the largest set with over 35 000 catchments. They also clustered their catchments in a PCA space of a range of hydrological signatures. To select the number of clusters, they used the elbow method (and two other methods to validate their results) and found that 10 or 11 clusters (depending on the method) were most appropriate for their data. Due to the similarity in the clustered data and the larger database of Kuentz et al. (2017), we also used 10 clusters (Berghuijs et al., 2014) also found that 10 clusters captured the distinct hydrological behaviors for the continental US. Those 10 clusters represent groups of catchments with distinctly different hydrological behavior.

Results and discussion

Catchment attribute correlations in the CAMELS dataset

Usually the 100th meridian is seen as the dividing climatic line in the US, splitting the country into a semiarid west and a humid east. We assume that this difference in climate also has implications for the hydrology and the overall catchment attributes in those regions. To quantify this we split the CAMELS dataset into a western and an eastern part, based on the 100th meridian (Figs. II.1 and II.4). This shows that many of the catchment attribute correlations do not differ much between the east and the west. In most cases (> 80 %), Spearman rank correlation coefficients vary by less than 0.4 (Fig. II.1c). Still, there are some catchment attributes with larger differences of up to 0.8 between both regions. Most striking are the mean elevation and the fraction of the precipitation falling as snow as well as the vegetation attributes LAI maximum and green vegetation fraction maximum. Even though these attributes are directly related to each other through temperature gradients, they differ substantially in both parts of the country. In the mountainous western US, elevation is highly correlated with the fraction of precipitation falling as snow (r=0.8), while it is not in the eastern US (r=0.4). This and the different correlations between vegetation and elevation are probably caused by the fact that the temperature gradients differ in both regions. The western US is much more mountainous and thus temperatures typically change with elevation. In the more level eastern US, the change in temperature is mainly linked to the latitude. Striking are also the changes of correlation with regard to the fraction of precipitation falling as snow. Here we find altered directions of the correlation; i.e., positive correlations with LAI maximum and frequency of high-precipitation events in the east turn to negative ones in the west. The change in the LAI maximum might be linked to the higher elevations in the west, as in higher elevations less vegetation is growing, but more snow falls. It also becomes obvious that all three measures of vegetation seem to track similar characteristics in the catchments, as they correlate highly with each other (especially in the eastern US with r=0.9). In addition, all vegetation attributes depict a large negative correlation with aridity. Hence, the vegetation attributes considered are likely good proxies for aridity. Overall, we see that the relations between the catchment attributes are quite similar for the eastern and western US, with the exception of the mean elevation, snow and the LAI maximum.

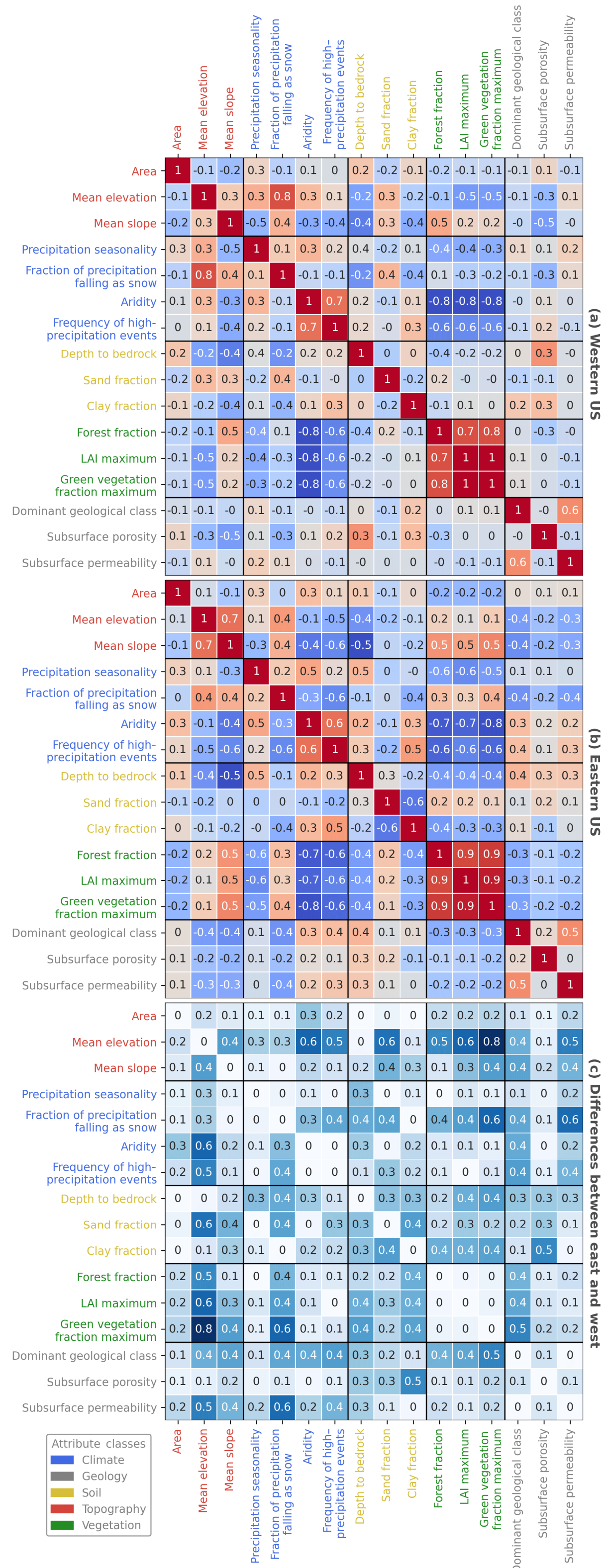

Figure II.1: *Spearman rank correlation coefficients given for all catchment attributes in the western* ***(a)*** *and eastern* ***(b)*** *US. Absolute differences of the correlation coefficients between the eastern and western US are given in* ***(c)****. Eastern and western is defined by the 100th meridian. Due to rounding effects, correlations with the same Spearman rank correlation coefficient might show slightly varying color codes.*

Impacts of catchment attributes on discharge characteristics in the whole dataset

Next we examined the weighted R^2 of the catchment attributes for the whole dataset. This analysis shows not only differences in their score between the single attributes, but also between the different classes of catchment attributes (Fig. II.2). Attributes related to climate (aridity) and vegetation (forest fraction) get the highest scores. However, it should be noted that all vegetation catchment attributes show a strong correlation with the aridity (Fig. II.1) and thus capture similar trends, in both the east and the west. With the exception of the mean slope, the first seven catchment attributes are all related to climate and vegetation. The last seven attributes on the other hand are all related to soil and geology, except the catchment area. They also show much lower scores of the weighted R^2. This indicates that soil and geology are less important for the chosen hydrological signatures. Similar patterns were also found by Yaeger et al. (2012). They stated climate as the most important driver for the hydrology. As the correlations between the catchment attributes showed that the climate and the vegetation attributes are highly correlated (Fig. II.1), it can be assumed that climate is the most important factor overall, with aridity and high-precipitation events being most important within the climate attributes.

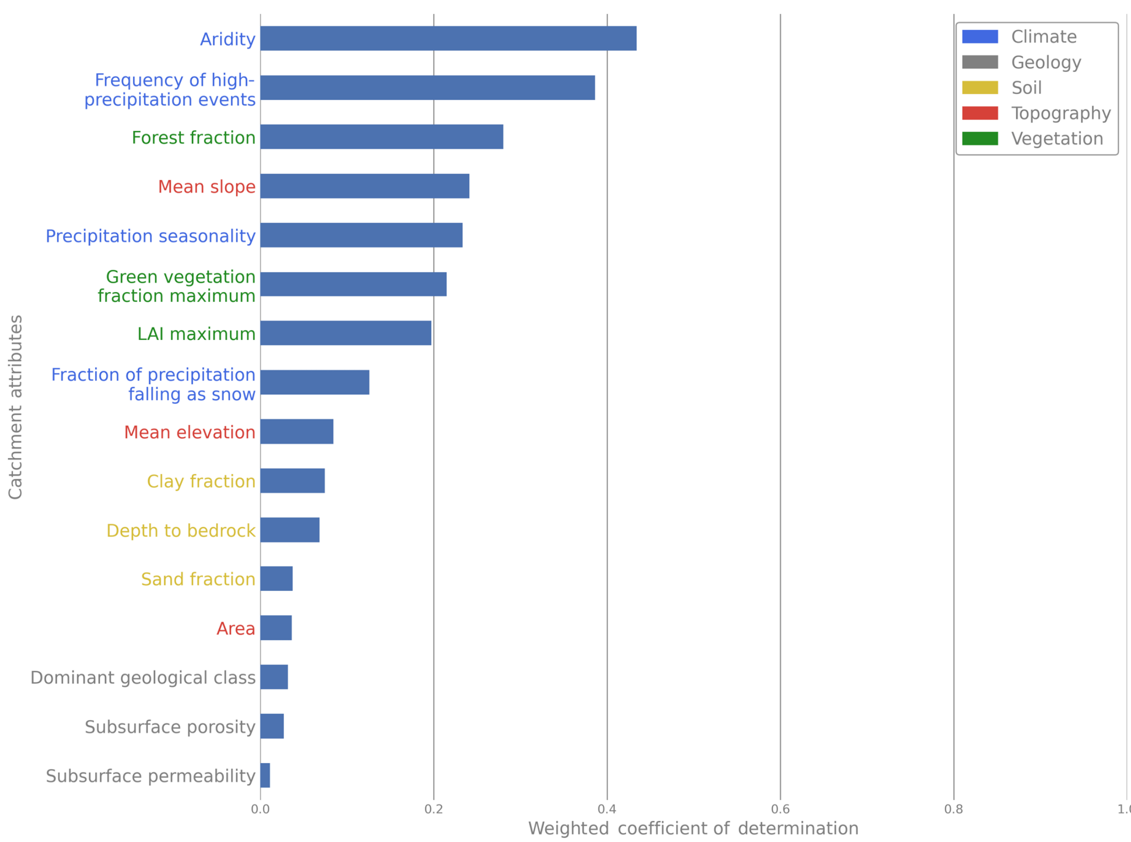

Figure II.2: *Importance of catchment attributes evaluated by quadratic regression for all considered catchments. Attributes colored according to their catchment attribute class.*

However, Yaeger et al. (2012) also unraveled that low flows are mainly controlled by soil and geology. The minor importance of soil and geology in our study might therefore be biased by the choice of hydrological signatures, which excluded low flow signatures due to their low predictability in space. Nevertheless, our study probably captures a more general trend as we used a larger dataset and hydrological signatures that vary more gradually in space (Addor et al., 2018). Addor et al. (2018) also explored the influence of different catchment attributes in the CAMELS dataset on discharge characteristics. They found that climate has the largest influence on discharge characteristics, well in agreement with Coopersmith et al. (2012). The latter also used a large group of catchments in the continental United States from the MOPEX dataset. They conclude that the seasonality of the climate is the most important driver of discharge characteristics. While the seasonality is still important in our analysis, the aridity is an even stronger factor. However, Coopersmith et al. (2012) only analyzed the flow duration curve, which has a mediocre predictability in space, and it is therefore less clear what it really depicts (Addor et al., 2018). Overall, this study here is in line with other literature in the field. Using the weighted R^2 reliably detects climatic forcing as the most important of the discharge characteristics for a large group of catchments.

Relation of the principal components and the hydrological signatures

The rivers considered in this study show a wide range of hydrological signatures. This is visible in the clusters of principal components of the hydrological signatures (Fig. II.3). Most of the rivers are opposite to the loading vectors (the loading vectors are shown as arrows). This shows that most rivers have relatively low values for all hydrological signatures and only some more extreme rivers have higher values for specific hydrological signatures. Most typical for the overall behavior of the river are the hydrological signatures mean annual discharge and Q95 (high flows), as they have a strong correlation with the first principal component. For the second principal component, the mean half-flow date has the highest correlation. Therefore, the first principal component can be seen as a measure of overall discharge and amount of high flows. Overall, it can also be seen that most of the rivers show a relatively similar behavior (Clusters 1, 2, 8, 9, 10), while smaller groups of rivers tend to deviate from that by having a more extreme behavior (Clusters 3, 5, 7). The remaining Clusters 4 and 6 are located between those extremes. This pattern also explains the different sizes of the clusters. While most catchments behave relatively similar, only some show extreme behavior and thus the clusters with extreme catchments are smaller.

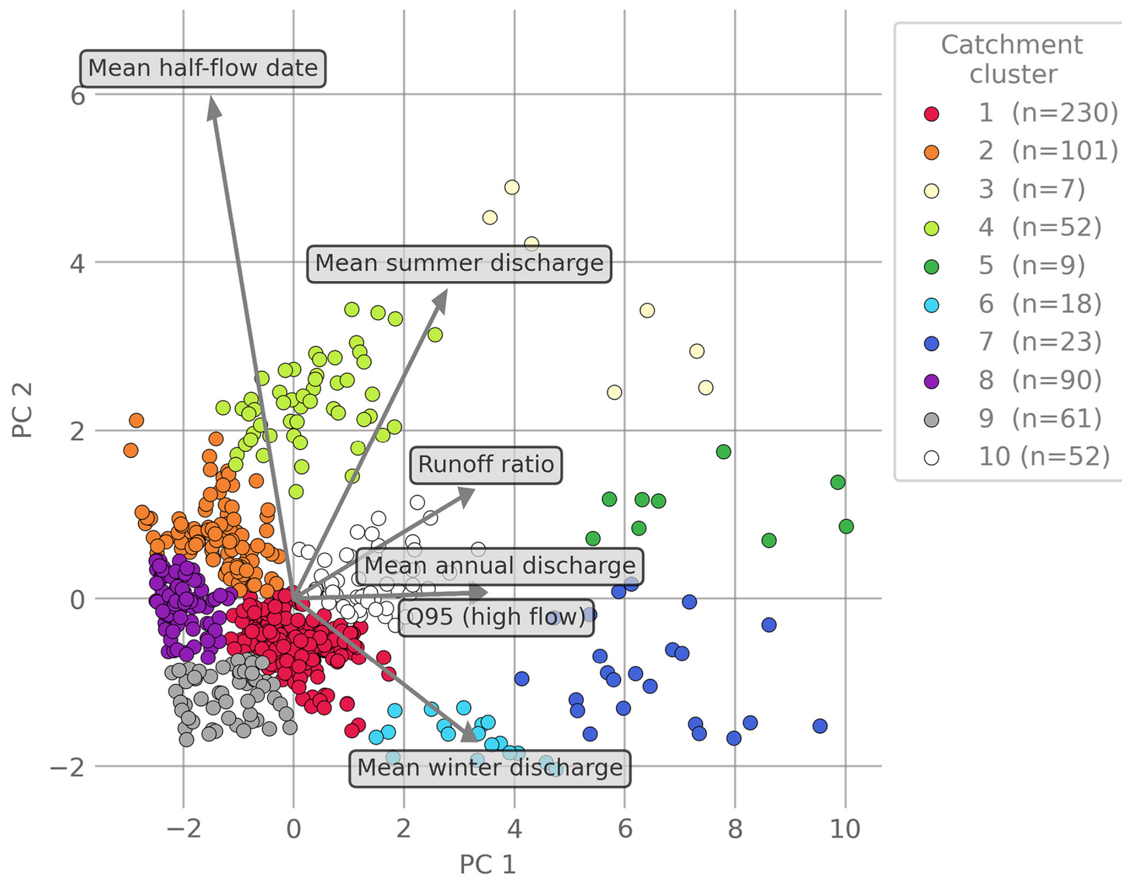

Figure II.3: *Biplot of the principal components (PCs). Colors indicate the cluster of the catchment. Grey arrows indicate the loadings of the original catchment attributes in the PCA space.*

Location and properties of the catchment clusters

The catchment attributes in the CAMELS and similar large-scale datasets often show a pattern that resembles climatic zones (Addor et al., 2018; Coopersmith et al., 2012; Yaeger et al., 2012). For the catchment clusters presented here, we can see that most of the clusters roughly follow ecoregions in the US (Fig. II.4). Clusters 1, 4, 6 and 7 in particular are almost entirely located within one ecoregion. Cluster 2, 8 and 9 on the other hand follow those ecological boundaries to a lesser degree.

We can see a split of the clusters along the 100th meridian. Clusters 3, 4, 5, 6 and 7 are located mainly in the west, while Clusters 1 and 10 are mainly found in the east. However, the remaining Clusters 2, 8 and 9 have roughly similar numbers of catchments in both regions. Overall, the catchments in the eastern half of the United States form large spatial patterns of similar behavior, while the catchments in the west are patchier. This same pattern can also be seen in some of the signatures used by Addor et al. (2018). In particular, the runoff ratio and mean annual discharge form very similar patterns to the clusters in this study.

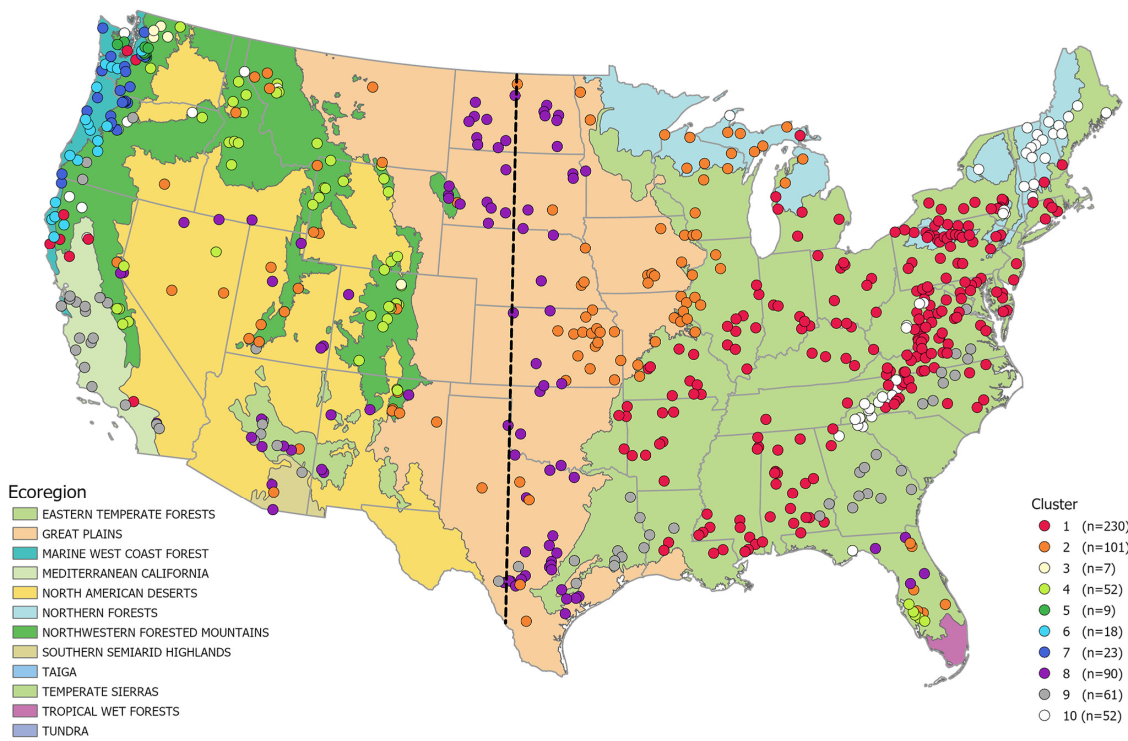

Figure II.4: *Locations of the clustered CAMELS catchments and level I ecoregions (Omernik and Griffith, 2014) in the continental US. Dotted line marks the 100th meridian. An interactive version of this map can be found at https://zutn.github.io/Catchment-Classification/map.html (last access: 26 February 2020).*

In addition, similar catchments can be quite far away from each other (Fig. II.5). Sometimes, the catchment with the most similar signature was found as far as 4000 km away (almost the entire longitudinal distance of the continental US). This explains why spatial proximity seems to be important in some studies that look into explanations of catchment behavior (Andréassian et al., 2012; Sawicz et al., 2014), but not in others (Trancoso et al., 2017). This also indicates that clustering by using spatial proximity might only work in regions like the eastern US, where the behavior of rivers changes only gradually, due to uniform climate that only changes gradually as well. The finding that the most similar catchment (based on their hydrological signatures) can be far away also explains the behavior of clusters that contain catchments quite distant from each other (e.g., Cluster 4). Even though the catchments might be far away from each other, the interplay of different catchment attributes and driving factors, including sometimes very different climates, can lead to similar (equifinal) discharge behavior, concerning the overall amount of discharge, its distribution in the year, the high flows and the runoff ratio. This was also found by several other studies (Berghuijs et al., 2014; Knoben et al., 2018; Kuentz et al., 2017a).

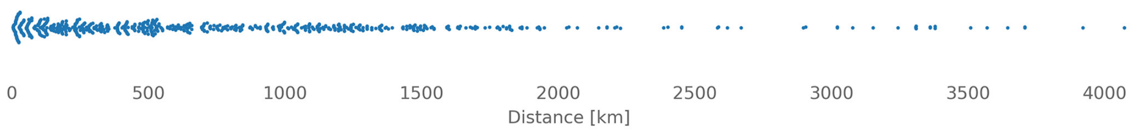

Figure II.5: *Swarm plot of the real-world distances of all catchments to the most hydrologically similar catchment (based on their distance in the PCA space of the hydrological signatures).*

In the following, we describe the catchment clusters in regard to their characteristics in meteorology (Fig. II.6), attributes (Fig. II.7), hydrology (Fig. II.8) and location (Fig. II.4). The main points of this description are summarized in Table II.2. A list of all catchments with index, position, cluster classification and climate indices is given in the Supplement.

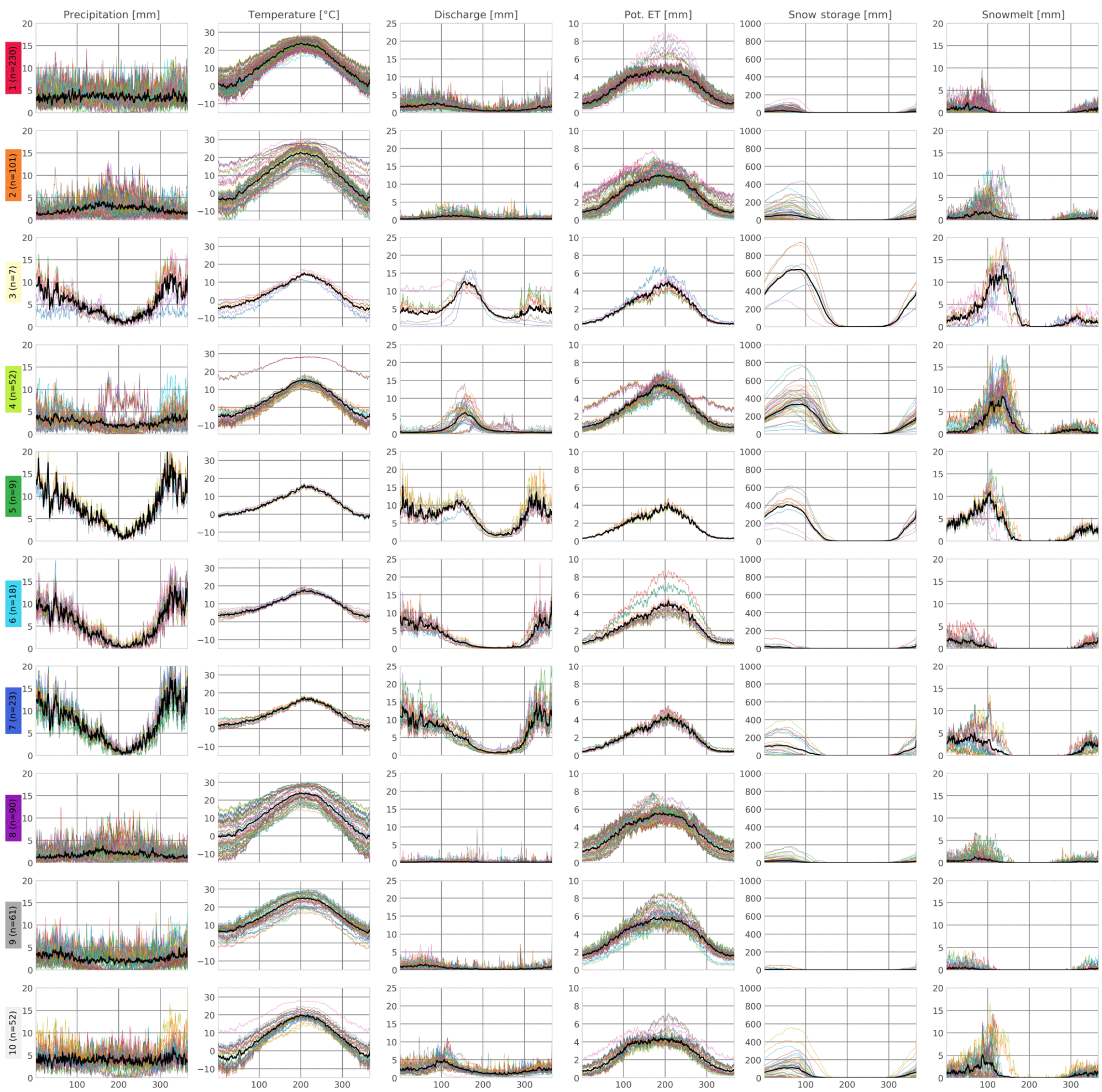

Figure II.6: *Meteorological attributes of the clustered CAMELS catchments averaged by day of the year. Potential evapotranspiration (Pot. ET) was calculated with Hargreaves–Samani (Samani, 2000). Snow storage and melting was calculated using a temperature-based approach described in*

Massmann (2019). Black lines indicate the mean of all cluster members. Colored lines represent the individual catchments.

Cluster 1 is defined by a dense vegetation cover (Fig. II.7). The low elevation of those catchments results in little annual snowfall. They are mainly located in the southeastern and central plains and therefore get relative high rainfall (> 1000 mm per year) (Fig. II.4), almost uniformly distributed over the year (Fig. II.6). Still, they produce only a small amount of discharge. This cluster contains the highest number of catchments (*n*=230). So over one-third of the catchments in CAMELS show a relatively similar behavior when it comes to the amount of water fluxes and their distribution throughout the year. This is particular visible when we look at the annual supply of discharge (Fig. II.6). Even though the cluster contains a large number of catchments that also partly differ a lot in their potential evapotranspiration, there is only a minor difference in the amount of discharge and its seasonality.

Cluster 2's most typical attribute is its high-precipitation seasonality. However, concerning most other catchment attributes, Cluster 2 is undefined as it contains catchments of most regions of the continental US (with a concentration in the eastern Great Plains) (Fig. II.4). The hydrological signatures on the other hand show a clearer pattern. Here, the mean winter discharge, Q95 and the mean annual discharge have a narrow range (Fig. II.8). This shows that catchments with very different attributes can produce similar discharge characteristics. The different attributes seem to cancel each other out in their influence on the discharge. This might be enhanced by the high-precipitation seasonality with higher precipitation in the summer, which creates a strong climatic forcing and thus a narrow range for the hydrological signatures (Fig. II.6). This cluster differs from the first one, by having even lower discharge, with almost no peaks and a higher influence of snowmelt.

Cluster 3 is the smallest cluster, with only seven catchments. Those are all located in the Northwestern Forested Mountains. Their most distinct feature is their strong negative precipitation seasonality (indicating a strong precipitation peak in the winter) (Figs. II.6, 7). They also experience high-precipitation events (mostly as snow). Hydrologically, their most distinct features is the very high mean summer discharge and high runoff ratio (Fig. II.8). This is probably caused by the large amounts of snowmelt in late spring and early summer. The catchments of Cluster 3 have the largest snow storage in the dataset, with a mean maximum value of over 600 mm. Overall, the catchments in this cluster seem to be, from a hydrological point of view, the most extreme in the overall CAMELS dataset. This can be seen in their varying discharge patterns. The uniting pattern is their large peak discharge during summer and their extreme values in the PCA space (indicating much higher values for the hydrological signatures in comparison with the other catchments) (Fig. II.3).

Cluster 4 is, like Cluster 3, located in the Northwestern Forested Mountains, with the exception of four catchments that are located in Florida (Fig. II.4). This cluster is another example of different catchment attributes being able to create similar discharge characteristics concerning the signatures used, while having very different catchment attributes (Fig. II.6). The catchments have overall low discharge and few high flow events, except one large peak in the middle of the summer, which is caused by melting snow in the northern catchments and strong rainfalls in Florida. Their catchment attributes vary widely, especially in all attributes that are related to elevation (e.g., fraction of precipitation falling as snow) (Fig. II.7), which is to be expected when some of the catchments are located close to the sea in the southeast, while others are mountainous.

Cluster 5 includes only few catchments (n=9), which are all located at regions in the northern part of the Marine West Coast Forests (Fig. II.4). This is the region in the continental US that receives the highest precipitation (> 2000 mm year), which is reflected in its discharge characteristics (Figs. II.6, II.8). These catchments have the highest discharge in the whole dataset, especially in the early summer, due to a combination of high precipitation and snowmelt. They also experience only few high-precipitation events as they receive large amounts of rain and snow most of the year, with a distinct very high peak in the winter months. They further depict an additional discharge peak in late spring–early summer that separates them from the other catchments found at the west coast. The catchments are almost 100 % covered by forest.

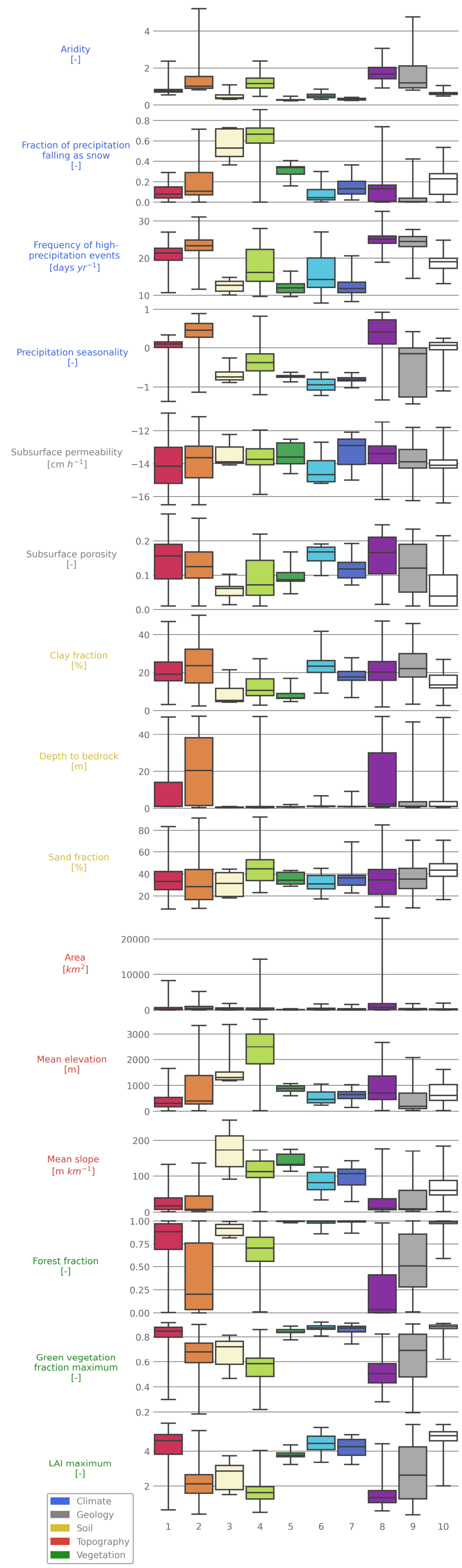

Figure II.7: *Boxplots of the catchment attributes of the clusters.*

Cluster 6 is located in the Marine West Coast Forest, but in contrast to Cluster 5, it covers the whole region and not only the northern part (Fig. 4). The catchments are very similar in their attributes and discharge characteristics to Cluster 5, with the exception of lower discharges and runoff ratios (Fig. 7, 8). This is caused by slightly lower precipitation in comparison with Cluster 5. Cluster 6 experiences the most negative precipitation seasonality across all clusters, with almost all precipitation falling in the winter month. Due to this seasonality and the lower precipitation in the summer, the catchments of this cluster uniformly dry out almost completely in late summer (Fig. II.6).

Cluster 7 is also located in the same region as Clusters 5 and 6 (Marine West Coast Forests) (Fig. II.4). In terms of the catchment attributes and the discharge characteristics, it is between Clusters 5 and 6. So, Clusters 5 to 7 all cover the same region and differ in their mean summer discharge, which is caused by variations in elevation and location (Fig. II.7). Cluster 7 has higher subsurface permeabilities than Cluster 6, which might explain the differences in hydrological behavior, even though the overall attributes of both clusters are rather similar. For example, Cluster 7 has an overall lower discharge than Cluster 5, but does not dry out during the summer as Cluster 6 does (Fig. II.6). This might be due to the larger amount of snow it receives in comparison with Cluster 6 and its lower evapotranspiration.

Cluster 8 is the most arid cluster (Fig. II.7). All of the catchments are located in western parts of the Great Plains and in the North American deserts (Fig. II.4). They are characterized by an overall low water availability and high evaporation, which is shown in the very low mean annual discharge and runoff ratio (Figs. II.6, II.8). This also results in low values for the LAI. Yet, the frequency of high-precipitation events is high. However, those high-precipitation events are only high in comparison with the mean precipitation for those catchments and not the overall range of precipitation in the entire CAMELS dataset.

Cluster 9 covers all southern states of the United States (Fig. II.4). The catchments here are quite similar to Cluster 8, but show a lower precipitation seasonality and a higher forest cover and green vegetation (Fig. II.7). In addition, all catchments of this cluster are in relative close proximity to the sea. The uniting factor in this cluster seems to be the very low snow fraction and the high evapotranspiration (Figs. II.6, II.7).

Cluster 10 catchments are all located in the Appalachian Mountains (Fig. II.4). The mean elevation is higher than that of most other clusters and the catchments have a low aridity and a very high forest cover (Fig. II.7). Their discharge characteristics are similar to that of the Marine West Coast Forests (Clusters 5 to 7; Figs. II.6, II.8). However, they receive less water than those catchments. Cluster 10 covers the same ecoregion as Cluster 1, but has a distinct behavior due to its mountainous character, which can be seen in the higher seasonality of the discharge. This is probably caused by the larger snow cover, with a discharge peak in spring due to snowmelt.

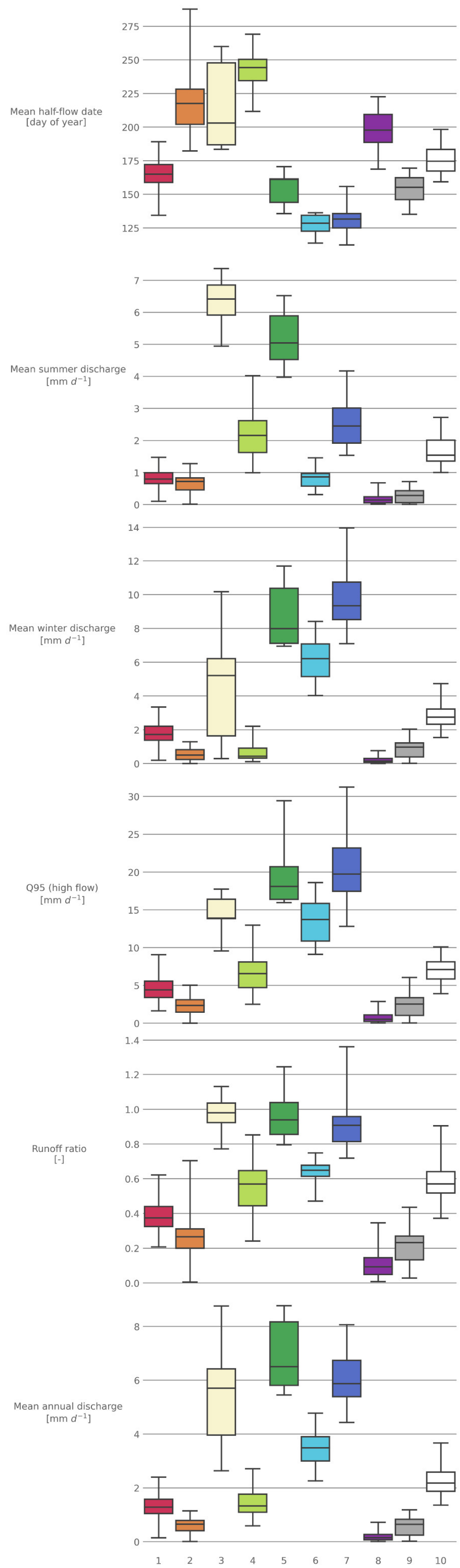

Figure II.8: *Boxplots of the hydrological signatures of the clusters.*

Overall, we can see similar trends for some of the clusters. The general similarities of the clusters are also represented by their distance and position in the PCA space (Fig. II.3). We identified four distinct groups:

- Group 1 (Clusters 1, 2, 8, 9): low seasonality in precipitation and discharge; located in the eastern US; due to low slope inclinations, water takes a long time to reach the outlet.
- Group 2 (Clusters 3, 4): dominant summer peak of discharge caused by rapid snowmelt; mostly located in the mountains of the western US; differ in precipitation inputs.
- Group 3 (Clusters 5, 6, 7): located in the Northwestern Forested Mountains; characterized by high-precipitation amount and seasonality, but more or less extreme versions.
- Group 4 (Cluster 10): located in the Appalachian mountains; share characteristics with Group 1, though influenced by higher elevations and steeper slopes.

Those groups of clusters are similar to the ones found by Berghuijs et al. (2014), even though they used a very different method to derive them. The main difference in the groups is probably caused by how we structure the clusters and groups in the eastern US, due our clusters being more influenced by the Appalachian Mountains. However, both approaches deliver similar results overall.

The question remains: what is the right numbers of clusters? Though we did find four distinct groups, having only four clusters would probably be too little, as the clusters in the groups show a wide range of behaviors (Figs. II.3, II.7, II.8, Table II.2). There are catchment attributes which we did not take into account but which could further split up the clusters (e.g., the shape of the catchments). However, this study considered the catchment attributes that are usually considered to be important. The fact that the clusters contain different numbers of catchments can be explained by their distances in the PCA space (Fig. II.3). Many of the catchments are rather similar. This produces some clusters which contain most of the catchments. However, we also have some extreme catchments (e.g., Clusters 3 and 5), which are very different to the bulk of the catchments in the CAMELS dataset. Thus, even though some of our presented clusters are quite small in number, they are needed to capture their extreme hydrological behavior. It can also be seen that for most of the clusters there is no clear dividing line to neighboring clusters. Therefore, it might be useful to use fuzzy clustering approaches in future research, to avoid those strict boundaries in a continuous space. Our results show that some of the clusters follow the boundaries of the ecoregions in the US very directly (Cluster 1), while others do not (Cluster 9). The worlds of ecology and hydrology are sometimes shaped by the same forcing, but not always.

Table II.1: *Properties of the catchment clusters. Typical signatures/attributes refers to the signature/attribute of the cluster with the lower coefficient of variation scaled by the mean coefficient of variation of the whole dataset. Dominating attribute refers to the catchment attribute that has the highest weighted R².*

Cluster	n	Main Region	Typical signature	Typical attribute and their manifestation	Dominating attribute
1	230	Southeastern and Central Plains	Low mean winter discharge	Low aridity	Aridity
2	101	Central Plains (with scattered catchments all over western US)	High mean half-flow date	High precipitation seasonality	Green vegetation fraction maximum
3	7	Northwestern Forested Mountains	High mean summer discharge	Low precipitation seasonality	Fraction of precipitation falling as snow
4	52	Northwestern Forested Mountains and Florida	High mean half-flow date	Mid frequency of high precipitation events	Precipitation seasonality
5	9	Northern Marine West Coast Forests	High mean summer discharge	Very high forest fraction	Forest fraction
6	18	Marine West Coast Forests	Mid runoff ratio	Low precipitation seasonality	Aridity
7	23	Western Cordillera (Part of Marine West Coast Forests)	High mean winter discharge	Low precipitation seasonality	Fraction of precipitation falling as snow
8	90	Great Plains and North American Deserts	Mid mean half-flow date	High frequency of high precipitation events	Precipitation Seasonality
9	61	All southernmost states of the US	Low mean half-flow date	High frequency of high precipitation events	Aridity
10	52	Appalachian Mountains	Low mean winter discharge	High forest fraction	Mean elevation

Importance of the catchment attributes in the clusters

The individual importance of the catchment attributes in the clusters is variable and partly deviates from the order of importance in the overall dataset (compare Figs. II.2 and II.9). For Clusters 1 (Southeastern and Central Plains), 6 (Marine West Coast Forests) and 9 (coastal states) aridity has the highest weighted coefficient of determination in the clusters. For Clusters 3 (Northwestern Forested Mountains) and 7 (Western Cordillera) the highest relevance is found for the fraction of precipitation falling as snow. For the remaining clusters it is precipitation seasonality (Cluster 4, Northwestern Forested Mountains, and Cluster 8, Great Plains and Deserts), the green vegetation fraction maximum (Cluster 2, Central Plains) and the mean elevation (Cluster 10, Appalachian Mountains). We can also see that some clusters have one dominating catchment attribute (investigated by the coefficient of determination, e.g., aridity in Cluster 1; see Fig. II.9), while for other clusters, all attributes seem equally important (e.g., Cluster II.8). Overall, the western clusters (west of the 100th meridian) display the highest weighted R^2 with the following:

- fraction of precipitation falling as snow (Clusters 3, 7),
- precipitation seasonality (Cluster 4),
- forest fraction (Cluster 5),
- aridity (Cluster 6).

Eastern clusters (east of the 100th meridian) display the highest weighted R^2 with the following:

- aridity (Cluster 1),
- mean elevation (Cluster 10).

Clusters equally present in west and east display the highest weighted R^2 with the following:

- green vegetation fraction maximum (Cluster 2),
- aridity (Cluster 9),
- precipitation seasonality (Cluster 8).

Keeping the correlation coefficients displayed in Fig. II.1 in mind, we see that climate is the most important factor in almost all clusters, as the vegetation attributes are highly correlated with the climate attributes. The only exception is Cluster 10, in which mean elevation is the most important catchment attribute. However, the catchment attributes in Cluster 10 have overall low R^2 values and the mean elevation is directly followed by the aridity. This again shows that climate seems to be the dominating factor for catchment behavior, as found in other large-sample studies (e.g., Berghuijs et al., 2014; Kuentz et al., 2017). Nevertheless, if one takes a closer look at the dataset, more detailed, regional correlations with regard to individual climate variables can be determined. For example, Cluster 1 is defined by the aridity, while Cluster 4 seems to be much more influenced by the precipitation seasonality. Overall, it is feasible to link dominating catchment attributes to the hydrological behavior. While it is straightforward in some regions of the US, it is more challenging in others. We link this to the signal of the climatic forcing being more superimposed by other catchment attributes, which results in a less clear connection between its hydrological behavior and the climate. This hints that climate and catchment attributes are more intertwined in those areas and indicates regions where different types of hydrological runoff generation processes exist. Furthermore, it indicates regions where hydrological predictions in ungauged basins (Hrachowitz et al., 2013a) can become very challenging, as the interplay of the available meteorological data and catchment attributes cannot sufficiently explain the hydrological characteristics. Those findings also highlight one current discrepancy between large-sample and single-catchment studies. While large-sample studies, especially the very large ones, identify climate as being most important for the hydrological behavior (Addor et al., 2018; Kuentz et al., 2017), smaller-sample studies (Chiverton et al., 2015; Pfister et al., 2017) and single-catchment studies (Floriancic et al., 2018) often identify the geology or soils as being very important. This might be linked to the overall problem of scales in hydrology, as different scales of soil/geology and climate have different effects and varying data accuracy (Addor et al., 2018; Blöschl, 2001). In addition to this, the overall scale might also come into play. Smaller sample studies often compare catchments that are not far away from each other and probably have similar climate forcings. Thus, the differences in hydrological behavior can only be caused by catchment attributes other than climate. Therefore, larger and smaller sample studies might be looking at different things. While very large-sample studies capture what drives catchments on large scales (the climate), smaller studies look at how this climatic signal is transferred to discharge by the catchment attributes.

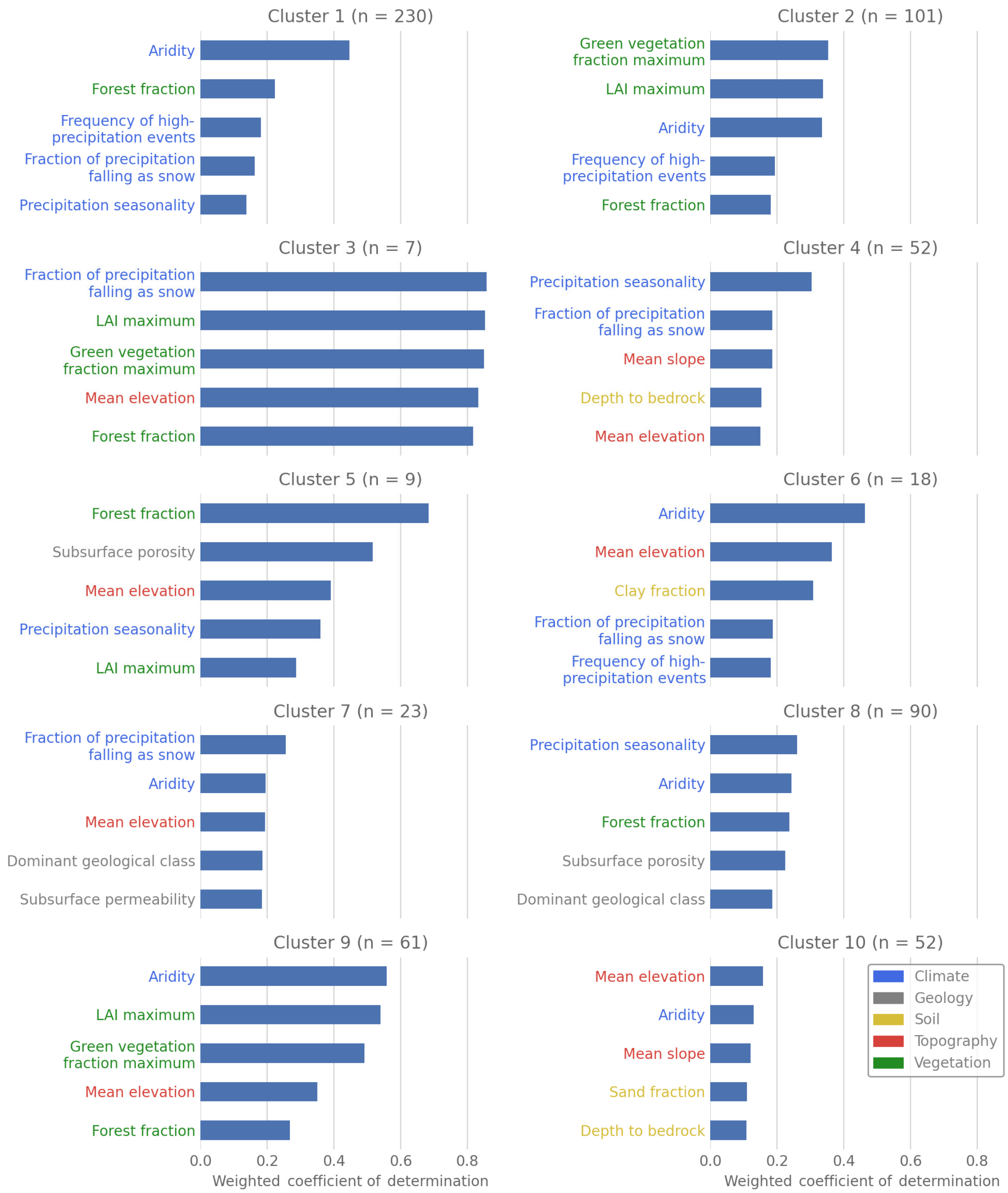

Figure II. 9: *Importance of the catchment attributes evaluated by the quadratic regression for the catchment clusters. Attributes colored according to their catchment attribute class.*

Differences in clusters in comparison with other hydrological clustering studies

The results of this study show some similarities with the clustering results of Kuentz et al. (2017), who derived their cluster from European catchments by an analogous method. Like them, this study here also found one cluster (Cluster 2) that does not have any distinct character. However, only around one-sixth of the CAMELS catchments belongs to this Cluster 2, while one-third of the catchments in the study by Kuentz et al. (2017) were in a cluster without distinct features. Therefore, our selection of hydrological signatures seems to allow a better identification of hydrological similarities. However, all catchments in CAMELS are mostly without human impact (Addor et al., 2017), while many catchments in the study of Kuentz et al. (2017) are under human influence. This human influence might mask otherwise apparent patterns. Kuentz et al. (2017) also found two clusters that contain mostly mountainous catchments. These show a similar behavior to Cluster 3 (Northwestern Forested Mountains) and Cluster 10 (Appalachian Mountains) (Fig. II.4). The main difference between their findings and this study here is Cluster 8, as it contains very arid catchments (with some being located in deserts). Obviously, this cluster cannot be found in Europe as Europe has no real deserts. Still, there is some similarity with their cluster of Mediterranean catchments as both are dominated by aridity. Summarizing, in their study and this study catchments are mainly clustered in groups of desert/arid catchments, mountainous catchments, medium-height mountains with high forest fraction, wet lowland catchments, and one cluster of catchments that does not show a very distinct behavior and therefore does not fit in the other clusters (Table II.2). One possible explanation for this unspecific behavior might be that many catchments have one or two important attributes that dictate most of their behavior, but which are different from other cluster members. For example, desert catchments are relatively easy to identify, as they are dominated by high energy and little precipitation. A European upland catchment on the other hand has several more influences such as snow in the winter, high energy in the summer, varying land use and a strong impact of seasonality. Here, many influences overlap each other and thus make it difficult to identify a single cause; see also the discussion by Trancoso et al. (2017) that goes in a similar direction. Those overlapping influences are probably also the reason why catchment classification studies often find one or two clusters that include a large number of catchments, while most other clusters only contain a few catchments (Coopersmith et al., 2012; Kuentz et al., 2017a). Therefore, it is quite difficult to confirm the "wish" of the hydrological community to have homogenous catchment groups with only a few outliers (e.g., Burn, 1997), because catchments are complex systems with a high level of self-organization arising from co-evolution of climate and landscape properties, including vegetation (Coopersmith et al., 2012). Accordingly, it requires many separate clusters to separate those multi-influence catchments into homogenous groups. This hints that for future research a fuzzy clustering approaches might provide less ambiguous results, as it respects the continuous nature of hydrological behavior. Still, the cluster found here might capture much of the variety present in the United States, as they roughly follow ecological regions (McMahon et al., 2001), which has been stated as a sign of a good classification (Berghuijs et al., 2014). In addition, this study shows that using clusters derived from principal components of hydrological signatures creates meaningful groups of catchments with similar attributes (Figs. II.6, II.7, II.8). Those clusters also show distinct spatial patterns (Fig. II.4). Similar results were also found in other studies that used the same method (Kuentz et al., 2017a; McManamay et al., 2014) but based them on partly different hydrological signatures. Therefore, the principal components of hydrological signatures can be used as a measure of similarity between catchments. They represent the "essence" of all hydrological signatures used. Our results also show that it is difficult to link those catchment clusters to simple averaged measures of catchment attributes. While some clusters have very clear connections to the attributes, others have no catchment attribute that could easily explain the behavior of the catchments. This hints that some catchments are easier to explain (in a hydrological sense) than others. Those difficulties might be an artifact of the averaged catchment attributes or be caused by a complex catchment reaction, forced by intertwined climate and catchment attributes, which in turn might indicate an equifinality of catchment response.

Comparing catchment clusters based on hydrological behavior and climate

Besides hydrological behavior, climate is often used to sort catchments into similar groups (e.g., Berghuijs et al., 2014; Knoben et al., 2018). Therefore, we are interested if both approaches deliver comparable results. To evaluate this, we contrasted our results to the commonly used Köppen–Geiger climate classification (Beck et al., 2018) (Fig. II.10) and recently published approach of Knoben et al. (2018), who sorted climate along three continuous axes of aridity, seasonality and fraction of precipitation falling as snow (Fig. II.11). The resulting clusters based on climate and hydrology should be the same, if climate is the dominating driver of hydrological behavior in every catchment. Yet, this is not the case for the Köppen–Geiger classification. In every hydrological cluster are at least two different climates regarding the Köppen–Geiger classification, ranging up to eight different climatic regions for Clusters 2 and 8 (those even include deserts and very cold regions). Thus, the Köppen–Geiger classification seems unable to capture the essential drivers of hydrological behavior, a critique also raised in other studies (Haines et al., 1988; Knoben et al., 2018).

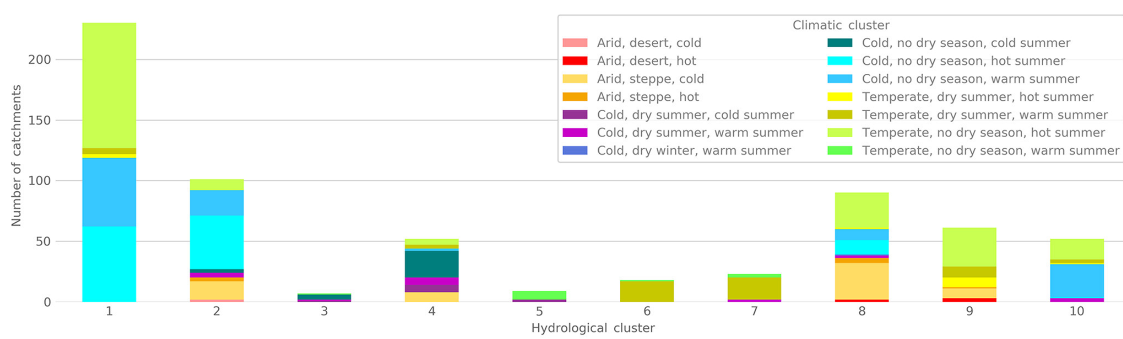

Figure II.10: *Membership of Köppen–Geiger clusters (Beck et al., 2018) in the hydrological clusters.*

The picture is less clear concerning the climatic index space of Knoben et al. (2018) (Fig. II.11a). Due to the continuous nature of the approach of Knoben et al. (2018), there are no clear boundaries as in the Köppen–Geiger classification. Still, there are some emerging patterns. For example, according to the approach of Knoben et al. (2018) Cluster 1 is mainly defined by a relatively arid climate, with some seasonal variability and little to no snow. This is in line with our analysis of the most influential catchment attributes for this cluster, as we identified aridity as the main driver. There seem to be regions where the forcing signal of the climate is transferred more directly to a streamflow response than in others. However, this does not mean that climate is unimportant in those regions. Either the climate forcing signal is changed more through other attributes of the catchment, or the mean values describing the climate do not properly reflect the variability of the climate in the single catchments. This leads to a less clear correlation between the climate and the hydrological behavior. Interestingly, when we look at the single hydrological signatures in the climate index space (Figs. II.11b, II.A2) we see a very clear connection between the single hydrological signatures and the climate. This direct connection of the signatures used was also found by Addor et al. (2018). Our results and the comparison show that the complex hydrological behavior, captured in a range of hydrological signatures, does not simply follow the climate only, even though the individual signatures do. Still, all signatures combined seem to capture a dynamic which is climatic in origin but is shaped through the attributes of the catchments (like vegetation and soils Berghuijs et al., 2014). Therefore, to find truly similar catchments, using climate characteristics only is probably not sufficient (see also Addor et al., 2018; Knoben et al., 2018; Kuentz et al., 2017).

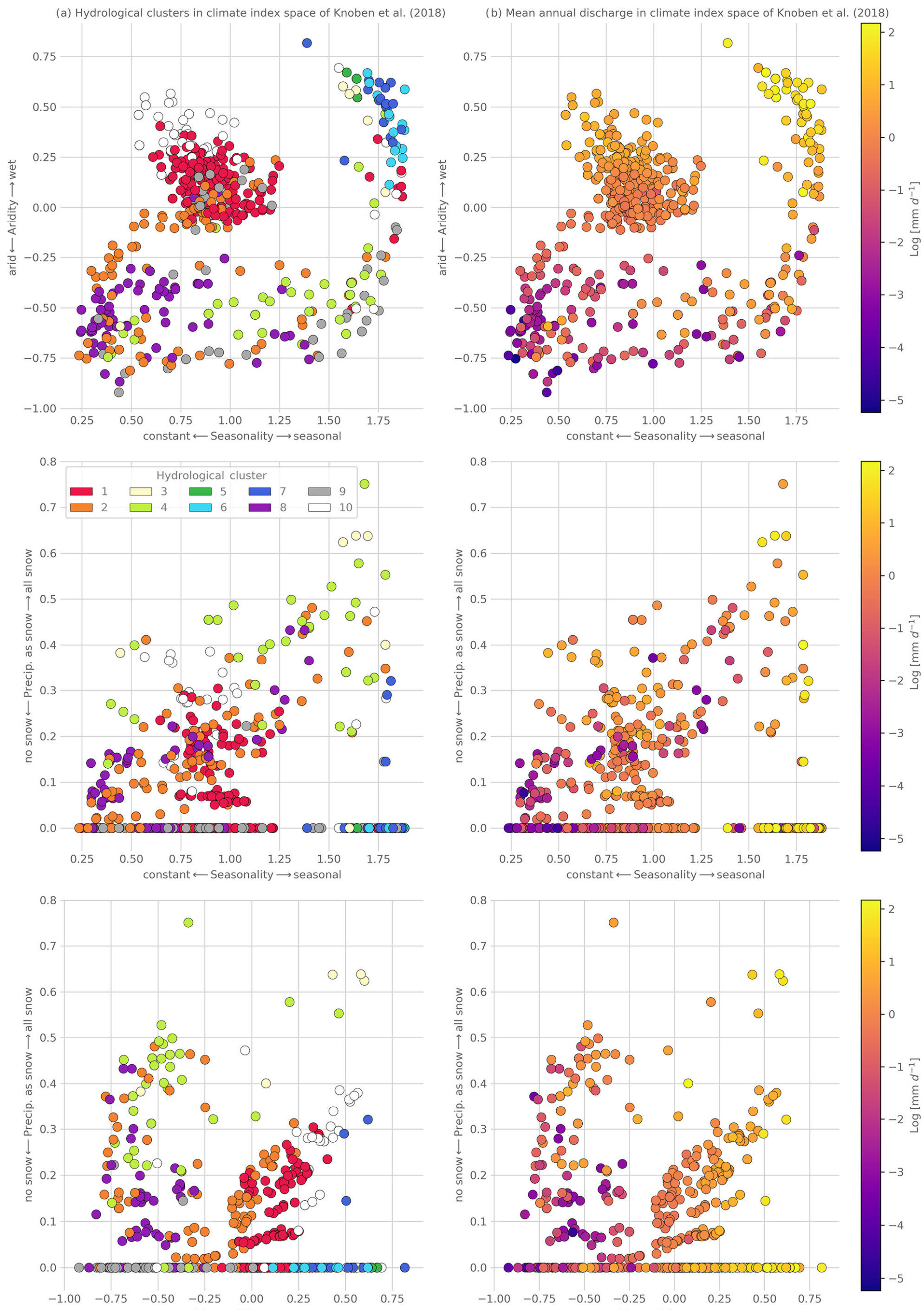

Figure II.11: (a) *Comparison of the hydrological clustering of this study with the climate index space of Knoben et al. (2018). Single dots show the catchments and are colored by their hydrological clusters.* ***(b)*** *Mean annual discharge for all catchments in the climate index space of Knoben et al. (2018). Single dots show the catchments and are colored according to the value of the mean annual discharge. The log of the mean annual discharge is used to show the relative differences between the catchments. For a depiction of all hydrological signatures used, see Fig. A2.*

Summary and conclusion

This study explored differences in the catchment characteristics between the eastern and western US, the properties and location of catchment clusters based on hydrological signatures, the importance of catchment attributes for those clusters, and how this study relates to other clustering studies and methods. We found that the correlations between catchment characteristics are quite similar for the eastern and western US with the exception of mean elevation, snow, geology and the leaf area index. For the overall CAMELS dataset climate seems to be the most important factor for the hydrological behavior. However, depending on the location either aridity, snow or seasonality were most important. The clusters derived from the hydrological signatures partly follow the ecological regions in the US and can combined into four groups of general behavior trends. Still, similar catchments can be quite far away from each other. We also found that most of the catchments have a rather similar discharge behavior, while only some more extreme catchments deviate from that main trend. This might be a hint as to why it is so difficult to cluster catchments, as those single extreme catchments are quite unique and do not fit together well with other catchments. We also found that there are differences of how directly the signal of forcing climate can be found again in the hydrological behavior. This explains why catchments often show a surprisingly similar behavior across many different climate and landscape properties (Troch et al., 2013) and why the most hydrologically similar catchment can be hundreds of kilometers away. Those findings also relate to the paradox that small-scale and single-catchment studies identify geology/soils as most important for the hydrological behavior, while large-sample studies usually find the climate to be most important. This might simply be influenced by spatial proximity. Small-scale studies look at catchments which all have a similar climatic forcing, and thus only the other catchment attributes can be the cause of differences in hydrological behavior. Large-sample studies on the other hand consider catchments with a wider area and thus attribute the differences in behavior to climate.

The aggregated data used in this study might level out the variability of the catchment attributes in the single catchment, but they also indicate that there is a kind of equifinality in the behavior of catchments. Different sets of intertwined climate forcing and catchment attributes could lead to a very similar overall behavior, not unlike hydrological models that produce the same discharge with different sets of parameters.

We acknowledge that the results are dependent on the amount and size of the clusters, the catchment attributes considered and the hydrological signatures used. Still, we think that the CAMELS dataset offers an excellent overview of different kinds of catchments in contrasting climatic and topographic regions. In addition, this study shows that using hydrological signatures with high spatial predictability results in hydrological meaningful clusters, which show consistent low flow behavior, even though those low flows were not explicitly considered. However, it seems that even a comprehensive dataset like CAMELS does not allow an easy way to find a conclusive set of clusters for catchments. For future research, we recommend including measures of spatial variability of the climate in the single catchments and to look into the single clusters in more depth. This might help to prove whether a less clear climatic signal is caused by intra-catchment variability of the climate or a larger influence from other catchment attributes.

.

III. Simple Catchments And Where To Find Them: The Storage-Discharge Relationship as a Proxy for Catchment Complexity

This chapter is published in *Frontiers in Water*:

Jehn, F. U., Breuer, L., Kraft, P., Bestian, K. and Houska, T.: Simple Catchments and Where to Find Them: The Storage-Discharge Relationship as a Proxy for Catchment Complexity, Front. Water, 3, 631651, https://doi.org/10.3389/frwa.2021.631651, 2021.

Introduction

What determines how a catchment reacts to a specific climatic forcing? A seemingly simple question, which is still hard to answer conclusively (Clark et al., 2016; Sivapalan, 2005b). This is mainly because every catchment is unique and thus slightly different to even the ones most similar to it (Beven, 2000). Still, we find similarities in catchment behavior, ranging from hillslope (Loritz et al., 2018) to continental scales (Kuentz et al., 2017a). We have even made progress in predictions of behavior in ungauged basins (Hrachowitz et al., 2013b). Still, we need to find more reliable ways to transfer our understanding of hydrological processes between catchments.

There are many different approaches in trying to quantify and understand the similarity between catchments. One approach is to take a large sample of catchments, sort them into groups of similar behavior and then examine which characteristics they share (Berghuijs et al., 2014; Jehn et al., 2020; Kuentz et al., 2017). This can also be done the other way around, thus starting from similar catchment characteristics and then study the catchment behavior (Knoben et al., 2018). Others derive understanding of catchment behavior from studies of experimental hillslopes or catchments like Tromp-van Meerveld and McDonnell (2006). However, there are also more theoretical approaches, such as using hydrological models to infer the underlying processes in the catchment (Clark et al., 2011; Fenicia et al., 2014). There are also approaches that try to link catchment behavior to thermodynamic theory (Loritz et al., 2019; Loritz et al., 2018) or to elegant mathematical approaches (Kirchner, 2009; Savenije, 2018).

This study is inspired by earlier works about recession and water balance that show that hydrological recession behavior can often be described with exponential functions, if no additional water is added. This implies that the outflow is proportional to storage and the underlying aquifer reacts like a single linear reservoir (Tallaksen, 1995; Wittenberg, 1999). However, it remains unclear how often this "simple" behavior really occurs in catchments and on what scales it is present. Those dynamics can be explored by examining a large sample of storage-discharge relationships. Using the storage-discharge relationship to explore catchment dynamics is not a new idea (e.g. (Kirchner, 2009; Sayama et al., 2011; Spence, 2010)), but is seen as a valuable way to improve the understanding of catchments (Tetzlaff et al., 2011). Especially dynamic storage behaviors of catchments provide a way for comparing catchments across landscapes (Buttle, 2016; Spence, 2010).

Studies of storage change often consider only few catchments (e.g. (Cheng et al., 2017; Floriancic et al., 2018; Geris et al., 2015)), or focus on a single catchment attribute class, like topography (Liu et al., 2016; Staudinger et al., 2017), geology (Creutzfeldt et al., 2014; Pfister et al., 2017; Sayama et al., 2011) or vegetation (Cheng et al., 2017; Geris et al., 2015). This makes it hard for generalization, as these investigations only capture a snapshot of catchment attributes and their effect on hydrological behavior. Therefore, studies using large sample sizes are needed that explore the storage-discharge relationship in complex landscapes that have similar climate conditions (Loritz et al., 2019). At best, the selection of catchments should consider similar climate conditions so that the boundary conditions are similar and catchment behavior is not governed by hydrometric differences. This will help connecting the knowledge gained of more theoretical approaches (Kirchner, 2009; Wittenberg, 1999) with experimental studies that examine single hillslopes in depth and highlight the importance of physical processes like preferential flow (Wienhöfer and Zehe, 2014). We use a dataset of 88 catchments from the federal state of Hesse, Germany, that features a wide range of catchment attributes, while having a relatively similar climate. To address all factors that are commonly attributed to influence hydrological behavior, we study catchment area, catchment shape, soil, geology, topography, and land use (Sivapalan, 2005b) and use the storage-discharge relationship as a proxy of catchment complexity. The more the storage-discharge relationship fits an exponential function, the simpler we view the behavior of that catchment.

The aim of this study is to scrutinize catchments with a varying complexity of their storage-discharge relationship and explore which catchment attributes are linked to this changing complexity using a large dataset. This focus on how much catchments obey a "simple" mode of behavior will highlight which processes are active or dominant in different locations and will help to understand causes of hydrological similarity.

Materials and Methods

Study area

We analyze a database of 88 catchments located in the state of Hesse, Germany (Fig. III.1) with discharge and climatic data over 26 years (1992-2018), resulting in 2314 separate catchment years. Rivers with major technical structures that obstruct the discharge by artificial impoundment (e.g. reservoir) are excluded from the analysis. However, some of the rivers have floodgates. As Hesse has a very diverse geology (HLNUG, 2007), it allows very different types of catchments to be considered under similar climatic conditions. Still, the climate has a considerable range, especially in the precipitation (Fig. III.2). This is a compromise between climatic similarity and sample size. We included climatic data in our analysis to determine the influence on the final results. Overall, the climate is humid and typical for Central Europe. To capture all factors that are usually attributed to influence the storage-discharge relationship, we investigate 15 attributes of climate (evapotranspiration, runoff-ratio, precipitation), land use , topography (slope, elongation ratio, area), soils (soil texture, soil type, soil depths) and geology (aquifer hydraulic conductivity, geology type, permeability) and water flow (discharge, ground water recharge). These attributes show a wide variety in the database (Fig. III.2). Snow was not explicitly considered in this study as Stoelzle et al. (2020) showed that in a similar set of catchments snow had only minor influence on streamflow if the catchments were below 800 m a.s.l., which is the case for the catchments in this study.

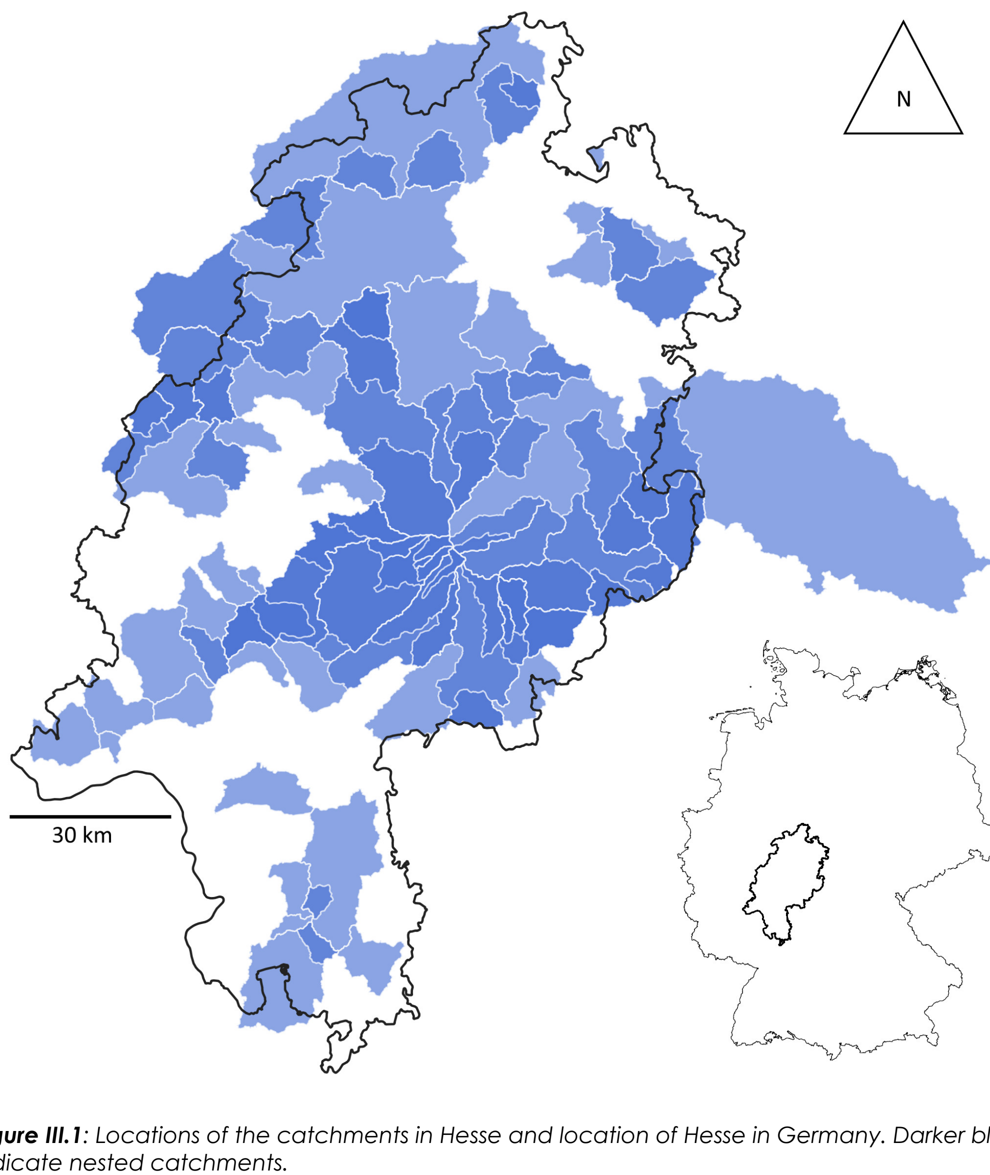

Figure III.1: *Locations of the catchments in Hesse and location of Hesse in Germany. Darker blues indicate nested catchments.*

Data sources

All soil and geology data are extracted from maps of the Federal Institute for Geosciences and Natural Resources (namely the HUEK 250 hydrogeology, GWN1000 groundwater, BOART 1000 soil texture, BK 500 soil typeand PHYSGRU 1000 soil depth maps). The values for the catchments were extracted from those maps with QGIS. Numerical attributes were averaged over the catchment. Categorical attributes used the dominant/highest value (e.g. a catchment with more than 50 % grassland was classified as grassland). The coarse resolution of the land cover data results in few cover types, so selecting the one with the highest portion is likely to distinguish the most extensive. Discharge is provided by the Hessian Agency for Nature Conservation, Environment and Geology (https://www.hlnug.de/static/pegel/wiskiweb2/). Further long-term data, which is not available online, can be obtained by contacting the Hessian Agency for Nature Conservation, Environment and Geology. Data on precipitation and evapotranspiration is obtained from the REGNIE project of the German Weather Service (https://www.dwd.de/DE/leistungen/regnie/regnie). The original raster datasets can be downloaded from the Climate Data Center of the German Weather Service (https://www.dwd.de/DE/klimaumwelt/cdc/cdc_node.html). The temporal resolution for discharge, evapotranspiration and precipitation is daily. The areas based values of water budget fluxes in mm per catchment can be found in the repository of this paper (Jehn, 2020). The elongation ratio (i.e., the ratio of the diameter of a circle of the same area as the basin to the maximum flow length) is assessed following (Sukristiyanti et al., 2018). Slope and catchment area are derived from a digital elevation map with a resolution of 40 x 40 m. The runoff-ratio is calculated from discharge and precipitation. All water fluxes (discharge, precipitation, evapotranspiration) are converted to mm.

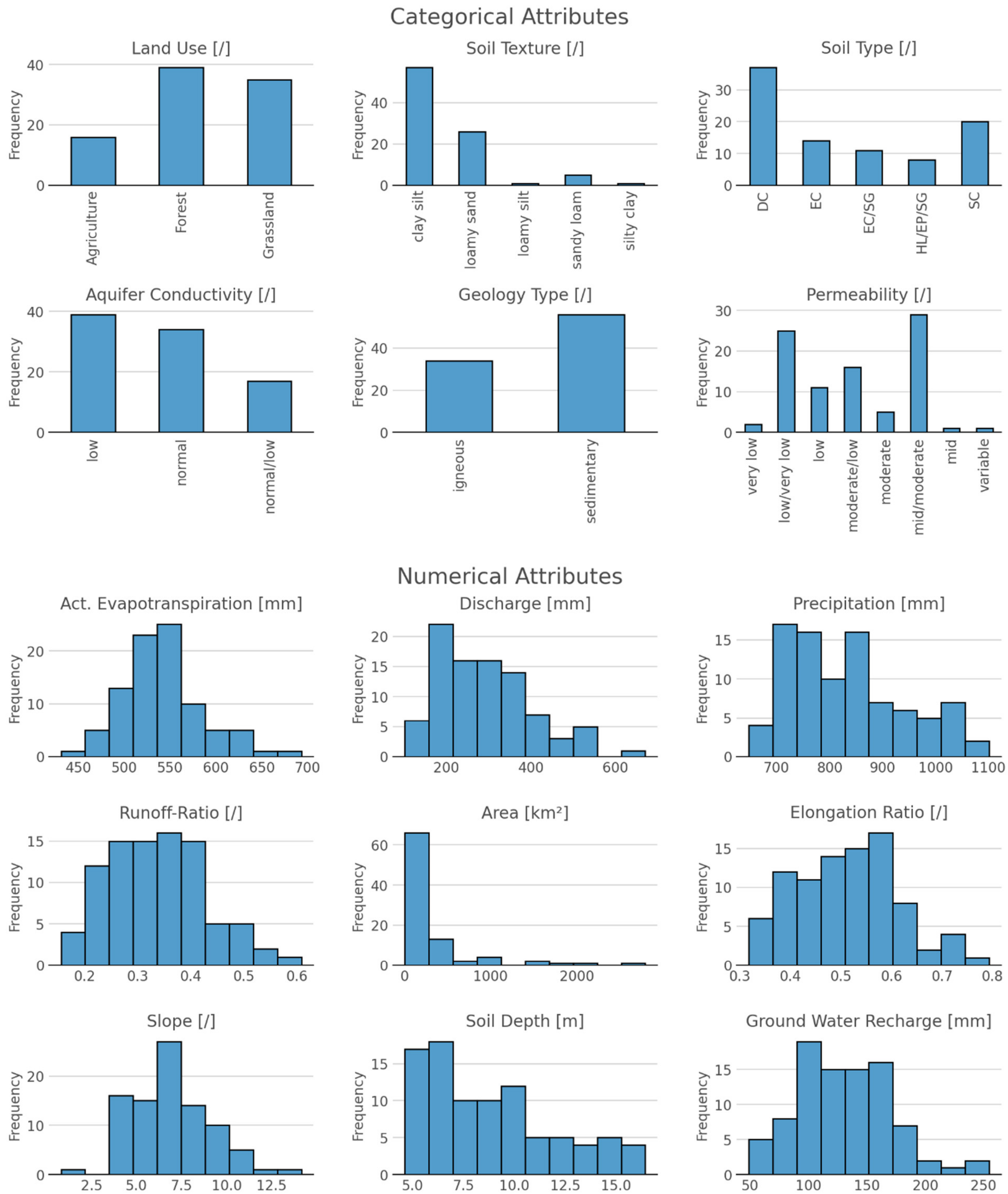

Figure III.2: *Attributes of the 88 catchments considered in this study. Climatic attributes of the catchment are the mean for all years. Soil type abbreviations: DC = Dystric Cambisol, EC = Eutric Cambisol, SC = Spodic Cambisol, SG = Stagnic Gleysol, HP = Haplic Luvisol, EP = Eutric Podzuluvisol.*

The REGNIE evapotranspiration data are calculated with the AMBAV model (Löpmeier, 1994), assuming a homogenous land cover of grass over sandy loam. As land uses and soils are often very different in the catchments considered, we correct the evapotranspiration accordingly. For this, we assume a storage change of zero over a long period, representing a closed water balance. Based on the uncorrected water balance and the total evapotranspiration over the whole time period we calculate the water balance error relative to the evapotranspiration. This results in catchment specific correction factors for the evapotranspiration, so that the storage change equals zero over the 26 years period. This procedure increased the mean actual evapotranspiration over all catchments from 431 to 541 mm per year, which is equivalent to the long term mean of 530 mm per year for Hesse (KLIWA et al., 2017). Both the start year 1992 and the end year 2018 have similar drought conditions as indicated by the standardized precipitation index (McKee et al., 1993) of -10 for 1992 and -12 for 2018 and therefore the long term change in storage can be assumed to be close to zero.

Annual cumulative storage change

As the storage of a catchment cannot be measured directly, we approximate the storage following the basic water balance equation Eq. III.1.

$$0 = P(t) - ET(t) - Q(t) - \Delta S(t) \qquad \text{(Eq. III.1)}$$

With daily precipitation P [mm], evapotranspiration ET [mm], discharge Q [mm] and storage change ΔS [mm]. We use Eq. III.2 to calculate the annual cumulative storage relative to the state at the beginning day (τ=0) of a hydrological year (1st Nov - 31st Oct) using daily data. This means storage change is calculated on daily basis. Based on this we determine the cumulative sum of the storage, resulting in the vector S'_{cum} that contains the cumulative sum of the storage change for every day.

$$S'_{cum}(t) = \sum_{\tau=0}^{t} P(\tau) - ET(\tau) - Q(\tau) \qquad \text{(Eq. III.2)}$$

This is done for every hydrological year separately to avoid the accumulation of errors in the measurements and to allow inter-annual comparisons. Note that this annual cumulative storage change does not capture the total storage of a catchment, but is a proxy of the active/dynamic storage changes as defined by Staudinger et al. (2017) and McNamara et al. (2011).

Complexity of storage-discharge relationship

After calculating the annual cumulative storage, we evaluate the relationship between the annual cumulative storage change and the discharge. As we want to test how well catchments follow an exponential storage-discharge relationship, we test how well the storage-discharge relationship can be fitted with an exponential function (Eq. III.3). As a functional approach we are using an exponential relationship of discharge with the storage as proposed in the baseflow component of TOPMODEL (Beven and Kirkby, 1979) (described in the supplement of Knoben et al. (2019)). However, this relationship could also be described power function (see e.g. Kirchner (2009).

$$Q = ae^{bS'_{cum}} \qquad \text{(Eq. III.3)}$$

with discharge Q [mm d-1], shape parameters a and b [-], as well as cumulative storage change ΔS' [mm d-1]. To fit the function to the data we used the curve_fit module of the Python package scipy (version 1.4.1), which uses a non-linear least squares fit. We only use days without precipitation to avoid a direct influence of precipitation on the discharge (Kirchner, 2009). This step allows us to estimate how much the real data deviate from the exponential function (Fig. III.3), by calculating the Kling-Gupta efficiency (KGE) (Gupta et al., 2009). A KGE of 1 describes a catchment with perfectly exponential behavior in a given year. This results in 27 (one per year) separate fits (as described in Fig. III.3), and therefore KGE values, for every catchment.

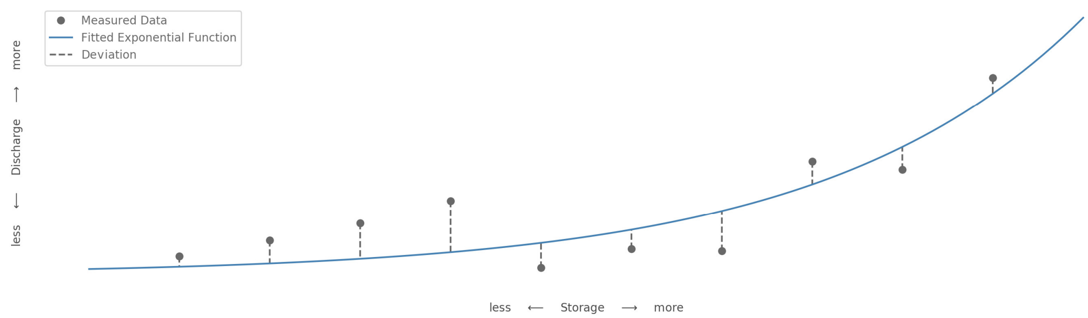

Figure III.3*: Visualization of the deviation of the measured and the idealized storage-discharge relationship (artificial data).*

This KGE is a proxy for the catchment's complexity. The lower the combined residuals are the simpler is the catchment. Therefore, catchments in this study that are described as simple refer to a storage-discharge relationship for a given year and catchment, which follow an exponential function without much deviation. Complex behavior on the other hand refers to a storage-discharge relationship which deviates substantially from an exponential function. We use the unbinned KGE in Tab. III.1 (section 3.2) to determine if there is a relationship among catchment attributes and complexity for the complete dataset.

To detect the effect of the catchment attributes on the storage-discharge relationship, we bin the 20 % of catchments (n = 18) together with the lowest/highest mean KGE and refer to them as being simple/complex. We use this in the analyses in Fig. III.6 (section 3.2) to highlight the differences between the most extreme catchments. We compare the mean KGE for all years of all catchments with catchment attributes using linear regression (for the numerical attributes) and ANOVA (for categorical attributes). All slopes of the linear regressions are tested, if they are significantly different from zero. To keep the rate of false positive results low, we corrected all p-values by multiplying them with the overall amount of statistical tests done in this study (Bonferroni adjustment (Haynes, 2013)). Significant in this study refers to a significance level of 1 %.

To delineate the differences between simple, complex catchments and the overall dataset, we calculate an ANOVA for the numerical attributes with the same correction as mentioned above. The categorical attributes are compared qualitatively.

Results and Discussion

Complexity of the storage-discharge relationship

The storage-discharge relationship shows four groups of different behaviors (Fig. III.4). Many storage-discharge relationships show almost perfect exponential behavior (Fig. III.4A). Those catchments are identified as simple. The other patterns (Fig. III.4 B,C,D) cannot be fitted with a simple function. Hence, we conclude them complex. Complex behavior comes in three distinct types.Simple behavior is more strictly defined than complex behavior, as it can only arise from the pattern seen in Fig. III.4A. The first complex type (Fig. III.4B) has a relationship where the catchment has a distinct peak discharge at the beginning of the year. After that peak it dries up and later refills. However, during refilling the catchment does not show an increase in discharge, even though it is also defined by storing more water than at the beginning of the year. The second complex type (Fig. III.4C) also has peak discharge at the beginning of the year and dries up after that. However, in contrast to the others, peak discharge can be found at low storage as well. The third complex type Fig. III.4D) shows erratic behavior with no clear pattern. In addition, the behavior of catchments often varies from year to year (Fig. III.5).

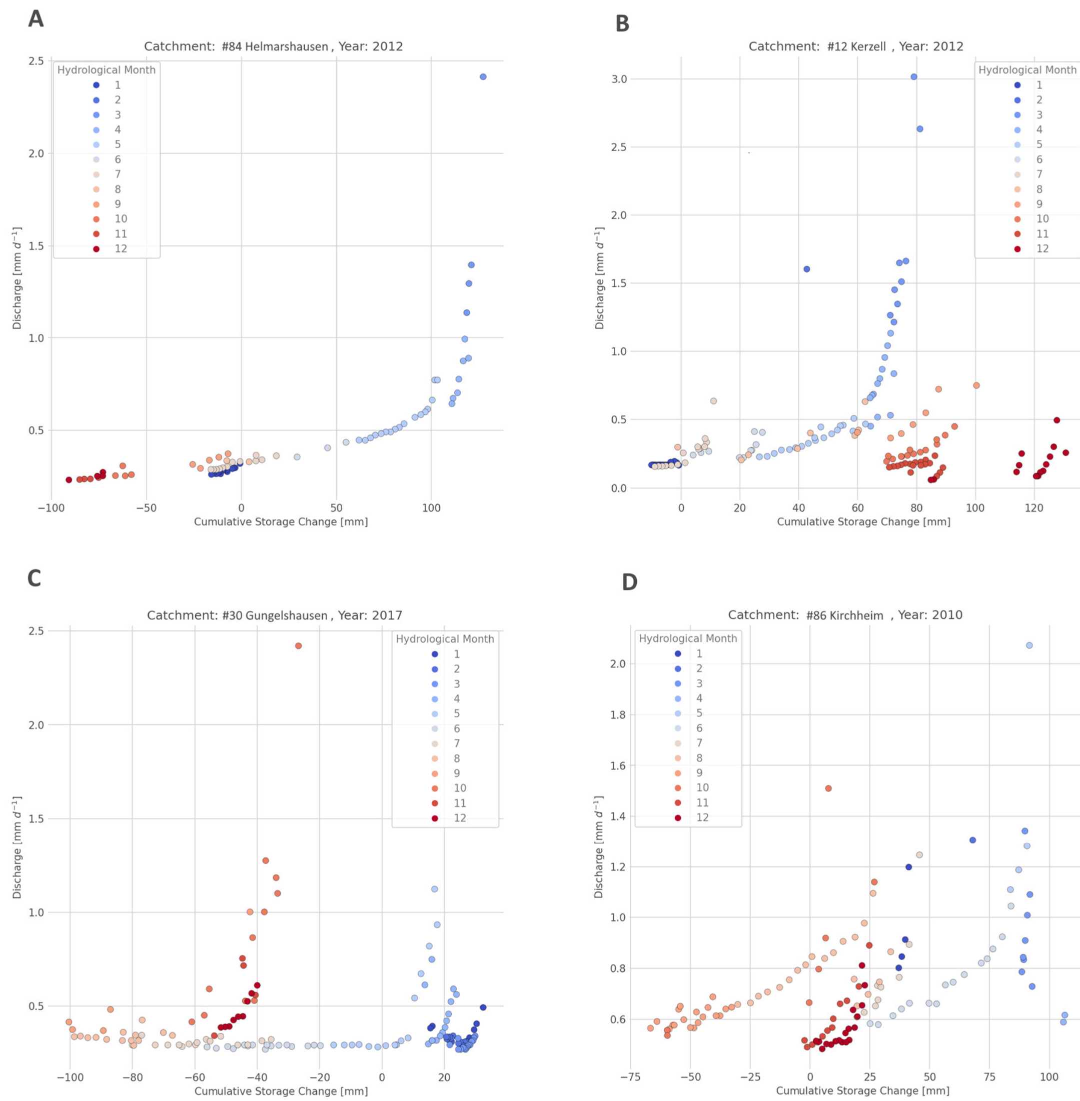

***Figure III.4**: Plots show cumulative change in storage against discharge during recession periods in a hydrological year. Color indicates the month of the hydrological year. Examples of the patterns in the storage-discharge relationship. A) an almost perfectly exponential storage-discharge relationship B) low discharge, even though the storage is higher than at previous peak discharge C) peak discharge, even though the storage is lower than at previous peak discharge D) erratic behavior.*

The complexity of a catchment's storage-discharge relationship as described in section 2.4 can show a wide range (Fig. III.5). Some of the catchments depict a low complexity (e.g. catchment #88) for the every year. Contrasting, none of the catchments are characterized by a high complexity during all years. Nevertheless, several catchments indicate a rather complex behavior most of the time (e.g. catchment #1). In general, the complexity (when measured as the catchment's mean KGE) changes by more than a factor of two between the most simple and the most complex catchments. This reflects the variability among catchments with an almost perfect exponential storage-discharge relationship and those of primarily erratic behavior. Even though we can order the catchments according to their complexity, we see that complexity varies greatly from year to year, even in the most complex catchments. For example, catchment #1, which has the highest complexity of all catchments, we still find years with very low complexity. We can also see the same pattern for the complexity of the years. For example, 2017 is the most complex of all years, but also has catchments with a very low complexity (e.g. #56, #69 or #76). We verified the reliability of our approach by using the mean sum of least squares and the Nash-Sutcliffe efficiency as alternative objective functions. This resulted in almost exactly the same results for both years and catchments.

The interannual variability suggests weather conditions play an important role in the complexity of the identified storage-discharge relationship. Both, years (histogram on the top of Fig. III.5) and catchments (histogram to the right of Fig. III.5), exhibit the same mean (meanyears = 0.013, meancatchments = 0.013) and similar standard deviation (std) (stdyears = 0.006, stdcatchments = 0.012) for their respective mean KGE. Several years show a very simple storage-discharge relationship across all catchments (e.g. 1999, 2001). As with the catchments, the factor between the most complex and most simple year is larger than two. Those most complex years are characterized by a negative standardized precipitation index (McKee et al., 1993), which is an indicator for drought conditions. This might be linked to processes analogous to fill and spill runoff generation (Tromp-van Meerveld and McDonnell, 2006) and more generally a function of the hydrological connectivity of the catchments (Bracken and Croke, 2007). In drier years, the catchments could be less connected and thus show more erratic behavior, while years with more precipitation allow more stable connections, both spatially and in time. This higher connectedness could lead to simpler behavior, as the amount of water in the river is more directly connected with the amount of water in the catchment. Interestingly, while the three most complex years (1996, 2013, 2017) all have a negative standardized precipitation index (McKee et al., 1993), none of them is considered a drought year (Erfurt et al., 2020). Thus, a severe drought cuts most connections in the catchment and only leaves groundwater as the main contributor to streamflow, which again results in simpler behavior. In addition, all three of the most complex years have at least one month with precipitation > 150 mm. This can also be linked to the approach of Loritz et al. (2018), which uses information theoretic and thermodynamic reasoning in combination with topographic information to study how the entropy of the hillslopes in a catchment changes of time. They show that the entropy and thus complexity increases strongly to large precipitation events after dry periods in the summer. They also highlight that this emerging behavior is caused by the interaction of different parts (in this case hillslopes) of the catchment.

This influence of large precipitation events on catchment complexity can also be found in other studies. For example (Capell et al., 2012) studied a catchment that was split between more mountainous uplands and lowlands with sand stone. The lowlands usually experienced linear discharge recession behavior, except when large precipitation events were recorded in the uplands. We conclude that this is probably the difference between the discharge being mainly baseflow provided from the valley bottom or stormflow generated in the upper stream reaches. The catchments cannot take up all the additional water due to large precipitation events, either because rainfall intensity exceeds infiltration capacity or catchments reach a storage capacity (Sayama et al., 2011; Teuling et al., 2010).

Another factor that might contribute to the complexity of the storage-discharge relationship is hysteresis. Hysteresis has been found to influence hydrological behavior at different scales (Zuecco et al., 2016). In this study, we use hydrological years to avoid cutting off a long time hysteresis process before it has ceased. However, hysteresis processes that start before or end after a hydrological year are not fully covered by our approach, but are important for catchments with low aquifer conductivity (Hellwig et al., 2020). Thus, more complex years and catchments might be an indicator of long term storage-discharge hysteresis that are triggered by reaching certain storage thresholds (Spence, 2010). It might also be the case that simple/complex behavior is a proxy for less/more hysteretic catchments. The more connected a catchment is, the more direct is the relationship between its storage and the discharge.

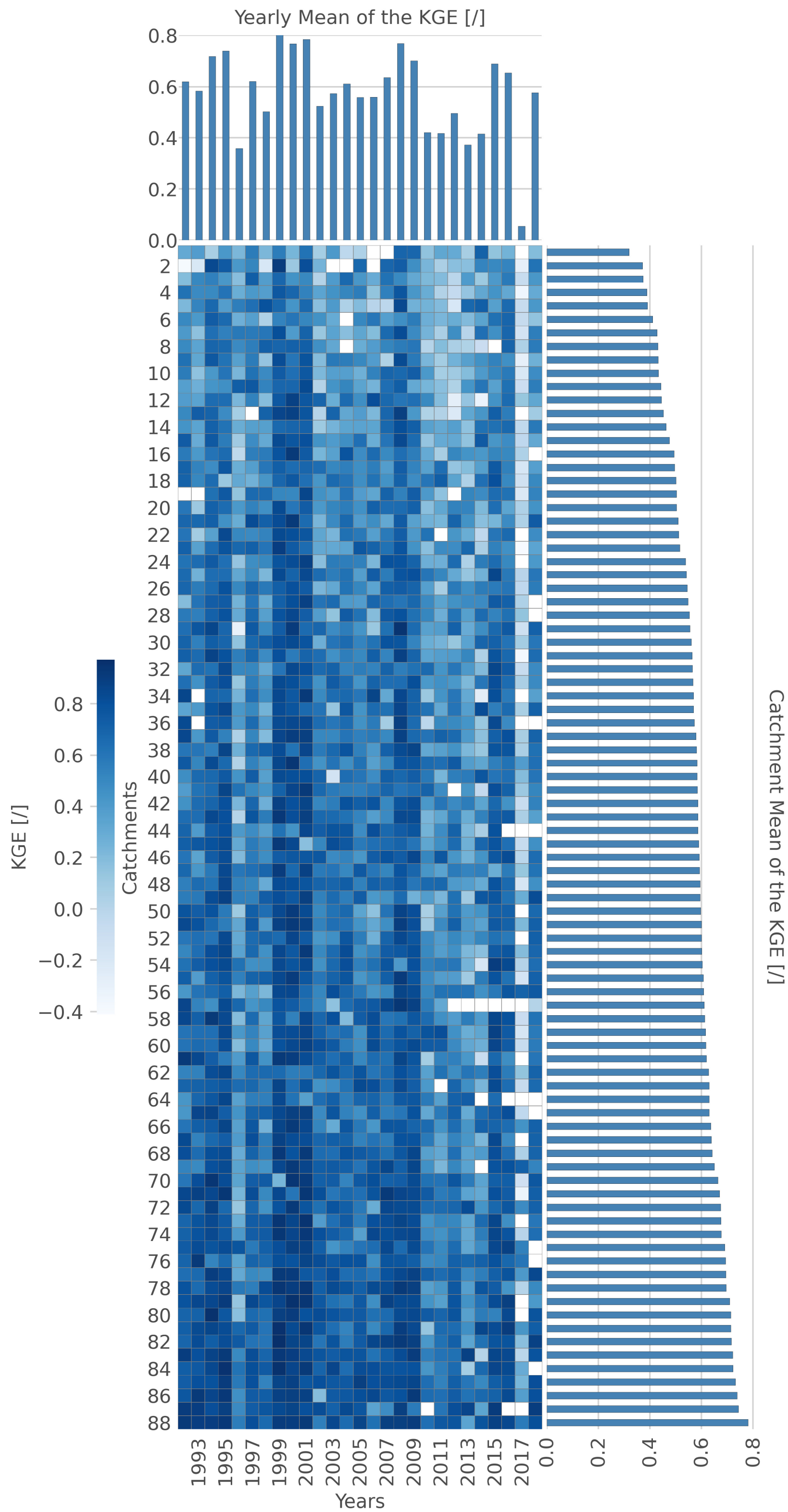

Figure III. 5: *Heatmap of the Kling-Gupta Efficiency (KGE) (measure of catchment complexity) for the 88 catchments of the Hesse dataset, separated by years. Darker colors indicate lower complexity. Bar charts depict the mean values for the rows and columns.*

Differences in catchment attributes between simple and complex catchments

To find the most consistently simple and complex catchments, we use 20 % of the catchments with the highest catchment mean KGE and the 20 % with the lowest catchment mean KGE (Fig. III.5). We only analyze those catchment attributes that show a significant relationship in the whole dataset (Tab. III.1). This removes the soil depth from the further analysis, as it shows no significant differences concerning the KGE. When we compare the remaining attributes of the simple and complex catchments with each other and the overall dataset, we can see clear differences, especially in the categorical catchment attributes (Fig. III.6):

- Simple catchments: Normal aquifer conductivity and permeability, regions with igneous geology, clay silt soil texture, wide range of soil types, more grassland and forest.
- Complex catchments: Low aquifer conductivity and permeability, regions with sedimentary geology, loamy sand soil texture, dystric cambisols, more agriculture.

Table III.1: *Differences in the Kling-Gupta Efficiency for the categorical and numerical catchment attributes for all years of all catchments. The p-values for the categorical attributes indicates if there is a significant relationship between the categories. The p-values for the numerical attributes indicate if the trend between the KGE and the attribute is significant.*

Numerical Attributes	Act. ET	Discharge	Precipitation	Runoff-Ratio	Area	Elongation Ratio	Slope	Soil Depth	GW Recharge
P-Values	< 0.001	< 0.0001	< 0.0001	< 0.0001	< 0.01	< 0.0001	< 0.0001	1	< 0.0001

Categorical Attributes	Land Use	Soil Texture	Soil Type	Aquifer Conductivity	Geology Type	Permeability
P-Values	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

The trend is less clear for the numerical catchment attributes. Simpler catchments tend to be steeper, while complex catchments are more flat. However, this trend is not significant. All other catchment attributes from Fig. III.2 do not influence the complexity of the storage-discharge relationship, at least not to an extend that is detectable with our method. As expected, climatic attributes, which are relatively similar for all catchments, are also not relevant. These results might also explain why it is easier to find the important drivers for the behavior of extreme 20 % most simple/complex catchments than for the remaining catchments (Singh et al., 2014). While the most simple/complex catchments have attributes that have a considerable influence on the hydrological behavior, the other catchments lie somewhere in between (Fig. III.5). These “in between” catchments have a compensating mix of attributes, which makes it very hard to disentangle the specific attributes that control complexity. This variation in catchment attribute influence has also been found for parts of a single catchment (Sun et al., 2014) and small sample studies (e.g. (Hoylman et al., 2019)).

Therefore, this study highlights what kind of catchment attributes are important when we have a large sample of catchments that all have a similar climate conditions. Our study found an influence of evapotranspiration and precipitation on the overall dataset (Tab. III.1), but not in the behavior of the most simple and complex catchments (Fig. III.6). The trend of the evaporation is significant for the complete dataset, but not significant when comparing the most simple with the most complex catchments. Contrasting, large sample studies conducted over larger scales usually that climate has a larger influence (Kuentz et al., 2017a; Oudin et al., 2010). This highlights that both climate and catchment attributes are important but on different scales. While the overall behavior is determined by the climate, this climatic signal is shaped by the catchment attributes, specifically soils and geology.

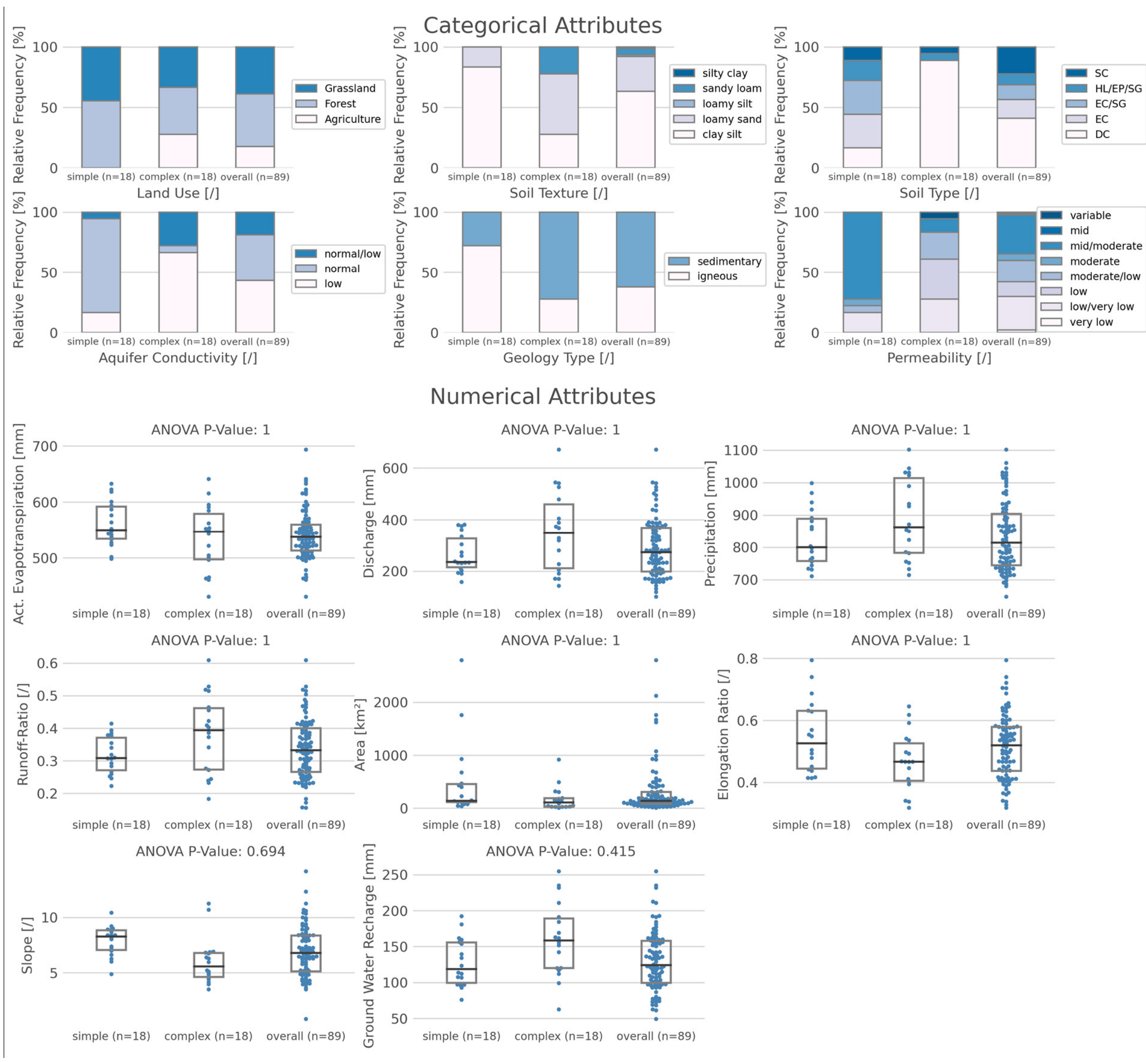

Figure III.6: *Differences in the categorical (top five panels) and numerical (bottom six panels) catchment attributes between simple and complex catchments and the overall dataset. Only those attributes are shown that have a significant trend over the whole dataset. Simple and complex refers to the 20 % of the catchments (n = 18), which have the lowest/highest catchment Kling-Gupta Efficiency (KGE) considered. The p-values on top of the box plots for the numerical attributes indicate significant differences between the simple, complex and all catchments. Black line is the median and grey lines show the interquartile range. Soil type abbreviations: DC = Dystric Cambisol, EC = Eutric Cambisol, SC = Spodic Cambisol, SG = Stagnic Gleysol, HP = Haplic Luvisol, EP = Eutric Podzuluvisol.*

Relationship of catchment complexity and hydrological processes.

Our results show that there are clear differences in the characteristics of simple and complex catchments. Interestingly, the attributes of both, simple and complex catchments show deviation in their attribute values of similar size from the overall dataset (Fig. III.6). This is unexpected, as there are several modes of complex behavior (Fig. III.4). Therefore, complex catchments should show a wider diversity of their attributes than simple catchments, if their complex behavior is caused by separate processes. This is not the case, which hints that the same underlying hydrological processes cause all the modes of complex behavior. Studies have highlighted that especially in humid and mountainous catchments subsurface stormflow is one of the main runoff generation processes (Chifflard et al., 2019; Wienhöfer and Zehe, 2014; Wittenberg, 1999) and that overall connectivity in a catchment defines its behavior (Jencso et al., 2009). This could also explains the results of this study. Simple catchments show high permeability and conductivity, while complex catchments show low permeability and conductivity. We therefore conclude that catchment complexity might simply be a proxy for catchment connectivity. Connectivity is to be understood herein as the combination of connections between hillslopes and riparian zone (Jencso et al., 2009) and the interconnectedness within hillslope (Tromp-van Meerveld and McDonnell, 2006) One possible explanation for the higher permeability and connectivity in the simpler catchment might be earthworms, as they prefer clay silt soils of those catchment (Curry, 2004) and increase preferential flow (Zehe et al., 2010). This also relates to the concept of thermodynamic equilibria discussed in Loritz et al. (2018). More permeable and connected catchments return quicker to their thermodynamic equilibria and thus show more simple behavior, while less connected and less permeable catchments need longer to return to their equilibrium and show more complex behavior.

This concept of connectivity can also be linked to threshold behavior, which has been identified as an important factor for catchment behavior (Spence, 2010; Tetzlaff et al., 2011) at both hillslope (Tromp-van Meerveld and McDonnell, 2006) and catchment scale (Jencso et al., 2009). Simple catchments might have lower thresholds, as they are always more interconnected due to their higher permeability and conductivity. Those lower thresholds could allow a more direct connection between the overall amount of water in the catchment and the discharge. For complex catchments, this is less the case, as they could have more isolated hillslopes and thus spill behavior happens more erratically. In essence, the amount of water in the catchment is less important without a connection to the river.

Summary and Conclusion

This study looks at the complexity of the storage-discharge relationship of 88 catchments in Hesse, Germany. The most simple and complex catchments show clear differences in their conductivity, permeability, geology and soils. The signal of weather patterns is transformed differently, depending on the catchment attributes. This leads to simple behavior for some catchment and to more complex for others. It is not uncommon for small and large scale studies to have contrasting results regarding the influence of climate on catchment response. The role of climate needs to be controlled in such studies in order to determine the influence of non-climatic factors, as we did in this study.

To further explore the importance of catchment attributes in relation to climate, additional studies in different climates are needed, as this study only focused on humid catchments in central Germany. What we finally need to understand is why certain catchments behave simple in one year and complex in another. Possible causes are extreme weather events, a complex interaction between the distribution of precipitation, the geology and the soils of a catchment and connecting and disconnecting of different stores in the catchment. All those attributes and processes ultimately define the catchment's active storage. We link the observed simple and complex behavior of catchments to the fill and spill hypotheses and the interconnectedness of spatial entities within a catchment. Simpler catchments have more preferential flow and more connected hillslopes and thus lower spill thresholds. High hydrological connectivity provides a more direct link from storage to discharge and implies that the simplicity of catchment is linked to certain hydrological processes being active or dominant. Further research should explore how this catchment simplicity can be linked to predictability of streamflow.

References

Addor, N., Newman, A. J., Mizukami, N. and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrology and Earth System Sciences, 21(10), 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N. and Clark, M. P.: A ranking of hydrological signatures based on their predictability in space, Water Resources Research, https://doi.org/10.1029/2018WR022606, 2018.

Ali, G., Tetzlaff, D., Soulsby, C., McDonnell, J. J. and Capell, R.: A comparison of similarity indices for catchment classification using a cross-regional dataset, Advances in Water Resources, 40, 11–22, https://doi.org/10.1016/j.advwatres.2012.01.008, 2012.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: Crop evapotranspiration: guidelines for computing crop water requirements, edited by Food and Agriculture Organization of the United Nations, Food and Agriculture Organization of the United Nations, Rome., 1998.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J. and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset, Hydrology and Earth System Sciences, 22(11), 5817–5846, https://doi.org/10.5194/hess-22-5817-2018, 2018.

Andréassian, V., Perrin, C. and Michel, C.: Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models, Journal of Hydrology, 286(1–4), 19–35, https://doi.org/10.1016/j.jhydrol.2003.09.030, 2004a.

Andréassian, V., Oddos, A., Michel, C., Anctil, F., Perrin, C. and Loumagne, C.: Impact of spatial aggregation of inputs and parameters on the efficiency of rainfall-runoff models: A theoretical study using chimera watersheds, Water Resources Research, 40(5), https://doi.org/10.1029/2003WR002854, 2004b.

Andréassian, V., Lerat, J., Le Moine, N. and Perrin, C.: Neighbors: Nature's own hydrological models, Journal of Hydrology, 414–415, 49–58, https://doi.org/10.1016/j.jhydrol.2011.10.007, 2012.

Bai, Y., Wagener, T. and Reed, P.: A top-down framework for watershed model evaluation and selection under uncertainty, Environmental Modelling & Software, 24(8), 901–916, https://doi.org/10.1016/j.envsoft.2008.12.012, 2009.

Bárdossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters, Hydrology and Earth System Sciences, 12(6), 1273–1283, https://doi.org/10.5194/hess-12-1273-2008, 2008.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A. and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, Scientific Data, 5, 180214, https://doi.org/10.1038/sdata.2018.214, 2018.

Berghuijs, W. R., Sivapalan, M., Woods, R. A. and Savenije, H. H. G.: Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales, Water Resources Research, 50(7), 5638–5661, https://doi.org/10.1002/2014WR015692, 2014.

Beven, K.: On undermining the science?, Hydrological Processes, 20(14), 3141–3146, https://doi.org/10.1002/hyp.6396, 2006a.

Beven, K. J.: Uniqueness of place and process representations in hydrological modeling, Hydrology and Earth System Sciences, 4(2), 203–213, 2000.

Beven, K. J.: Searching for the Holy Grail of scientific hydrology: Qt = (S, R, delta t)A as closure, Hydrology and Earth System Sciences, 10(5), 609–618, https://doi.org/10.5194/hess-10-609-2006, 2006b.

Beven, K. J.: Chapter 2: Evolution of Rainfall-Runoff Models: Survival of the Fittest?, in Rainfall-runoff modelling: the primer, pp. 25–50, Wiley, Chichester, , 2008a.

Beven, K. J.: Chapter 5: Predicting Hydrographs Using Distributed Models Based on Process Descriptions, in Rainfall-runoff modelling: the primer, Wiley, Chichester, , 2008b.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrological Sciences Bulletin, 24(1), 43–69, https://doi.org/10.1080/02626667909491834, 1979.

Blöschl, G.: Scaling in hydrology, Hydrol. Process., 15(4), 709–711, https://doi.org/10.1002/hyp.432, 2001.

Blume, H.-P., Brümmer, G. W., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretzschmar, R., Stahr, K., Wilke, B.-M., Scheffer, F. and Schachtschabel, P.: Kapitel 9: Böden als Pflanzenstandorte, in Scheffer/Schachtschabel Lehrbuch der Bodenkunde, Springer Spektrum, Berlin Heidelberg, , 2016.

Boyle, D. P., Gupta, H. V., Sorooshian, S., Koren, V., Zhang, Z. and Smith, M.: Toward improved streamflow forecasts: value of semidistributed modeling, Water Resources Research, 37(11), 2749–2759, https://doi.org/10.1029/2000WR000207, 2001.

Bracken, L. J. and Croke, J.: The concept of hydrological connectivity and its contribution to understanding runoff-dominated geomorphic systems, Hydrological Processes, 21(13), 1749–1763, https://doi.org/10.1002/hyp.6313, 2007.

Breuer, L., Eckhardt, K. and Frede, H.-G.: Plant parameter values for models in temperate climates, Ecological Modelling, 169(2–3), 237–293, https://doi.org/10.1016/S0304-3800(03)00274-6, 2003.

Breuer, L., Huisman, J. A., Willems, P., Bormann, H., Bronstert, A., Croke, B. F. W., Frede, H.-G., Gräff, T., Hubrechts, L., Jakeman, A. J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D. P., Lindström, G., Seibert, J., Sivapalan, M. and Viney, N. R.: Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use, Advances in Water Resources, 32(2), 129–146, https://doi.org/10.1016/j.advwatres.2008.10.003, 2009.

Burn, D. H.: Catchment similarity for regional flood frequency analysis using seasonality measures, Journal of Hydrology, 202(1–4), 212–230, https://doi.org/10.1016/S0022-1694(97)00068-1, 1997.

Buttle, J. M.: Dynamic storage: a potential metric of inter-basin differences in storage properties: Dynamic Storage: Metric of Inter-Basin Differences in Storage, Hydrological Processes, 30(24), 4644–4653, https://doi.org/10.1002/hyp.10931, 2016.

Buytaert, W., Reusser, D., Krause, S. and Renaud, J.-P.: Why can't we do better than Topmodel?, Hydrological Processes, 22(20), 4175–4179, https://doi.org/10.1002/hyp.7125, 2008.

Capell, R., Tetzlaff, D., Hartley, A. J. and Soulsby, C.: Linking metrics of hydrological function and transit times to landscape controls in a heterogeneous mesoscale catchment, Hydrological Processes, 26(3), 405–420, https://doi.org/10.1002/hyp.8139, 2012.

Cheng, L., Zhang, L., Chiew, F. H. S., Canadell, J. G., Zhao, F., Wang, Y.-P., Hu, X. and Lin, K.: Quantifying the impacts of vegetation changes on catchment storage-discharge dynamics using paired-catchment data: VEGETATION CHANGES ON STORAGE-DISCHARGE, Water Resources Research, 53(7), 5963–5979, https://doi.org/10.1002/2017WR020600, 2017.

Chifflard, P., Blume, T., Maerker, K., Hopp, L., Meerveld, I., Graef, T., Gronz, O., Hartmann, A., Kohl, B., Martini, E., Reinhardt-Imjela, C., Reiss, M., Rinderer, M. and Achleitner, S.: How can we model subsurface stormflow at the catchment scale if we cannot measure it?, Hydrological Processes, 33(9), 1378–1385, https://doi.org/10.1002/hyp.13407, 2019.

Chiverton, A., Hannaford, J., Holman, I., Corstanje, R., Prudhomme, C., Bloomfield, J. and Hess, T. M.: Which catchment characteristics control the temporal dependence structure of daily river flows?: TEMPORAL DEPENDENCE IN RIVER FLOW, Hydrological Processes, 29(6), 1353–1369, https://doi.org/10.1002/hyp.10252, 2015.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T. and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models: Differences between hydrological models, Water Resources Research, 44(12), https://doi.org/10.1029/2007WR006735, 2008.

Clark, M. P., Kavetski, D. and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling: Hypothesis testing in hydrology, Water Resources Research, 47(9), https://doi.org/10.1029/2010WR009827, 2011.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger, G. N. and Marks, D. G.: A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies: A unified approach for process-based hydrologic modeling, Water Resources Research, 51(4), 2515–2542, https://doi.org/10.1002/2015WR017200, 2015.

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Freer, J. E., Arnold, J. R., Moore, R. D., Istanbulluoglu, E. and Ceola, S.: Improving the theoretical underpinnings of process-based hydrologic models: Narrowing the gap between hydrologic theory and models, Water Resources Research, 52(3), 2350–2365, https://doi.org/10.1002/2015WR017910, 2016.

Clausen, B. and Biggs, B. J. F.: Flow variables for ecological studies in temperate streams: groupings based on covariance, Journal of Hydrology, 237(3–4), 184–197, https://doi.org/10.1016/S0022-1694(00)00306-1, 2000.

CMF: Catchment Modelling Framework Repository, <https://github.com/philippkraft/cmf>, accessed February 2018, 2018.

Coopersmith, E., Yaeger, M. A., Ye, S., Cheng, L. and Sivapalan, M.: 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(11), 4467–4482, https://doi.org/10.5194/hess-16-4467-2012, 2012.

Costanza, R., Wainger, L., Folke, C. and Mäler, K.-G.: Modeling Complex Ecological Economic Systems: Toward an Evolutionary, Dynamic Understanding of People and Nature, in Ecosystem Management, pp. 148–163, Springer New York, New York, NY, https://doi.org/10.1007/978-1-4612-4018-1_15, , 1993.

Craig, J. R. and Raven Development Team: Raven user's and developer's manual (Version 2.8) http://www.civil.uwaterloo.ca/jrcraig/Raven/Main.html (Accessed May, 2018). http://www.civil.uwaterloo.ca/jrcraig/Raven/Main.html, 2018.

Creutzfeldt, B., Troch, P. A., Güntner, A., Ferré, T. P. A., Graeff, T. and Merz, B.: Storage-discharge relationships at different catchment scales based on local high-precision gravimetry: WATER STORAGE AND DISCHARGE DYNAMICS: A GRAVIMETER PERSPECTIVE, Hydrological Processes, 28(3), 1465–1475, https://doi.org/10.1002/hyp.9689, 2014.

Curry, J. P.: Chapter 6: Factors Affecting the Abundance of Earthworms in Soils, in Earthworm ecology, edited by C. A. Edwards, CRC Press, Boca Raton, Fla, , 2004.

Das, T., Bárdossy, A., Zehe, E. and He, Y.: Comparison of conceptual model performance using different representations of spatial variability, Journal of Hydrology, 356(1–2), 106–118, https://doi.org/10.1016/j.jhydrol.2008.04.008, 2008.

Dawdy, D. R. and O'Donnel, T.: Mathematical Models of Catchment Behaviors, Journal of the Hydraulics Division, 91(4), 123–137, 1965.

Djabelkhir, K., Lauvernet, C., Kraft, P. and Carluer, N.: Development of a dual permeability model within a hydrological catchment modeling framework: 1D application, Science of The Total Environment, 575, 1429–1437, https://doi.org/10.1016/j.scitotenv.2016.10.012, 2017.

Erfurt, M., Skiadaresis, G., Tijdeman, E., Blauhut, V., Bauhus, J., Glaser, R., Schwarz, J., Tegel, W. and Stahl, K.: Exploring the added value of a long-term multidisciplinary dataset in drought research – a drought catalogue for southwestern Germany dating back to 1801, preprint, Hydrological Hazards., 2020.

van Esse, W. R., Perrin, C., Booij, M. J., Augustijn, D. C. M., Fenicia, F., Kavetski, D. and Lobligeois, F.: The influence of conceptual model structure on model performance: a comparative study for 237 French catchments, Hydrology and Earth System Sciences, 17(10), 4227–4239, https://doi.org/10.5194/hess-17-4227-2013, 2013.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S. and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, Hydrology and Earth System Sciences, 17(5), 1893–1912, https://doi.org/10.5194/hess-17-1893-2013, 2013.

Fenicia, F., Savenije, H. H. G., Matgen, P. and Pfister, L.: Understanding catchment behavior through stepwise model concept improvement, Water Resources Research, 44(1), https://doi.org/10.1029/2006WR005563, 2008.

Fenicia, F., Kavetski, D. and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development: Flexible framework for hydrological modeling, 1, Water Resources Research, 47(11), https://doi.org/10.1029/2010WR010174, 2011.

Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L. and Freer, J.: Catchment properties, function, and conceptual model representation: is there a correspondence?, Hydrological Processes, 28(4), 2451–2467, https://doi.org/10.1002/hyp.9726, 2014.

Fink, G. S. M. and Koch, M.: Climate change effects on the water balance in the Fulda catchment, Germany, during the 21st centruy, conference paper at Symposium on sustainable water ressource management and climate change adaption, Nakon Pathom, , 2010.

Floriancic, M. G., Meerveld, I., Smoorenburg, M., Margreth, M., Naef, F., Kirchner, J. W. and Molnar, P.: Spatio-temporal variability in contributions to low flows in the high Alpine Poschiavino catchment, Hydrological Processes, 32(26), 3938–3953, https://doi.org/10.1002/hyp.13302, 2018.

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S. and Savenije, H. H. G.: Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China, Hydrology and Earth System Sciences, 18(5), 1895–1915, https://doi.org/10.5194/hess-18-1895-2014, 2014.

Gauch, M., Mai, J. and Lin, J.: The Proper Care and Feeding of CAMELS: How Limited Training Data Affects Streamflow Prediction, arXiv:1911.07249 [cs, stat] http://arxiv.org/abs/1911.07249, last access: 26 August 2020, 2020.

Geris, J., Tetzlaff, D., McDonnell, J. and Soulsby, C.: The relative role of soil type and tree cover on water storage and transmission in northern headwater catchments: SOIL AND VEGETATION EFFECTS ON WATER STORAGE AND TRANSMISSION, Hydrological Processes, 29(7), 1844–1860, https://doi.org/10.1002/hyp.10289, 2015.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H. and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, Hydrology and Earth System Sciences, 18(12), 4839–4859, https://doi.org/10.5194/hess-18-4839-2014, 2014.

Gupta, H. V., Wagener, T. and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrological Processes, 22(18), 3802–3813, https://doi.org/10.1002/hyp.6989, 2008.

Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377(1–2), 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M. and Andréassian, V.: Large-sample hydrology: a need to balance depth with breadth, Hydrology and Earth System Sciences, 18(2), 463–477, https://doi.org/10.5194/hess-18-463-2014, 2014.

Gustard, A., Bullock, A. and Dixon, J. M.: Low flow estimation in the United Kingdom, Institute of Hydrology, Wallingford., 1992.

Haines, A. T., Finlayson, B. L. and McMahon, T. A.: A global classification of river regimes, Applied Geography, 8(4), 255–272, https://doi.org/10.1016/0143-6228(88)90035-5, 1988.

Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, Applied Engrg. in Agric., 1, 96–99, 1985.

Haynes, W.: Bonferroni Correction, in Encyclopedia of Systems Biology, edited by W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, pp. 154–154, Springer, New York, NY, https://doi.org/10.1007/978-1-4419-9863-7_1213, , 2013.

Hellwig, J., Stoelzle, M. and Stahl, K.: Stress-testing groundwater and baseflow drought responses to synthetic climate change-informed recharge scenarios, preprint, Groundwater hydrology/Modelling approaches., 2020.

Her, Y. and Chaubey, I.: Impact of the numbers of observations and calibration parameters on equifinality, model performance, and output and parameter uncertainty: Parameters, Observations, and Uncertainty, Hydrological Processes, 29(19), 4220–4237, https://doi.org/10.1002/hyp.10487, 2015.

Hindmarsh, A. C., Brown, P., Grant, K. E., Lee, S. L., Serban, R., Shumaker, D. E. and Woodward, C. S.: SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers, ACM Transactions on Mathematical Software (TOMS), 31(3), 363–396, 2005.

HLNUG: „Geologische Übersichtskarte von Hessen“, in: Umweltatlas Hessen <http://atlas.umwelt.hessen.de/atlas/>, accessed February 2017, 2007.

Houska, T., Kraft, P., Chamorro-Chavez, A. and Breuer, L.: SPOTting Model Parameters Using a Ready-Made Python Package, edited by D. Hui, PLOS ONE, 10(12), e0145180, https://doi.org/10.1371/journal.pone.0145180, 2015.

Houska, T., Kraft, P., Liebermann, R., Klatt, S., Kraus, D., Haas, E., Santabarbara, I., Kiese, R., Butterbach-Bahl, K., Müller, C. and Breuer, L.: Rejecting hydro-biogeochemical model structures by multi-criteria evaluation, Environmental Modelling & Software, 93, 1–12, https://doi.org/10.1016/j.envsoft.2017.03.005, 2017.

Hoylman, Z. H., Jencso, K. G., Hu, J., Holden, Z. A., Martin, J. T. and Gardner, W. P.: The Climatic Water Balance and Topography Control Spatial Patterns of Atmospheric Demand, Soil Moisture, and Shallow Subsurface Flow, Water Resources Research, 55(3), 2370–2389, https://doi.org/10.1029/2018WR023302, 2019.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E. and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrological Sciences Journal, 58(6), 1198–1255, https://doi.org/10.1080/02626667.2013.803183, 2013a.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E. and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrological Sciences Journal, 58(6), 1198–1255, https://doi.org/10.1080/02626667.2013.803183, 2013b.

Hublart, P., Ruelland, D., Dezetter, A. and Jourde, H.: Reducing structural uncertainty in conceptual hydrological modelling in the semi-arid Andes, Hydrology and Earth System Sciences, 19(5), 2295–2314, https://doi.org/10.5194/hess-19-2295-2015, 2015.

Jehn, F.: Zutn/Incremental_Breakdown: Reupload, , https://doi.org/10.5281/zenodo.1067939, 2017.

Jehn, F.: zutn/model_complexity: First citable version of the model code for the model complexity paper, 2018a.

Jehn, F. U.: zutn/Catchment-Classification: Release for Zenodo, , https://doi.org/10.5281/zenodo.2203639, 2018b.

Jehn, F. U.: zutn/Simple-Catchments-Hesse: Updated version for Frontiers in Water (Version 1.5), , https://doi.org/10.5281/zenodo.3532653, 2020.

Jehn, F. U., Breuer, L., Houska, T., Bestian, K. and Kraft, P.: Incremental model breakdown to assess the multi-hypotheses problem, Hydrology and Earth System Sciences, 22(8), 4565–4581, https://doi.org/10.5194/hess-22-4565-2018, 2018.

Jehn, F. U., Chamorro, A., Houska, T. and Breuer, L.: Trade-offs between parameter constraints and model realism: a case study, Sci Rep, 9(1), 10729, https://doi.org/10.1038/s41598-019-46963-6, 2019.

Jehn, F. U., Bestian, K., Breuer, L., Kraft, P. and Houska, T.: Using hydrological and climatic catchment clusters to explore drivers of catchment behavior, Hydrol. Earth Syst. Sci., 24(3), 1081–1100, https://doi.org/10.5194/hess-24-1081-2020, 2020.

Jehn, F. U., Breuer, L., Kraft, P., Bestian, K. and Houska, T.: Simple Catchments and Where to Find Them: The Storage-Discharge Relationship as a Proxy for Catchment Complexity, Front. Water, 3, 631651, https://doi.org/10.3389/frwa.2021.631651, 2021.

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E. and Marshall, L. A.: Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale, Water Resour. Res., 45(4), https://doi.org/10.1029/2008WR007225, 2009.

Jowett, I. G. and Duncan, M. J.: Flow variability in New Zealand rivers and its relationship to in-stream habitat and biota, New Zealand Journal of Marine and Freshwater Research, 24(3), 305–317, https://doi.org/10.1080/00288330.1990.9516427, 1990.

Kannan, N., White, S. M., Worrall, F. and Whelan, M. J.: Sensitivity analysis and identification of the best evapotranspiration and runoff options for hydrological modelling in SWAT-2000, Journal of Hydrology, 332(3–4), 456–466, https://doi.org/10.1016/j.jhydrol.2006.08.001, 2007.

Kavetski, D. and Clark, M. P.: Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing, Hydrological Processes, 25(4), 661–670, https://doi.org/10.1002/hyp.7899, 2011.

Kavetski, D. and Fenicia, F.: Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights: Flexible framework for hydrological modeling, 2, Water Resources Research, 47(11), https://doi.org/10.1029/2011WR010748, 2011.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resources Research, 42(3), https://doi.org/10.1029/2005WR004362, 2006.

Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, Water Resources Research, 45(2), https://doi.org/10.1029/2008WR006912, 2009.

KLIWA, Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg, Bayerisches Landesamt für Umweltschutz, Rheinland-Pfalz, Hessisches Landesamt für Naturschutz, Umwelt und Geologie and Deutscher Wetterdienst, Eds.: Entwicklung von Bodenwasserhaushalt und Grundwasserneubildung in Baden-Württemberg, Bayern, Rheinland-Pfalz und Hessen (1951-2015): KLIWA-Projekt A 2.3.4: Langzeitverhalten des Grundwasserhaushalts, KLIWA-Projekt B 3.3.2: Simulation der Grundwasserneubildung mit weiteren regionalen Klimaprojekten, März 2017., LUBW, Karlsruhe., 2017.

Knoben, W. J. M., Woods, R. A. and Freer, J. E.: A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data, Water Resources Research, 54(7), 5088–5109, https://doi.org/10.1029/2018WR022913, 2018.

Knoben, W. J. M., Freer, J. E., Fowler, K. J. A., Peel, M. C. and Woods, R. A.: Modular Assessment of Rainfall–Runoff Models Toolbox (MARRMoT) v1.2: an open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations, Geoscientific Model Development, 12(6), 2463–2480, https://doi.org/10.5194/gmd-12-2463-2019, 2019.

Kraft, P., Vaché, K. B., Frede, H.-G. and Breuer, L.: CMF: A Hydrological Programming Language Extension For Integrated Catchment Models, Environmental Modelling & Software, 26(6), 828–830, https://doi.org/10.1016/j.envsoft.2010.12.009, 2011.

Krysanova, V., Bronstert, A. and MüLler-Wohlfeil, D.-I.: Modelling river discharge for large drainage basins: from lumped to distributed approach, Hydrological Sciences Journal, 44(2), 313–331, https://doi.org/10.1080/02626669909492224, 1999.

Kuentz, A., Arheimer, B., Hundecha, Y. and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrology and Earth System Sciences, 21(6), 2863–2879, https://doi.org/10.5194/hess-21-2863-2017, 2017a.

Kuentz, A., Arheimer, B., Hundecha, Y. and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrology and Earth System Sciences, 21(6), 2863–2879, https://doi.org/10.5194/hess-21-2863-2017, 2017b.

Li, C., Cao, J., Nie, S.-P., Zhu, K.-X., Xiong, T. and Xie, M.-Y.: Serum metabolomics analysis for biomarker of Lactobacillus plantarum NCU116 on hyperlipidaemic rat model feed by high fat diet, Journal of Functional Foods, 42, 171–176, https://doi.org/10.1016/j.jff.2017.12.036, 2018.

Li, Z., Shao, Q., Xu, Z. and Cai, X.: Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method: A case study of SWAT model applied to Yingluoxia watershed in northwest China, Journal of Hydrology, 385(1–4), 76–83, https://doi.org/10.1016/j.jhydrol.2010.01.025, 2010.

Liu, J., Han, X., Chen, X., Lin, H. and Wang, A.: How well can the subsurface storage-discharge relation be interpreted and predicted using the geometric factors in headwater areas?: Interpreting Storage-Discharge Relations Using Geometric Factors, Hydrological Processes, 30(25), 4826–4840, https://doi.org/10.1002/hyp.10958, 2016.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P. and Loumagne, C.: When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events, Hydrology and Earth System Sciences, 18(2), 575–594, https://doi.org/10.5194/hess-18-575-2014, 2014.

Löpmeier, F.-J.: Berechnung der Bodenfeuchte und Verdunstung mittels agrarmeteorologischer Modelle., Zeitschrift f. Bewaesserungswirtschaft, 29, 157–167, 1994.

Loritz, R., Gupta, H., Jackisch, C., Westhoff, M., Kleidon, A., Ehret, U. and Zehe, E.: On the dynamic nature of hydrological similarity, Hydrology and Earth System Sciences, 22(7), 3663–3684, https://doi.org/10.5194/hess-22-3663-2018, 2018.

Loritz, R., Kleidon, A., Jackisch, C., Westhoff, M., Ehret, U., Gupta, H. and Zehe, E.: A topographic index explaining hydrological similarity by accounting for the joint controls of runoff formation, , https://doi.org/10.5445/IR/1000098521, 2019.

Lu, J., Sun, G., McNulty, S. G. and Amatya, D. M.: A comparison of six potential evapotranspiration methods for regional use in the southeastern United States, Journal of the American Water Resources Association, 41(3), 621–633, https://doi.org/10.1111/j.1752-1688.2005.tb03759.x, 2005.

Maier, N., Breuer, L. and Kraft, P.: Prediction and uncertainty analysis of a parsimonious floodplain surface water-groundwater interaction model: Uncertainty analysis of floodplain model, Water Resources Research, 53(9), 7678–7695, https://doi.org/10.1002/2017WR020749, 2017.

McGlynn, B. L. and Seibert, J.: Distributed assessment of contributing area and riparian buffering along stream networks: TECHNICAL NOTE, Water Resources Research, 39(4), https://doi.org/10.1029/2002WR001521, 2003.

McKee, T. B., Doesken, N. J. and Kleist, J.: The Relationship of Drought Frequency and Duration to Time Scales, Proceedings of the Eighth Conference on Applied Climatology, American Meteorological Society, 179–184, 1993.

McMahon, G., Gregonis, S. M., Waltman, S. W., Omernik, J. M., Thorson, T. D., Freeouf, J. A., Rorick, A. H. and Keys, J. E.: Developing a Spatial Framework of Common Ecological Regions for the Conterminous United States, Environmental Management, 28(3), 293–316, https://doi.org/10.1007/s0026702429, 2001.

McManamay, R. A., Bevelhimer, M. S. and Kao, S.-C.: Updating the US hydrologic classification: an approach to clustering and stratifying ecohydrologic data: US HYDROLOGIC CLASSIFICATION, Ecohydrology, 7(3), 903–926, https://doi.org/10.1002/eco.1410, 2014.

McNamara, J. P., Tetzlaff, D., Bishop, K., Soulsby, C., Seyfried, M., Peters, N. E., Aulenbach, B. T. and Hooper, R.: Storage as a Metric of Catchment Comparison, Hydrological Processes, 25(21), 3364–3371, https://doi.org/10.1002/hyp.8113, 2011.

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M. and Uijlenhoet, R.: Representation of spatial and temporal variability in large-domain hydrological models: case study for a mesoscale pre-Alpine basin, Hydrology and Earth System Sciences, 20(6), 2207–2226, https://doi.org/10.5194/hess-20-2207-2016, 2016.

Mulvaney, T.: On the use of self registering rain and flood gauges, Institution of Civil Engineers Proceedings, 4, 1–8, 1851.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10(3), 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Newman, A., Sampson, K., Clark, M., Bock, A., Viger, R. and Blodgett, D.: A large-sample watershed-scale hydrometeorological dataset for the contiguous USA, , https://doi.org/10.5065/D6MW2F4D, 2014.

Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, River Research and Applications, 19(2), 101–121, https://doi.org/10.1002/rra.700, 2003.

Omernik, J. M. and Griffith, G. E.: Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework, Environmental Management, 54(6), 1249–1266, https://doi.org/10.1007/s00267-014-0364-1, 2014.

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J. and Zappa, M.: Does model performance improve with complexity? A case study with three hydrological models, Journal of Hydrology, 523, 147–159, https://doi.org/10.1016/j.jhydrol.2015.01.044, 2015.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F. and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?, Journal of Hydrology, 303(1–4), 290–306, https://doi.org/10.1016/j.jhydrol.2004.08.026, 2005.

Oudin, L., Kay, A., Andréassian, V. and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, Water Resources Research, 46(11), https://doi.org/10.1029/2009WR008887, 2010.

Patil, S. D., Wigington, P. J., Leibowitz, S. G., Sproles, E. A. and Comeleo, R. L.: How does spatial variability of climate affect catchment streamflow predictions?, Journal of Hydrology, 517, 135–145, https://doi.org/10.1016/j.jhydrol.2014.05.017, 2014.

Perrin, C., Michel, C. and Andréassian, V.: Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, Journal of Hydrology, 242(3–4), 275–301, https://doi.org/10.1016/S0022-1694(00)00393-0, 2001.

Pfister, L., Martínez-Carreras, N., Hissler, C., Klaus, J., Carrer, G. E., Stewart, M. K. and McDonnell, J. J.: Bedrock geology controls on catchment storage, mixing, and release: A comparative analysis of 16 nested catchments, Hydrological Processes, 31(10), 1828–1845, https://doi.org/10.1002/hyp.11134, 2017.

Potter, N. J., Zhang, L., Milly, P. C. D., McMahon, T. A. and Jakeman, A. J.: Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments: WATER BALANCE OF AUSTRALIAN CATCHMENTS, Water Resources Research, 41(6), https://doi.org/10.1029/2004WR003697, 2005.

Quan, Z., Teng, J., Sun, W., Cheng, T. and Zhang, J.: Evaluation of the HYMOD model for rainfall–runoff simulation using the GLUE method, Proceedings of the International Association of Hydrological Sciences, 368, 180–185, https://doi.org/10.5194/piahs-368-180-2015, 2015.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J. and DMIP Participants, and: Overall distributed model intercomparison project results, Journal of Hydrology, 298(1–4), 27–60, https://doi.org/10.1016/j.jhydrol.2004.03.031, 2004.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M. and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors: IDENTIFIABILITY OF INPUT AND STRUCTURAL ERRORS, Water Resour. Res., 46(5), https://doi.org/10.1029/2009WR008328, 2010.

Rouhier, L., Le Lay, M., Garavaglia, F., Le Moine, N., Hendrickx, F., Monteil, C. and Ribstein, P.: Impact of mesoscale spatial variability of climatic inputs and parameters on the hydrological response, Journal of Hydrology, 553, 13–25, https://doi.org/10.1016/j.jhydrol.2017.07.037, 2017.

Samadi, S., Tufford, D. L. and Carbone, G. J.: Assessing Parameter Uncertainty of a Semi-Distributed Hydrology Model for a Shallow Aquifer Dominated Environmental System, JAWRA Journal of the American Water Resources Association, 53(6), 1368–1389, https://doi.org/10.1111/1752-1688.12596, 2017.

Samani, Z.: Estimating solar radiation and evapotranspiration using minimum climatological data, Journal of Irrigation and Drainage Engineering, 126(4), https://doi.org/10.1061/(ASCE)0733-9437(2000)126:4(265), 2000.

Savenije, H. H. G.: HESS Opinions: Linking Darcy's equation to the linear reservoir, Hydrology and Earth System Sciences, 22(3), 1911–1916, https://doi.org/10.5194/hess-22-1911-2018, 2018.

Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M. and Carrillo, G.: Characterizing hydrologic change through catchment classification, Hydrology and Earth System Sciences, 18(1), 273–285, https://doi.org/10.5194/hess-18-273-2014, 2014.

Sayama, T., McDonnell, J. J., Dhakal, A. and Sullivan, K.: How much water can a watershed store?, Hydrological Processes, 25(25), 3899–3908, https://doi.org/10.1002/hyp.8288, 2011.

Schaake, J., Cong, S. Z. and Duan, Q. Y.: The US MOPEX data set, IAHS-AISH, 307, 9–28, 2006.

Seibert, J.: Estimation of Parameter Uncertainty in the HBV Model, Hydrol Res, 28(4–5), 247, https://doi.org/10.2166/nh.1998.15, 1997.

Seiller, G. and Anctil, F.: How do potential evapotranspiration formulas influence hydrological projections?, Hydrological Sciences Journal, 61(12), 2249–2266, https://doi.org/10.1080/02626667.2015.1100302, 2016.

Shen, Z. Y., Chen, L. and Chen, T.: Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China, Hydrology and Earth System Sciences, 16(1), 121–132, https://doi.org/10.5194/hess-16-121-2012, 2012.

Sherman, L. K.: Streamflow from rainfall by the unit graph method, Engineering News Record, 108, 501–505, 1932.

Sikorska, A. E. and Seibert, J.: Appropriate temporal resolution of precipitation data for discharge modelling in pre-alpine catchments, Hydrological Sciences Journal, 63(1), 1–16, https://doi.org/10.1080/02626667.2017.1410279, 2018.

Singh, R., Archfield, S. A. and Wagener, T.: Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments – A comparative hydrology approach, Journal of Hydrology, 517, 985–996, https://doi.org/10.1016/j.jhydrol.2014.06.030, 2014.

Singh, V. P.: Is hydrology kinematic?, Hydrological Processes, 16(3), 667–716, https://doi.org/10.1002/hyp.306, 2002.

Singh, V. P. and Xu, C.-Y.: Sensitivity of mass transfer-based evaporation equations to errors in daily and monthly input data, Hydrological Processes, 11(11), 1465–1473, https://doi.org/10.1002/(SICI)1099-1085(199709)11:11<1465::AID-HYP452>3.0.CO;2-X, 1997.

Sivapalan, M.: Prediction in ungauged basins: a grand challenge for theoretical hydrology, Hydrological Processes, 17(15), 3163–3170, https://doi.org/10.1002/hyp.5155, 2003.

Sivapalan, M.: Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, in Encyclopedia of Hydrological Sciences, edited by M. G. Anderson and J. J. McDonnell, John Wiley & Sons, Ltd, Chichester, UK, https://doi.org/10.1002/0470848944.hsa012, , 2005a.

Sivapalan, M.: Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, in Encyclopedia of Hydrological Sciences, edited by M. G. Anderson and J. J. McDonnell, John Wiley & Sons, Ltd, Chichester, UK, https://doi.org/10.1002/0470848944.hsa012, , 2005b.

Son, K. and Sivapalan, M.: Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data: Improving model structure through auxiliary data, Water Resources Research, 43(1), https://doi.org/10.1029/2006WR005032, 2007.

Spence, C.: A Paradigm Shift in Hydrology: Storage Thresholds Across Scales Influence Catchment Runoff Generation: Paradigm shift in hydrology, Geography Compass, 4(7), 819–833, https://doi.org/10.1111/j.1749-8198.2010.00341.x, 2010.

Staudinger, M., Stoelzle, M., Seeger, S., Seibert, J., Weiler, M. and Stahl, K.: Catchment water storage variation with elevation, Hydrological Processes, 31(11), 2000–2015, https://doi.org/10.1002/hyp.11158, 2017.

Stoelzle, M., Schuetz, T., Weiler, M., Stahl, K. and Tallaksen, L. M.: Beyond binary baseflow separation: a delayed-flow index for multiple streamflow contributions, Hydrol. Earth Syst. Sci., 24(2), 849–867, https://doi.org/10.5194/hess-24-849-2020, 2020.

Sugawara, M.: The flood forecasting by a series storage type model, IAHS Publication, 85, 1–6, 1967.

Sukristiyanti, S., Maria, R. and Lestiana, H.: Watershed-based Morphometric Analysis: A Review, IOP Conference Series: Earth and Environmental Science, 118, 012028, https://doi.org/10.1088/1755-1315/118/1/012028, 2018.

Sun, Y., Tian, F., Yang, L. and Hu, H.: Exploring the spatial variability of contributions from climate variation and change in catchment properties to streamflow decrease in a mesoscale basin by three different methods, Journal of Hydrology, 508, 170–180, https://doi.org/10.1016/j.jhydrol.2013.11.004, 2014.

Tallaksen, L. M.: A review of baseflow recession analysis, Journal of Hydrology, 165(1–4), 349–370, https://doi.org/10.1016/0022-1694(94)02540-R, 1995.

Tetzlaff, D., McNamara, J. P. and Carey, S. K.: Measurements and modelling of storage dynamics across scales: Measurements and modelling of storage dynamics across scales, Hydrol. Process., 25(25), 3831–3835, https://doi.org/10.1002/hyp.8396, 2011.

Teuling, A. J., Lehner, I., Kirchner, J. W. and Seneviratne, S. I.: Catchments as simple dynamical systems: Experience from a Swiss prealpine catchment, Water Resources Research, 46(10), https://doi.org/10.1029/2009WR008777, 2010.

Teweldebrhan, A. T., Burkhart, J. F. and Schuler, T. V.: Parameter uncertainty analysis for an operational hydrological model using residual-based and limits of acceptability approaches, Hydrology and Earth System Sciences, 22(9), 5021–5039, https://doi.org/10.5194/hess-22-5021-2018, 2018.

Todini, E.: Hydrological catchment modelling: past, present and future, Hydrology and Earth System Sciences, 11(1), 468–482, https://doi.org/10.5194/hess-11-468-2007, 2007.

Toth, E.: Catchment classification based on characterisation of streamflow and precipitation time series, Hydrology and Earth System Sciences, 17(3), 1149–1159, https://doi.org/10.5194/hess-17-1149-2013, 2013.

Trancoso, R., Phinn, S., McVicar, T. R., Larsen, J. R. and McAlpine, C. A.: Regional variation in streamflow drivers across a continental climatic gradient, Ecohydrology, 10(3), e1816, https://doi.org/10.1002/eco.1816, 2017.

Tromp-van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis: THRESHOLD FLOW RELATIONS, 2, Water Resour. Res., 42(2), https://doi.org/10.1029/2004WR003800, 2006.

Turner, H. M. and Burdoin, A. S.: The flood hydrograph, Boston Society of Civil Engineering, 28, 232–256, 1941.

Tyralis, H., Papacharalampous, G. and Tantanee, S.: How to explain and predict the shape parameter of the generalized extreme value distribution of streamflow extremes using a big dataset, Journal of Hydrology, 574, 628–645, https://doi.org/10.1016/j.jhydrol.2019.04.070, 2019.

Vázquez, R. F.: Effect of potential evapotranspiration estimates on effective parameters and performance of the MIKE SHE-code applied to a medium-size catchment, Journal of Hydrology, 270(3–4), 309–327, https://doi.org/10.1016/S0022-1694(02)00308-6, 2003.

Wagener, T., Sivapalan, M., Troch, P. and Woods, R.: Catchment Classification and Hydrologic Similarity, Geography Compass, 1(4), 901–931, https://doi.org/10.1111/j.1749-8198.2007.00039.x, 2007.

Wang, J., Endreny, T. A. and Hassett, J. M.: A flexible modeling package for topographically based watershed hydrology, Journal of Hydrology, 314(1–4), 78–91, https://doi.org/10.1016/j.jhydrol.2005.03.030, 2005.

Ward, J. H.: Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association, 58(301), 236–244, https://doi.org/10.1080/01621459.1963.10500845, 1963.

Westerberg, I. K. and Birkel, C.: Observational uncertainties in hypothesis testing: investigating the hydrological functioning of a tropical catchment: Observational Uncertainties in Hypothesis Testing, Hydrological Processes, 29(23), 4863–4879, https://doi.org/10.1002/hyp.10533, 2015.

Westerberg, I. K. and McMillan, H. K.: Uncertainty in hydrological signatures, Hydrology and Earth System Sciences, 19(9), 3951–3968, https://doi.org/10.5194/hess-19-3951-2015, 2015.

Wienhöfer, J. and Zehe, E.: Predicting subsurface stormflow response of a forested hillslope – the role of connected flow paths, Hydrol. Earth Syst. Sci., 18(1), 121–138, https://doi.org/10.5194/hess-18-121-2014, 2014.

Wilby, R. L.: Uncertainty in water resource model parameters used for climate change impact assessment, Hydrological Processes, 19(16), 3201–3219, https://doi.org/10.1002/hyp.5819, 2005.

Winsemius, H. C., Schaefli, B., Montanari, A. and Savenije, H. H. G.: On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information: Integrating hard and soft information, Water Resources Research, 45(12), https://doi.org/10.1029/2009WR007706, 2009.

Wittenberg, H.: Baseflow recession and recharge as nonlinear storage processes, Hydrological Processes, 13(5), 715–726, 1999.

Wittmann, S.: Tritiumgestützte Wasserbilanzierung im Einzugsgebiet von Fulda und Werra, <http://www.hydrology.uni-freiburg.de/abschluss/Wittmann_S_2002_DA.pdf>, Diploma-Thesis at the Institut for Hydrology,Albert-Ludwigs-University Freiburg, 2002.

Xu, C.-Y. and Singh, V. P.: Cross comparison of empirical equations for calculating potential evapotranspiration with data from Switzerland, Water Resources Management, 16(3), 197–219, 2002.

Yadav, M., Wagener, T. and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Advances in Water Resources, 30(8), 1756–1774, https://doi.org/10.1016/j.advwatres.2007.01.005, 2007.

Yaduvanshi, A., Srivastava, P., Worqlul, A. and Sinha, A.: Uncertainty in a Lumped and a Semi-Distributed Model for Discharge Prediction in Ghatshila Catchment, Water, 10(4), 381, https://doi.org/10.3390/w10040381, 2018.

Yaeger, M., Coopersmith, E., Ye, S., Cheng, L., Viglione, A. and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves – Part 4: A synthesis of empirical analysis, process modeling and catchment classification, Hydrology and Earth System Sciences, 16(11), 4483–4498, https://doi.org/10.5194/hess-16-4483-2012, 2012.

Yeung, K. Y. and Ruzzo, W. L.: Principal component analysis for clustering gene expression data, Bioinformatics, 17(9), 763–774, https://doi.org/10.1093/bioinformatics/17.9.763, 2001.

Zehe, E., Blume, T. and Blöschl, G.: The principle of 'maximum energy dissipation': a novel thermodynamic perspective on rapid water flow in connected soil structures, Phil. Trans. R. Soc. B, 365(1545), 1377–1386, https://doi.org/10.1098/rstb.2009.0308, 2010.

Zhang, J., Li, Y., Huang, G., Chen, X. and Bao, A.: Assessment of parameter uncertainty in hydrological model using a Markov-Chain-Monte-Carlo-based multilevel-factorial-analysis method, Journal of Hydrology, 538, 471–486, https://doi.org/10.1016/j.jhydrol.2016.04.044, 2016.

Zhao, F., Wu, Y., Qiu, L., Sun, Y., Sun, L., Li, Q., Niu, J. and Wang, G.: Parameter Uncertainty Analysis of the SWAT Model in a Mountain-Loess Transitional Watershed on the Chinese Loess Plateau, Water, 10(6), 690, https://doi.org/10.3390/w10060690, 2018.

Zhao, F., Wu, Y., Sivakumar, B., Long, A., Qiu, L., Chen, J., Wang, L., Liu, S. and Hu, H.: Climatic and hydrologic controls on net primary production in a semiarid loess watershed, Journal of Hydrology, 568, 803–815, https://doi.org/10.1016/j.jhydrol.2018.11.031, 2019.

Zuecco, G., Penna, D., Borga, M. and van Meerveld, H. J.: A versatile index to characterize hysteresis between hydrological variables at the runoff event timescale: A Hysteresis Index for Variables at the Runoff Event Timescale, Hydrol. Process., 30(9), 1449–1466, https://doi.org/10.1002/hyp.10681, 2016.

Additional Publications

Published

Drahorad, S. L., Jehn, F. U., Ellerbrock, R. H., Siemens, J. and Felix-Henningsen, P.: Soil organic matter content and its aliphatic character define the hydrophobicity of biocrusts in different successional stages, Ecohydrology, https://doi.org/10.1002/eco.2232, 2020.

Jehn, F. U., Breuer, L., Houska, T., Bestian, K. and Kraft, P.: Incremental model breakdown to assess the multi-hypotheses problem, Hydrology and Earth System Sciences, 22(8), 4565–4581, https://doi.org/10.5194/hess-22-4565-2018, 2018.

Submitted

Houska, T., Kraft, P., Jehn, F.U., Bestian K., Kraus, D., Breuer, L., Detection of hidden model errors by combining single and multi-criteria calibration, Submitted to Science of the Total Environment

Jehn, F.U., Schneider, M., Wang, J.R., Kemp, L., Breuer, L., Betting on the best case: Extreme warming is underrepresented in research, Submitted to Nature Climate Change

Acknowledgments

This dissertation was funded by money which Lutz conjured somehow. Therefore, first I would like to acknowledge all the luck I had during my time writing this dissertation. So many little things had to happen to give me the opportunity to work on those topics for three (or more like four) years. Many (if not all) of those lucky events, I owe to the great people in my life (listed in random order):

- Thank you Lutz for being one of the main reasons I even started to do a PhD. Without your friendly and welcoming character, I probably would not even have considered doing a PhD. You allowed me to freely follow the path were my ideas took me, but you always had an open ear for my problems if I needed it.
- Thank you Philipp and Tobias. You had to answer so many questions from me and still you did not get annoyed. I really appreciated all the conversations about science and all the other topics we had.
- Thank you Paula for always giving me support when I struggled. You had to listen to so many PhD related complains over the years, but still you always had a kind word for me. In addition, all the evenings with great food and your company allowed me to keep my batteries charged even in stressful times. Thank you for making my life better!
- Thank you Captain Janeway and Major Kira for being great cats!
- Thank you Konrad for being a great office neighbor, a great flat mate and great person overall.
- Than you Jutta and Hans-Werner for being great parents. You always encouraged me in what I did, even if it was sometimes hard for you to understand why I did it.
- Thank you Johannes and Anna-Lena for all your support. I am glad to have you as my siblings.
- Thank you Marius for being a great friend for so long. Our Age of Empires 2 matches always calm me down and you inspire me with your curiosity.
- Thank you Luise for all the nice conversation and all the wine we drank together. Talking to you helped me figure things out and I hope that we will do so again someday in the future.
- Thanks to Michael and Sabine for supporting me.
- Thank you Ekaterina for talking with me about what to do after my PhD. Our conversations helped to decrease my anxiety around this topic.
- Thank you to all the nice people in my institute.
- Thank you to the EA Darmstadt group for sparking interesting thoughts about many topics.
- Thank you to Leonie and Barbara for giving me such a warm welcome in Zürich.
- Thank you to all the people that talked with me about my dissertation and made it better (this list is quite incomplete): Ina Pohle, Marc Vis, Jan Seibert, Wouter Knoben, Andrew Newmann, Sandra Pool, Diana Spieler, Stefanie Lutz and many more.
- Thank you for all the other people in my life. I am really lucky to have such fine company ☺

I hope no one reads this list and is sad they did not make it by name. I assure you this is a mistake on my part and I am grateful for you!

Erklärung gemäß der Promotionsordnung des Fachbereichs 09 vom 07. Juli 2004 § 17 (2)

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis" niedergelegt sind, eingehalten.

Darmstadt, 10th February 2021

____

Florian Ulrich Jehn