Baseflow generation at the catchment scale – An investigation using comparative hydrology

By

SEBASTIAN JOHANNES GNANN

Department of Civil Engineering
UNIVERSITY OF BRISTOL

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

FEBRUARY 2021

Word count: 34927

Abstract

Baseflow is typically defined as flow that comes from delayed sources and thus sustains streamflow during dry periods. Baseflow is both an important water resource and an important aspect of catchment functioning, and thus a key element of hydrological theory. Yet we still lack a thorough understanding of how climate and landscape properties control baseflow generation. This thesis investigates baseflow generation at the catchment scale through comparative hydrology, in catchments that are mostly free from human impacts. The thesis is centred around three inter-related aspects: (1) the quantification of baseflow through hydrological signatures, (2) the exploration of climate and landscape attributes that control baseflow generation, and (3) the search for simple models that adequately simulate baseflow. First, we show that mean annual baseflow is controlled both by climate aridity and by catchment storage capacity. In humid catchments, baseflow is limited by a catchment's storage capacity, while in arid catchments, baseflow is limited by high amounts of evapotranspiration. Second, we show how the seasonal translation of climatic forcing into streamflow can be used to define signatures of catchment response that complement typically used baseflow signatures. We show that these signatures vary systematically with climate and landscape attributes, and we use them in a model evaluation experiment. Third, we show that typically used (subsurface) catchment attributes lack hydrologically relevant information, but that region-specific knowledge can improve baseflow signature predictions. We also show that multiple baseflow signatures are needed to better distinguish between different baseflow sources. Finally, we discuss overarching questions and future research needs that we ultimately need to address to advance baseflow hydrology, particularly if we want to predict baseflow under change. Overall, this thesis advances our knowledge in relation to all three above-mentioned aspects and thereby improves our understanding of baseflow generation at the catchment scale.

ACKNOWLEDGEMENTS

This work was funded as part of the Water Informatics Science and Engineering Centre for Doctoral Training (WISE CDT) under a grant from the Engineering and Physical Sciences Research Council (EPSRC), grant number EP/L016214/1.

I am thankful to the many people who have supported me during the journey that led to this work. First and foremost, I want to thank my supervisors Nicholas Howden and Ross Woods, who guided me through this big project, left me with enough space to explore my own ideas, and enabled me to experience how exciting research can be. Thanks to Nicholas, for always being supportive and for asking me how I am, and not just how my work is. Thanks to Ross, for all the good hints and for providing me with the big picture view. Thanks to Hilary McMillan, for hosting me in San Diego, letting me be part of her lab throughout the mostly virtual last year, and for being an exceptional additional supervisor. Thanks to Thorsten Wagener and the Water Group at Bristol, for creating such a stimulating and comfortable environment. Thanks to Gemma Coxon for always having some data at hand and for promoting open science. Thanks to Wouter Berghuijs for inviting me to contemplate a little bit on the Budyko framework.

Thanks to my WISE colleagues, many of whom have become great friends and who made this time so enjoyable: Lina, Charlie, Simbi, Gwyn, Levke, Jack, Lola, and all the others. I don't know half of you half as well as I should like; and I like less than half of you half as well as you deserve (this is, of course, from Tolkien). Thanks – or perhaps rather grazie mille – to my office mates and the whole Woodland Road office, for all the great coffee breaks, cakes, lunches, wannabe gardening, and pub sessions. What a pity that we (mostly) missed out on that in the last year. Special thanks to Wouter Knoben, for many discussions about hydrology and lots of useful feedback on drafts I wrote, including many parts of this thesis. Thanks to Vale and Anna for keeping the green rebellion up and running. Thanks to all the other people I met along the way, in Bristol, the UK (still European at heart), San Diego, at conferences, summer schools, or elsewhere. Thanks to my friends from Tübingen and from back home, who in one way or another, helped me to get ready for this adventure and stayed with me all along.

Most importantly, thanks to my family – to Mum, Dad, and my sister Linda – for their unconditional support throughout my entire life. Danke für alles!

Those who know nothing must believe everything.

— Marie von Ebner-Eschenbach

AUTHOR'S DECLARATION

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

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

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1 Background

1.1.1 Baseflow: relevance and definition

Baseflow is typically defined as flow that comes from delayed sources and thus sustains streamflow during dry periods (Hall, 1968). Baseflow is an important freshwater resource that provides water for human consumption, agriculture, industry, and hydropower (Smakhtin, 2001; Price, 2011; Beck et al., 2013). Understanding baseflow processes is essential for low flow forecasting, as baseflow is the main source of streamflow during drought conditions (Smakhtin, 2001; Stoelzle et al., 2014). Baseflow influences water quality by diluting effluent returns from wastewater treatment plants and by transporting contaminants present in soils and groundwater (Price, 2011). Baseflow is also crucial for ecosystem functioning. Baseflow provides habitat for stream biota, and by affecting water temperature and water quality it affects the ecological status of a stream (Poff et al., 1997; Price, 2011). Managing water resources in a way that benefits both society and nature, thus requires an understanding of the processes that generate baseflow.

Baseflow is an old yet still debated concept in hydrology as "flow that comes from delayed sources" is a rather loose definition (e.g. Hewlett and Hibbert, 1967; Hall, 1968; Freeze, 1972; Pelletier and Andréassian, 2020). First, we have to define what counts as "delayed". This is typically done through a time scale. Second, there are various delayed sources and their contributions vary in space and time. Therefore, baseflow can be associated with very different hydrological processes, both within and between catchments (e.g. Hall, 1968; Cartwright et al., 2014; Stoelzle et al., 2020). Nonetheless, baseflow hydrology remains a very active research area as baseflow is an important water resource and an important aspect of catchment response. Yet for baseflow to be a useful concept, we have to be clear about its definition (Hall, 1968).

So how to define baseflow? We could define baseflow purely based on the hydraulic response of a catchment, independent of the processes involved. The slowly varying component of streamflow is then called baseflow and the remaining component fast flow. Most hydrograph separation methods aim at such a separation into slow and fast streamflow components (e.g. Lyne and Hollick, 1979; Institute of Hydrology, 1980). While a binary separation will always be imperfect in a naturally continuous system (Stoelzle et al., 2020; Pelletier and Andréassian, 2020), it can still be a useful starting point to classify catchment response. If we instead focus on the processes that generate baseflow, we could use different water sources as our definition. We could say that certain – typically slowly varying – streamflow sources contribute baseflow, such as groundwater discharge, lake outflow, or snowmelt. Understanding the processes involved in baseflow generation is particularly important if we want to understand how environmental and human induced change will affect baseflow. Yet in practice, we often lack the necessary information to distinguish between different baseflow sources. We therefore need ways to quantify baseflow based on widely available observations such as streamflow measurements.

1.1.2 Quantifying baseflow

1.1.2.1 Baseflow separation

Direct hydrograph-based estimations of baseflow aim at separating the hydrograph into several (typically two) components. If we assume streamflow Q to consist of fast flow Q_f and baseflow Q_b, we can write:

$$Q = Q_b + Q_f \tag{1.1}$$

Such a separation into Q_b and Q_f can be performed with various baseflow separation methods, two of which are compared in Figure 1.1. Automated graphical methods systematically connect minima within a certain interval and the flow below these lines is considered baseflow (Institute of Hydrology, 1980; Sloto and Crouse, 1996). Digital filters aim at filtering out high-frequency streamflow components so that only a smoothed baseflow hydrograph remains (Lyne and Hollick, 1979; Chapman, 1991; Boughton, 1993; Eckhardt, 2005). The degree of smoothing depends on one or more parameters that need to be specified. Conceptual methods view baseflow as the outflow from a conceptual reservoir fed by recharge (Wittenberg, 1999; Pelletier and Andréassian, 2020). The resulting baseflow estimate depends on the form and parameter values of that reservoir and on the way recharge is approximated. Other more process-based baseflow separation methods have been developed (e.g. digital filters based on groundwater theory; Huyck et al., 2005), but they usually require more parameters, more data, and make more assumptions.

All hydrograph-based baseflow separation methods have two main shortcomings. First, we cannot evaluate these methods as "true" baseflow is not measurable. Second, such methods by themselves do not allow us to make any inferences about baseflow generating processes. They have, therefore, been widely criticised for their sensitivity to subjective decisions (e.g. parameter

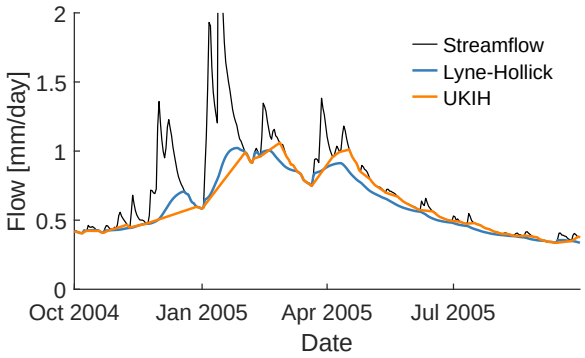

Figure 1.1: Comparison of two common baseflow separation methods: the Lyne-Hollick filter (Lyne and Hollick, 1979) and the UKIH method (Institute of Hydrology, 1980). Note that only a subperiod of the whole time series is shown and that the y-axis is capped. BFIs for the whole 34 year time series are 0.75 and 0.77 for the Lyne-Hollick filter and the UKIH method, respectively.

values) and for their weak hydrological basis (e.g. Hewlett and Hibbert, 1967; Hall, 1968; Klemeš, 1986; Beven, 2012).

The baseflow separation methods introduced so far only require streamflow and sometimes precipitation data, which are widely available. There are, however, other data sources that can help to gain more process insights. Groundwater level observations allow to explicitly estimate the groundwater contribution to baseflow (Peters and van Lanen, 2005; Gonzales et al., 2009; Costelloe et al., 2015). Tracer data such as stable isotope data (Klaus and McDonnell, 2013) or geochemical data (Cartwright et al., 2014) are probably the most widely used additional data source and they can help to identify sources of streamflow. Since tracer-based methods investigate transit times (which are a function of velocities), their baseflow estimates are not directly comparable to baseflow defined as the slow hydraulic response of a catchment (which is a function of celerities; Cartwright et al., 2014; McDonnell and Beven, 2014; Hrachowitz et al., 2016; Harman, 2019). Stored water can be released quickly in response to rainfall inputs (Kirchner, 2003), and slowly released water can have a geochemical signature similar to surface water (Cartwright et al., 2014). Tracer data can, however, help to identify the sources of water and therefore enable important process-based insights into baseflow generating mechanisms that cannot be gained from streamflow-based methods alone.

1.1.2.2 Hydrological signatures describing baseflow

Rather than using a time series of estimated baseflow, we can look at summary metrics that describe aspects of baseflow response, that is hydrological signatures (McMillan et al., 2017). The probably most common baseflow signature is the baseflow index (BFI), the ratio between mean baseflow $\bar{Q}_b$ and mean total streamflow $\bar{Q}$.

$$\text{BFI} = \frac{\bar{Q}_b}{\bar{Q}} \tag{1.2}$$

The BFI varies between 0 and 1, where 1 implies that streamflow is fully made up of baseflow and vice versa. Since the BFI is estimated by means of a baseflow separation method, it has the same shortcomings as the underlying method. Its value cannot be compared to a "true" BFI and it – by itself – does not allow for process interpretations. The BFI has, however, still proven useful as a metric for catchment comparison as it quantifies roughly how smooth or flashy a stream is. For example, many studies have shown that the BFI relates to soil types or geology associated with productive aquifers (e.g. Boorman et al., 1995; Bloomfield et al., 2009), or to other baseflow sources such as snow and surface water bodies (Beck et al., 2013). Figure 1.1b compares BFIs obtained with two baseflow separation methods for a large sample of catchments in the US (for a description of the data sources, see e.g. Chapter 2). The catchments rank very consistently, indicating that a comparative approach can lead to robust results independent of the baseflow separation method.

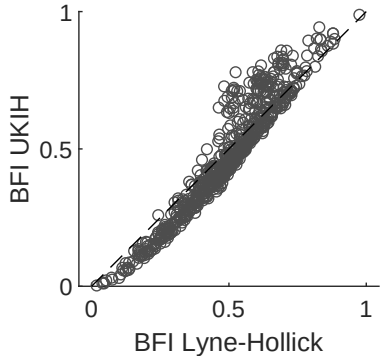

Figure 1.2: Comparison of BFI obtained with two common baseflow separation methods: the Lyne-Hollick filter (Lyne and Hollick, 1979) and the UKIH method (Institute of Hydrology, 1980). The Spearman rank correlation ρ_s is 0.96.

Besides the BFI there are various other metrics that target baseflow processes. Recessions – the falling limbs of a hydrograph when catchment storage is depleted – can be analysed to better understand the drainage behaviour of a catchment (Brutsaert and Nieber, 1977; Tallaksen, 1995). It is often assumed that the relationship between the rate of change of streamflow and streamflow follows a power law:

$$-\frac{dQ}{dt} = \alpha Q^{\beta} \tag{1.3}$$

where α and β ($\beta = 1$ implies a linear reservoir) are parameters that can be obtained by fitting Eq. (1.3) to recession data. Recession analysis provides insights into the temporal dynamics of streamflow and its relation to storage, and it has been used to specify parameters used in baseflow separation methods (e.g. Eckhardt, 2005) and hydrological model building (e.g. Clark et al., 2009; McMillan et al., 2011).

The flow duration curve (FDC) shows the percentage of time that a certain flow is equalled or exceeded in a given record (Winter, 2007). A FDC with a low slope indicates a stable regime,

typically attributed to the influence of groundwater or other baseflow sources (Winter, 2007). This can be quantified by using the slope of the middle segment of the FDC (Yadav et al., 2007). Low flow signatures can also be used to study baseflow as low flows are typically made up of baseflow (Smakhtin, 2001). This includes different percentiles of the FDC such as the 5% flow percentile Q_5 (flow that is *not* exceeded 5% of the time), the "lowest streamflow for seven consecutive days that would be expected to occur once every 10 years" (7Q10; Price, 2011), or the the mean annual 10-day minimum (MAM(10); Institute of Hydrology, 1980). Each metric might also be calculated for different periods of time, for example, for each season or year, which can provide further insights into catchment processes (e.g. Zimmer and Gannon, 2018).

Hydrological signatures have been used for various purposes, such as catchment classification (Wagener et al., 2007; Sawicz et al., 2011), hydrological process exploration (McMillan et al., 2014; Zimmer and Gannon, 2018), and model building and evaluation (Gupta et al., 2008; Euser et al., 2013; Hrachowitz et al., 2014). Yet while various baseflow signatures exist, many of them do not explicitly provide information about processes (e.g. the BFI as discussed above), are difficult to compare between catchments (e.g. non-normalised flow percentiles; Price et al., 2011), or are associated with methodological uncertainty (e.g. recession characteristics; Stoelzle et al., 2013; Dralle et al., 2017; Jachens et al., 2020). Hence, we still need to identify the most informative and robust baseflow signatures. And we also need to better understand the relationship between signatures and catchment characteristics which would enable a clearer linkage to baseflow generation processes (Addor et al., 2018; McMillan, 2020).

1.1.3 Baseflow generating processes – controls and mechanisms

As a way to organise our knowledge of catchment hydrological processes, we adapt the classification framework by Wagener et al. (2007) and distinguish between:

- forcing, that is the climatic input a catchment receives;
- catchment form, that is (mostly) static catchment characteristics such as geology;
- catchment function, that is the actions of a catchment on the water and energy it receives.

Catchment functions are further divided into partition, storage, and release (Wagener et al., 2007). Figure 1.3 shows a conceptual diagram that applies this framework to baseflow generating processes. In the next paragraphs, we will explain the different catchment functions with a focus on baseflow, and review our current understanding of how form and forcing control catchment function.

When water falls onto a catchment, it will be partitioned depending on forcing characteristics and catchment form. It might be intercepted (and eventually evaporated), stored as snow, run off to a stream or a lake, or infiltrate. This partitioning can be seen as the first control on baseflow generation, as it determines how much water is going into each store. Especially infiltration characteristics of a catchment have a key influence on baseflow generation as they determine

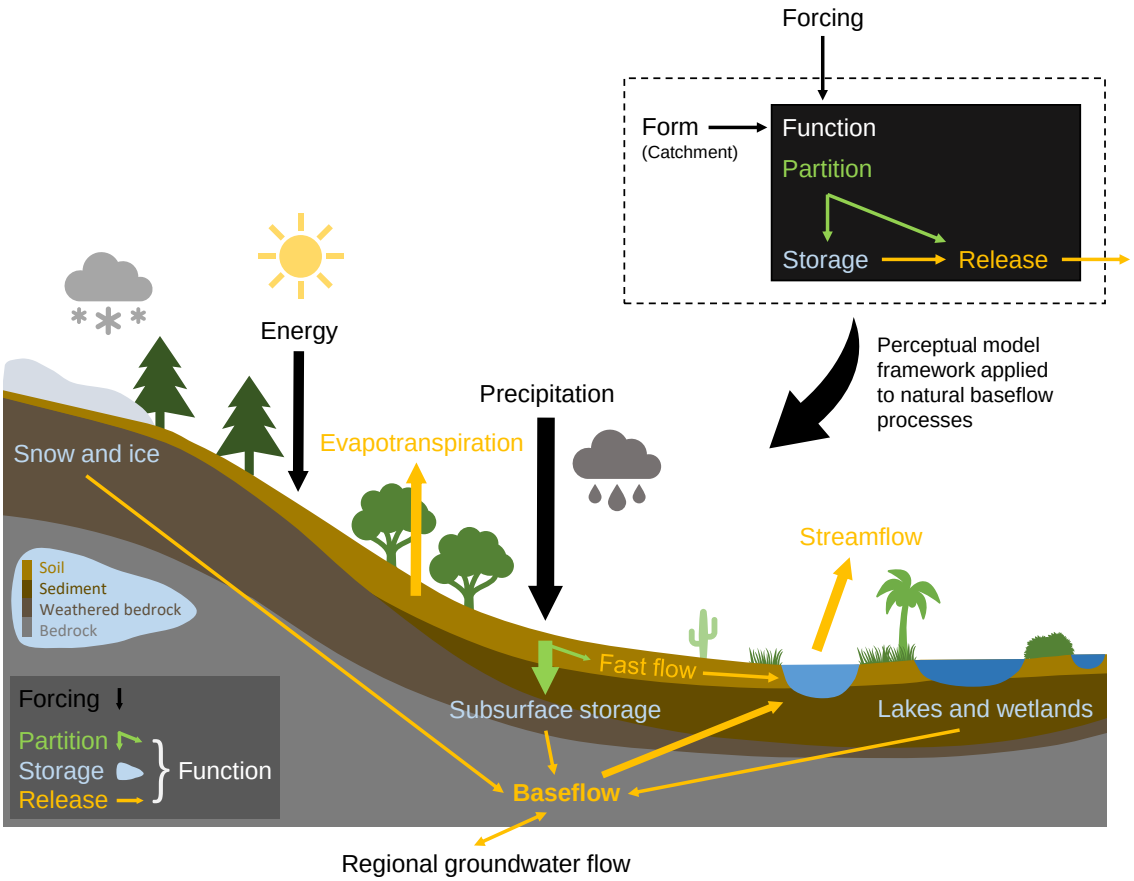

Figure 1.3: Conceptual diagram showing natural baseflow processes (taken from Chapter 4). The terminology follows Wagener et al. (2007) who distinguish between forcing, catchment form, and catchment function in the form of partition, storage, and release.

recharge (Neff et al., 2005; Naylor et al., 2016). This is supported by various studies that found soils (Boorman et al., 1995; Schneider et al., 2007; Santhi et al., 2008) and land use (Zhang and Schilling, 2006) to be important controls on baseflow generation.

Storage in different parts of a catchment is a key control on catchment functioning in general, and baseflow generation in particular (McNamara et al., 2011; Buttle, 2016). Following McDonnell and Woods (2004) we can distinguish between water that is stored (a) in surface water bodies, (b) as snow or ice, and (c) in the subsurface. Here we mostly focus in storage in the subsurface, in particular groundwater. Storage in the subsurface can be in soils (including saprolite), unconsolidated or consolidated sediments, weathered bedrock (regolith), and porous or fractured bedrock (Pelletier et al., 2016; Rempe and Dietrich, 2018). Water might be stored preferentially in certain parts of a catchment, such as in the proximity of the river where flow paths converge (e.g. in the alluvium; Price, 2011; Cartwright et al., 2014; Käser and Hunkeler, 2016). Various studies could link baseflow to different storage sources, such as surface water bodies (Beck et al., 2013; McLaughlin et al., 2014; Ameli and Creed, 2017), snowmelt (Beck

et al., 2013; Safeeq et al., 2013; Addor et al., 2018), and also to various catchment characteristics related to subsurface storage capacity, such as soils (Boorman et al., 1995; Schneider et al., 2007; Santhi et al., 2008), geology (Farvolden, 1963; Tague and Grant, 2004; Bloomfield et al., 2009; Pfister et al., 2017; Carlier et al., 2018), geology-vegetation groups (Lacey and Grayson, 1998), or topography (Santhi et al., 2008; Sayama et al., 2011; Price et al., 2011).

Eventually, water will be released from a catchment. Some water is released quickly after a storm, for example as overland flow. Most water is released via evapotranspiration which exceeds global continental runoff (Rodell et al., 2015). Since plants transpire water stored in soil and groundwater, transpiration is in direct competition to recharge and baseflow. The remaining stored water will be released via different flow paths. Even the water that is released quickly during storms is often made up of stored water (Kirchner, 2003; Berghuijs et al., 2017). Water released between rainfall events is entirely made up of stored water, but the relationship between the age of the released water (i.e. the transit time) and the hydraulic response time is not well understood (Harman, 2019). But also the hydraulic response itself is a matter of active debate, in particular the nature of the storage-discharge relationship (i.e. recession characteristics, see Eq. (1.3)). Recession characteristics have been related to hillslope groundwater hydraulics and thus hydraulic subsurface properties (e.g. Brutsaert and Nieber, 1977; Troch et al., 2013), landscape heterogeneity (Harman et al., 2009; Harman and Sivapalan, 2009), continuous groundwater recharge (Fenicia et al., 2006), and variability in antecedent wetness or evapotranspiration (Tashie et al., 2019, 2020). In some catchments, substantial amounts of water might also come from or leave to neighbouring catchments as regional groundwater flow, which is difficult to quantify but crucial to close the water balance (Schaller and Fan, 2009; Fan, 2019).

While different aspects of catchment form have shown to influence baseflow generation (e.g. soils and geology as described above), a lot of that knowledge is fragmented and place-specific (Beck et al., 2013). Partly this is due to our limited ability to observe the subsurface (Beven, 2012), resulting in limited information about subsurface properties, particularly at larger scales. But we also lack a thorough understanding of how heterogeneous subsurface properties upscale to the catchment scale, how multiple catchment characteristics (and processes) interact, and how that affects catchment response (Sivapalan, 2005; McDonnell et al., 2007). For example, differences in topography can affect recharge and hydraulic gradients. This can alter the hydrological response even if the hydraulic properties of the subsurface are the same (Carlier et al., 2019). Flow at the catchment scale also tends to be preferential, and hence a small part of the subsurface might have a disproportionally large impact on the catchment response (Beven and Germann, 1982).

The strong influence of climatic forcing on catchment response at long-term, seasonal, and event scales (Budyko, 1974; Berghuijs et al., 2014; Knoben et al., 2018) further complicates catchment comparison, particularly at large (continental to global) scales. The same geology might respond differently if it is forced by a different climate. Many studies have shown that climate is a strong control on average baseflow behaviour at larger scales (Van Dijk, 2010; Peña-

Arancibia et al., 2010; Beck et al., 2013). Furthermore, seasonal climatic variability can affect baseflow (Zimmer and Gannon, 2018) and recession characteristics (Tashie et al., 2019, 2020), and thus complicate the linkage between (static) catchment attributes and baseflow response. One way to tackle this problem is to analyse sub-climates (and seasons) individually. Ultimately, however, theories that integrate climate and landscape characteristics to explain catchment response would be worthwhile. Such theories would both serve as evidence of a thorough scientific understanding and be of practical use as they would allow for transferability in space and time (Dooge, 1986; Sivapalan, 2005; Wagener et al., 2007).

1.1.4 Modelling baseflow

Modelling is mostly carried out for two reasons. First, models are useful scientific tools that can help to test our theories and hypotheses, and to explain and understand hydrological phenomena. Second, models are practical tools that enable us to predict unknowns in space and time, and thereby inform water resources management and decision making (Kirchner, 2006; Beven, 2012). While a good fit to the data might be deemed enough for some purposes, we generally want models to be a realistic representation of hydrological processes (Kirchner, 2006; Clark et al., 2017). From a scientific modelling perspective this is fairly obvious (even though our conception of a realistic representation of hydrological processes might differ). But also practical applications require models that credibly represent processes, particularly if we want to extrapolate in space and time.

Models vary hugely in their complexity and type. There are various terminologies describing models, many of them rather ill-defined. There are purely data-based models that predict long-term system signatures such as the BFI based on catchment characteristics (Beck et al., 2013). Then there are spatially explicit, three-dimensional models based on small scale physical principles (e.g. Richards equation; Carlier et al., 2019). Many hydrological models are of intermediate complexity and describe the dominant aspects of catchment response (e.g. groundwater discharge) in a spatially lumped or semi-distributed, and temporally varying, way using interconnected storage elements (Fenicia et al., 2011). These models are often called conceptual models. Yet despite terminological differences, all models are to some degree conceptual and contain empirical elements (Hrachowitz and Clark, 2017).

All models – independent of their type – face some key challenges that repeatedly appear in the literature. They should be consistent with observations (which is perhaps the broadest definition of physically based; see Beven, 2002), they should adequately represent the dominant hydrological processes and their heterogeneity, and both their structure and their parameters should be based on (or at least constrained by) process understanding rather than calibration (Sivapalan, 2005; Kirchner, 2006; Clark et al., 2017; Hrachowitz and Clark, 2017). Furthermore, models – already imperfect themselves – are always fed with imperfect data and consequently result in uncertain predictions (Beven, 2012; McMillan et al., 2012; Hrachowitz and Clark, 2017).

So how to choose an appropriate hydrological model? We can build (and calibrate) models in a step-wise fashion based on our process understanding and diagnostic signatures (McMillan et al., 2011; Clark et al., 2011; Hrachowitz et al., 2014). This might allow us to constrain the model structure (in particular the number and arrangement of its subsurface stores and their parameterisations) to match recession characteristics (McMillan et al., 2011; Clark et al., 2011), or to put prior constraints on BFI and groundwater losses to neighbouring catchments (Hrachowitz et al., 2014). The advantage of such an approach is that we explicitly test our hypotheses, allowing for falsification and targeted model improvement (Clark et al., 2016). The disadvantage is the relatively large effort (data collection, data analysis, model implementation) which hampers its application to large samples of catchments.

Alternatively, we can compare multiple (existing) model structures and parameter sets and choose the one(s) that perform best according to some performance criteria. We might then check whether certain model structural elements and parameter values relate to certain catchment characteristics and our perceptual understanding of these catchments (Fenicia et al., 2014; Stoelzle et al., 2015; Knoben et al., 2020). For example, Stoelzle et al. (2015) found that certain models worked better for certain geologies, and that there is some conceptual agreement between model structure and calibrated parameter values and aquifer type. Fenicia et al. (2014) compared multiple model structures and could relate the best fitting model structure to bedrock permeability and the presence of multiple aquifers. While the tested models might be seen as multiple hypotheses (Fenicia et al., 2014), such an approach is less explicit. It can happen that no model or multiple models are adequate, and it is not necessarily clear why certain models perform well (or poorly).

A lot of effort has gone into transferring model structures and parameter values to ungauged catchments (i.e. regionalisation; Sivapalan, 2003; He et al., 2011; Wagener and Montanari, 2011; Hrachowitz et al., 2013). Modelling ungauged catchments remains a challenge and is closely related to our limited ability to specify model structures and estimate parameter values a priori (Duan et al., 2006; Clark et al., 2017). Partly, this is because the translation of physiographic similarity into hydrological similarity is not straightforward (He et al., 2011) and we currently lack the necessary understanding of how heterogeneous landscape properties and climate interact to form catchment response (Sivapalan, 2005; Wagener et al., 2007). Partly, this is because to date we cannot observe catchment characteristics well enough, particularly subsurface characteristics (Beven, 2012). As a result, correspondence between physiographic and hydrological similarity has often been limited to small (well studied) regions (e.g. Fenicia et al., 2014). In summary, advancements in baseflow modelling are impeded by similar challenges as our understanding of baseflow generation processes (outlined in Section 1.1.3). There remains a need for process-based theories and models that can be transferred to other places and to the future (McDonnell et al., 2007; Wagener et al., 2010).

1.2 Research objectives

The overall aim of this thesis is to better understand baseflow generation at the catchment scale ($\approx$ 10–10000 km^2) across different climate zones and landscapes. We adopt a comparative hydrology approach (Falkenmark and Chapman, 1989), which has been advocated by many authors as a means to find new theories of hydrology at the catchment scale (Sivapalan, 2005; McDonnell et al., 2007; Wagener et al., 2007; Harman and Troch, 2014). Investigating large samples of catchments allows us to explore catchment (dis-)similarity and patterns that would remain hidden if we would only study catchments in isolation. Comparative hydrology hence presents a powerful approach to generate findings that can be regionalised and thereby transcend – to some extent – the uniqueness of place (Beven, 2000).

With the advent of more and more large sample datasets (for a recent overview see Addor et al., 2020) we are given the opportunity to study large scale patterns in an unprecedented manner. Most large sample datasets – including the ones we use here – contain precipitation, potential evapotranspiration, and streamflow data, as well as a wide range of catchment attributes. Large sample tracer data or groundwater level data are far less widespread, and hence we focus on the study of streamflow data. We also limit our analyses to catchments that are mostly free from human impacts to avoid a conflation of natural and human-induced processes. We argue that before we can assess human impacts, we should first understand mostly natural systems.

Specifically, we are motivated by the following inter-related questions:

(1) How can we quantify baseflow in a meaningful way?

(2) Which climate and catchment attributes influence baseflow generation?

(3) How can we model baseflow adequately for large samples of catchments?

These three questions are not addressed individually in each research chapter, but to a different extent in all chapters throughout this thesis. We think that none of these questions has a definite answer, nor can they be answered separately. Thus, this thesis does not aim to provide specific answers to these questions. Rather, we aim at improving our understanding of baseflow generation at the catchment scale by letting us guide by these questions.

With regard to (1), we use and compare multiple hydrological signatures that focus on baseflow. As each baseflow signature aims at describing a certain aspect of baseflow response, it will be useful in a different context. While a complete description of baseflow based on streamflow data alone is impossible, using a variety of signatures can give us a more complete picture of the baseflow response. In particular, the relationship between different baseflow signatures might help us to better distinguish between hydrologically different catchments.

When addressing (2) for many catchments across different climate zones and landscapes, we face some key challenges. First, we need to better disentangle the roles of climatic forcing and catchment form (see Section 1.1.3). This might be done by using signatures that explicitly relate

streamflow response to climatic forcing, as they focus more explicitly on the transformation of the climate input into catchment output. Alternatively, we can analyse regions with a similar climate individually, and then try to combine the regional results through the use of a generic framework. Second, we need to make sure that the climate and catchment attributes contain information that is hydrologically relevant. Only if an attribute describes the hydrologically relevant aspects of a catchment, will it be possible to link it to catchment response.

Regarding (3), we focus on relatively simple conceptual models. From a practical point of view, simple models allow us to keep the computational demand manageable when studying large samples of catchments. Simple models allow us to focus on general patterns, rather than place-based details. That is, they allow us to balance depth with breadth (Gupta et al., 2014). This leads to more robust and hence potentially transferable conclusions. Focusing on simple models also allows us to avoid using (potentially) over-parameterised models that rely heavily on calibration and that are hard to falsify. Such models in combination with calibration often offer limited hydrological insight and their deficiencies are hard to diagnose (e.g. Kirchner, 2006).

1.3 Thesis outline

The thesis is divided into three research chapters, a conclusion chapter, and several appendices. Below we give an overview of the thesis structure and short descriptions of the three research chapters. Figure 1.4 shows how the three research chapters relate to the three research questions outlined in Section 1.2.

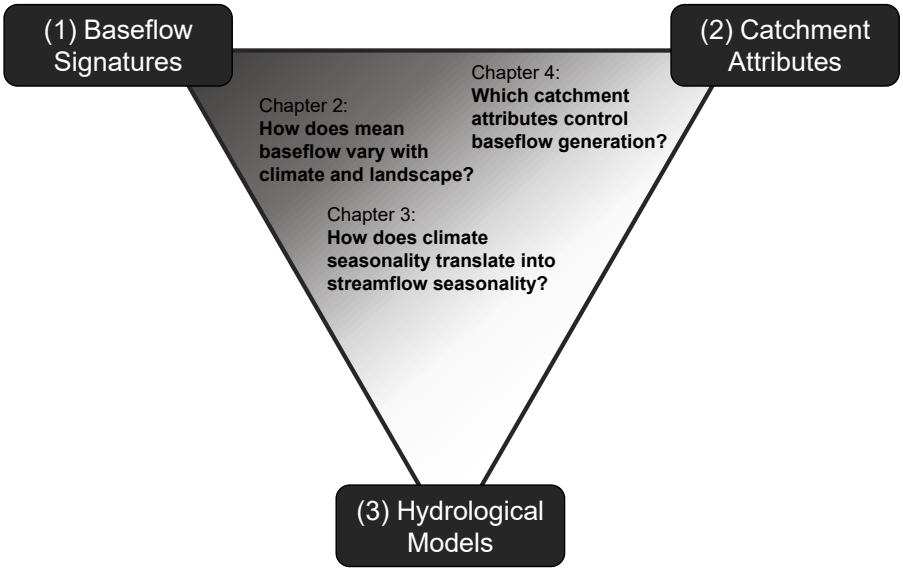

Figure 1.4: Overview that roughly illustrates how the research chapters relate to the three research questions outlined in Section 1.2. The grey-scale indicates the approximate extent to which each question is covered in this thesis; darker colours indicate stronger coverage.

Chapter 2: How does mean baseflow vary with climate and landscape?
The Budyko hypothesis (Budyko, 1974) states that the ratio between mean actual evapotranspiration and precipitation is primarily a function of the aridity index (the ratio between mean potential evapotranspiration and precipitation). In Chapter 2, we investigate whether the baseflow fraction (the ratio between mean baseflow and precipitation) is similarly strongly controlled by aridity. We compare hundreds of (non-snow dominated) catchments from the United States and the United Kingdom to explore how climate and landscape factors control baseflow fraction. We explain the observed patterns using a simple annual water balance model.

Chapter 3: How does climate seasonality translate into streamflow seasonality?
In Chapter 3, we introduce hydrological signatures describing the translation of climate seasonality into streamflow seasonality. We represent climate (precipitation minus potential evapotranspiration) and streamflow time series by their annual Fourier mode, that is by sine waves. We then use the amplitude ratio and the phase shift between these two sine waves as seasonal signatures. We calculate the seasonal signatures for hundreds of catchments in the UK and the US and explore how they vary with climate and catchment attributes. We also use the signatures to evaluate two conceptual hydrological models. The seasonal signatures offer an alternative view at seasonal streamflow dynamics and complement baseflow signatures such as the BFI.

Chapter 4: Which catchment attributes control baseflow generation?
In Chapter 4, we investigate the relationship between catchment attributes and several baseflow signatures for hundreds of catchments in the US. Specifically, we ask why catchment attributes such as geology descriptors have shown little explanatory power when used to predict baseflow signatures in recent large sample studies. By reviewing literature from several US regions, we compile regional knowledge that helps to explain regional baseflow patterns. We contrast this regional knowledge with information contained in often used large scale maps to explore if these maps contain sufficient information to predict baseflow response. We also explore how the use of multiple baseflow signatures can help to better identify different baseflow sources. To organise the findings from diverse regions, we propose and apply a framework based on perceptual models.

Chapter 5: Summary and outlook
In Chapter 5 we summarise the findings of this thesis, discuss their overall connection, and point at future research directions.

Appendices
Appendices A, B, and C contain Supporting Information for Chapters 2, 3, and 4, respectively.
Appendix D contains information on TOSSH: A toolbox for streamflow signatures in hydrology.
Appendix E contains information on an opinion paper about the Budyko framework.
Appendix F contains the author's CV.

Note on equations

Some equations appear in multiple chapters of this thesis. They are labelled separately so that each chapter is self-contained.

CHAPTER 2

IS THERE A BASEFLOW BUDYKO CURVE?

This chapter has been published as a research article in Water Resources Research. Slight modifications have been made to better fit the general layout of this thesis. We thank Gemma Coxon and Jim Freer for discussions and for assisting with the UKBN2 data. We acknowledge the comments from the associate editor and three anonymous reviewers, that helped to clarify and improve this chapter.

SJG, RAW, and NJKH conceptualised the research project. SJG performed the formal analysis. SJG prepared the manuscript with contributions from all co-authors.

Citation: Gnann, S. J., Woods, R. A. and Howden, N. J. K. (2019). Is there a baseflow Budyko curve? Water Resources Research, 55(4):2838–2855. https://doi.org/10.1029/2018WR024464

2.1 Introduction

Baseflow is defined as flow derived from groundwater and other delayed sources and thus sustains streamflow also during dry periods (Hall, 1968; Smakhtin, 2001). Understanding how baseflow varies with changing climate and landscape properties is crucial for various issues related to water quantity and quality (e.g. Smakhtin, 2001; Price, 2011; Beck et al., 2013; Buttle, 2018). Population growth is linked to an increase in freshwater demand for agriculture, industry and human consumption, and water shortages pose a threat even in humid regions (Price, 2011). Baseflow is essential for ecosystem functioning and provides habitat for stream biota (Poff et al., 1997; Price, 2011). Furthermore, baseflow is important with respect to water quality issues (chemistry, temperature) such as effluent-load from wastewater treatment plants (Smakhtin,

2001; Ficklin et al., 2016). If we want to understand how humans impact baseflow, we need to understand what determines baseflow under (near-)natural conditions.

Many studies found that baseflow is correlated with climate and landscape properties such as soils, geology, topography, and vegetation, but a universal relationship or general theory is yet to be found (Price, 2011). Geology was found to be the key variable in various regional studies (e.g. Neff et al., 2005; Longobardi and Villani, 2008; Bloomfield et al., 2009). Similarly, soil classes (which are correlated with geology) were used to explain baseflow variability in the UK (Boorman et al., 1995) and Europe (Schneider et al., 2007). Schneider et al. (2007) found that soils were less influential towards southern Europe, which might be attributed to differences in topography and climate. Van Dijk (2010) explored catchments in Australia and concluded that climate was the most important control on baseflow, while Lacey and Grayson (1998) found that for southeastern Australia vegetation-geology groups were the main influence. In summary, the studies that found landscape properties to be most influential were usually of regional nature and thus investigated catchments with relatively similar climates. Continental studies and the first global study by Beck et al. (2013) led to somewhat inconclusive results. While some key landscape and climate characteristics could be identified, the underlying processes remain to be explained. The influence of lakes (Neff et al., 2005) and snow (Beck et al., 2013), that is, baseflow generating mechanisms different than groundwater discharge, further complicates the analysis.

Baseflow is usually quantified by the baseflow index (BFI), the long-term ratio between baseflow and total streamflow. Alternatively, we can use the baseflow fraction K_B (Sivapalan et al., 2011), defined as the ratio between mean annual baseflow $\bar{Q}_b$ and precipitation $\bar{P}$ (cf. to the runoff ratio, the ratio between total streamflow $\bar{Q}$ and precipitation $\bar{P}$). K_B has the advantage that it relates baseflow to precipitation, a climate input that is (mostly) independent of catchment form. The similarity to the runoff ratio allows us to investigate K_B in the context of the Budyko hypothesis. A disadvantage of K_B is that we need both streamflow and rainfall data.

The Budyko hypothesis (Budyko, 1974) is a widely applied empirical top-down approach in catchment hydrology (Wang et al., 2016). It hypothesises that the ratio between mean annual actual evapotranspiration $\bar{E}_a$ and precipitation $\bar{P}$ is primarily a function of the ratio between mean annual potential evapotranspiration $\bar{E}_p$ and precipitation $\bar{P}$, that is, the aridity index $\varphi = \bar{E}_p/\bar{P}$. As $\bar{E}_a$ is usually not available, $\bar{Q}$ might be used instead (Andréassian and Perrin, 2012). Figure 2.1a shows a Budyko-type plot for catchments in the US and the UK (data sources will be explained in Section 2.2.2). The catchments fall relatively close to a single curve, the so called the Budyko curve, for which various model equations exist (see e.g. review by Wang et al., 2015). Is there a similar behaviour for baseflow, that is, a baseflow Budyko curve? That is, is the aridity index the primary control on baseflow fraction? Wang and Wu (2013) modelled the relationship between baseflow fraction and aridity by means of a Budyko-type curve that approaches unity for increasing humidity. Similarly, Sivapalan et al. (2011) reported "that the fraction of precipitation partitioned to slow flow is highest in wet catchments (as high as 0.7) and

decreases with increasing aridity". Both studies analysed MOPEX data (Duan et al., 2006), that is data from the contiguous US. Redoing this analysis with data from the US and the UK reveals a different behaviour. We can see from Figure 2.1b that the fraction of precipitation that becomes baseflow does not always increase with decreasing aridity index but decreases for many humid catchments.

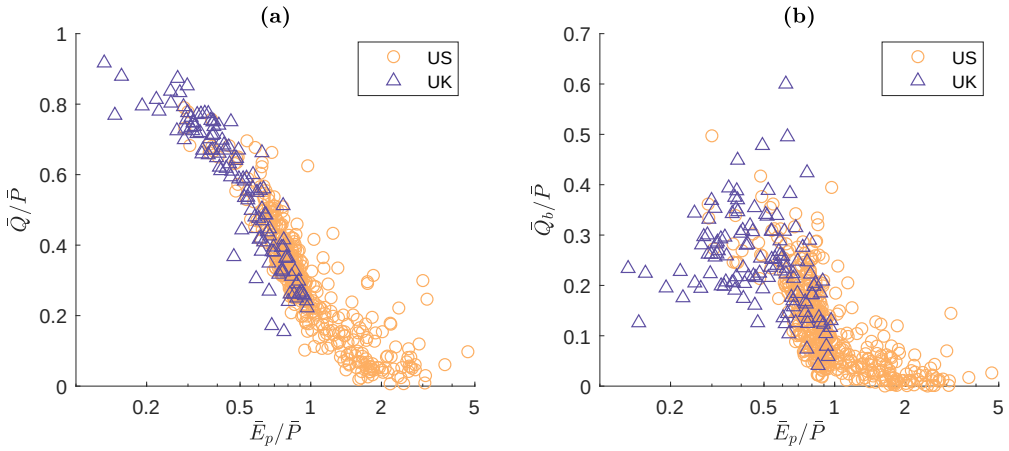

Figure 2.1: Budyko-type curves relating **(a)** mean annual runoff ratio $\bar{Q}/\bar{P}$ to mean aridity index $\bar{E}_p/\bar{P}$ and **(b)** mean annual baseflow fraction $\bar{Q}_b/\bar{P}$ to mean aridity index $\bar{E}_p/\bar{P}$. US catchments are denoted by orange circles, UK catchments are denoted by purple triangles. Catchments with significant snow fractions were removed.

The data presented in Figure 2.1 suggest that the influence of climate aridity on baseflow fraction is not straightforward or universal. This reinforces the variability in the literature on the relative importance of climate and landscape characteristics. Is there a way to quantify and/or parameterise these relative importances? Can we disentangle the influences of different causal factors such as forcing and catchment form? How can we model baseflow variability in a process-based way? As a framework for addressing these questions, we will use the revised Ponce-Shetty model (Ponce and Shetty, 1995a,b; Sivapalan et al., 2011) to model the catchment water balance at the annual scale. The Ponce-Shetty model has been described as a functional model (Sivapalan et al., 2011) as it focuses on how water is partitioned, stored, and released, that is, a catchment's functions (Black, 1997; Wagener et al., 2007). This approach is promising as it goes beyond mere empiricism by representing processes such as the partitioning of water at the annual scale. The processes and the respective parameters are arguably highly abstracted and connecting emergent parameters to catchment characteristics remains challenging (Sivapalan et al., 2011). This approach, however, allows us to investigate large samples of catchments and thus enables us to explore catchment (dis-)similarity and patterns which eventually might be synthesised to new catchment-scale theory (Sivapalan, 2005; McDonnell et al., 2007; Wagener et al., 2007; Harman and Troch, 2014). In the face of environmental change (Milly et al., 2008), process-based models that allow for extrapolation are more needed than ever (Wagener et al.,

2010).

We will use the revised Ponce-Shetty annual water balance model to obtain and investigate a theoretical model of baseflow fraction (and baseflow index) as a function of mean annual climate variables (Sivapalan et al., 2011). We will fit the Ponce-Shetty model to catchments in the US and the UK to obtain catchment-scale parameter values defining how water is partitioned at the annual scale (Ponce-Shetty parameters; described in Section 2.2). We will then assess whether this approach has the potential to explain the variability in baseflow fraction (and baseflow index) shown in Figure 2.1b and the apparently differing behaviour exhibited by the catchments in the UK.

2.2 Theory and data

2.2.1 Theory

2.2.1.1 Annual water balance model

The revised Ponce-Shetty model (Sivapalan et al., 2011) is a functional approach to water balance modelling following Horton (1933), L'vovich (1979) and Ponce and Shetty (1995a,b). A catchment's annual water balance is conceptualised as a two-stage partitioning process. First, precipitation P is partitioned into fast flow Q_f (direct runoff and fast subsurface flow) and wetting W (water that is being stored). The stored water is then further partitioned into vaporisation V (water returned to the atmosphere) and baseflow (slow flow) Q_b. Fast flow and baseflow combined yield total streamflow Q. Inter-annual water storage change and other water gains or losses such as inter-catchment groundwater flows are assumed to be negligible. Figure 2.2 shows a schematic of the model.

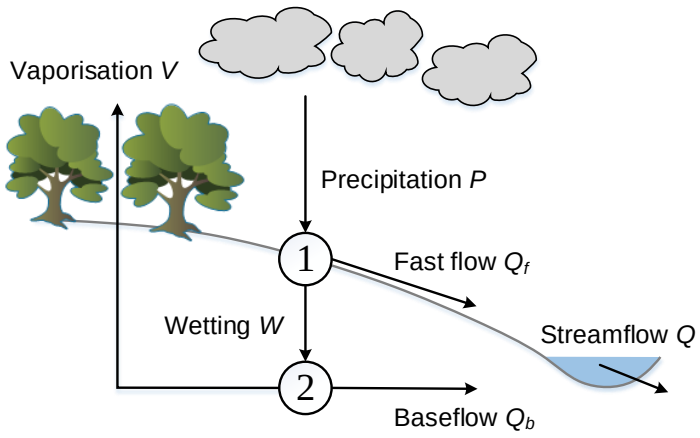

Figure 2.2: Schematic of the Ponce-Shetty model indicating the two partitioning stages (2.1) and (2.2).

The balance equations for the two partitioning stages are:

$$P = Q_f + W \tag{2.1}$$

$$W = Q_b + V \tag{2.2}$$

The balance equations for the whole catchment are:

$$P = V + Q \tag{2.3}$$

$$Q = Q_f + Q_b \tag{2.4}$$

These balance equations are used to determine V (from Eq. (2.4)) and W (from Eq. (2.1)). Data sources for Q and P and the estimation of Q_f and Q_b are described in the following subsections.

2.2.1.2 Baseflow estimation

To obtain an estimate of fast flow and baseflow we perform a hydrograph separation using digital filtering techniques. Following Troch et al. (2009) who reported that the choice of the filter has no significant influence on annual water balance metrics (they analysed the Horton index), many subsequent studies used only one hydrograph separation technique (e.g. Sivapalan et al., 2011; Harman et al., 2011). Since in the original Troch et al. (2009) paper only 33 catchments were analysed, we perform a comparative analysis of baseflow separation methods for all the catchments investigated here. We use the one-parameter Lyne-Hollick digital filter (Lyne and Hollick, 1979) which is applied forwards, backwards and forwards again using a filter parameter of 0.925. As an alternative, we test the UK Institute of Hydrology (UKIH) smoothed minima method (Institute of Hydrology, 1980). Both filters have the advantage of being only minimally parameterised (one parameter) and thus being easily applied to a large sample of catchments. Knowing P, Q (both measured), Q_f, Q_b (both estimated), we can then calculate V and W.

2.2.1.3 Ponce-Shetty equations

Based on empirical observations Ponce and Shetty (1995a) presented a mathematical model of the two-stage partitioning which was re-introduced by Sivapalan et al. (2011). The form of the equations follows the curve number runoff equation (NRCS, 2004), which is an empirical equation that satisfies conservation of mass. The idea of two competing processes (here: fast flow vs. wetting and baseflow vs. vaporisation) was later generalised by means of the so called proportionality hypothesis and the Maximum Entropy Production (MEP) principle was identified as a possible thermodynamic basis for this mathematical form (Wang and Tang, 2014; Wang et al., 2015; Zhao et al., 2016).

The first partitioning stage is modelled as follows:

$$Q_f = \begin{cases} 0, & \text{if } P \leq \lambda_P W_\mathrm{p} \\ \frac{(P-\lambda_P W_p)^2}{P+(1-2\lambda_P)W_p}, & \text{if } P > \lambda_P W_\mathrm{p} \end{cases} \tag{2.5}$$

$$W = \begin{cases} P, & \text{if } P \leq \lambda_P W_\mathrm{p} \\ P - \frac{(P-\lambda_P W_p)^2}{P+(1-2\lambda_P)W_p}, & \text{if } P > \lambda_P W_\mathrm{p} \end{cases} \tag{2.6}$$

$$P \to \infty,\ Q_f \to P - W_p,\ W \to W_p \tag{2.7}$$

where λ_P is the fast flow initial abstraction coefficient and W_p is the wetting potential. Their product $\lambda_P W_p$ is the fast flow generation threshold. This form is convenient as λ_P ranges between zero and unity (Ponce and Shetty, 1995a). The second partitioning stage is modelled as follows:

$$Q_b = \begin{cases} 0, & \text{if } W \leq \lambda_W V_\mathrm{p} \\ \frac{(W-\lambda_W V_p)^2}{W+(1-2\lambda_W)V_p}, & \text{if } W > \lambda_W V_\mathrm{p} \end{cases} \tag{2.8}$$

$$V = \begin{cases} W, & \text{if } W \leq \lambda_W V_\mathrm{p} \\ W - \frac{(W-\lambda_W V_p)^2}{W+(1-2\lambda_W)V_p}, & \text{if } W > \lambda_W V_\mathrm{p} \end{cases} \tag{2.9}$$

$$W \to \infty,\ Q_b \to W - V_p,\ V \to V_p \tag{2.10}$$

where λ_W is the baseflow initial abstraction coefficient and V_p is the vaporisation potential. Their product $\lambda_W V_p$ is the baseflow generation threshold.

Figure 2.3 shows curves derived from the Ponce-Shetty model equations. Both the P-W-plot (Figure 2.3a) and the W-V-plot (Figure 2.3c) start at the origin and approach a limit (their potentials). The wetting potential W_p can be seen as some sort of storage capacity of a catchment. The vaporisation potential V_p can be seen as some sort of energy limit (somewhat analogous to potential evapotranspiration). The P-Q_f-plot (Figure 2.3b) and the W-Q_b-plot (Figure 2.3d) start to rise after a certain threshold and then rise without a (theoretical) limit. The precipitation threshold is a minimum amount of rainfall required to generate fast slow. The baseflow threshold is a minimum amount of wetting required to generate baseflow. This reflects the idea that if there is only little rain (or wetting), the water will not reach the stream and evaporate (e.g. interception). The physical meaning of these parameters is somewhat ambiguous as they are emergent parameters representing processes over a large area (catchment) and over a long time (years). Links to physical (observable) catchment characteristics remain to be explored, but will be discussed qualitatively in Section 2.4.

2.2.1.4 Rescaled form of the Ponce-Shetty equations

In order to compare between catchments the (mean annual) Ponce-Shetty variables can be normalised using the Ponce-Shetty parameters (Sivapalan et al., 2011). We define two rescaled

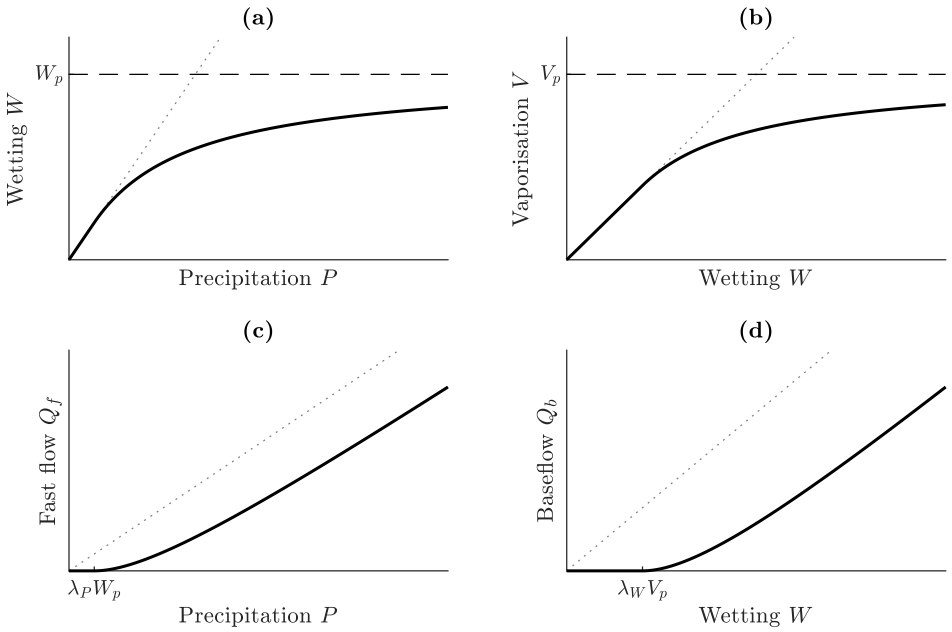

Figure 2.3: Example L'vovich-type curves: **(a)** precipitation-wetting curve (Eq. (2.6)), **(b)** wetting-vaporisation curve (Eq. (2.9)), **(c)** precipitation-fast flow curve (Eq. (2.5)), **(d)** wetting-baseflow curve (Eq. (2.8)). The dotted lines indicate the lines through the origin, which (in theory) cannot be exceeded. The dashed lines indicate the potentials. The ticks indicate the thresholds.

driving variables: rescaled (mean annual) precipitation $\tilde{P}$ and a rescaled vaporisation potential $\tilde{V}_p$.

$$\tilde{P} = \frac{\bar{P} - \lambda_P W_p}{(1-\lambda_P)W_p} \tag{2.11}$$

$$\tilde{V}_p = \frac{V_p - \lambda_W V_p}{(1-\lambda_P)W_p} \tag{2.12}$$

2.2.1.5 Catchment indices

We define two catchment indices: the baseflow fraction K_B (note that this definition is slightly different from the usual definition as it includes the parameter $\lambda_W V_p$) and the baseflow index BFI.

$$K_B = \frac{\bar{Q}_b}{\bar{P} - \lambda_W V_p} \tag{2.13}$$

$$\mathrm{BFI} = \frac{\bar{Q}_b}{\bar{Q}} \tag{2.14}$$

We can approximate these indices using the rescaled driving variables (Eq. (2.11) and Eq. (2.12)) (for the full derivation of K_B see Sivapalan et al., 2011, and for the derivation of BFI see

Appendix A):

$$K_B = \frac{\tilde{P}}{(1+\tilde{P})(\tilde{P}+\tilde{V}_p+\tilde{V}_p\tilde{P})} \tag{2.15}$$

$$\mathrm{BFI} = \frac{1}{(1+\tilde{P})(1+\tilde{V}_p)} \tag{2.16}$$

These expressions can be used to model the observed catchment indices (Eq. (2.13) and Eq. (2.14)). These equations are functions of two variables ($\tilde{V}_p$ and $\tilde{P}$) and not just a single variable such as aridity (which might be defined here as rescaled aridity index $\tilde{\varphi} = \frac{\tilde{V}_p}{\tilde{P}}$). Note that in the derivation of these equations we assume a parameter $K = \frac{\lambda_P W_p - \lambda_W V_p}{(1-\lambda_P)W_p}$ (not presented here for brevity) to be zero. This assumption led to insignificant differences which is consistent with Sivapalan et al. (2011).

2.2.2 Data

We use data from the contiguous US and Great Britain. CAMELS (Newman et al., 2015; Addor et al., 2017) includes daily precipitation, potential evapotranspiration and streamflow data, as well as a wide range of catchment attributes for 671 catchments in the contiguous US. The UK Benchmark Network (UKBN2) (Harrigan et al., 2017) describes catchments in the UK that are near-natural. It consists of 146 catchments whereof 8 catchments in Northern Ireland are not considered. The data is obtained from different sources. Daily streamflow data, catchment characteristics and catchment boundaries are obtained from the NRFA (National River Flow Archive, 2019), precipitation data from CEH-GEAR (Tanguy et al., 2016), and potential evapotranspiration data from CHESS-PE (Robinson et al., 2016). We trim the daily data to contain only full water years (starting 1 October). We then aggregate daily data to obtain annual data, which are used to calibrate the Ponce-Shetty model for each catchment. For all other calculations we use mean annual data, that is, data averaged over all full water years. To obtain a suitable dataset we remove some of the catchments according to the following criteria:

- Catchments with areas smaller than 10 km^2 as measurement errors and catchment delineation errors tend to be significant for very small catchments.

- Catchments with records shorter than 15 years as calibrating the Ponce-Shetty model requires many annual values. This threshold is chosen to remove some rather short and thus potentially unreliable records, while trying to keep enough catchments for the ongoing analysis.

- Catchments where snow and lakes are influential, as these processes are not considered in the Ponce-Shetty model. We remove catchments with fractions of precipitation falling as snow > 0.2 and catchments with significant surface water bodies. The latter is done by removing UKBN2 catchments with FARL < 0.8 (a parameter quantifying the influence of lakes and reservoirs) and CAMELS catchments with frac_water > 0.05.

- Catchments with runoff ratios larger than unity in any year of record ($Q/P > 1$), resulting in negative vaporisation values ($V < 0$), as this indicates significant water balance issues and thus violates the assumptions of the Ponce-Shetty model.

The final dataset consists of 571 out of 817 catchments.

2.3 Results

2.3.1 Baseflow estimation

Table 2.1 shows several metrics comparing results obtained using the Lyne-Hollick filter (Lyne and Hollick, 1979) and the UKIH method (Institute of Hydrology, 1980). The two methods show good agreement. While the choice of filter might have a significant impact on individual catchments, it does not alter the overall results. We continue using the baseflow estimates obtained by using the Lyne-Hollick filter.

Table 2.1: Comparison of mean annual baseflow $\bar{Q}_b$, Ponce-Shetty parameters, K_B and BFI using different baseflow separation techniques (Lyne-Hollick filter and UKIH method). The relative error (RE) is defined as RE = $\left|1 - \frac{x_a}{x_b}\right|$. The absolute error (AE) is defined as AE = $|x_a - x_b|$.

	$\bar{Q}_b$ [mm]	W_p [mm]	λ_P [-]	V_p [mm]	λ_W [-]	K_B [-]	BFI [-]
Pearson correlation	1.00	0.84	0.98	0.95	0.97	0.99	0.93
Spearman correlation	1.00	0.99	0.96	0.99	0.95	0.99	0.96
Median RE	0.07	0.05	0.17	0.05	0.31	0.07	0.07
Median AE	11	159	0.01	147	0.00	0.01	0.03

2.3.2 Parameter estimation and uncertainty

The Ponce-Shetty parameters are fitted to each individual catchment by means of a non-linear least squares fitting algorithm, whereby λ_P and λ_W are restricted to be between zero and unity (their theoretical limits), and W_p and V_p are restricted to be between 0 mm and an upper limit. We choose an (arbitrary) upper limit of 50000 mm which is deemed high enough to not affect the parameter estimation. An even higher limit does not affect the estimated parameter values except for very few catchments with W_p and/or V_p values which are (almost) at the limit. The problem that some of the obtained parameter values are at the upper limit is discussed in the next paragraph. We can use two values for the wetting W to fit the second partitioning stage. Either the observed W obtained from Eq. (2.1) or the modelled W following from the fitted model for the first partitioning stage (Eq. (2.6)). Following (Sivapalan et al., 2011) we use the modelled W to obtain an internally consistent water balance.

To fit a meaningful parameter set, the catchments should exhibit their functional behaviour (Sivapalan et al., 2011). If the vaporisation values (wetting values) are far away from the

vaporisation potential (wetting potential), we will have a roughly linear relationship and hence fitting the functional form is not possible (see Figure 2.4a). This can be seen especially for V_p in arid catchments (e.g. in the middle of the US). In these catchments, the obtained potentials are at the specified upper limit (50000 mm). Similarly, being at the potential all the time does not allow us to fit a functional relationship either; this can be seen especially for V_p in humid catchments (e.g. along the west coast of the UK). In these catchments the obtained initial abstraction coefficient is unity (see Figure 2.4b). We remove these catchments from the analysis because the Ponce-Shetty model is unable to describe them adequately.

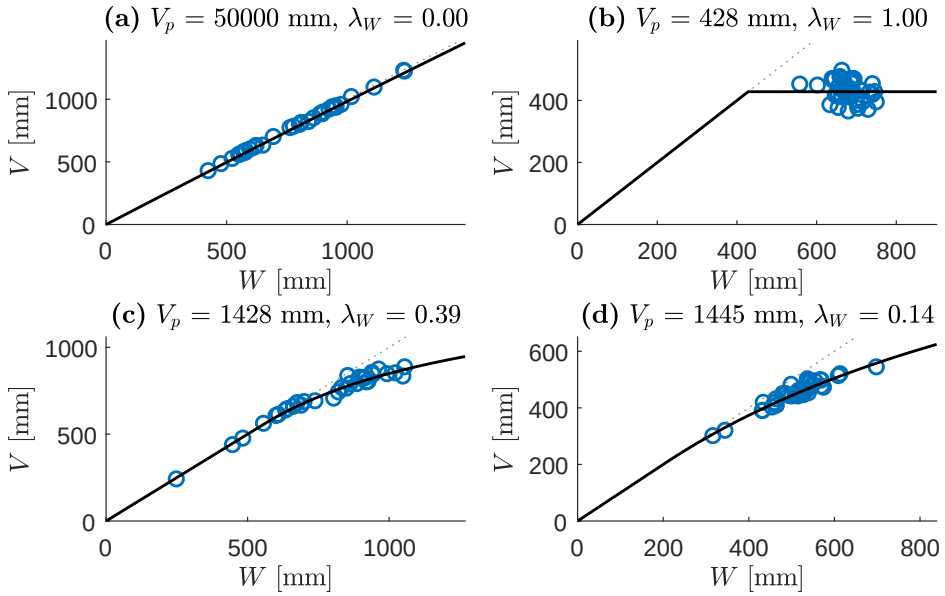

Figure 2.4: Examples of catchments (station numbers in brackets) with fitted W-V-curves. **(a)** Coleto Creek, Texas (08176900): extremely high V_p, V_p not identifiable. **(b)** Aire, Yorkshire (27035): V always approximately equal to V_p, λ_W not identifiable. **(c)** Bear Creek, Texas (08158810): V_p and λ_W identifiable. **(d)** Pincey Brook, Essex (38026): V_p and λ_W identifiable.

Table 2.2 shows overall statistics for the parameter estimation after having removed the catchments described in the last paragraph. The parameter uncertainty (in the form of 95% confidence intervals) is particularly high for extremely large values for either of the potentials ($\gg$ 10000 mm). These large values are consistently uncertain, which coincides with Sivapalan et al. (2011) who found that for some catchments the (apparently very high) potentials could not be properly identified. The confidence intervals for λ_P and λ_W need careful interpretation, as these two parameters have heavily skewed distributions (most catchments have parameter values close to zero). We do not remove catchments with high uncertainty from the analysis as a threshold would necessarily be subjective, which leaves us with 545 catchments for the ongoing analysis.

Table 2.2: Parameter statistics and uncertainty for all catchments used in the analysis. *CI 95%* denotes the 95% confidence interval. *Rel. CI 95%* denotes the relative confidence limits, that is, the confidence limits normalised by the parameter values. *Spearman* denotes the Spearman correlation of the relative confidence limits with the parameter values.

	Min	Median	Max	Median CI 95%	Median Rel. CI 95%	Spearman
W_p [mm]	756	3044	42857	1591	0.50	0.32
λ_P [-]	0	0.05	0.64	0.12	>1	-0.91
V_p [mm]	316	2911	44652	2264	0.74	0.49
λ_W [-]	0	0.02	0.91	0.13	>1	-0.91

2.3.3 Maps of Ponce-Shetty parameters and baseflow metrics

Figure 2.5 shows maps of the fitted parameters for CAMELS catchments. The patterns agree well with Sivapalan et al. (2011) who used MOPEX catchments. High wetting potentials W_p can be seen in the middle of the US (Great Plains), in the east (southern parts of the Appalachians), south east (around Florida) and in parts of the central north (Michigan). High vaporisation potentials V_p can be seen in the middle of the US (Great Plains) and in all southern regions. The fast flow thresholds $W_p\lambda_P$ are high in the south, the south east and in the middle of the US except for the north. The baseflow thresholds $V_p\lambda_W$ are similarly high in most of these areas, but also in some catchments along the west coast. The spatial similarity of the thresholds is reflected by a significant rank correlation of 0.61 between $W_p\lambda_P$ and $V_p\lambda_W$.

Figure 2.6 shows maps of the fitted parameters for UKBN2 catchments. On average, the values are lower than for the CAMELS catchments, especially for V_p, which is consistent with generally lower vaporisation intensities (cf. to E_p). High wetting potentials W_p can be found in the south west, the south, the middle (the Midlands) and along the south eastern coast. The vaporisation potentials V_p are high in the south, especially in the south east. High $W_p\lambda_P$ can be found in the south east and for a few catchments in the north. High $V_p\lambda_W$ can be found in catchments scattered throughout the UK, most notably all along the west coast and in the south east.

Figure 2.7 shows maps of K_B and BFI for CAMELS and UKBN2 catchments. Generally, K_B is lower than BFI as it compares $\bar{Q}_b$ to $\bar{P}$ rather than $\bar{Q}$, which is always lower than $\bar{P}$. This is reflected in the ranges of values shown in Figure 2.7. While in some regions both K_B and BFI are rather high (e.g. in the eastern US and in the south west of the UK), in other regions BFI can be high while K_B is rather low (e.g. in the southern US and in the middle of the US and in the south east of the UK), which broadly agrees with Santhi et al. (2008) who found that catchments with high BFI can still have low baseflow volumes. This coincides with the maps showing the Ponce-Shetty parameters (Figures 2.5 and 2.6). Catchments with high K_B generally have a high W_p, low $W_p\lambda_P$ and low V_p. Catchments with high BFI also occur in areas with high V_p.

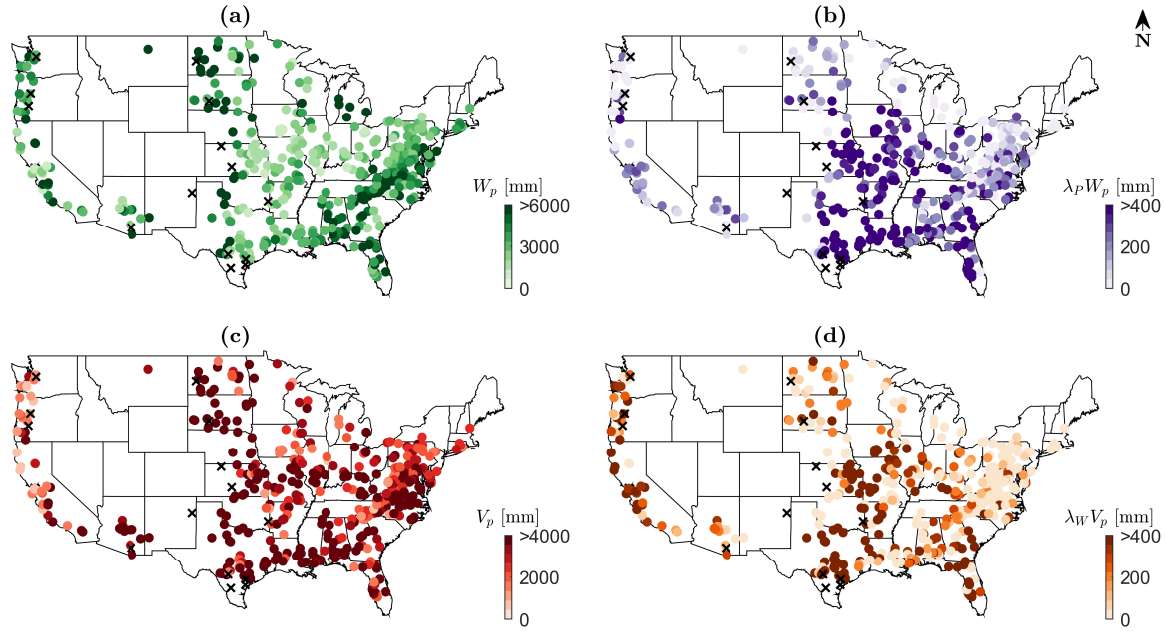

Figure 2.5: The fitted parameters for CAMELS catchments: wetting potential **(a)**, fast flow threshold **(b)**, vaporisation potential **(c)**, and baseflow threshold **(d)**. Crosses denote catchments where some of the parameters could not be identified properly.

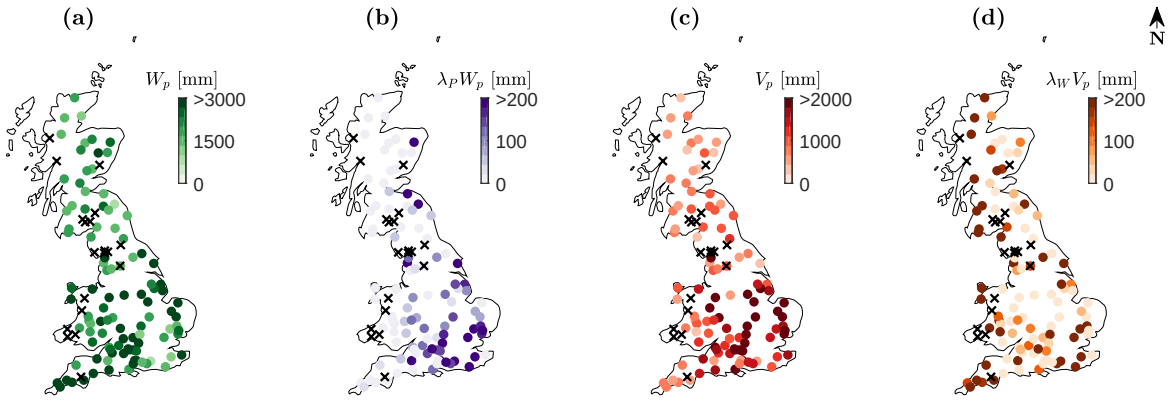

Figure 2.6: The fitted parameters for UKBN2 catchments: wetting potential **(a)**, fast flow threshold **(b)**, vaporisation potential **(c)**, and baseflow threshold **(d)**. Crosses denote catchments where some of the parameters could not be identified properly.

2.3.4 Baseflow variability with climate variables

Figure 2.8 shows how the baseflow fraction varies with the rescaled climate variables. To show the dependence of K_B on both $\tilde{P}$ and $\tilde{V}_p$ we make use of a contour plot (see Figure 2.8a). We plot $\tilde{P}$ and $\tilde{V}_p$ on the x- and y- axes, respectively, and use contours to represent the model for K_B (Eq. (2.15)) and coloured dots to represent the observed K_B values (Eq. (2.13)). Figure 2.8b,

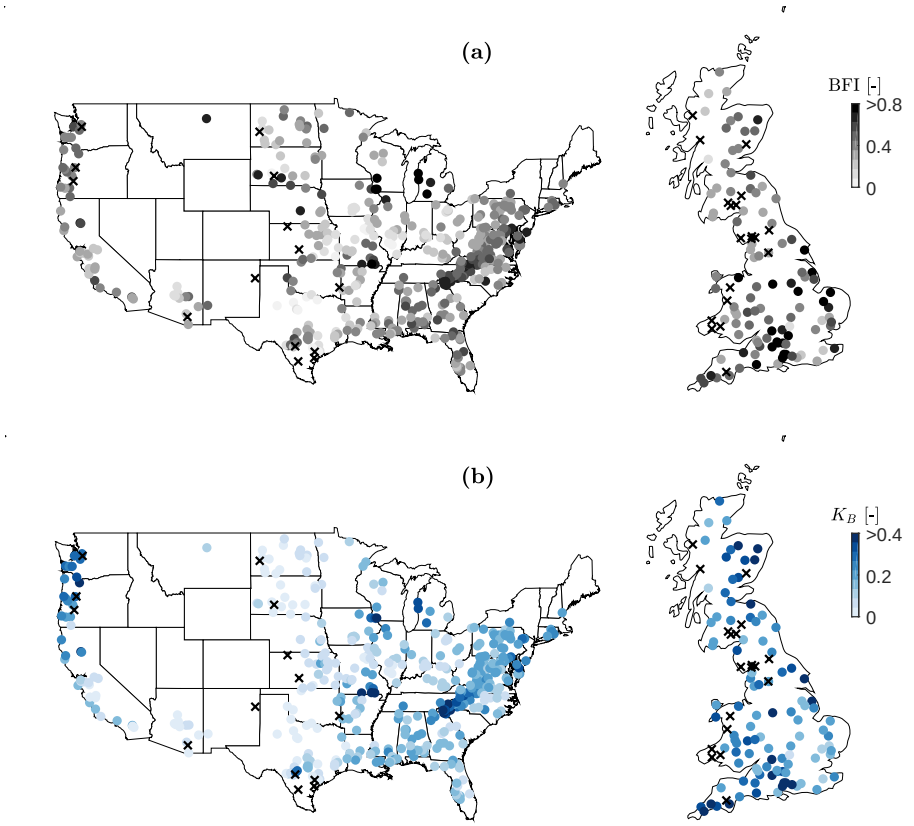

Figure 2.7: K_B **(a)** and BFI **(b)** for CAMELS and UKBN2 catchments. Note that the colour scales are different to reflect the range of the values. Crosses denote catchments where some of the parameters could not be identified properly. Note that the maps of the US and the UK are not to the same scale.

shows an equivalent plot using the ratio between $\tilde{P}$ and $\tilde{V}_p$ (rescaled aridity index $\tilde{\varphi}$) with some example model curves with either fixed $\tilde{P}$ or $\tilde{V}_p$, respectively – this is comparable to common Budyko-type plots. To get a better understanding it is useful to recall how a contour plot of the rescaled aridity index would look like, which is shown in Figure 2.8c. The line through the origin represents a rescaled aridity index of unity, above that line (top left) are humid catchments, below that line (bottom right) are arid catchments. Note that we are using rescaled variables and hence we are not looking at the common aridity index. $\tilde{P}$ is a relative rainfall amount and $\tilde{V}_p$ is a relative vaporisation potential, both rescaled by their thresholds and the wetting potential of the catchment. The general notion that low $\tilde{\varphi}$ indicates humid (energy-limited) catchments and that high $\tilde{\varphi}$ indicates arid (water-limited) catchments is still valid.

The contours in Figure 2.8a start parallel to the line through the origin and thus parallel to the rescaled aridity index. They start to bend for higher values of $\tilde{P}$ (humid side of the plot) and become perpendicular to the rescaled aridity index. This demonstrates that a catchment having a certain rescaled aridity index can have very different values of K_B. Roughly, if both $\tilde{P}$ and $\tilde{V}_p$

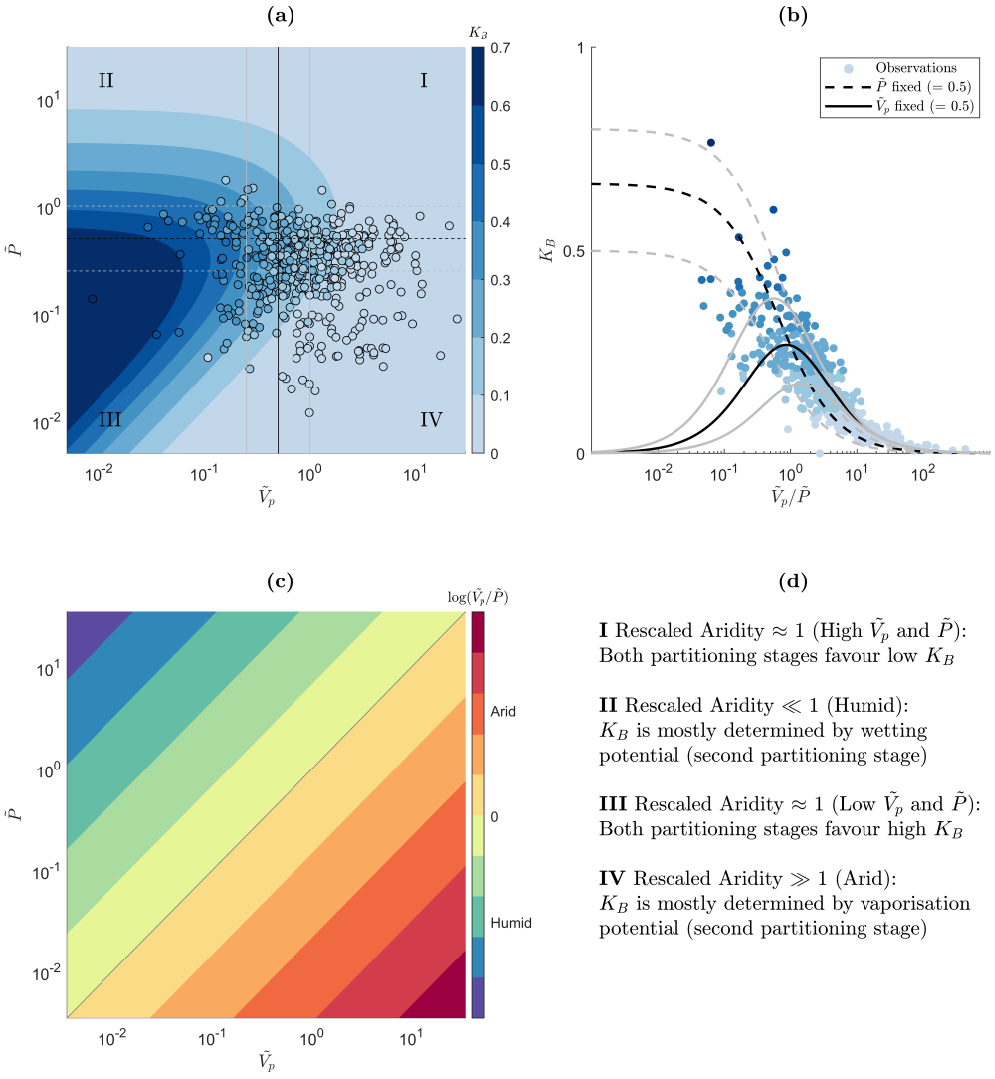

Figure 2.8: **(a)** Contour plot of K_B as a function of the rescaled vaporisation potential $\tilde{V}_p$ and rescaled precipitation $\tilde{P}$ (Eq. (2.15)). The dots indicate the observed values (Eq. (2.13)). **(b)** K_B as function of the ratio between $\tilde{V}_p$ and $\tilde{P}$ (i.e. rescaled aridity index $\tilde{\varphi}$). The black and grey lines (solid and dashed) are example model curves with either fixed $\tilde{V}_p$ or $\tilde{P}$. The dots indicate the observed values. **(c)** Logarithm of the rescaled aridity index $\tilde{\varphi}$ as a function of $\tilde{V}_p$ and $\tilde{P}$. The grey line denotes a rescaled aridity index of unity (log equals zero). **(d)** Different regions of the K_B contour plot are annotated, a more detailed explanation is given in Section 2.4.

are low, we get a rather high K_B and if both are high, we get a rather low K_B. The contours are not just bending on the humid side (top left), they are also indicating higher gradients and thus a high variability in K_B. In contrast, there is relatively little variation on the arid side (bottom right), that is, most of the catchments have a similar K_B. The observed values (represented by coloured dots) agree well with the model contours (median absolute error = 0.02, median relative error = 0.14). This can be expected, since the model has sufficient degrees of freedom to fit the data well (the Ponce-Shetty model is fitted to each individual catchment). The Budyko-type plot shown in Figure 2.8b reflects these observations with a tight ensemble of curves for arid catchments and a spread out ensemble of curves for humid catchments. The observed values agree with this general behaviour, they are tight for arid catchments and scattered for humid catchments.

Figure 2.9 shows how BFI varies with $\tilde{P}$ and $\tilde{V}_p$. The contours shown in Figure 2.9a are symmetric around the line through the origin. The BFI is highest for low values of both $\tilde{P}$ and $\tilde{V}_p$ and gets lower for both higher $\tilde{P}$ and $\tilde{V}_p$. The observed values agree well with the model contours (median absolute error = 0.05, median relative error = 0.14). Again, this can be expected, since the model has sufficient degrees of freedom to fit the data well. Figure 2.9b shows that there is no clear relationship between BFI and the rescaled aridity index. This is in agreement with the observed values, which are scattered over most areas of the plot.

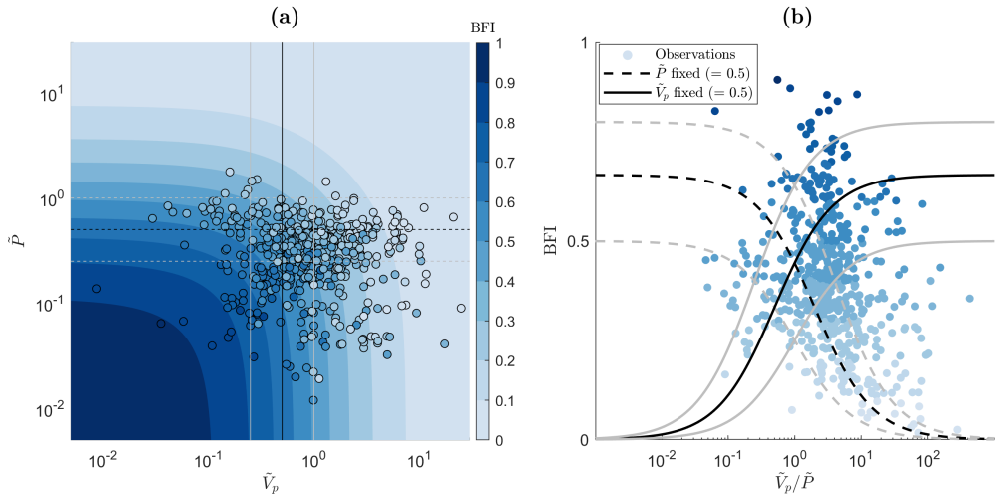

Figure 2.9: **(a)** Contour plot of BFI as a function of the rescaled vaporisation potential $\tilde{V}_p$ and rescaled precipitation $\tilde{P}$ (Eq. (2.16)). The dots indicate the observed values (Eq. (2.14)). **(b)** BFI as function of the ratio between $\tilde{V}_p$ and $\tilde{P}$ (i.e. rescaled aridity index $\tilde{\varphi}$). The black and grey lines (solid and dashed) are example model curves with either fixed $\tilde{V}_p$ or $\tilde{P}$. The dots indicate the observed values.

2.4 Discussion

The ranges of the parameter values (see Table 2.2) are in general agreement with Sivapalan et al. (2011) who also used a non-linear least squares method, and Harman et al. (2011) who used a Bayesian framework. The high parameter uncertainty for some catchments and problems in parameter identifiability might have two reasons. As described before, it could simply be a consequence of not having sufficient data to meaningfully fit the Ponce-Shetty model. It could, however, also indicate that the Ponce-Shetty model is not adequate for certain catchments. Even a good fit does not necessarily mean that the model is correctly representing the processes, which are arguably very simplified. We assume inter-annual water storage change as well as other water gains and losses to be negligible. This might not be a valid assumption for every catchment investigated here, and hence adds uncertainty to the parameter estimation. To assess the influence of inter-annual water storage change we alternatively calculated 3-year averages and calibrated the Ponce-Shetty model to these. This leads to overall similar parameter values (Pearson correlations: W_p: 0.86, $\lambda_P W_p$: 0.81, V_p: 0.79, $\lambda_W V_p$: 0.67). There are, however, problems associated with averaging. Extreme years, which are especially important to fit the Ponce-Shetty model, are averaged out and thus information is lost. Furthermore, by averaging and fitting a non-linear function, we introduce some bias ("the average of the function will not be the function of the average inputs", see Rouholahnejad Freund and Kirchner, 2017). This makes it difficult to tell whether inter-annual water storage change is the cause for the deviations in the parameter values. For now we argue that the model fits our data sufficiently well for the purpose of this work. Being capable of explaining the observed variations in baseflow further corroborates the model's suitability. For specific places, however, the uncertainty might be very large and conclusions or predictions should therefore be made with care. It would be interesting to see whether more detailed modelling approaches would lead to the emergent behaviour inherent in the Ponce-Shetty theory and/or similar parameter values.

From Figure 2.8 we can see how K_B varies with $\tilde{P}$ and $\tilde{V}_p$. Generally, K_B cannot be described by a single Budyko-type curve, but by a continuum of curves that depend on the catchment's (Ponce-Shetty) parameters. K_B is consistently low for high rescaled aridity values, which can be attributed to relatively high amounts of vaporisation (K_B is dominated by the second partitioning stage, i.e. V_p). The behaviour of K_B is more complicated for humid catchments. Starting at the origin of Figure 2.8a and moving along the y-axis towards more humid catchments, K_B first increases, then reaches a peak and decreases again. This decrease can be attributed to an exhausted wetting potential leading to "saturation excess fast flow" (K_B is dominated by the first partitioning stage, i.e. W_p). This was already recognised by Milly (1994) who stated that finite water storage capacity and finite permeability are possible causes for runoff. In such humid catchments, an increase in precipitation thus mainly leads to an increase in fast flow, which agrees with Harman et al. (2011) who found that fast flow elasticities are clearly larger than baseflow elasticities in humid catchments. Similarly, Trancoso et al. (2017) found that "higher

precipitation in tropical regions may be generating more overland flow, which tends to reduce the slow component [...]". Baseflow fraction can hence be low for both arid and humid catchments, but for different reasons. This may help to explain the diversity of results from empirical studies on controls on baseflow.

Figure 2.9 shows how the BFI varies with $\tilde{P}$ and $\tilde{V}_p$. The magnitude of $\tilde{P}$ and $\tilde{V}_p$ rather than the ratio between them determines the BFI. If both $\tilde{P}$ and $\tilde{V}_p$ are low, BFI is high. That means that at the first partitioning stage precipitation becomes mainly wetting, and at the second partitioning stage this wetting becomes mainly baseflow. If either $\tilde{P}$ and $\tilde{V}_p$ are high, we obtain a lower BFI. In the first case, most of the precipitation becomes fast flow and thus the BFI is low. In the second case, most of the precipitation becomes wetting, but most of that wetting evaporates, so that Q_b and thus the BFI will be rather low. In comparison to K_B, BFI is highly variable also for high rescaled aridity. Low amounts of baseflow (compared to precipitation) can lead to a high BFI if the amount of fast flow is even lower. This explains most of the differences between K_B and BFI (see Figures 2.5 and 2.6 and the description in Section 2.3.3).

The results show that K_B (and BFI) is influenced by the magnitude of $\tilde{P}$ and $\tilde{V}_p$ and not just their ratio. This explains the scatter especially for humid catchments (see Figure 2.8b). While an aridity index is certainly useful, it can be restrictive in cases where the magnitude of precipitation is important. This agrees for example with Berghuijs et al. (2017) who found that runoff is most sensitive to changes in precipitation and this sensitivity is not captured by only looking at the aridity index. Similarly, the ratio between precipitation and the wetting potential ($\approx \tilde{P}$) explains most of the variability in baseflow fraction which the aridity index could not explain (see Figure 2.8a, especially region II, and Figure 2.10).

Especially in humid catchments, the ratio of precipitation to a catchment's wetting potential can be a major control on baseflow. Given the same climate, a catchment with a higher wetting potential will have a higher baseflow fraction and BFI. This is a possible explanation for the partly inconclusive results found in studies before. Regional studies with similar climate could relate the amount of baseflow to a catchment's form, mostly soils (Boorman et al., 1995) and geology (Neff et al., 2005; Longobardi and Villani, 2008; Bloomfield et al., 2009). These attributes are parametrised by the Ponce-Shetty parameters (especially W_p), yet in a rather abstract way which so far eludes a quantitative linking to landscape characteristics. Continental (Schneider et al., 2007; Van Dijk, 2010; Trancoso et al., 2017) and global studies (Beck et al., 2013, 2015) found catchment form to be less influential and often couldn't come to conclusive results, as it is neither climate nor form alone that lead to a certain catchment response, but their interaction.

Figure 2.10 shows the $\bar{Q}_b/\bar{P}$ vs. $\bar{E}_p/\bar{P}$ plot (from Figure 2.1) with catchments stratified and coloured according to their wetting and vaporisation potentials, respectively. Three different ranges of W_p are shown and they form three somewhat distinct point clouds. The remaining variation can be attributed to differences in the thresholds, the rather broadly defined categories and differences in the magnitude of $\bar{E}_p$ and $\bar{P}$. The cloud with the lowest W_p exhibits the lowest

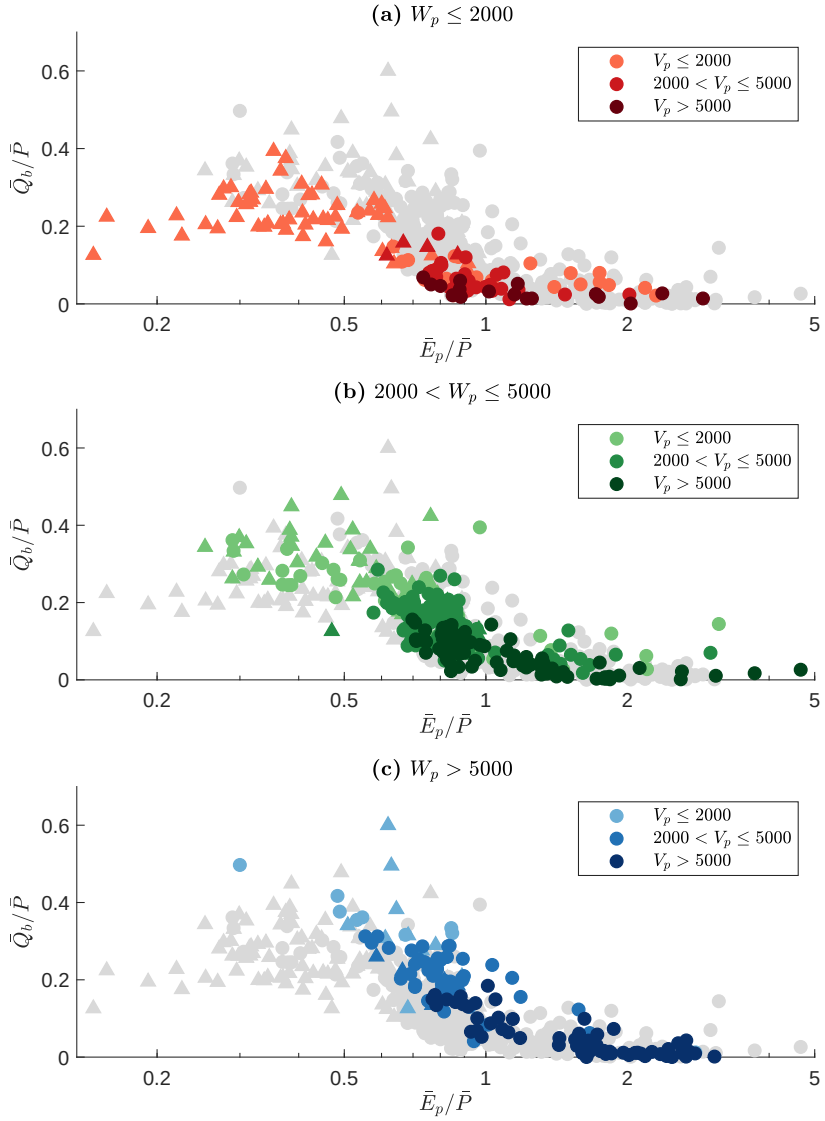

Figure 2.10: Scatter plots of mean annual baseflow fraction $\bar{Q}_b/\bar{P}$ vs. mean aridity index $\bar{E}_p/\bar{P}$. CAMELS catchments are denoted by circles, UKBN2 catchments are denoted by triangles. Catchments are highlighted according to their wetting potential W_p: **(a)** low wetting potentials, **(b)** medium wetting potentials, and **(c)** high wetting potentials. Darker shading indicates higher vaporisation potential V_p. All units are in mm.

baseflow fraction and vice versa. High values of K_B are usually associated with low values of V_p (indicated by the lightness of the colours). We can also see that CAMELS and UKBN2 catchments do not generally behave differently, but since certain catchment types occur predominantly in the US or the UK, the CAMELS and UKBN2 point clouds appear to be different. Very humid catchments with rather low W_p are mostly located in the UK and they are most clearly deviating from the point cloud representing CAMELS catchments (see also Figure 2.1).

We did not include catchments with significant snow fraction or lakes. While these catchments might be seen as having an "extended" wetting potential (storage), they represent conceptually different processes, for which additional explanatory variables might be needed. These processes might be added as an additional partitioning stage to the model to make it more universal. Especially the snowy catchments show an increase in K_B for increasing humidity almost up to unity (not shown here), which could explain, for example, why Wang and Wu (2013) used a baseflow Budyko model that approaches unity. Snowy catchments might be considered to have virtually unlimited storage potential as the snowpack can grow continuously, and thus baseflow fractions in these catchments can get very high.

The Ponce-Shetty parameters are emergent, rather abstract properties and relating them to catchment characteristics might not be straightforward. The Ponce-Shetty parameters are lumping a variety of processes and characteristics, notably soils, geology, vegetation, topography, and climate seasonality. This means that for now, the presented model can only explain and predict annual baseflow variability in gauged catchments where the model was calibrated. It might be used to investigate the effects of a changing climate (e.g. changing precipitation) on baseflow in different types of (gauged) catchments (cf. Buttle, 2018). A transfer to ungauged catchments requires a regionalisation procedure. Qualitatively, links between parameters and catchment characteristics can be seen. V_p is correlated with energy availability (comparable to potential evapotranspiration), yet it rather emerges from the interaction of the available energy with vegetation and other catchment characteristics. Large wetting potentials can be seen in moorland and wetland areas (e.g. south west UK, Florida) and in the presence of major aquifers (e.g. Chalk in southern England, Great Plains aquifer). A quantitative linking of the Ponce-Shetty parameters to landscape properties or other regionalisation approaches are, however, beyond the scope of this work.

2.5 Conclusions

The present work shows that there is no single baseflow Budyko curve, that is, in general baseflow fraction cannot be modelled as a function of an aridity index alone. Even if samples of catchments seem to form a single curve, this might be misleading as many of them might actually sit on different curves (see Figure 2.9b). The influence of catchment water storage on long-term water balance has long been recognised (e.g. Milly, 1994). The approach employed

here incorporates that in a simple way by modelling baseflow fraction as a function of two variables. A rescaled precipitation, that is the ratio between precipitation and a catchment's wetting potential, and a rescaled vaporisation potential. These two variables reflect the two-stage partitioning underlying the Ponce-Shetty model, namely the partitioning between fast flow and wetting, and the subsequent partitioning between slow flow and vaporisation. Depending on the climatic regime, one of these partitioning stages dominates. In arid catchments, baseflow fraction is mainly limited by high amounts of vaporisation. In humid catchments, baseflow fraction is mainly limited by the storage capacity of a catchment.

The differences between CAMELS (US) and UKBN2 (UK) catchments shown in Figure 2.1b and Figure 2.10 have two main causes. Firstly, using aridity as a ratio is restrictive. Catchments with a similar aridity index usually have lower precipitation and vaporisation intensities in the UK than in the US. Secondly, the wetting potentials in the UK differ from the ones in the US. Most of the very humid catchments in the UK have rather low wetting potentials, that is, they are (almost) fully saturated and a large fraction of precipitation runs off quickly to the stream. This difference is, however, not a clear distinction as it can be seen from Figure 2.10. Catchments in the US and the UK do not behave fundamentally differently, they rather happen to have predominantly different characteristics.

Baseflow (a catchment function) can be seen as the result of climate interacting with landscape (forcing acting on form, cf. Wagener et al., 2007). To explain baseflow variability in a process-based way, we should try to disentangle forcing and form, knowing that this might only be partially possible as catchment form (and function) may reflect a co-evolution with climate forcing. The Ponce-Shetty approach partly disentangles forcing and form, yet in a rather abstract way. Furthermore, the parameters still lump together a variety of processes that are not only reflecting catchment form (e.g. topography, geology, vegetation, etc.), but also climate (e.g. seasonality, storminess). Intra-annual climate variability can have a significant impact on such lumped parameters (Roderick and Farquhar, 2011; Berghuijs and Woods, 2016).

Using large samples of catchments allows us to detect and explain (dis-)similarities and patterns and to synthesise already available data (Falkenmark and Chapman, 1989; Sivapalan, 2005; Harman and Troch, 2014). While large sample hydrology arguably neglects many details, synthesising data to find new theory has proven to be a fruitful approach that – besides improved understanding – might help to constrain models (Shafii et al., 2017), to transfer knowledge to ungauged catchments (Hrachowitz et al., 2013) and to deal with predictions under change (Wagener et al., 2010; Ehret et al., 2014). It is essential to include a variety of catchments, both in terms of climate and landscape characteristics, which is exemplified by the "unexpected behaviour" of UK catchments in this work. Even more data are needed to corroborate the theory, to understand more of the details (e.g. Ponce-Shetty parameters) or to detect limitations of the presented approach, which eventually advances our understanding.

Simple approaches such as the Ponce-Shetty model are useful as they are easily applied to

large samples. They also allow us to better understand the model's dynamics and stop us from being lost in the calibration stage. We acknowledge that there is a danger in being too simple or simple due to lack of understanding (cf. Schwartz et al., 2017), which might partly be true for the hydrograph separation approach and the Ponce-Shetty model here. We are confident, however, that the chosen methods are appropriate for the present work as they are capable of explaining the observed phenomena and thus help to improve our understanding of how baseflow varies with climate and landscape.

CHAPTER 3

HYDROLOGICAL SIGNATURES DESCRIBING THE TRANSLATION OF CLIMATE SEASONALITY INTO STREAMFLOW SEASONALITY

This chapter has been published as a research article in Hydrology and Earth System Sciences. Slight modifications have been made to better fit the general layout of this thesis. We thank Wouter Knoben for help with MARRMoT, helpful discussions, and helpful comments on an earlier version of this chapter. We also thank Gemma Coxon for assisting with the data. We acknowledge the comments from the associate editor and three anonymous reviewers, which helped to clarify and improve this chapter. Parts of this work were carried out using the computational facilities of the Advanced Computing Research Centre of the University of Bristol (`http://www.bris.ac.uk/acrc/`).

SJG, NJKH, and RAW conceptualised the research project. SJG performed the formal analysis. SJG prepared the manuscript with contributions from all co-authors.

Citation: Gnann, S. J., Howden, N. J. K. and Woods, R.A. (2020). Hydrological signatures describing the translation of climate seasonality into streamflow seasonality. Hydrology and Earth System Sciences, 24(2):561–580. `https://doi.org/10.5194/hess-24-561-2020`

3.1 Introduction

The annual course of the Earth around the Sun leads to seasonal cycles in climate in many places. Seasonal patterns in precipitation, evapotranspiration, and snowfall, as well as the characteristics of the catchment a stream drains, often result in a distinct seasonal streamflow regime (Cayan et al., 1993; Regonda et al., 2005; Berghuijs et al., 2014). The seasonal flow regime is closely linked to water chemistry and water quality (DeWalle et al., 1997; Vega et al., 1998).

Streamflow seasonality plays a crucial role for biological systems and ecosystems (Colwell, 1974; Poff et al., 1997; Poff and Zimmerman, 2010). Low flows are typically seasonal, and droughts – albeit a more general phenomenon than low flows – often occur during the low flow season and thus are to some degree predictable (Smakhtin, 2001; Peters et al., 2003). From a more applied point of view, the seasonal streamflow regime is crucial for water resources management, agriculture, and hydropower generation (Weingartner et al., 2013; Laaha et al., 2013; Svensson, 2016; Harrigan et al., 2018). This is reflected in the increased application and development of seasonal forecasting methods (Shi et al., 2008; Svensson, 2016; Harrigan et al., 2018). In summary, for many applications the mean seasonal regime is of high importance and thus deserves attention.

In this work we focus on the average seasonal hydrological response of snow-free catchments. We do not focus, for instance, on the seasonality of events (e.g. storms), noting, however, that the seasonal water balance can have an impact at event scales (Berghuijs et al., 2014). In snow-free areas, the seasonality of the flow regime is primarily driven by the incoming forcing, that is, the seasonality of precipitation (water) and potential evapotranspiration (energy). Given a certain forcing, the flow regime of a catchment is determined by a catchment's form and function, that is, by how much water can infiltrate, how much water can be stored, and how slowly that water is being released. Since groundwater recharge and thus groundwater discharge are often very seasonal (Jasechko et al., 2014), many hydrogeological studies focus on seasonality, or more specifically on how seasonal recharge is propagated through an aquifer system (Townley, 1995; Erskine and Papaioannou, 1997; Peters et al., 2003; Obergfell et al., 2019). Slowly responding, groundwater-dominated catchments closely resemble the aquifer system feeding the stream. Understanding the seasonal streamflow regime is therefore particularly important for understanding slow (groundwater-driven) dynamics in catchments.

Different aspects of hydrological behaviour, such as streamflow seasonality, can be quantified by summarising metrics now mostly called hydrological signatures (McMillan et al., 2017). The use of such summarising metrics is not new, and they have been used extensively in ecohydrological studies (e.g. Clausen and Biggs, 2000; Olden and Poff, 2003) and hydrological studies (e.g. Jothityangkoon et al., 2001; Farmer et al., 2003). Hydrological signatures offer a way to quantify hydrological similarity. This makes them useful for catchment classification (Wagener et al., 2007; Sawicz et al., 2011), for hydrological process exploration (McMillan et al., 2014), and for predictions in ungauged basins (Yadav et al., 2007; Hrachowitz et al., 2013; Westerberg et al., 2016). Hydrological signatures can also be used to guide diagnostic model evaluation (Gupta et al., 2008; Peel and Blöschl, 2011; Euser et al., 2013; Hrachowitz et al., 2014; Shafii and Tolson, 2015), as they offer a potentially more meaningful and fit-for-purpose alternative to the typically used statistical metrics such as the Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970) or the Kling-Gupta efficiency (KGE; Gupta et al., 2009).

There are many hydrological signatures and we therefore need guidelines for signature selec-

tion (McMillan et al., 2017; Addor et al., 2018). Some of these guidelines refer to more technical aspects: the uncertainty in a signature should not be larger within a catchment than between catchments (identifiability), a signature should be insensitive to the data sources (robustness), and a signature should be comparable across (heterogeneous) catchments (consistency; McMillan et al., 2017). When using combinations of signatures, the different signatures should also contain different information, that is, they should not be redundant (Olden and Poff, 2003; Addor et al., 2018). From a more hydrological perspective, a signature should be meaningful at the relevant scale (representativeness) and a signature should relate to and increase our knowledge of hydrological function (discriminatory power; McMillan et al., 2017). Since (hydro-)climatic signatures such as the mean flow are already well understood, we should try to explain and use signatures that tell us more about catchment functioning (Addor et al., 2018), such as signatures that relate climate input to catchment output.

There is a multitude of hydrological signatures focusing on seasonality. Climate seasonality is accounted for by (hydro-)climatic signatures such as the (co-)seasonality of precipitation and potential evapotranspiration (Milly, 1994; Knoben et al., 2018). Streamflow seasonality can be characterised by the Pardé coefficients (Weingartner et al., 2013) or the regime curve, which is related to the slow flow component of the flow duration curve (FDC; Yokoo and Sivapalan, 2011). Seasonal signatures related to streamflow timing are the half flow date and the half flow interval (Court, 1962), and the date of each annual one-day maximum (or minimum; Richter et al., 1996). Seasonal streamflow signatures focusing on low flows are for example the seasonality index, which measures the mean day of low flow occurrence and the intensity of seasonality, or the seasonality histogram, which shows the occurrence of low flows in each month (Laaha and Blöschl, 2006). Colwell's predictability is another measure describing periodic signals (Colwell, 1974), mostly used in ecological studies. It consists of constancy (how variable the intra-annual flow regime is) and contingency (how persistent the inter-annual flow regime is). All of these signatures describe (parts of) the seasonality of either climate or streamflow, yet none of them look at how climate seasonality translates into streamflow seasonality. As the transformation of climate input into streamflow is, after all, what we are trying to understand, investigating the seasonal aspect of that seems worthwhile. Relating streamflow to climate input also removes the arbitrariness of picking a start date (e.g. by defining a water year), which is a limitation of many signatures that relate flows to a date (e.g. the half flow date).

In this work, we propose the use of hydrological signatures based on how catchments attenuate the seasonal climate input (forcing). We approximate the input signal to a catchment (the forcing) by precipitation minus potential evapotranspiration and the output signal from a catchment by streamflow. We quantify the seasonal components of both signals by fitting sine waves to them, that is, we extract their (annual) Fourier modes. As the period is fixed (one year), the incoming sine wave and the outgoing sine wave differ only in their amplitude, their phase and their mean. As the mean is rather a measure of the annual water balance, we are primarily interested in

amplitude and phase. The differences in amplitude and phase are used as signatures describing the steady-state response of a catchment to periodic forcing. This idea is similar to the approach of Peters et al. (2003) who investigated drought propagation through groundwater using sinusoidal recharge, and to the approach of Obergfell et al. (2019) who used the seasonal behaviour as an additional signature in recharge estimation. The approach is also similar to approaches in transit time modelling (e.g. McGuire and McDonnell, 2006; Kirchner, 2016). Instead of focusing on the velocity of water particles, we, however, focus on the hydraulic response to periodic forcing, that is the celerity of the input "wave" of hydraulic potential (Harman, 2019). The proposed signatures are essentially also spectral domain signatures (Montanari and Toth, 2007), focusing only on a certain meaningful period – the annual period.

While there are other methods that quantify input-output relations, we propose the use of the seasonal signatures for several reasons. The seasonal signatures can be related to conceptual linear reservoirs (this will be outlined in Section 3.2), that is, they can be interpreted in terms of simple conceptual model structures and parameter values (the reservoir time constants or response times). This gives them some hydrological interpretability (cf. discriminatory power; McMillan et al., 2017). Furthermore, by quantifying the delay between seasonal climate input and catchment output, we obtain a time scale that focuses on seasonal and thus rather slow dynamics. This might make it a valuable addition to methods focusing on event scales (e.g. recession analysis) and to other slow flow signatures such as the baseflow index (BFI), or the flow duration curve and parts thereof (e.g. Q_{95}), which focus on volumes and frequencies, respectively. Lastly, the signatures do not require any parameters, they can be estimated directly from precipitation, potential evapotranspiration and streamflow data, which makes it straightforward to apply them to large samples of catchments.

In the following, we will first define the seasonal signatures, and we will present analytical solutions describing the response of linear reservoirs to periodic forcing (Section 3.2). Second, we will calculate the seasonal signatures for a range of catchments in the UK and in the US (Section 3.4, the data sources are presented in Section 3.3). We will explore how they relate to hydro-climatic forcing and catchment form, and we will interpret the underlying hydrological processes as well as limitations of the approach (Section 3.5). Finally, we will present an example application, in which we test whether two commonly used hydrological models (IHACRES, GR4J) can reproduce the observed ranges of seasonal signatures in the UK. This modelling experiment aims at exploring whether the signatures can be used as an additional source of information in model evaluation (Section 3.5.4).

3.2 Methods

3.2.1 Extracting seasonal components from time series

3.2.1.1 Quantification of periodic components

To analyse the periodic components (Fourier modes) of time series we first need to quantify these components. While we could investigate the whole frequency spectrum of our time series and see how this is altered by a catchment (Montanari and Toth, 2007), we will focus on a period T of one year. The annual period has a clear physical meaning as it is the period the Earth moves in its orbit around the Sun, which is directly linked to the energy input to the Earth system. Furthermore, the annual mode is the strongest mode in the vast majority of catchments investigated here (see Appendix B for further details). The input to a catchment, the forcing F, is approximated by precipitation P minus potential evapotranspiration E_p ($F = P - E_p$). We use E_p to avoid the need for a model or additional data which would be needed to obtain actual evapotranspiration E_a. This might be particularly problematic in water-limited catchments, where actual evapotranspiration is much smaller than potential evapotranspiration, and in catchments where precipitation and potential evapotranspiration are out of phase. We will discuss that in Section 3.5. The seasonal component of the forcing $F_{\sin}$ is given by (Milly, 1994):

$$F_{\sin} = \bar{F}\left(1 + \delta_F \sin\left(\frac{2\pi}{T}t + \phi_F\right)\right) \tag{3.1}$$

where $\bar{F}$ is the mean, δ_F is the ratio between the amplitude and the mean (the dimensionless amplitude), and ϕ_F is the phase (with respect to a reference date) of the seasonal forcing component. The output from a catchment is approximated by streamflow Q. The seasonal component of streamflow $Q_{\sin}$ is given by:

$$Q_{\sin} = \bar{Q}\left(1 + \delta_Q \sin\left(\frac{2\pi}{T}t + \phi_Q\right)\right) \tag{3.2}$$

where $\bar{Q}$ is the mean, δ_Q is the ratio between the amplitude and the mean, and ϕ_Q is the phase (with respect to the same reference date) of the seasonal streamflow component.

Since we know the period T of interest, we need to quantify the mean, the amplitude and the phase of the periodic components. There are different methods to fit a sine curve of a certain period to data, that is, to extract Fourier modes. We have compared two sine curve fitting, namely multiple linear regression and a method that makes use of the cross-covariance of two sine waves. Both methods lead to virtually the same results. A description and a comparison of the methods is shown in Appendix B. For the rest of the analysis, we will use results obtained by means of multiple linear regression (details on the fitting method can be found in Appendix B).

3.2.1.2 Calculation of seasonal signatures

Once we have extracted the seasonal components from our time series (precipitation minus potential evapotranspiration, streamflow), we can quantify how the outgoing sine wave $Q_{\sin}$ has

been altered by the catchment by comparing it to the incoming sine wave $F_{\sin}$. We define two metrics, the amplitude ratio and the phase shift, which together we call seasonal signatures. The amplitude ratio A is the ratio between the seasonal streamflow amplitude $\delta_Q \bar{Q}$ and the seasonal forcing amplitude $\delta_F \bar{F}$:

$$A = \frac{\delta_Q \bar{Q}}{\delta_F \bar{F}} \tag{3.3}$$

Given a closed long-term water balance, the amplitude ratio should theoretically always be between zero and unity, that is, the streamflow amplitude cannot be larger than the forcing amplitude. The phase shift ϕ is the difference between the phase of the seasonal streamflow component ϕ_Q and the phase of the seasonal forcing component ϕ_F:

$$\phi = \phi_Q - \phi_F \tag{3.4}$$

The phase shift should theoretically always be positive (the input should lead the output) and smaller than one year.

3.2.2 Linear reservoir theory

The derivations presented here all rely on the assumption of a linear time-invariant system (see e.g. Dooge, 1973, for an overview of linear theory of hydrological systems). This implies that forcings of different wavelengths are not influencing each other. The assumption of linearity is invalid for most real systems, yet it is still widely made as it can yield useful insights.

A linear reservoir is described by:

$$Q = \frac{S}{\tau} \tag{3.5}$$

where Q [mm d^{-1}] is the outflow from the reservoir, S is storage [mm] and τ [d] is a time constant describing how fast (slow) the reservoir responds. Conservation of mass requires:

$$\frac{dS}{dt} = Q_{\text{in}} - Q \tag{3.6}$$

where Q_{in} is the inflow to the reservoir.

3.2.2.1 Periodic forcing of a linear reservoir

If we approximate the seasonal input to a linear reservoir by a sine wave of period T (e.g. one year), we can combine Eq. (3.1), Eq. (3.5) and Eq. (3.6) to obtain:

$$\frac{dQ_{\sin}}{dt} = \frac{\bar{F}}{\tau}\left(1 + \delta_F \sin\left(\frac{2\pi}{T}t + \phi_F\right)\right) - \frac{Q_{\sin}}{\tau} \tag{3.7}$$

We might neglect the (initial) phase if we choose a starting time t that is aligned with the seasonal forcing component ($\phi_F = 0$). It can be shown that the steady state response of a linear reservoir to

a sinusoidal input signal is a damped and phase shifted version of the input signal (see Appendix B for a more detailed derivation; or Eriksson, 1971; Peters et al., 2003):

$$Q_{\sin}(t) = \bar{F}\left(1 + \delta_F A \sin\left(\frac{2\pi}{T}t + \phi\right)\right) \tag{3.8}$$

where A is the amplitude ratio and ϕ is the phase shift induced by a single linear reservoir.

$$A = \frac{1}{\sqrt{1 + (2\pi\frac{\tau}{T})^2}} \tag{3.9}$$

$$\phi = \arccos\left(\frac{1}{\sqrt{1 + (2\pi\frac{\tau}{T})^2}}\right) = \arccos(A) \tag{3.10}$$

We can rewrite Eq. (3.8) as follows:

$$Q(t) = \bar{Q}\left(1 + \delta_Q \sin\left(\frac{2\pi}{T}t + \phi\right)\right) \tag{3.11}$$

In a mass conserving system in steady-state, the mean of the output should equal the mean of the input. If the means obtained from data are different, either the forcing term is inaccurate (e.g. due to differences between actual and potential evapotranspiration) or the streamflow term is inaccurate (e.g. due to other losses or gains). The product of input amplitude and amplitude ratio equals the output amplitude ($\delta_F \bar{F} A = \delta_Q \bar{Q}$).

From Eq. (3.9) and Eq. (3.10), we can see that the amplitude ratio and the phase shift are given by A and $\arccos(A)$, respectively. Since A is fully defined by the ratio between τ and T, and T is usually known (e.g. one year), we can theoretically use A to determine the time constant τ of the reservoir. This requires the identification of both the seasonal components of the input and output signal of that period (see Section 3.2.1), and assumes the system to behave as a single linear reservoir. Theoretically, we could also apply the theory to other periods than one year, but for the reasons stated above we only investigate the annual period.

The amplitude ratio A and the phase shift $\arccos(A)$ can be plotted against each other for various values of τ as shown in Figure 3.1. This results in a characteristic curve which captures the response of all single linear reservoirs. Different time constants τ (as proportions of the period, here one year) lead to different positions on the curve. For very fast reservoirs, the phase shift is close to 0 days and the amplitude ratio is close to unity (that is, the signal is not attenuated at all). For very slow reservoirs, the signal is phase shifted up to 91 days and the amplitude ratio approaches 0. The maximum phase shift of about 91 days corresponds to a quarter of a period (90 degrees). Mathematically, this can be explained by Eq. (3.10), as the arccosine of a quantity between 0 and unity (such as A) ranges between 0 and 90 degrees.

Note the similarity of Figure 3.1 to Figure 3c in Kirchner (2016), which shows the relationship between phase shift and amplitude ratio for gamma-distributed catchment transit time distributions. An exponential transit time distribution (a special case of the gamma distribution)

corresponds to a linear reservoir describing the velocity of particles. Similarly, a linear reservoir describing the impulse response (the linear reservoir from Eq. (3.5)), that is, the celerity of the incoming wave of hydraulic potential, corresponds to an exponential response time distribution or an exponential unit hydrograph (cf. Eriksson, 1971; Dooge, 1973).

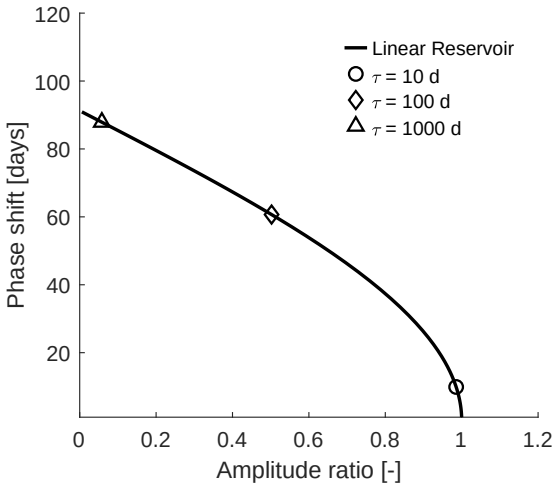

Figure 3.1: Amplitude ratio against phase shift for a single linear reservoir for varying time constants τ. Three example time constants are indicated by the symbols.

3.2.2.2 Combinations of linear reservoirs

Linear systems (Dooge, 1973) have the advantage that it is relatively straightforward to add more components, that is, reservoirs. It is quite common to have serial and/or parallel combinations in rainfall-runoff models. In theory, we can find analytical solutions for the amplitude ratio and phase shift for all combinations of linear reservoirs (cf. to the transfer function approach of Young, 1998, who identifies combinations of reservoirs that fit the data best in an inductive way). There are two basic arrangements, a serial arrangement of reservoirs and a parallel arrangement of reservoirs.

3.2.2.3 Linear reservoirs in series

Linear reservoirs in series can be conceptualised as follows. Every outflow is the inflow to the next reservoir. Hence, if the i-th reservoir has a time constant τ_i, the amplitude ratios A_i are multiplied and the phase shifts ϕ_i are added (see Appendix B for a more detailed derivation):

$$A_{\text{tot}} = \prod_{i=1}^{n} A_i \tag{3.12}$$

$$\phi_{\text{tot}} = \sum_{i=1}^{n} \phi_i = \sum_{i=1}^{n} \arccos(A_i) \tag{3.13}$$

Figure 3.2 shows the amplitude ratio plotted against the phase shift similar to Figure 3.1, but now with two linear reservoirs in series. The different lines are examples with fixed time constants of the first reservoir. They all start from the black line (from the points marked by the symbols in Figure 3.1), the characteristic curve for a single linear reservoir, which is the lower limit. Then, as the time constant of the second reservoirs increases, the lines "move" left and upwards, which corresponds to a decrease in amplitude ratio and an increase in phase shift. For example, the red line (τ_1 = 10 d) starts out with a phase shift of about 10 days, and ends at a phase shift of about 101 days, which is an increase of about 91 days, the maximum phase shift of the second reservoir. The lines cross each other as we allow τ_2 to be larger than τ_1. This implies that sometimes a faster reservoir is followed by a slower one, and sometimes a slower reservoir is followed by a faster one. The grey shaded area contains all possible combinations for two reservoirs in series. The lower limit is a single linear reservoir. The upper limit corresponds to two reservoirs with the same time constant (a two-reservoir Nash cascade), which equals a gamma distribution with a shape parameter equal to 2 (Nash, 1957).

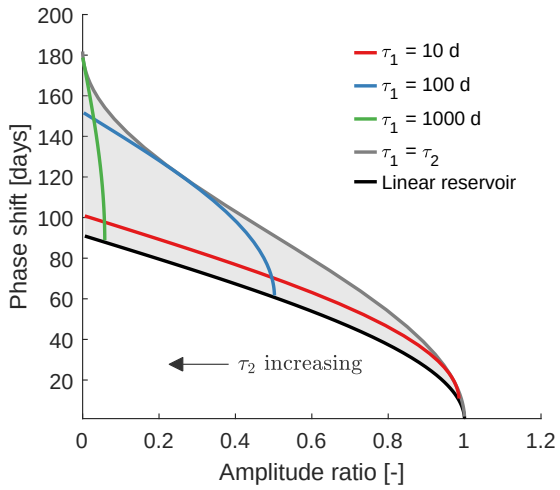

Figure 3.2: Amplitude ratio against phase shift for two linear reservoirs in series. Each line corresponds to a fixed time constant for the first reservoir (τ_1), while the time constant of the second reservoir varies (1 d $\leq \tau_2 \leq$ 10000 d; it is increasing from right to left). The black line indicates a single linear reservoir (the lower boundary). The grey line indicates the upper boundary where $\tau_1 = \tau_2$. The shaded area contains all possible combinations of amplitude ratio and phase shift for two linear reservoirs in series.

3.2.2.4 Linear reservoirs in parallel

Linear reservoirs in parallel result in a "mixture" of the outflows from each reservoir. The resulting flow is a combination of sine waves of the same period, weighted by the fraction p_i going into each reservoir. For the sake of simplicity, we only consider two reservoirs in parallel. We denote the fraction going into the second reservoir by p, and therefore the fraction going into the

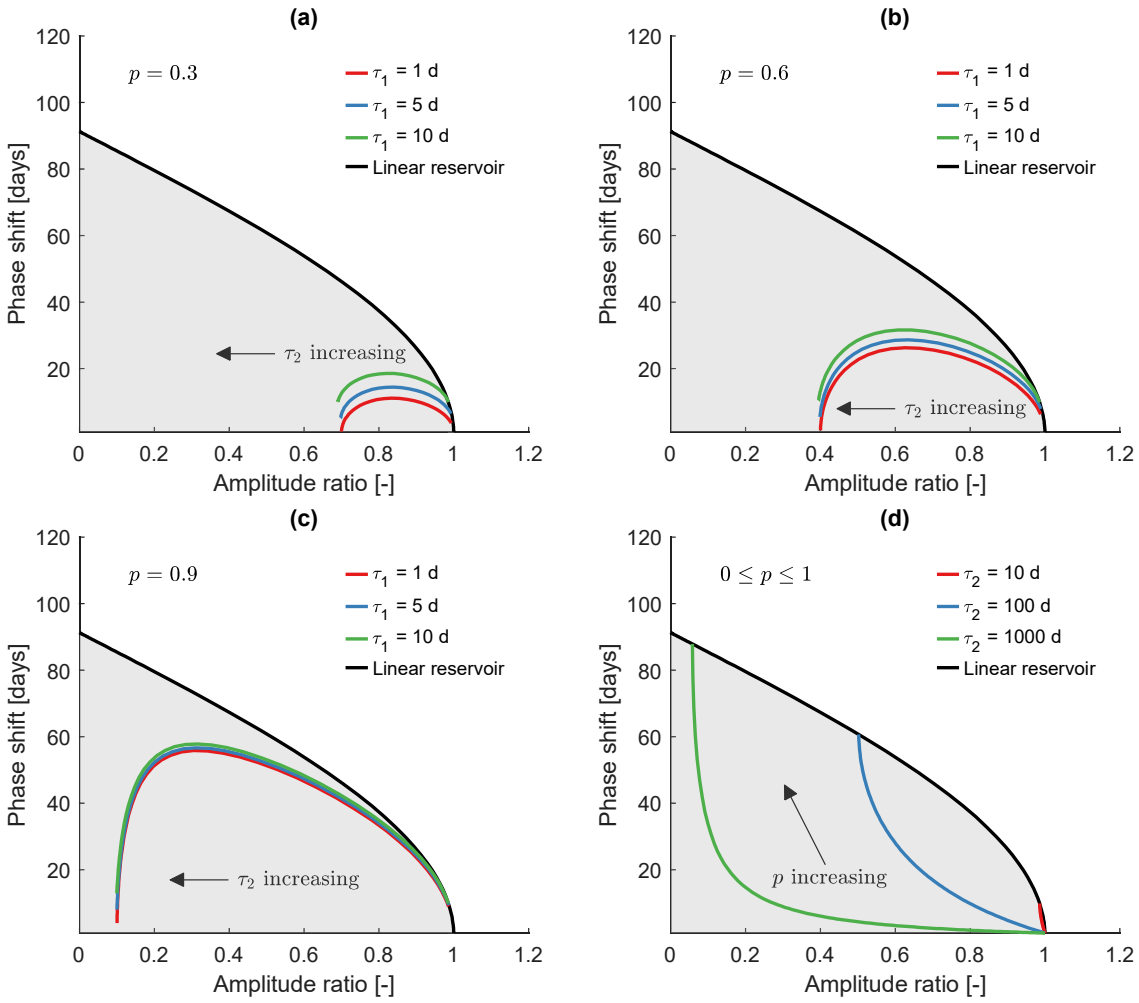

Figure 3.3: Amplitude ratio against phase shift for two linear reservoirs in parallel. **(a)** Each line has a fixed time constant for the first reservoir (τ_1), while the time constant of the second reservoir varies (10 d $\leq \tau_2 \leq$ 10000 d; it is increasing from right to left). The fraction p going into the second reservoir is 0.3. **(b)** Same as **(a)** with $p = 0.6$. **(c)** Same as **(a)** with $p = 0.9$. **(d)** Each line has a fixed time constant for the first reservoir ($\tau_1 = 1$ d), and for the second reservoir (τ_2). The fraction p going into the second reservoir is varied (it is increasing from right to left). The shaded area contains all the possible combinations of amplitude ratio and phase shift for two linear reservoirs in parallel.

first reservoir by $1 - p$. Thinking of the second reservoir as the slow one, p might be compared to the idea of the baseflow index (BFI), the volumetric ratio between baseflow and total streamflow (Institute of Hydrology, 1980). For two reservoirs in parallel we get (see Appendix B for a more

detailed derivation):

$$A_{\text{tot}} = \sqrt{\left[(1-p)A_1\cos\phi_1 + pA_2\cos\phi_2\right]^2 + \left[(1-p)A_1\sin\phi_1 + pA_2\sin\phi_2\right]^2} \tag{3.14}$$

$$\phi_{\text{tot}} = \arctan\left(\frac{(1-p)A_1\sin\phi_1 + pA_2\sin\phi_2}{(1-p)A_1\cos\phi_1 + pA_2\cos\phi_2}\right) \tag{3.15}$$

Figure 3.3 shows the amplitude ratio plotted against the phase shift similar to Figure 3.1, but now with two linear reservoirs in parallel. We show multiple plots to highlight the three degrees of freedom: the two reservoir time constants and the fraction going into each reservoir. The latter is highlighted in Figure 3.3d, but also visible in Figures 3.3a-c. The grey shaded area contains all the possible combinations for two reservoirs in parallel. The upper limit is a single linear reservoir. The lower limit is effectively given by the x- and the y-axis.

As an example, Figure 3.3a can be explained as follows: τ_1 is always 1 d, the fraction p going into the second reservoir is 0.3, and τ_2 starts with a value of 10 d and then increases. So at first, both reservoirs are rather fast and we get a high amplitude ratio and a small phase shift for the combined sine wave (see Eq. (3.14) and Eq. (3.15)). Then, the second reservoir gets slower, leading to a decrease in amplitude ratio and an increase in phase shift. As the second reservoirs gets slower and slower, it will contribute less and less to the overall sine wave. For very high values of τ_2 (e.g. 10000 d), the sine wave coming out of the second reservoir is almost a straight line, so the combined sine wave is primarily consisting of the sine wave coming out of the first reservoir. Since only a fraction of $1 - p = 0.7$ of the total input has gone into the first reservoir, the amplitude of the combined sine wave is approximately 0.7 times the input amplitude with a very small phase shift, as the first reservoir hardly attenuates the signal.

3.2.3 Seasonal signatures as a diagnostic tool for evaluating hydrological models

We use two conceptual rainfall-runoff models and we test whether the seasonal signatures can be used as a diagnostic tool to assess model performance (Gupta et al., 2008). In particular, we test whether the models are capable of reproducing the range of observed signatures without calibrating them to streamflow data (cf. Vogel and Sankarasubramanian, 2003). This modelling experiment is intended to test whether the proposed signatures have the potential to be a useful additional source of information in model building and evaluation. We do not intend (or suggest) that the presented evaluation approach can replace existing model evaluation methods. We limit the analysis to two models and 40 catchments to keep the computational demand manageable. We also limit the model evaluation to catchments in the UK, as the seasonal signatures are unreliable in arid catchments (see Section 3.5). The subset of catchments is described in Section 3.3 and in Appendix B.

The first model is the IHACRES model (Identification of unit Hydrographs and Component flows from Rainfall, Evaporation and Streamflow). It is conceptually relatively similar to the

considerations in Section 3.2. It has a soil moisture store (non-linear deficit store), and two parallel linear stores for fast flow and for slow flow (Croke and Jakeman, 2004). It has been used in many modelling studies in Australia (Post and Jakeman, 1999) and also in the UK (Sefton and Howarth, 1998). The second model is the GR4J model (modèle du Génie Rural à 4 paramètres Journalier). It also has a parallel flow structure, but the internal parameterisation is different. It contains more non-linearities and it has fixed internal parameters. Additionally, it has a groundwater exchange parameter aimed at representing inter-catchment groundwater flows. It has been used in many modelling studies in France (Perrin et al., 2003), in the UK (Smith et al., 2019; Harrigan et al., 2018), and in the US (Oudin et al., 2018). We use the implementations of the two models in the MARRMoT toolbox v1.2 (Knoben et al., 2019a), a Matlab toolbox containing many hydrological models aimed at model comparison studies. The pure delay function in the MARRMoT implementation of IHACRES is set to 0, making it (conceptually, not necessarily numerically) equal to the version used by Croke and Jakeman (2004). In our modelling experiment, IHACRES has therefore 6 parameters, and GR4J has 4 parameters. Detailed information on the parameter ranges and on model warm-up periods can be found in Appendix B.

To test which ranges of seasonal signatures the two models can reproduce, we run a Monte-Carlo sampling experiment. We sample parameter sets for both models using Latin Hypercube sampling, an efficient sampling method (Cheng and Druzdzel, 2000) that assumes uniform prior parameter distributions. With the parameter sets obtained, we run both models for each of the 40 catchments, that is, we use the same parameter sets for each catchment. To test for robustness, we sample an increasing number of parameter sets (20000 parameter sets are considered sufficient, see Appendix B for more information). We then use the modelled streamflow time series to calculate three hydrological signatures per parameter set: the two seasonal signatures presented here and the baseflow index (BFI). The resulting modelled signatures are compared to observed signatures and explored in a rather general way, as we want to examine what the models can do without actually calibrating them to streamflow data (cf. Vogel and Sankarasubramanian, 2003). That is, we are not interested in finding the "best" parameter set, but in whether a certain model (given certain parameter ranges) is generally capable of reproducing the signatures we observe.

3.3 Data

3.3.1 Data sources

We use catchment data from Great Britain and the United States. The data for the UK are obtained from different sources. Daily streamflow data, catchment characteristics and catchment boundaries are obtained from the NRFA (National River Flow Archive, 2019), precipitation data from CEH-GEAR (Tanguy et al., 2016), and potential evapotranspiration data from CHESS-PE (Robinson et al., 2016). For the model evaluation we select catchments that are part of the UK Benchmark Network (Harrigan et al., 2017), which describes catchments in the UK that are near-

natural. The subset of catchments is chosen to be representative of the UK, details are shown in Appendix B. The data for the US are obtained from the CAMELS dataset (Newman et al., 2015; Addor et al., 2017). CAMELS includes daily precipitation, potential evapotranspiration (we use Daymet forcing data) and streamflow data, as well as a wide range of catchment attributes for 671 catchments in the contiguous US. We trim the daily data to contain only full water years (starting 1 October) and we analyse data from 1989 to 2009. We also remove catchments with missing records during that time period. While we need to pick a start date for the analysis, this date does not influence the results (e.g. using 1 January as starting date would result in the same phase shift).

3.3.2 Hydrological signatures and catchment attributes

We calculate different hydrological signatures and we use different catchment attributes, all summarised in Table 3.1. The climate indices from Knoben et al. (2018) are based on monthly averages and they need to be interpreted as follows. A moisture index I_m of 1 indicates the most humid (energy-limited) catchments, a moisture index of -1 indicates the most arid (water-limited) catchments. A moisture index seasonality $I_{m,r}$ of 0 indicates catchments where the climate stays constant throughout the year, a moisture index seasonality of 2 indicates catchments where the climate switches between fully arid and fully humid within the year.

Table 3.1: Hydrological signatures and catchment attributes used in this study. [1]Should in theory be smaller than unity. [2]Should theoretically always be positive and in practice be smaller than one year. Further discussions on the possible ranges of the seasonal signatures can be found in the text.

Name	Description	Unit	Range	Reference
Hydrological signatures				
BFI	Baseflow index	[-]	$[0,1]$	Institute of Hydrology (1980)
A	Amplitude ratio	[-]	$[0,1]$[1]	Eq. (3.3)
ϕ	Phase shift	[d]	$[0,365]$[2]	Eq. (3.4)
Catchment attributes				
I_m	Moisture index	[-]	$[-1,1]$	Knoben et al. (2018)
$I_{m,r}$	Moisture index seasonality	[-]	$[0,2]$	Knoben et al. (2018)
f_s	Snow fraction	[-]	$[0,1]$	Knoben et al. (2018)
PROPWET	Catchment wetness index	[-]	$[0,1]$	National River Flow Archive (2019)
% fractured aquifer	Fraction of highly productive fractured aquifer	[%]	$[0,100]$	National River Flow Archive (2019)
% carbonate rock	Fraction of carbonate sedimentary rock	[%]	$[0,100]$	Addor et al. (2017)

3.4 Results

3.4.1 Extracting seasonal components from time series

First, we extract seasonal components from $P - E_p$ (forcing) and Q (streamflow) for all catchments. The resulting sine wave parameters are then used to calculate the amplitude ratios (Eq. (3.3)) and phase shifts (Eq. (3.4)), respectively. Figure 3.4 shows $P - E_p$ and Q for two catchments alongside their seasonal (sinusoidal) components. Both catchments experience a similar forcing, but their response is very different. The Ericht at Craighall, a rather responsive catchment, shows a seasonal streamflow component that is very similar to the seasonal forcing component. In contrast, the East Avon at Upavon, a groundwater-dominated catchment, shows a strongly attenuated seasonal streamflow component. For our seasonal signatures this would mean (a) that the responsive catchment has a high amplitude ratio, that is, the streamflow amplitude is almost as large as the forcing amplitude, while the groundwater-dominated catchment has a low amplitude ratio. And (b) that the responsive catchment has a small phase shift, that is, it responds quickly to the (seasonal) forcing, while the groundwater-dominated catchment has a large phase shift.

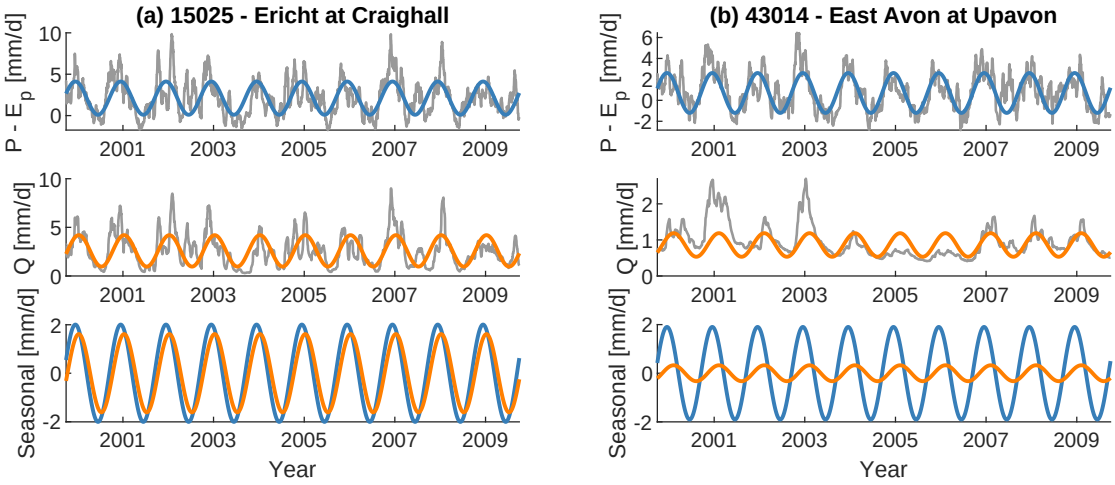

Figure 3.4: Climate input (P - E_p; blue) and catchment output (Q; orange) for two example catchments in the UK, and their respective seasonal components. The time series are smoothed using a 30-day moving mean. The Ericht is a rather responsive catchment (BFI = 0.47), while the East Avon has a large baseflow component (BFI = 0.89). Note that for the bottom plots ("Seasonal") the mean values of the sine curves are set to zero.

3.4.2 Seasonal signatures of observed catchment data

To visualise the seasonal signatures, we plot the amplitude ratios and phase shifts in a similar way as in Figures 3.1, 3.2, and 3.3. This is shown in Figure 3.5a for all UK catchments. These include

catchments with human influences, such as groundwater abstractions, man-made reservoirs or water transfers. The overall pattern in Figure 3.5a is very similar to the pattern using benchmark catchments alone (grey dots). We therefore use all of the catchments, noting that a few catchments might be unsuitable for individual analyses.

Figure 3.5a shows that most of the catchments fall below the solid grey line, the line which indicates the type of response that could be simulated by a single linear reservoir (see Figure 3.5b). The area below the solid line can be simulated by two reservoirs in parallel. This would be the most parsimonious way to reproduce the observed behaviour if we decide to construct our model using linear reservoirs only. A few catchments plot above the solid line. For these catchments, the most parsimonious way to reproduce the pair of observed amplitude ratio and phase shift would therefore be two reservoirs in series. Very few catchments have an amplitude ratio larger than unity. While this could be caused by various errors in the data, it is likely due to erroneous catchment areas and/or the presence of inter-catchment groundwater flows or water transfers. If a catchment receives more net rainfall than the surface catchment area suggests (runoff ratio > 1), the amplitude in the output signal (streamflow) can be larger than the amplitude in the (erroneous) input signal.

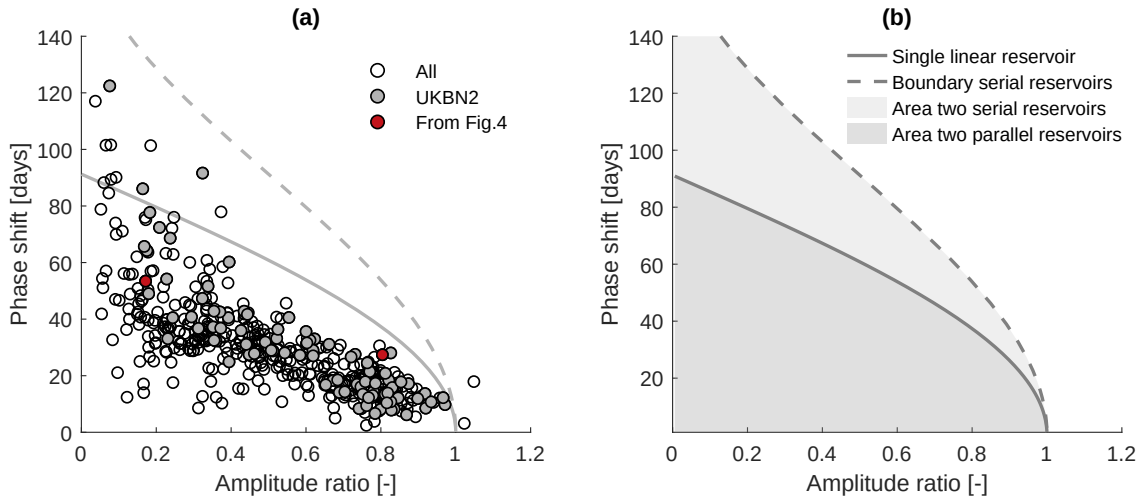

Figure 3.5: **(a)** Amplitude ratio against phase shift for UK catchments. Grey dots indicate benchmark catchments, red dots indicate the two catchments shown in Figure 3.4. Grey solid line indicates a single linear reservoir, grey dashed line indicates the outer envelope for two reservoirs in parallel. Note that both axes are limited (two catchments are not shown). **(b)** Theoretical areas and limits for a single linear reservoir, two reservoirs in series, and two reservoirs in parallel.

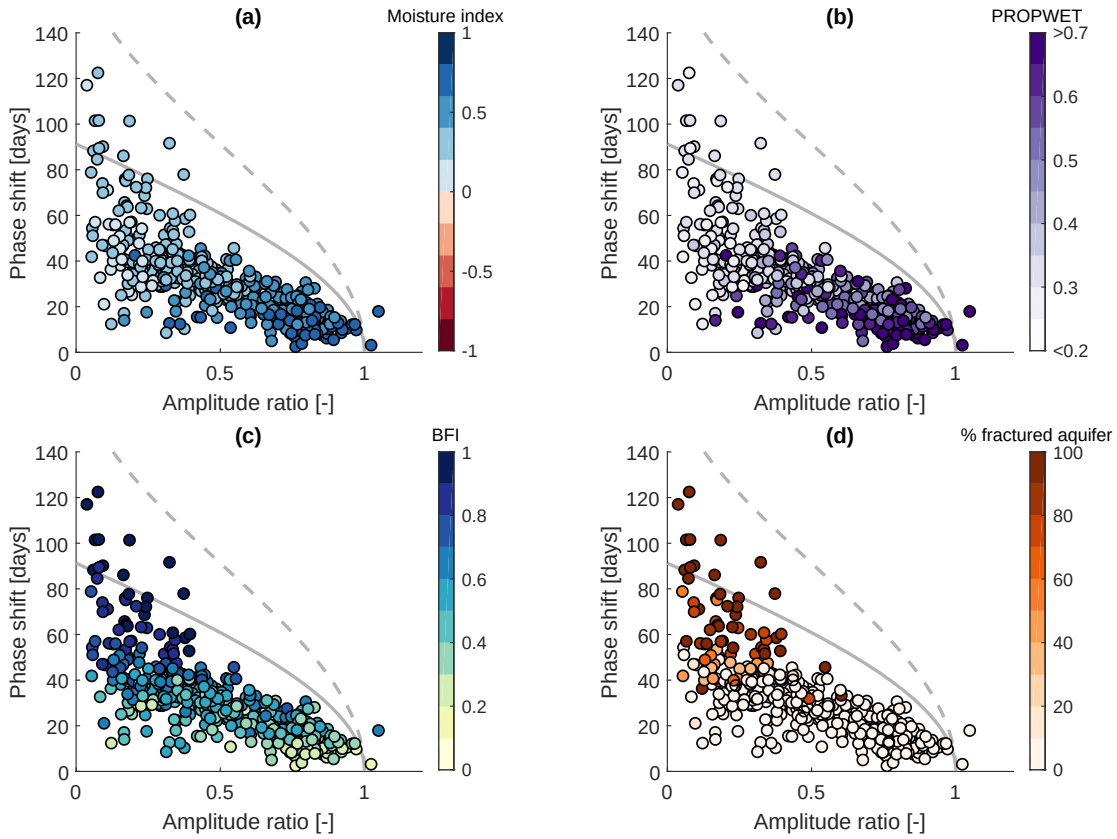

Figure 3.6: Amplitude ratio against phase shift for UK catchments. Grey solid line indicates a single linear reservoir, grey dashed line indicates the outer envelope for two reservoirs in parallel. Colours indicate **(a)** the moisture index, **(b)** the catchment wetness index, **(c)** the baseflow index, and **(d)** the fraction of highly productive fractured aquifer. Note that both axes are limited (two catchments are not shown).

Table 3.2: Pearson and Spearman correlation coefficients between seasonal signatures and catchment attributes for UK catchments.

Pearson	Amplitude ratio	Phase shift	I_m	PROPWET	BFI	% fractured aquifer
Amplitude ratio	1.00	-0.60	0.80	0.74	-0.58	-0.49
Phase shift	-0.60	1.00	-0.49	-0.50	0.66	0.58
Spearman	Amplitude ratio	Phase shift	I_m	PROPWET	BFI	% fractured aquifer
Amplitude ratio	1.00	-0.80	0.82	0.78	-0.58	-0.51
Phase shift	-0.80	1.00	-0.76	-0.75	0.77	0.60

3.4.3 Relationship between seasonal signatures and catchment attributes – UK

Figure 3.6 shows pairs of amplitude ratios and phase shifts, coloured according to different hydrological signatures and catchment attributes, respectively (explained in Table 3.1). Corresponding correlation coefficients are shown in Table 3.2. Figure 3.6a shows a clear pattern between the

moisture index and the seasonal signatures. Generally, the less humid the catchments, the lower the amplitude ratio and the larger the phase shift. In other words, drier catchments attenuate the incoming forcing signal more strongly. This might partly be because we use potential evapotranspiration as our forcing. Lower actual evapotranspiration than potential evapotranspiration leads to a decreased input amplitude and thus to a higher amplitude ratio. Most of the very humid catchments plot close together and the relationship between amplitude ratio and phase shift seems to be almost linear. Less humid catchments (note that in the UK none of the catchments are actually water-limited at the annual scale) show a larger spread, especially regarding the phase shift. Figure 3.6b shows a very similar pattern between the catchment wetness index and the seasonal signatures. Wetter catchments exhibit higher amplitude ratios and lower phase shifts, and vice versa. The catchment wetness index is strongly correlated with the moisture index (Spearman rank correlation of 0.94). Figure 3.6c shows a clear pattern between the baseflow index and the seasonal signatures. In contrast to the moisture index, where the stratification follows mostly the x-axis (amplitude ratio), the stratification follows mostly the y-axis (phase shift). Catchments with high BFIs exhibit low amplitude ratios and large phase shifts, and vice versa. Finally, in Figure 3.6d we can see that catchments underlain by highly productive fractured aquifers exhibit (with a few exceptions) low amplitude ratios and large phase shifts.

3.4.4 Relationship between seasonal signatures and catchment attributes – US

Figure 3.7 shows pairs of amplitude ratios and phase shifts for the US, coloured according to different hydrological signatures and catchment attributes, respectively (explained in Table 3.1). Corresponding correlation coefficients are shown in Table 3.3. Catchments with significant snow fraction ($f_s > 0.001$) are removed, as snow presents another hydrological process which is not our focus here. Generally, snow adds another storage process, and this is reflected in large phase shifts observed in snowy catchments (see Appendix B). The non-snowy catchments in the US show a similar trend to the catchments in the UK. Yet generally, the amplitude ratios are lower and the phase shifts larger compared to the UK (note that the y-axes in Figure 3.7 differ in their range from the y-axes in Figure 3.6). Humid catchments tend to have higher amplitude ratios and smaller phase shifts (Figure 3.7a). Climate seasonality, indicated by the moisture index seasonality (see Figure 3.7b), also influences the seasonal signatures. Catchments with a larger moisture index seasonality, that is, a more variable monthly moisture index over the year, tend to have smaller phase shifts. The BFI (Figure 3.7c) does not show such a clear pattern as for the UK catchments (Figure 3.6c). Similarly, subsurface properties such as the fraction of carbonate sedimentary rock (Figure 3.7d; and other attributes not shown here) only show a weak relationship with the seasonal signatures. Catchments with larger fractions of carbonate sedimentary rocks tend to have lower amplitude ratios and larger phase shifts. The overall pattern, however, is rather scattered. Contrary to the UK, some of the catchments in the US plot

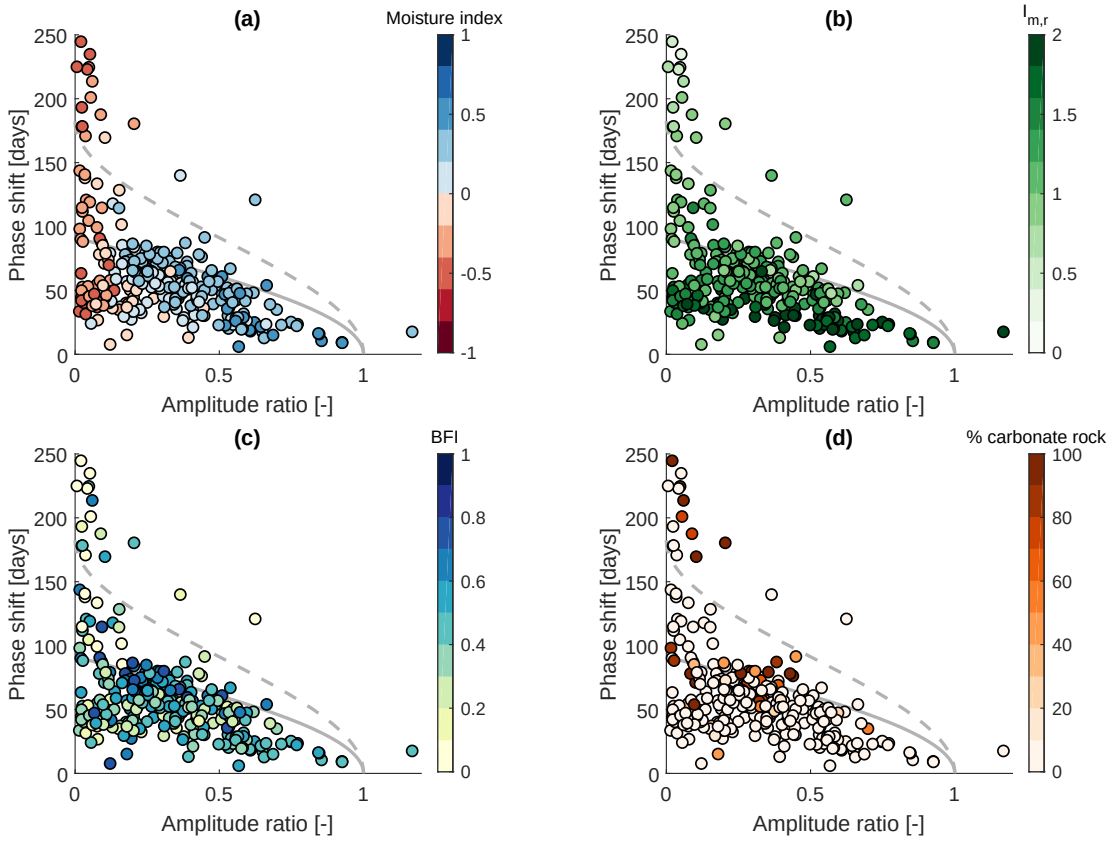

Figure 3.7: Amplitude ratio against phase shift for CAMELS catchments. Catchments with snow fraction $f_s > 0.001$ are removed from the analysis. Grey solid line indicates a single linear reservoir, grey dashed line indicates the outer envelope for two reservoirs in parallel. Colours indicate **(a)** the moisture index, **(b)** the moisture index seasonality, **(c)** the baseflow index, and **(d)** the fraction of carbonate sedimentary rock. Note that both axes are limited (12 catchments are not shown) and that the range of the phase shift-axis is different from Figure 3.6.

Table 3.3: Pearson and Spearman correlation coefficients between seasonal signatures and catchment attributes for CAMELS catchments.

Pearson	Amplitude ratio	Phase shift	I_m	$I_{m,r}$	BFI	% carbonate rock
Amplitude ratio	1.00	-0.26	0.75	0.31	0.06	-0.16
Phase shift	-0.26	1.00	-0.39	-0.51	-0.14	0.26
Spearman	Amplitude ratio	Phase shift	I_m	$I_{m,r}$	BFI	% carbonate rock
Amplitude ratio	1.00	-0.46	0.78	0.23	0.04	-0.15
Phase shift	-0.46	1.00	-0.32	-0.64	0.06	0.36

outside the area that can be modelled by either two reservoirs in series or in parallel and some catchments have phase shifts larger than 182 days, the approximate limit for two reservoirs in series. These catchments are very arid and the low moisture seasonality index indicates that most of the precipitation falls when potential evapotranspiration is highest (i.e. in summer).

3.4.5 Seasonal signatures as a diagnostic tool for evaluating hydrological models

In a similar fashion as for the observed catchment data, we now investigate the model runs using IHACRES and GR4J. Figure 3.8 shows the resulting amplitude ratios and phase shifts for all model runs, that is for 20000 parameter sets using data from a subset of 40 catchments in the UK. These plots show which combinations of seasonal signatures (and BFI) can be obtained with each model, given the forcing of 40 different catchments covering most of the hydro-climatic variability of the UK, and given the parameter ranges chosen. They hence show the "signature space" of a model in the dimensions given by amplitude ratio and phase shift (and BFI).

IHACRES (Figure 3.8a) shows a pattern that covers the area that can be modelled by two reservoirs in parallel, and a large fraction of the area that can be modelled by two reservoirs in series (see Figures 3.2 and 3.3). The BFI spans the whole range from 0 to 1. IHACRES can reproduce the observed amplitude ratios and phase shifts, although one catchment sits just at the boundary of the point cloud. GR4J (Figure 3.8b) covers a different signature space. The phase shift never exceeds 110 days, the amplitude ratio often exceeds unity, and the BFI tends to be high. GR4J can reproduce most of the observed amplitude ratios and phase shifts, except for catchments with very large phase shifts. Furthermore, it struggles to simultaneously reproduce the observed phase shifts and BFIs.

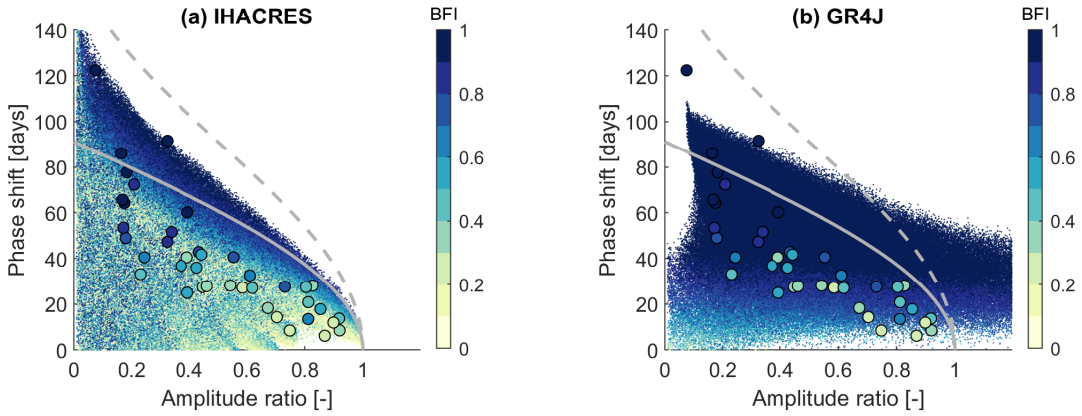

Figure 3.8: Amplitude ratio against phase shift for 40 catchments in the UK using 20000 parameter sets each for **(a)** IHACRES and **(b)** GR4J. The large dots show the observed signatures of the 40 catchments used for the modelling experiment. Colours indicate the BFI. Note that both axes are limited.

Both models sometimes yield phase shifts that are close to one year (not shown here), which are effectively negative phase shifts. Negative implies that the periodic component of Q leads the periodic component of $P - E_p$. This can happen if actual evapotranspiration E_a differs considerably from potential evapotranspiration E_p, and hence most of the input seasonality stems from P (and not E_p). This can be observed in a few catchments in the US (not shown here).

It is only observed once in the UK (in a catchment with a man-made reservoir, not shown here), and therefore we do not investigate these model runs further.

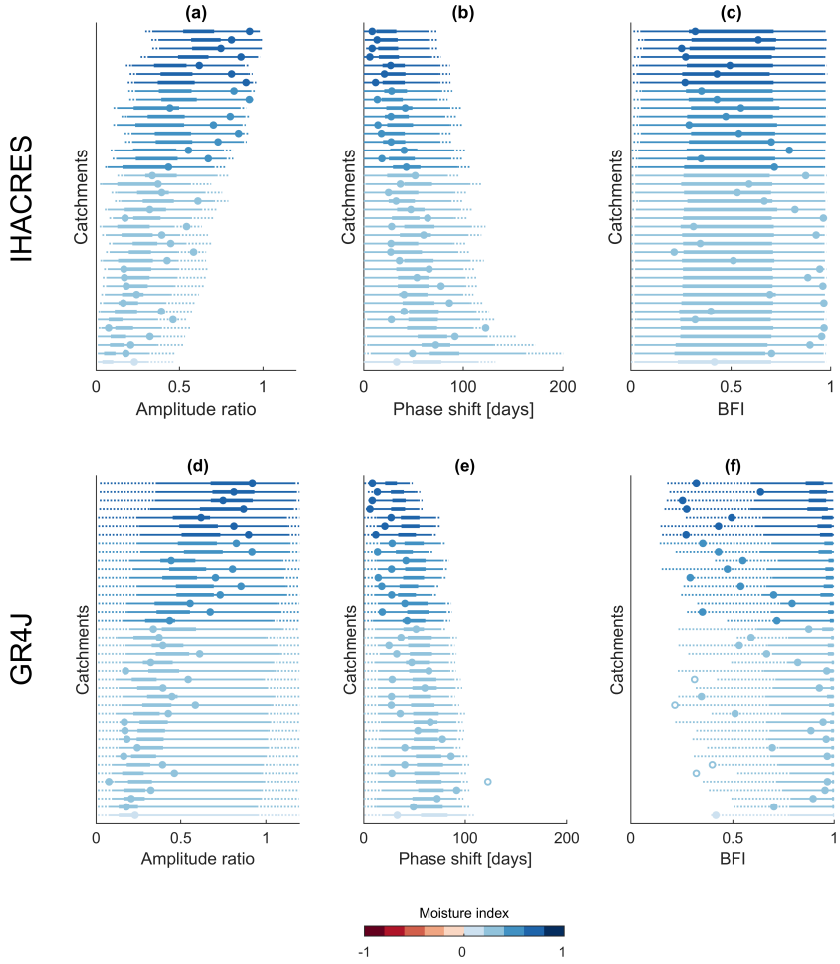

Figure 3.9: Distributions of different hydrological signatures resulting from the modelling experiment. Each line stands for one of the 40 catchments and the colours indicate the corresponding moisture index. The distributions of the modelled signatures are indicated by box-whisker-type plots. The thick line spans from the 25th to the 75th percentile. The thin line spans from the 1st (75th) to the 25th (99th) percentile. The dotted line indicates values below (above) the 1st (99th) percentile. The circles indicate the observed signature values, while filled circles indicate that the observed signature is inside the modelled signature space and vice versa. **(a)**, **(b)**, and **(c)** show results for IHACRES. **(d)**, **(e)**, and **(f)** show results for GR4J. Model runs with amplitude ratios lower than 0.01, amplitude ratios larger than 1.2, or phase shifts larger than 200 days have been removed.

Figure 3.9 shows distributions ("one-dimensional signatures spaces") for three hydrological signatures for the 40 catchments investigated here. These plots tell us which signature values a model tends to produce (given a certain sampling scheme), the ranges of signatures a model can

reproduce (given the parameter ranges chosen), and how (much) a signature varies with varying forcing.

Figure 3.9 displays similar information as Figure 3.8, yet without considering interactions between the three signatures. IHACRES can produce amplitude ratios from 0 to 1 and phase shifts up to 182 days (which is the limit for two reservoirs in series) and larger. GR4J can produce amplitude ratios that clearly exceed one, and cannot model phase shifts larger than 110 days (given the parameter ranges chosen). For both models, more arid forcing leads to lower amplitude ratios and larger phase shifts, and vice versa. For the BFI (Figures 3.9c,f) we can see that IHACRES covers the whole possible space (0 to 1) relatively evenly. GR4J tends to produce very high BFIs for almost every parameter set. BFIs smaller than 0.5 are possible with GR4J, but rather rare (or unlikely).

3.5 Discussion

3.5.1 Representation of seasonal components by sine waves and limitations of the approach

A sine wave is a simple way of describing the seasonality of a signal. The results suggest that for most of the catchments investigated here, this approach is reasonable and efficient. Figure 3.4 shows that the average seasonal pattern is captured by the fitted sine waves. Differences between years cannot be captured by our approach, as we fit a single sine wave to describe the average seasonal behaviour. To robustly capture the average seasonal behaviour, we need relatively long time series. Comparing results from two different 10 year periods shows that the signatures are robust for the majority of catchments, that is, their values do not differ substantially from one time period to the other (details are shown in Appendix B).

The UK catchments and most of the US catchments exhibit a relatively strong unimodal (climate) seasonality (see e.g. Knoben et al., 2018). In other climates with a less distinct seasonal pattern, or with two seasons per year (Knoben et al., 2019b), our approach will not work. Semi-arid and arid catchments also tend to have a less smooth seasonal input, as water availability is more fragmented (Peters et al., 2003). Water-limited catchments can show a strong difference between potential evapotranspiration and actual evapotranspiration, which limits the applicability of our approach (we will discuss that later in more detail). We exclude catchments where precipitation is falling as snow. While snowy catchments are typically also strongly seasonal (Schaefli, 2016), this seasonality is mostly a climate phenomenon. It is rather related to temperature seasonality and not to the response of a catchment to periodic forcing.

3.5.2 A perceptual model of the seasonal response of catchments in the UK

The results, in particular Figures 3.5 and 3.6 and Table 3.2, show clear patterns in the seasonal signatures. We can see that the seasonal response in the UK can be simulated by either two

reservoirs in series or two reservoirs in parallel. This does not mean that there are no other configurations of more reservoirs leading to the same pairs of amplitude ratio and phase shift. Rather, two reservoirs in series and in parallel, respectively, are the most parsimonious reservoir configuration to reproduce the observed seasonal behaviour. Of course, two reservoirs in parallel and two reservoirs in series, respectively, might be seen as "special cases" of a soil reservoir followed by a fast and a slow reservoir, that is, a three-reservoir arrangement. Furthermore, there might be concepts other than reservoirs which can explain the observed behaviour. Still, the observed patterns, both where the catchments plot in the amplitude ratio vs. phase shift plot (Figure 3.5) and how the catchment attributes relate to that (Figure 3.6), suggest that the seasonal signatures are indeed a window into catchment functioning (Berghuijs et al., 2014) and thus have discriminatory power (McMillan et al., 2017; Addor et al., 2018).

Figures 3.6a and 3.6b show how climate aridity and catchment wetness influence amplitude ratio and phase shift. The observation that more humid catchments respond more quickly to forcing (Figures 3.6a and 3.6b) concurs with our understanding of these catchments. Wetter and therefore more saturated catchments partition the incoming water mostly into fast flow. The hydrograph closely resembles the forcing, which can also be seen in Figure 3.4 for the responsive Ericht river. The drier the catchments become, the more water is able to infiltrate and subsurface properties become more important. This might explain why the spread becomes larger for less humid and hence less saturated catchments. In less humid catchments, actual evapotranspiration is more likely to deviate from potential evapotranspiration. This might be another reason for the greater attenuation in drier catchments, as the actual input ($P - E_a$) is lower than the theoretical one we compare to ($P - E_p$). In the UK, the assumption that $E_a = E_p$ seems reasonable (see Appendix B for further information). In more arid regions, such as parts of the US (see Section 3.5.3), this assumption is invalid.

The variability among UK catchments that cannot be explained by catchment wetness can mostly be explained by subsurface properties and the associated response time of a catchment. Catchments with high BFIs and thus large baseflow components show lower amplitude ratios and larger phase shifts, that is a more damped and lagged response (Peters et al., 2003). This can also be seen in Figure 3.4 for the groundwater-dominated East Avon river. The relationship between BFI and the seasonal signatures (Figure 3.6c) is not surprising, yet since the relationship is not unique, the seasonal signatures add another piece of information. In particular, the phase shift adds a time scale, which quantifies how long – on average – the seasonal input is delayed to become the seasonal output. While the phase shift is only a few days for the most responsive catchments, in the slowest catchments the seasonal signal is shifted up to four months. Since the BFI is rather a consequence of a catchment's hydrological behaviour (as are the seasonal signatures) than an attribute of a catchment, the BFI cannot be seen as a cause for the observed patterns in the seasonal signatures. It cannot be used, for example, as a predictor in ungauged catchments. A qualitative attribute that could theoretically be available in ungauged catchments,

the fraction of highly productive fractured aquifer, reinforces the influence of the subsurface (Figure 3.6d). Except for a few catchments, catchments underlain by such an aquifer exhibit very large phase shifts. In fact, all the catchments above the single reservoir line are underlain by highly productive aquifers. In these catchments, mostly underlain by Chalk, almost all the incoming water infiltrates into the aquifer, and the fast flow component often is negligible. This might explain why they do not behave like reservoirs in parallel, but rather like reservoirs in series, for example, a soil reservoir (recharge) and a very slow groundwater reservoir. The few catchments which are underlain by highly productive aquifers, but do not exhibit large phase shifts, are typically overlain by rather impermeable drift, which stops water from infiltrating into the aquifer below.

Many models frequently used (and some of them developed) in the UK have a parallel flow structure, and catchments are usually conceptualised as having a fast and a slow component. While parameterisations and model structures vary between models, an overall parallel flow structure following a soil moisture module can be found in the PDM model (Moore, 2007), the TOPMODEL modelling concept (consisting of two fast flow responses; Beven and Kirkby, 1979), the IHACRES model (Croke and Jakeman, 2004), the GR4J model (Perrin et al., 2003), and many others. These or similar models have been applied to many catchments in the UK by various authors (e.g. Smith et al., 2019; Lane et al., 2019; Coxon et al., 2019). The seasonal signatures suggest that for most of the catchments, particularly if they are not underlain by a highly productive aquifer, a parallel model structure is a reasonable choice (at least for reproducing the response to seasonal forcing). For some groundwater-dominated catchments, however, the fast flow component seems to be rather unimportant. Many of these catchments, typically catchments underlain by Chalk, could only be poorly modelled in national-scale modelling studies (Smith et al., 2019; Lane et al., 2019; Coxon et al., 2019). While this might partly be due to water balance problems (inter-catchment groundwater flows), it might also be due to an inadequate model structure or inadequate parameter ranges. The most parsimonious reservoir configuration to explain the seasonal behaviour of these catchments (phase shifts > 91 days) would be two reservoirs in series, for example, a soil or unsaturated zone reservoir transforming the incoming forcing into recharge, and a (linear) groundwater reservoir. At least one of these reservoirs would need to be very slow to obtain such large phase shifts (cf. Figure 3.2). For these groundwater-dominated catchments, a serial structure as it is also used in simple lumped groundwater models (e.g. Peters et al., 2003; Obergfell et al., 2019), seems to be a reasonable choice (at least for reproducing the response to seasonal forcing). As mentioned before, two reservoirs in parallel and two reservoirs in series, respectively, might be seen as "special cases" of a soil reservoir followed by a fast and a slow reservoir. For example, some of the catchments underlain by a highly productive aquifer fall in the area that can be simulated by two reservoirs in parallel (see Figure 3.6d). Their large phase shifts and their proximity to the "single reservoir line" suggest, however, that the slow flow component is of particular importance and that large time constants

(> 100 days) are required to model their behaviour.

In summary, the first control on the attenuation of the seasonal signal in the UK is the partitioning between fast flow and slow flow. More saturated catchments partition more rainfall into fast flow and hence lead to a higher amplitude ratio and to a smaller phase shift. The second control are catchment subsurface properties, which determine the available storage and how slowly water leaves the system. The slower the catchment responds, the larger the phase shift and the lower the amplitude ratio. The Chalk catchments in the UK might be seen as an extreme case where almost all the water infiltrates, and hence the response time of a single slow reservoir (or perhaps two reservoirs in series) is the main control on the propagation of a periodic signal. On the other end of the spectrum, there are fully saturated, very responsive catchments mostly along the west coast of the UK, which behave almost like a single fast reservoir. Using conceptual reservoirs is only one way to interpret the seasonal signatures. It is useful as many hydrological models are built in that way. There might be, however, other possible ways of interpretation which we do not consider here.

3.5.3 A hydro-meteorologically more diverse set of catchments – the contiguous US

From Figure 3.7 and Table 3.3 it can be seen that for CAMELS catchments (US) the climate indices explain most of the variability in the seasonal response. Again, more humid catchments tend to create more fast flow, and hence they have high amplitude ratios and small phase shifts. Catchments with a larger moisture index seasonality tend to have smaller phase shifts. In these catchments precipitation and potential evapotranspiration are mostly out of phase. Therefore, precipitation falls in more humid months, which might lead to a flashier response. That means that both precipitation falling on wetter catchments and precipitation falling in wetter months will be attenuated less. The influence of catchment form is much less pronounced than in the climatically more homogeneous UK. Continental or global studies tend to identify climate as the dominant hydrological driving force (Van Dijk, 2010; Beck et al., 2015), yet regional studies often show other attributes such as geology to be important (for baseflow, see e.g. Longobardi and Villani, 2008; Bloomfield et al., 2009). Our findings highlight anew that generalising from global to regional scales, or from regional to global scales, is not straightforward. Such scaling should ideally be done in a process-based way, or by analysing sub-climates, as the dominance of climate might mask the influence of other factors at large scales. We can also see that the attribute "fraction of highly productive fractured aquifers" (Figure 3.6d), which is a hydrogeological classification available for the UK, shows a much clearer pattern than any soil or geology attributes in the US (see e.g. Figure 3.7d which shows the fraction of carbonate sedimentary rock; the same is true if we use for example, soil permeability for the UK). This might partly be due to the more heterogeneous US climate which masks the influence of subsurface properties to some degree. But it might also indicate that the soil or geology data used do not

contain the hydrologically relevant soil or geology information. The hydrogeological classification based on expert judgement available for the UK, even though it is only categorical, might be more representative of the actual hydro(geo-)logical processes at the scale of interest. We therefore cannot conclude that in the US catchment form does not play a role. We can merely say that the catchment attributes used do not show clear patterns at the continental scale.

Some of the rather arid catchments in the US plot outside the area that can be modelled by two reservoirs in series or in parallel (Figure 3.7). This either indicates that we would need another reservoir in series to model the observed phase shift (three reservoirs in series would result in a maximum phase shift of approximately 273 days), that (linear) reservoirs are not a good description of the hydrological processes, or that the proposed signatures are unreliable for these arid catchments. Since in water-limited catchments, actual evapotranspiration is typically much smaller than potential evapotranspiration, the input signal we use is very likely a poor proxy for the actual input signal. In very arid catchments ($I_m < -0.5$, dark red dots in Figure 3.7a), particularly with low moisture seasonality index (Figure 3.7b), the results should therefore be interpreted with care. It is unclear to what extent these large phase shifts are the result of a poorly approximated input signal or actual catchment function. This compromises the consistency (McMillan et al., 2017) of the seasonal signatures and makes them most suitable for energy-limited catchments. A way to overcome this limitation would be the use of modelled or measured actual evapotranspiration as input data. As this would require another modelling step or additional data, we leave this for future work (see Appendix B for further information).

3.5.4 Can two common hydrological models reproduce the observed seasonal signatures?

The ensemble of IHACRES simulations covers the observed range of amplitude ratios and phase shifts, although one catchment sits just at the boundary of the point cloud (Figure 3.8a, Figures 3.9a-c). The BFI pattern also roughly resembles the observed pattern (Figure 3.6c). Catchments with low BFIs tend to have high amplitude ratios and small phase shifts and vice versa. To explain the signature space of IHACRES, it is useful to recall the structure of the model. IHACRES consists of a soil moisture deficit store, followed by two parallel linear reservoirs. It thus approximately features the two examples introduced in Section 3.2, namely two reservoirs in series or in parallel.

If one of the parallel reservoirs in IHACRES receives very little water (due to an extremely high or low fraction p going into the slow reservoir), the whole system acts like two reservoirs in series. The only difference is that the first reservoir is not a single linear reservoir. It is a non-linear deficit store and thus different from the idealised linear reservoir. This might explain why the upper boundary looks similar to the grey dashed line indicating two linear reservoirs in series, yet not exactly the same. We did explore how non-linear reservoirs behave in terms of amplitude ratio and phase shift and they seem to behave similar to linear reservoirs (see

Appendix B). Another reason for IHACRES not covering the whole area might be the parameters ranges (see Appendix B for details). The parameters ranges used are intended to be wide, yet especially the fast reservoir is (to be indeed fast) limited to 10 days, which limits the theoretical space to be smaller than shown in Figure 3.2.

If the soil moisture reservoir transmits water relatively quickly without much attenuation, the whole system acts like two reservoirs in parallel. In summary, IHACRES is very similar to the idealised arrangement we introduced in Section 3.2 and this can be seen in the model output. It is therefore likely that IHACRES is capable of reproducing the observed seasonal signatures for catchments in the UK (Figure 3.6) and for most of the catchments in the US (Figure 3.7). Whether IHACRES can reproduce the seasonal signatures, other hydrological signatures and achieve satisfactory statistical performance metrics simultaneously is to be explored and beyond the scope of this chapter.

The ensemble of GR4J simulations covers most of the amplitude ratios and phase shifts observed in the UK (Figure 3.8b, Figures 3.9d-f). Many of the model runs lead to amplitude ratios higher than unity, which is caused by the groundwater exchange parameter, which allows the model to import water in addition to incoming P. While this is possible (and can in fact be observed; e.g. in Figure 3.6c the blue dot outside the grey boundaries is a catchment with water transfer from a neighbouring catchment), it is observed very rarely in the catchments investigated. Furthermore, a non-zero groundwater exchange parameter should ideally be associated with actual water inputs or outputs (e.g. inter-catchment groundwater flows), and these inputs or outputs are usually unknown. It is worth noting that many model runs that lead to signature values at the boundaries of the signature space (e.g. low BFIs or large phase shifts) are associated with large (positive or negative) values for the groundwater exchange parameter (not shown here). This might further reduce the "realistic" signature space, as, for example, obtaining a low amplitude ratio by removing water might be seen as "the right answer for the wrong reason". No model run leads to a phase shift larger than about 110 days. GR4J also has a soil moisture store followed by two parallel routing stores, that is, the overall model structure is similar to IHACRES. The stores are, however, not linear reservoirs. In addition to that, GR4J has fixed internal parameter values, such as the fraction of water going through the slow routing store, which is set to 0.9. This might explain why the BFI tends to be very high, as it can be seen from Figure 3.9f. Despite the tendency towards large BFIs, GR4J cannot produce phase shifts larger than about 110 days given the parameter ranges used here. This might be due to a too narrow range of the flow delay parameter (maximum 15 d). So, to model both the phase shift and the BFI correctly, we might require a more flexible splitting between fast and slow routing and a means to produce larger phase shifts (e.g. via a wider range for the flow delay parameter).

Figure 3.9 also shows how the seasonal signatures (and the BFI) vary with different input (forcing). For both models, more humid catchments lead to higher amplitude ratios and smaller phase shifts, and vice versa. This trend, not necessarily the values themselves, agrees with the

observed behaviour shown in Figures 3.6a and 3.7a.

This analysis is necessarily incomplete for (at least) two reasons. First, we only looked at 40 catchments in the UK to limit the computational demand. Therefore, the conclusions are not necessarily transferable to catchments outside the UK. More arid catchments (e.g. in the US) might show a different behaviour (e.g. the catchments showing phase shifts larger than 182 days, see Figure 3.7). Second, the sampling scheme (Latin Hypercube sampling) explores only a subspace of the actual parameter values (both because of the parameter ranges and because of the finite amount of parameter sets). We also made an a priori decision of how to sample by choosing Latin Hypercube sampling in the first place. This is inevitably subjective, and other sampling schemes might lead to different results. This might especially affect the distributions of the modelled signatures shown in Figure 3.9. Wider parameter ranges might change the ranges of the resulting signature spaces. As we use rather wide ranges based on recent literature (see Appendix B for details), our results should (at least) be representative of current modelling practice. This kind of analysis and the seasonal signatures can therefore help to select (or not select) models a priori, without calibrating them to streamflow data (cf. Vogel and Sankarasubramanian, 2003). This might be particular helpful for large sample studies where often a certain model structure is chosen a priori, even if it might be inadequate for the catchment sample investigated (Addor and Melsen, 2019).

3.6 Conclusions and outlook

We have tested seasonal hydrological signatures aimed at representing how climate seasonality is translated into streamflow seasonality, both approximated by sine waves. The damping (the amplitude ratio) and the phase shift of the incoming sine wave have been used to quantify how catchments respond to seasonal forcing. The presented signatures follow the guidelines of McMillan et al. (2017). The signatures are identifiable, robust, and consistent (see Appendix B for further information). They are representative and have discriminatory power as they exhibit explicable, hydrologically interpretable patterns, particularly for energy-limited catchments (Figures 3.6 and 3.7). They can be related to conceptual model structures (arrangements of linear reservoirs, Figure 3.5), and the model evaluation (Figure 3.8) has shown that we can indeed observe this theoretical behaviour in model outputs. As we use precipitation minus potential evapotranspiration as a proxy for the input to a catchment, the seasonal signatures are unreliable for water-limited catchments. To use the seasonal signatures in water-limited catchments we would need to estimate actual evapotranspiration. The current approach is therefore only suitable for energy-limited, non-snowy catchments with a distinct unimodal seasonality, such as catchments in the UK.

We have found that the propagation of the seasonal input through a catchment depends both on climate and catchment form. Climate aridity and seasonality, and corresponding annual

and seasonal catchment wetness, drive the partitioning of the incoming forcing into fast and slow flow. Catchment form, such as subsurface properties, influences how strongly the seasonal input gets attenuated. This is particularly visible in the UK, where the hydrogeological classification available (fraction of highly productive aquifer) can explain the very slow response of some catchments. The seemingly more dominant (and less clear) role of climate in the US highlights that scaling from regional to continental (or global) scales is not straightforward and requires thoughtful, ideally process-based approaches. Or in the words of Turner (1989), "conclusions or inferences regarding landscape patterns and processes must be drawn with an acute awareness of scale". Nonetheless, the clear link to climate and aquifer characteristics in the UK suggests that the signatures might be useful for catchment classification and for predictions in ungauged catchments, as long as potential evapotranspiration is an adequate proxy for actual evapotranspiration.

The model evaluation has shown that the signatures have the potential to be used as a diagnostic tool. GR4J could not reproduce the observed combinations of phase shift and BFI, pointing towards structural deficiencies of the model for certain catchments. As the seasonal signatures are relatable to conceptual model structures (arrangements of reservoirs), we could – given sufficient data – also build models based on inference from observed values of the signatures, and not just test existing model structures. This could be done in a stepwise fashion, starting with the seasonal time scale and then adding more complexity if needed (Jothityangkoon et al., 2001; Farmer et al., 2003; McMillan et al., 2011). It would be a step towards model structure identification based on hydrological reasoning, that is, getting the right answers for the right reasons (Kirchner, 2006). If we decide on a certain model structure (e.g. two reservoirs in series), we can then use the presented theory to estimate time constants of the reservoirs (the parameters). This could be used as additional constraint in the calibration process. If the time constants obtained from the seasonal signatures differ from time constants obtained by other means, for example, by calibrating the model using a metric such as KGE, this might be indicative of limitations of typical modelling approaches (Fowler et al., 2018). It might be that the slower annual signal is exciting different parts of the catchments than events (individual peaks or recessions) do, which we typically calibrate to.

The idea of exploring a model's signature space (following the approach of Vogel and Sankarasubramanian, 2003) perhaps deserves more attention. It allows to explore models systematically and it can reveal whether a model can simulate the ranges of hydrological signatures we obtain by analysing catchment data. Similar to sensitivity analysis, it allows us to explore and to better understand how a model works, which parameters are important for which signature, and what output behaviour a model can generate in general – without (and before) calibration. While we limited this analysis to a few signatures, in future studies we should focus on testing whether a model can simultaneously reproduce multiple signatures focusing on different aspects of the hydrological system (Euser et al., 2013; Hrachowitz et al., 2014).

CHAPTER 4

INCLUDING REGIONAL KNOWLEDGE IMPROVES BASEFLOW SIGNATURE PREDICTIONS IN LARGE SAMPLE HYDROLOGY

This chapter has been published as a research article in Water Resources Research. Slight modifications have been made to better fit the general layout of this thesis. This chapter is the result of a research visit to San Diego State University, supervised by Dr. Hilary McMillan. We thank Ryoko Araki for many discussions and Gemma Coxon for help with the extraction of catchment attributes. We acknowledge the comments from the editor, Michael Stoelzle, and one anonymous reviewer, that helped to clarify and improve this chapter.

SJG, HKM, RAW, and NJKH conceptualised the research project. SJG performed the formal analysis. SJG prepared the manuscript with contributions from all co-authors.

Citation: Gnann, S. J., McMillan, H. K., Woods, R. A. and Howden, N. J. K. (2020). Including regional knowledge improves baseflow signature predictions in large sample hydrology. Water Resources Research, 56:e2020WR028354. https://doi.org/10.1029/2020WR028354

4.1 Introduction

A stream reflects the catchment it drains. Its mean discharge is mostly controlled by climatic forcing (Budyko, 1974), and so are many response characteristics at shorter time scales (Berghuijs et al., 2014; Knoben et al., 2018). Yet we see striking differences in the hydrological response from catchments forced by a very similar climate (Farvolden, 1963; Tague and Grant, 2004; Pfister et al., 2017). These differences are typically attributed to differences in a catchment's form, such as the underlying geology (e.g. Price, 2011). Especially the slow response of a catchment (e.g. baseflow, recessions) is thought to carry the signature of the subsurface in which water is stored

and from which it is eventually released.

Many studies could relate baseflow signatures to catchment attributes, such as soils (Boorman et al., 1995; Schneider et al., 2007; Santhi et al., 2008), geology (Farvolden, 1963; Tague and Grant, 2004; Bloomfield et al., 2009; Pfister et al., 2017; Kuentz et al., 2017; Carlier et al., 2018), geology-vegetation groups (Lacey and Grayson, 1998), land use (Zhang and Schilling, 2006), or topography (Santhi et al., 2008). A lot of that knowledge is, however, fragmented and place-specific (Beck et al., 2013). This is reflected in results from recent large sample studies (Beck et al., 2013, 2015; Addor et al., 2018); while climate indices were the dominant predictors of most hydrological signatures, baseflow signatures were harder to predict, and non-climatic catchment attributes (e.g. geology attributes) could not significantly improve these predictions.

So, why is it so difficult to link catchment attributes (catchment form) to hydrological response (catchment function), despite extensive evidence that these attributes are important? We might argue that every place is unique (Beven, 2000) and that synthesising the diversity of catchments around the globe is impossible. There are, however, examples of hydrological similarity (e.g. Budyko, 1974; Berghuijs et al., 2014) which suggest that we can transfer knowledge across places through a comparative hydrology approach (Falkenmark and Chapman, 1989). When we compare many catchments, it is important to balance "depth with breadth" (Gupta et al., 2014), and to acknowledge place-specific processes (uniqueness) within general theories (similarity). Bridging this gap between the local and global scale is not just important for the advancement of our scientific understanding, but also for practical applications that require knowledge at regional scales (e.g. water resources management; Wagener et al., 2010).

The main aim of this chapter is to investigate the following question. Why have non-climatic catchment attributes shown limited explanatory power in recent large sample studies, even for hydrological signatures that are generally thought to be controlled by these catchment attributes (e.g. baseflow index; see Beck et al., 2013, 2015; Addor et al., 2018)? We hypothesise that this is due to limitations in:

(a) the catchment attributes we use to inform our analyses, and

(b) the hydrological signatures we use to describe the hydrological response.

The input data (a), in particular non-climatic catchment attributes, might not reflect the catchment characteristics that are regionally important, thus limiting their explanatory power. This might be because the resolution of the data is too coarse to capture the relevant spatial variability, or because of imperfect upscaling methods (Addor et al., 2018). While some catchment attributes nominally represent soils or geology, they might not represent the relevant hydrological aspects of soils or geology (Beck et al., 2013). As discussed by Addor et al. (2018), sometimes catchment attributes are simply not (yet) available, even though they have shown to be important. Lastly, data uncertainty might complicate a linkage to the hydrological response even if an attribute is theoretically relevant (Beck et al., 2013, 2015; Addor et al., 2018, 2020).

Hydrological signatures (b) that have limited discriminatory power (McMillan et al., 2017), or are highly uncertain (Westerberg et al., 2016), will be difficult to link to catchment attributes and hydrological processes (see also McMillan, 2020). For example, the baseflow index is not only associated with methodological uncertainty, but also with conceptual uncertainty as it lumps together various processes, such as lake outflow, snowmelt, and groundwater discharge (e.g. Parry et al., 2016; Stoelzle et al., 2020). Therefore, it is possible that catchment attributes, even if they were hydrologically relevant, will not be good predictors of such a signature.

To address hypotheses (a) and (b) we review regionally relevant literature which we contrast with information contained in a large sample dataset. We use the CAMELS dataset (Newman et al., 2015; Addor et al., 2017) in our analysis, which consists of several hundred catchments in the contiguous US (for a brief description see Section 4.2.3). The CAMELS dataset has been used in many recent studies (e.g. Addor et al., 2018; Kratzert et al., 2019; Jehn et al., 2020) and we deem it representative of many large sample datasets (for a recent review see Addor et al., 2020).

As a way to better synthesise regionally relevant knowledge, we propose the use of standardised perceptual models of catchment function (see Black, 1997; Wagener et al., 2007). Standardised perceptual models offer a qualitative yet systematic way to communicate our understanding of hydrological systems. We view these perceptual models as a first step to formalise the relationship between catchment attributes and hydrological signatures. Developing a perceptual model of a region might point at datasets worth collecting and allows us to synthesise and communicate soft information (e.g. expert knowledge) in a more systematic way. These perceptual models will evolve continuously and may be updated (or rejected) as we learn about processes and places (see e.g. McGlynn et al., 2002; Shanley et al., 2015). The perceptual model framework is introduced in more detail in Section 4.2.2.

In summary, the aim of this chapter is to demonstrate how limitations in input data and hydrological signatures can obscure relationships between catchment attributes and hydrological signatures. To organise the findings from different regions, we propose a framework based on perceptual models that enables a systematic comparison of attribute-signature relationships.

4.2 Methods and datasets

4.2.1 Literature review and case study regions

We argue that large scale datasets of catchment attributes must reflect deep, region-specific knowledge. Therefore, we selected eight contrasting US regions where an initial literature review has indicated that non-climatic catchment attributes influence the streamflow response (Neff et al., 2005; Zimmer and Gannon, 2018; Tague and Grant, 2004; Adamski et al., 1995; Woodruff and Abbott, 1979; Winter, 1999), shown in Figure 4.1. In each region we explore regionally relevant literature, field knowledge and availability of datasets that characterise this knowledge but that have not previously been used in US-wide approaches such as the CAMELS dataset.

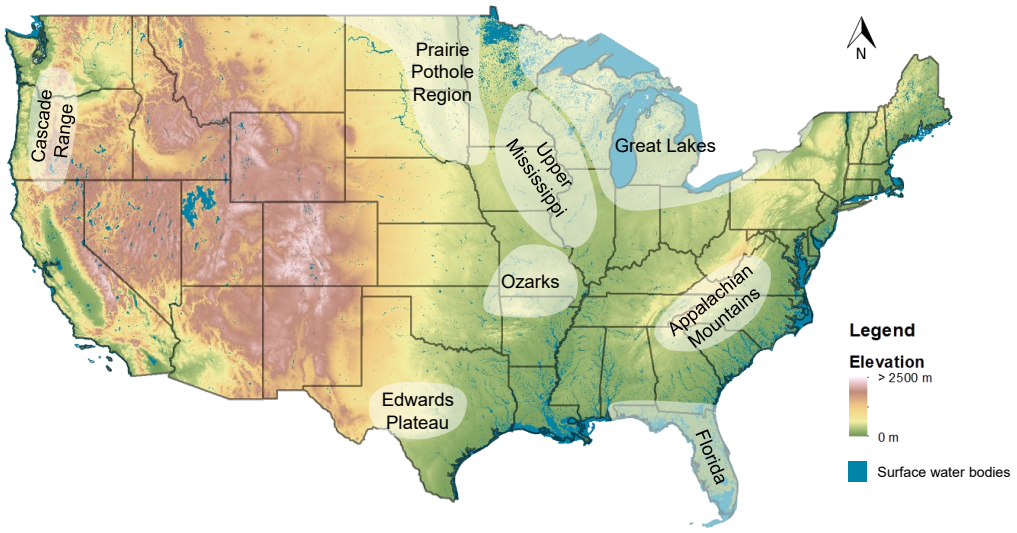

Figure 4.1: Map of the contiguous US indicating the approximate regions of the case studies. Note that some regions might be different to the whole region of the same name (e.g. Appalachian Mountains). The map shows elevations and surface water bodies (data sources are described in Section 4.2.3).

The literature review will be the basis of both our perceptual models (described in Section 4.2.2) and the catchment attributes (described in Section 4.2.3) that are used to better understand several baseflow signatures (described in Section 4.2.4). We found many references that have – to our knowledge – rarely been considered in this context; possibly due to their local or regional scope, because they do not directly stem from hydrology (but from related fields such as geomorphology), or because they are scientific reports rather than journal papers. In particular, reports and datasets from the United States Geological Survey (USGS) or State Agencies contain useful information about the places we investigate here. Figure 4.2 outlines our methodological approach, which is described in more detail in the upcoming sections.

4.2.2 Perceptual models

As a way to formalise the relationship between catchment attributes and hydrological signatures we propose to use standardised perceptual models based on the framework of Wagener et al. (2007). Wagener et al. (2007) distinguish between forcing (incoming water and energy), catchment form (e.g. soils and geology), and catchment function (the actions of the catchment on the incoming water and energy). Catchment functions are further divided into partition, storage, and release. As water is partitioned into different stores, and these stores release water in different ways, partition, storage, and release depend upon each other and cannot be viewed in isolation. Nevertheless, they provide a useful framework to organise our knowledge of catchment hydrological processes. Figure 4.3 shows a general perceptual model that gives an overview of

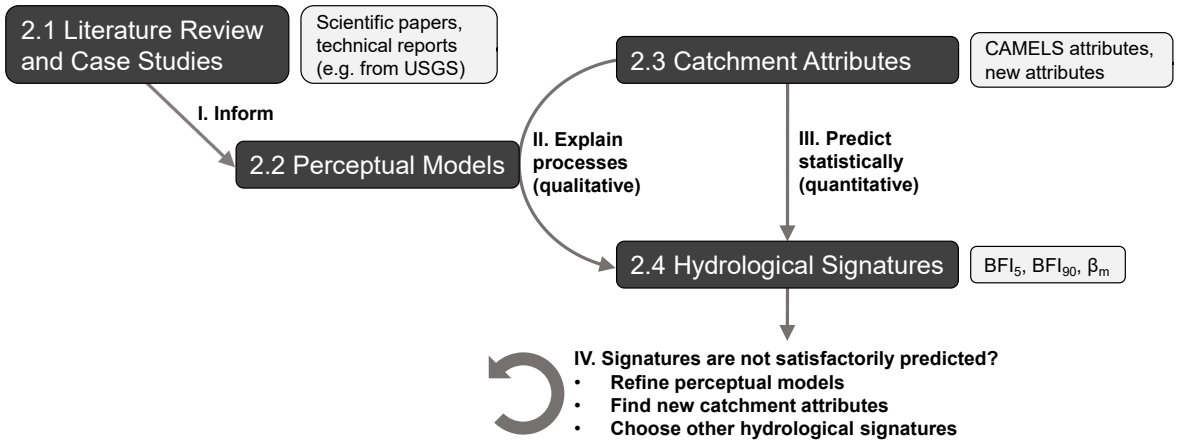

Figure 4.2: Overview of our methodological approach. The boxes correspond to Sections 4.2.1-4.2.4, where the notations are defined. The Roman numerals indicate the order in which the steps are carried out.

the catchment functions we explore in this chapter. This serves as a standard model that is adapted for each of the case studies shown in Figure 4.1 – an approach similar to the concept of hydrological landscapes (Winter, 2001). Drawing from the diagrammatic concepts of Falkenmark and Chapman (1989), we also try to approximately quantify the relative magnitude of the fluxes associated with the different catchment functions (e.g. release in the form of baseflow).

4.2.3 Datasets

4.2.3.1 CAMELS

Hydro-meteorological data, catchment shapefiles, and catchment attributes are obtained from the CAMELS dataset (Newman et al., 2015; Addor et al., 2017). CAMELS includes daily precipitation P, potential evapotranspiration E_p (catchment-averaged forcing data are based on the Daymet dataset, one of three gridded precipitation products used in CAMELS; see Newman et al., 2015) and streamflow data Q, a wide range of catchment attributes, and catchment shapefiles for 671 mostly natural catchments (i.e. minimal land use changes or disturbances, minimal human water withdrawals; Newman et al., 2015) in the contiguous United States. The catchment attributes from CAMELS that are used in this chapter are summarised in Table 4.1.

4.2.3.2 Additional catchment attributes

We use several datasets that are not (yet) contained in CAMELS. They are summarised in Table 4.1. We use these datasets to calculate new catchment attributes which are provided with this chapter. Details on the calculation of catchment attributes can be found in Appendix C.

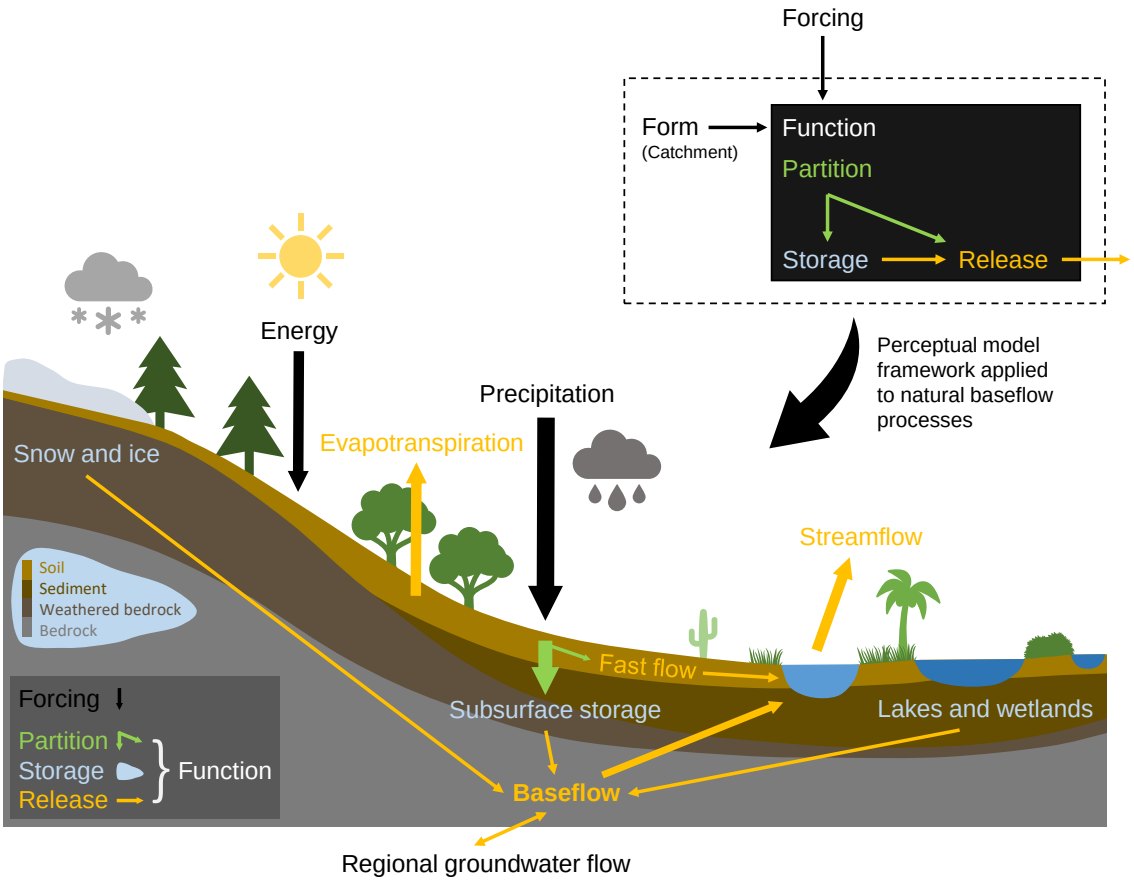

Figure 4.3: Perceptual model framework following Wagener et al. (2007) applied to natural baseflow processes, illustrating the catchment functions that control baseflow generation. The width of the arrows indicates the amount of water partitioned into and released from different stores. Note that this is not intended to represent any real catchment, but to serve as a general overview. We show refined perceptual models for each of the case studies in Section 4.3.

4.2.4 Baseflow signatures

We use three baseflow signatures to characterise the slow response of a catchment: two different baseflow indices (BFIs), and the median recession exponent β_m. These three signatures are correlated, but do provide independent information (see Appendix C for details).

4.2.4.1 Baseflow indices

Baseflow Q_b is defined as the portion of streamflow Q that is derived from groundwater and other delayed sources (Hall, 1968; Smakhtin, 2001). Baseflow is typically quantified by the baseflow index (BFI), the ratio between mean baseflow $\bar{Q}_b$ and mean total streamflow $\bar{Q}$.

$$\text{BFI} = \frac{\bar{Q}_b}{\bar{Q}} \tag{4.1}$$

Table 4.1: Datasets used in this chapter, both for visualisation and analysis. "Datasets in CAMELS" refers to datasets in CAMELS that we use or refer to in this chapter. Links to the datasets are provided in Appendix C.

Dataset name	Attributes	Reference
CAMELS	Hydro-meteorological data Catchment shapefiles Catchment attributes	Newman et al. (2015); Addor et al. (2017)
Datasets in CAMELS		
STATSGO	Soil texture, soil depth	Miller and White (1998)
GLiM	Geological classes	Hartmann and Moosdorf (2012)
GLHYMPS	Geol. permeability, porosity	Gleeson et al. (2014)
Additional datasets		
HydroSHEDS	Digital elevation model	Lehner et al. (2008)
Generalized Glacial Limit Lines	Glacial areas	National Atlas of the United States (2005)
Physiographic Divisions of the US	Physiographic provinces	Fenneman and Johnson (1946)
USGS Geological Map	Geological classes, age	Horton et al. (2017)
Principal Aquifers of the US	Aquifer extents	U.S. Geological Survey (2003)
MGS Sinkhole Points	Sinkhole locations	Missouri Geological Survey (2018)
TWDB Major Aquifers	Major aquifer extents	Texas Water Development Board (2020)
National Wetlands Inventory	Surface water bodies	U.S. Fish and Wildlife Service (2020)

We estimate baseflow with the help of the smoothed minima method (UKIH method; Institute of Hydrology, 1980). The method is particularly sensitive to one parameter, the time window N over which the streamflow minima are calculated (default: $N = 5$ days). To address this problem, Stoelzle et al. (2020) calculated the BFI for a continuous range of time window values. They then used the obtained range of BFIs (which they termed Delayed Flow Index; DFI) to distinguish between different baseflow sources. We follow this idea and calculate two BFIs. A "standard" BFI_5 using a baseflow estimate $Q_{b,5}$ obtained with a time window of 5 days; and a BFI_{90} using a baseflow estimate $Q_{b,90}$ obtained with a time window of 90 days. BFI_5 aims at separating events from inter-event baseflow and BFI_{90} aims at separating seasonal variations from more stable (multi-annual) baseflow. Increasing the value beyond 90 days has relatively little effect on the resulting BFI for most of the catchments analysed here. Note that BFI_{90} is strongly correlated with the normalised 5% flow quantile $Q_5/\bar{Q}$ (Spearman rank correlation $\rho_s = 0.95$).

4.2.4.2 Recession exponent

Recession analysis has been used extensively to quantify the drainage behaviour of catchments (Brutsaert and Nieber, 1977; Roques et al., 2017; Jachens et al., 2020; Tashie et al., 2020). It is often assumed that the relationship between the rate of change of streamflow and streamflow follows a power law.

$$-\frac{dQ}{dt} = \alpha Q^{\beta} \tag{4.2}$$

where α and β_m are parameters that can be obtained by fitting Eq. (4.2) to recession data. There are numerous methodological choices that can impact the resulting parameter values (e.g. Stoelzle et al., 2013; Dralle et al., 2017; Jachens et al., 2020). We extract recession segments

Table 4.2: Overview of catchment functions, corresponding regions, key catchment characteristics, associated hydrological processes, and relevant datasets (see Table 4.1 for details on the datasets). N/A indicates that we did not find suitable datasets. *Datasets contained in CAMELS.

Function	Regions	Catchment characteristics	Hydrological processes	Datasets
Partition	Great Lakes Region, Upper Mississippi Valley	Soil and sediment texture, glacial history	Infiltration, groundwater discharge	STATSGO*, Generalized Glacial Limit Lines
	Appalachian Mountains	Soil stratigraphy	Infiltration	N/A
Storage	Oregon Cascades	Subsurface maturity (volcanic rock)	Groundwater storage	USGS Geological Map
	Ozarks Plateau	Subsurface maturity (carbonate rock)	Groundwater storage	USGS Geological Map, MGS Sinkhole Points
	Edwards Plateau	Weathering characteristics	Groundwater storage	TWDB Major Aquifers
Release	Ozarks Plateau, Edwards Plateau	Losing/gaining streams	Regional groundwater flow	N/A
	Prairie Pothole Region, Florida	Lakes and wetlands	Discharge from surface water bodies	National Wetlands Inventory
	The contiguous US	Baseflow source (e.g. snow)	Snowmelt, discharge from surface water bodies	Snow fraction*, National Wetlands Inventory

that are strictly decreasing ($\frac{dQ}{dt} < 0$), remove the first day, and only keep recession segments of 5 days or longer (Jachens et al., 2020). We calculate the derivative $\frac{dQ}{dt}$ by using the exponential time stepping scheme proposed by Roques et al. (2017). We then use a weighted least square regression approach to fit a line in log-log space to individual recession segments (for details see Roques et al., 2017). We use the median exponent β_m to describe a catchment's average recession behaviour. We do not use the parameter α as it is strongly influenced by seasonal variations in catchment wetness and evapotranspiration (e.g. Dralle et al., 2015; Tashie et al., 2020).

4.2.4.3 Visual inspection of hydrographs

For each region, we show hydrographs to contrast catchments with a different hydrological response. We use the two baseflow estimates $Q_{b,5}$ and $Q_{b,90}$ to divide the hydrograph into fast flow and two baseflow components. Note that while we divide the hydrograph into three parts, the value of BFI_5 "contains" BFI_{90}, that is, it resembles the commonly used BFI (Institute of Hydrology, 1980). These two baseflow components do not necessarily relate to any single baseflow source (or hydrological process), but they are rather meant to emphasise differences in baseflow response between catchments. These hydrographs are complemented by perceptual models, as outlined in Section 4.2.2.

4.3 Results

In Section 4.2.2 we have introduced three catchment functions: partition, storage, and release. In the next sections, we explore the processes that control these functions in the regions shown in Figure 4.1. A summary is given in Table 4.2.

4.3.1 Partition

4.3.1.1 Soil and sediment texture control partitioning: regions covered by glacial deposits

Extensive parts of the north and north eastern US were covered by ice during past glaciations. Glacial erosion and deposition have resulted in thick (tens to hundreds of meters) sediment layers covering the underlying bedrock (e.g. Larson and Schaetzl, 2001). We can distinguish between areas glaciated during the most recent glaciation (Wisconsin) and areas glaciated during earlier glaciations (Pre-Wisconsin; see Figure 4.4a). The border between these two areas (Wisconsin and Pre-Wisconsin) roughly aligns with the border between the Great Lakes Region and the Upper Mississippi Valley (see Figure 4.1). Comparing these two regions shows that soil and sediment texture – rather than bedrock properties – control baseflow generation in glacial regions.

The US part of the Great Lakes Region is dominated by glacial deposits such as till and unconsolidated sediments which often mask the underlying geology (Larson and Schaetzl, 2001). The hydrology of the region is strongly influenced by the composition of soils and sediments (i.e. the soil parent material; Neff et al., 2005; Zhang et al., 2013; Naylor et al., 2016). Soils and sediments in the Great Lakes Region tend to be coarse, particularly in the regions that were located deep within the glaciated area (e.g. Michigan).

While most parts of the Upper Mississippi Valley were glaciated in the past, they were not glaciated during the Wisconsin glaciation (see Figure 4.4a). During this ice-free period, meltwater and precipitation draining via the Upper Mississippi created a fluvial landscape (Bettis et al., 2008) with a more developed surface drainage network than in the Great Lakes Region. Soils and sediments in the Upper Mississippi Valley are finer than in the Great Lakes Region, with larger clay and silt contents and less sand.

Soil and sediment texture are a key control on the hydraulic properties of the subsurface, and thus affect recharge (Naylor et al., 2016) and baseflow (Neff et al., 2005). Sandy soils enable high infiltration rates and thus allow for a lot of recharge. Sandy aquifers provide a lot of groundwater discharge which can sustain continuous baseflow, but also allows for continuous recharge as subsurface saturation is less likely to occur. A sand-rich catchment is illustrated in Figure 4.4d,f which shows a perceptual model and a hydrograph of a typical Great Lakes catchment. Finer soils with higher clay content limit infiltration as well as groundwater discharge, leading to a flashier response. A clay-rich catchment is illustrated in Figure 4.4c,e which shows a perceptual model and a hydrograph of a typical Upper Mississippi Valley catchment. Figure 4.4b shows that clay and sand fraction (STATSGO data contained in CAMELS) are a strong control on the hydrological response in catchments that were glaciated in the past. Since soils are strongly related to their parent material (Naylor et al., 2016), the soil classification will also reflect sediment texture and thus also characterises deeper layers in these regions. Therefore, to predict baseflow signatures across the US, we should include catchment attributes that delineate previous glacial extents. If we want to characterise or model catchments in glacial areas, we should include information

about soils and sediments rather than bedrock.

4.3.1.2 Soil stratigraphy controls partitioning: the Appalachian Mountains in North Carolina

The Appalachian Mountains in North Carolina consist of the Blue Ridge Mountains in the west, which transition into the lower Piedmont in the east (see Figure 4.5a). Both regions are underlain by a relatively old, complex mixture of different lithologies (predominantly metamorphic and classified accordingly in GLiM and thus CAMELS). Soils and bedrock are deep and highly weathered (Zimmer and Gannon, 2018). As the topography transitions from steep (Blue Ridge) to shallow (Piedmont), soils and unconsolidated sediments become thicker. Yet despite having a deeper critical zone, Piedmont catchments generate less baseflow and Zimmer and Gannon (2018) hypothesised that this is due to continuous shallow impeding layers.

In the Piedmont, continuous clay-rich impeding layers can lead to perched water tables and thus to a more flashy response. In the Blue Ridge Mountains, these impeding layers are less continuous and thus allow for more recharge. This is illustrated in Figure 4.5d,f which shows perceptual models for both regions (following Zimmer and Gannon, 2018). The corresponding hydrographs (Figure 4.5b,c) show a similar seasonal $Q_{b,5}$ for both catchments, but the more stable baseflow component $Q_{b,90}$ is almost absent in the Piedmont catchment, indicating a lack of or disconnection from deeper storage. This agrees with Zimmer and Gannon (2018) who found that baseflow amounts in the Blue Ridge are larger and seasonally more stable. The hypothesised dominance of soil stratigraphy over soil texture in this region is supported by the fact that none of the soil textural attributes in CAMELS are strongly correlated with any of the baseflow signatures ($\rho_s(\mathrm{BFI}_5, f_{\mathrm{clay}}) = -0.18$; $\rho_s(\mathrm{BFI}_5, f_{\mathrm{sand}}) = 0.15$).

In-depth regional studies such as Zimmer and Gannon (2018) can help to bridge the gap between the local and continental scale, and they can point out potentially useful datasets such as datasets that describe soil stratigraphy. The importance of soil stratigraphy (e.g. impeding layers) and soil structure (e.g. macropores) has also been highlighted elsewhere (e.g. Price, 2011; Naylor et al., 2016; Fatichi et al., 2020), but there are currently no readily available large scale datasets describing soil stratigraphy.

4.3.2 Storage

4.3.2.1 Subsurface maturity of volcanic rock: the Oregon Cascades

The western slopes of the Oregon Cascades can be divided into two main geological units, the Western Cascades and the High Cascades (Tague and Grant, 2004). While both are underlain primarily by volcanic rock, and classified accordingly in CAMELS, they differ markedly in their appearance and hydrology. The High Cascades consist of young and highly permeable volcanic rock. They have a poorly developed surface drainage system and drain primarily via

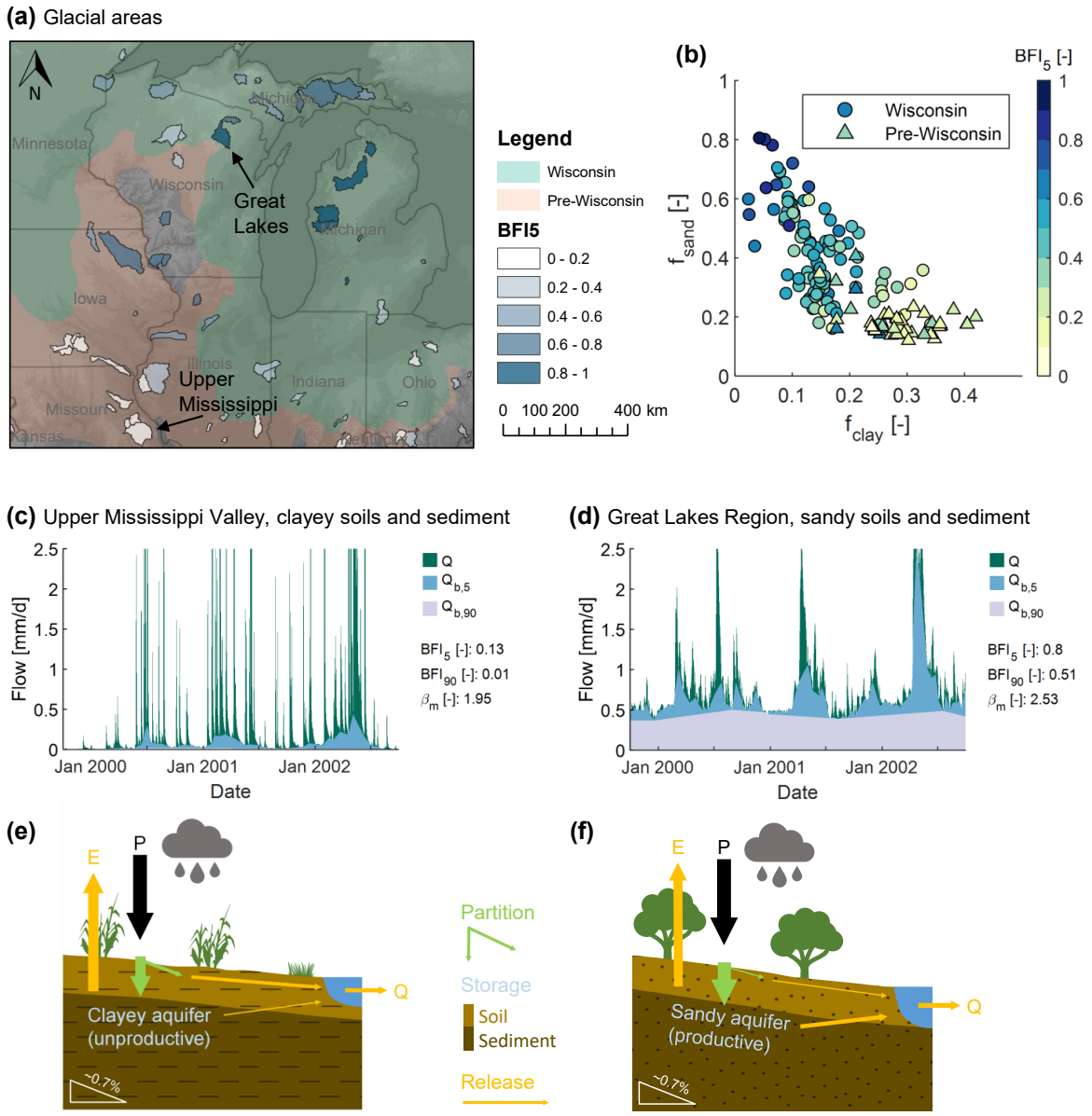

Figure 4.4: **(a)** Map of the glacial areas showing CAMELS catchments coloured according to BFI_5 and two example catchments. **(b)** Scatter plot showing BFI_5 as a function of clay and sand fraction ($\rho_s(BFI_5, f_{clay}) = -0.70$; $\rho_s(BFI_5, f_{sand}) = 0.68$). Hydrographs of the two example catchments with estimated baseflow components for **(c)** Cuivre River near Troy (Upper Mississippi Valley; HU 5514500) and **(d)** Wolf River at Langlade (Great Lakes Region; HU 4074950). Note that the y-axis is capped. Perceptual models for **(e)** catchments with high clay fractions and **(f)** catchments with high sand fractions. The width of the arrows indicates the amount of water relative to a normalised precipitation input.

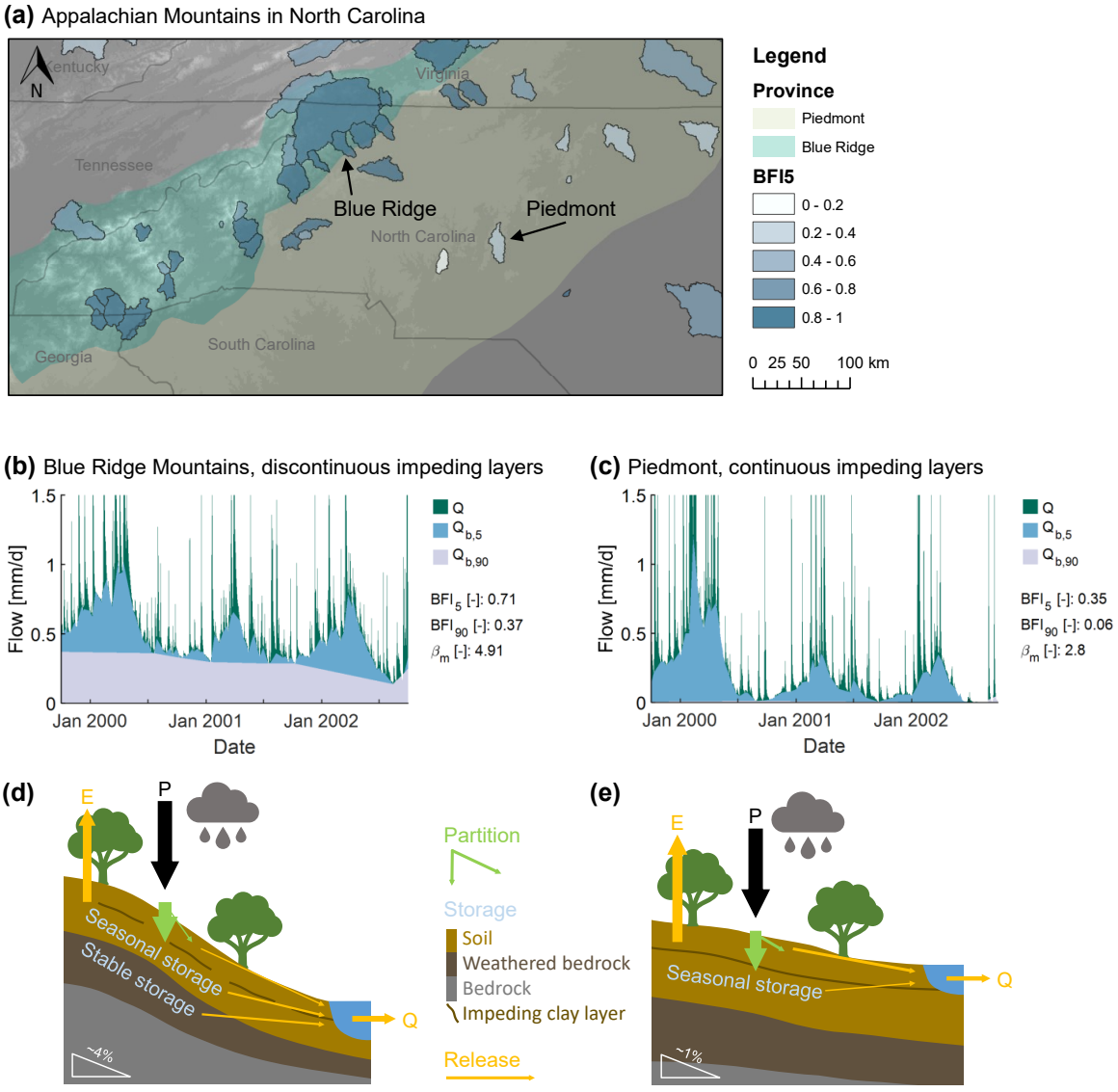

Figure 4.5: **(a)** Map of the Appalachian Mountains in North Carolina divided into physiographic provinces showing CAMELS catchments coloured according to BFI_5 and two example catchments. Hydrographs of the two example catchments with estimated baseflow components for **(b)** Reddies River at North Wilkesboro (Blue Ridge; HU 2111500) and **(c)** Little River near Star (Piedmont; HU 2128000). Note that the y-axis is capped. Perceptual models for **(d)** Blue Ridge catchments and **(e)** Piedmont catchments. The width of the arrows indicates the amount of water relative to a normalised precipitation input.

the subsurface and springs. The Western Cascades are much older and deeply weathered. The landscape is steep, dissected, and there is an extensive surface drainage network fed by shallow subsurface stormflow (Tague and Grant, 2004; Jefferson et al., 2010). The general lithological category (volcanic igneous rock) is therefore not enough to understand the regional hydrology, and we need to understand the geomorphological evolution of the region and the maturity of the subsurface.

The differences between Western and High Cascades are reflected in the hydrology of the streams draining them, with a flashier response in Western Cascade streams and a more damped response with sustained summer low flows in High Cascade streams (Tague and Grant, 2004; Tague et al., 2008; Jefferson et al., 2010). This can be seen in Figure 4.6c-f, which shows perceptual models and hydrographs for two catchments primarily located in either the Western or the High Cascades. Note that both streams show two annual peaks, one in winter when precipitation is highest, and one in late spring due to snowmelt.

We can classify the Oregon Cascades similar to Tague and Grant (2004) by using geological age data contained in the USGS geology map (more details can be found in Appendix C). We classify volcanic (igneous) rocks younger than 2 Ma (million years) as High Cascades, volcanic rocks older than 8 Ma as Western Cascades, and volcanic rocks between 2 Ma and 8 Ma as mixed. The resulting map is shown in Figure 4.6a. Catchments in the High Cascades show higher BFI_{90} values, indicating sustained low flows. To show quantitatively how geological age influences low flows, we extracted the mean age of each catchment's geology from the USGS geology map, which is plotted against BFI_{90} in Figure 4.6b. We also show the corresponding snow fractions to point out that they do not cause the differences in BFI_{90}. While the overall sample size is small ($n = 12$), particularly for the High Cascades, our results agree with many other studies (e.g. Tague and Grant, 2004; Tague et al., 2008; Jefferson et al., 2010; Safeeq et al., 2013). This shows that a simple classification as volcanic rock is insufficient to characterise these catchments, but that accounting for the maturity of the landscape by means of geological age data can help to better link catchment geological attributes to baseflow signatures.

4.3.2.2 Subsurface maturity of carbonate rock: the Ozarks

The Ozarks are located primarily in Missouri, with smaller parts in Arkansas, Kansas, and Oklahoma. The Ozarks are underlain by different types of carbonate and other sedimentary rock (Adamski et al., 1995), and they are classified primarily as carbonate rock in CAMELS. Literature about the Ozarks shows, however, that the region consists of different carbonatic units which differ in their age, composition, and degree of karstification, and thus their hydrology (Harvey, 1981; Adamski et al., 1995; Hays et al., 2016). To differentiate between the different aquifer units we make again use of the geological age data from the USGS geology map. We can divide the Ozark Plateaus aquifer system (delineated from the USGS Aquifer Map) into two units, one being older than 360 Ma (the end of the Devonian, roughly resembling the Ozark aquifer) and

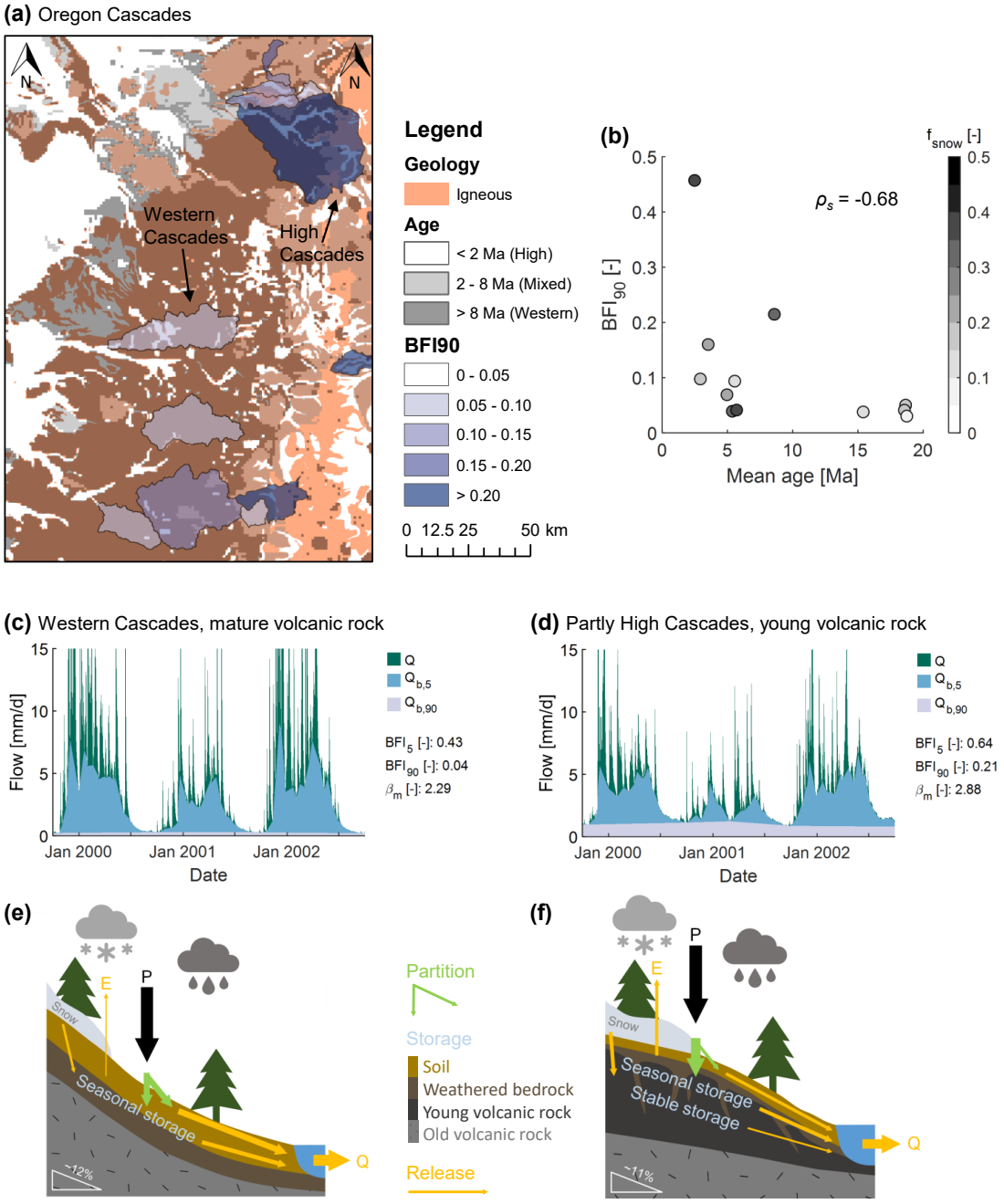

Figure 4.6: **(a)** Map of the Oregon Cascades showing CAMELS catchments coloured according to BFI_{90} and two example catchments. Areas composed of igneous rock are overlain by shades of grey indicating geological age. **(b)** Scatter plot showing BFI_{90} vs. mean geological age ($\rho_s = -0.68$) with dots coloured according to the snow fraction f_{snow}. Hydrographs of the two example catchments with estimated baseflow components for **(c)** Quartzville Creek near Cascadia (HU 14185900) and **(d)** Sandy River near Marmot (HU 14137000). Note that the y-axis is capped. Perceptual models for **(e)** Western Cascade catchments and **(f)** High Cascades catchments. The width of the arrows indicates the amount of water relative to a normalised precipitation input.

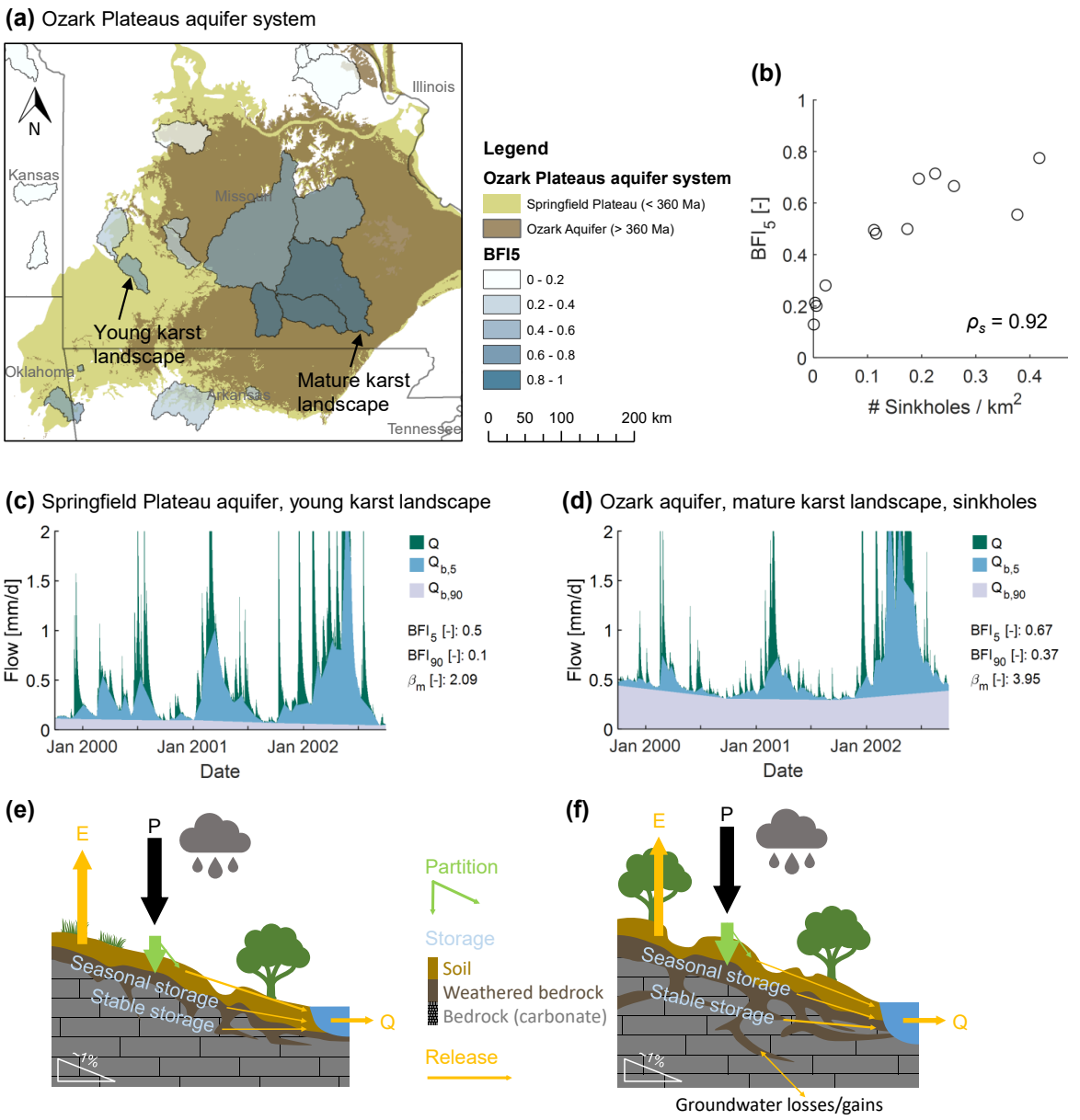

Figure 4.7: **(a)** Map of the Ozarks showing CAMELS catchments coloured according to BFI_5 and two example catchments. **(b)** Scatter plot showing BFI_5 vs. sinkhole density ($\rho_s = 0.92$). Hydrographs of the two example catchments with estimated baseflow components for **(c)** Turnback Creek above Greenfield (HU 6918460) and **(d)** Current River at Van Buren (HU 7067000). Note that the y-axis is capped. Perceptual models for **(e)** Springfield Plateau catchments and **(f)** Ozark aquifer catchments. The width of the arrows indicates the amount of water relative to a normalised precipitation input.

one being younger than 360 Ma (roughly resembling the Springfield Plateau aquifer; Adamski et al., 1995; Hays et al., 2016), shown in Figure 4.7a.

Catchments inside the aquifer system (coloured areas in Figure 4.7a) generate more baseflow than catchments outside the aquifer system. Within the aquifer system, catchments underlain by the Ozark aquifer (the dark brown area in Figure 4.7a) generate the highest amounts of baseflow. This agrees with other studies which state that the dissolution of rocks and hence the degree of karstification is greater in the Ozark aquifer than in the Springfield Plateau aquifer (Harvey, 1981; Adamski et al., 1995; Hays et al., 2016). This difference is illustrated in Figure 4.7c-f, which shows hydrographs and perceptual models for two catchments underlain by the Springfield Plateau aquifer and the Ozark aquifer, respectively. The catchment underlain by the Ozark aquifer (Figure 4.7d,f) has a more stable baseflow component stemming from an extensive subsurface flow network. Figure 4.7f indicates another typical karst feature, namely groundwater flow between (surface) catchments. This is common in the Ozarks (Kleeschulte, 2000; Mugel et al., 2009) and will be discussed in Section 4.3.3.1.

Distinguishing between the different aquifer units allows us to better explain the hydrological response in this area. But we can go a step further by looking at typical features of mature karst landscapes such as springs and sinkholes (Harvey, 1981; Adamski et al., 1995). To assess the degree of karstification we extracted the number of sinkholes per catchment from a map of the Missouri Geological Survey. Figure 4.7b shows that sinkhole density strongly correlates with BFI_5 for catchments in the Ozarks in Missouri. Sinkholes are therefore a useful and measurable surface feature that indicate subsurface maturity, which might be particularly useful in ungauged catchments. However, while other sinkhole datasets exist (e.g. for Florida), limited availability of good quality sinkhole data might limit this approach to certain regions (here Missouri).

4.3.2.3 Erosion of rocks with different weathering characteristics: the Edwards Plateau

The Edwards Plateau region in central Texas can be divided into the Edwards Plateau proper and the Texas Hill Country (Wilcox et al., 2007). They are bounded to the south-east by the Balcones Fault Zone which gave rise to high relief and has resulted in a complex geological structure. These regions roughly align with the aquifers of the Edwards-Trinity aquifer system obtained from the Texas Water Development Board (TWDB), which are shown in Figure 4.8a. The Edwards-Trinity aquifer is the principal aquifer in the Edwards Plateau, the Trinity aquifer is the principal aquifer in the Hill Country, and the Edwards aquifer is the principal aquifer in the Balcones Fault Zone (Barker and Ardis, 1996). The regional climatic gradient (more humid in the east), differences in relief (higher in the east), as well as regional groundwater flows towards the east, have led to increased erosion towards the east, resulting in the dissected landscape of the Texas Hill country (Woodruff and Abbott, 1979; Barker and Ardis, 1996), shown in Figure 4.8a. This hydrogeological diversity is not reflected in CAMELS, which classifies the whole region

primarily as carbonate rock.

The Edwards-Trinity aquifer provides baseflow even during periods with little rainfall. This is illustrated in Figure 4.8c,e which shows a hydrograph and a perceptual model for a catchment in the Edwards Plateau proper. In the Texas Hill country, the upper parts of the Edwards-Trinity aquifer have been eroded, exposing the Glen Rose formation which consists of a sequence of limestone and dolomitic beds with varying weathering potentials (Wilcox et al., 2007; Woodruff and Wilding, 2008). This leads to a stepped topography consisting of steep risers and flat treads. Wilcox et al. (2007) and Woodruff and Wilding (2008) showed that the steep risers have deeper soils and weathered regolith and thus act as stores and zones of subsurface flow, whereas the treads create more fast flow. This is illustrated in Figure 4.8d,f which shows a hydrograph and a perceptual model for a catchment in the Texas Hill Country. Storage in the steep risers only provides intermittent baseflow, leading to an ephemeral flow regime.

The difference between the Edwards Plateau proper and the Texas Hill Country can be shown more quantitatively when the catchment fraction underlain by the Edwards-Trinity aquifer (delineated from the TWDB aquifer map) is plotted against BFI_{90} (Figure 4.8b). Catchments outside the Edwards-Trinity aquifer have low to zero BFI_{90}, whereas most catchments underlain by the Edwards-Trinity aquifer have a high BFI_{90}. A few catchments that have a very low BFI_{90} also have a particularly low runoff ratio (indicated by light colours in Figure 4.8b), likely because they lose water in the Balcones Fault Zone. The Balcones Fault Zone acts as a major recharge zone for the confined aquifer in the south (Woodruff and Abbott, 1979; Schaller and Fan, 2009), which might explain the low BFI_{90} values of some catchments that extend into it (see Figure 4.8a). We therefore also need to account for groundwater losses and gains, which is discussed in Section 4.3.3.1. While the aquifer map of Texas contains useful information, it is also unique to the region and needs to be interpreted with the help of regional knowledge. A next step would therefore be the integration of this knowledge into a more widely applicable classification (see discussion in Section 4.4.4).

4.3.3 Release

4.3.3.1 Losing and gaining catchments: the Ozarks and the Edwards Plateau

Catchments are often regarded as closed systems, where incoming water leaves either via evapotranspiration or stream discharge. Groundwater discharge from or to neighbouring (topographic) catchments is, however, common (Schaller and Fan, 2009; Fan, 2019). This is especially true for karst landscapes, such as the Ozarks Plateau (Kleeschulte, 2000; Mugel et al., 2009) or the Edwards Plateau (Woodruff and Abbott, 1979; Schaller and Fan, 2009). Since groundwater losses and gains can affect baseflow signatures (see Figure 4.8b), we tried to estimate regional groundwater flows via the water balance (see Schaller and Fan, 2009) using actual evapotranspiration estimates from two different products: MODIS (Mu et al., 2011) and GLEAM (Miralles et al., 2011; Martens et al., 2017); details can be found in Appendix C. We did not use the resulting

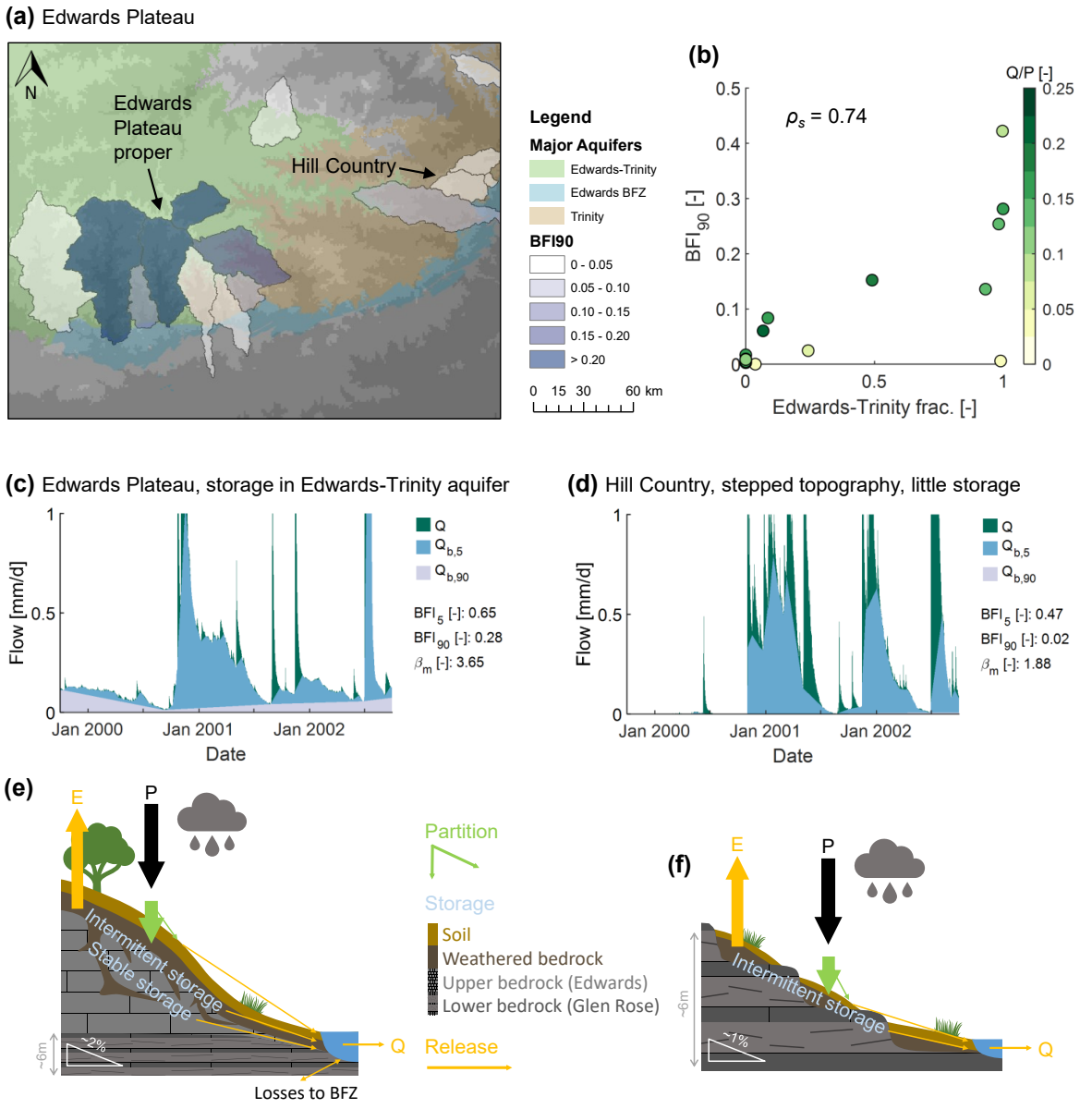

Figure 4.8: **(a)** Map of outcrop areas of the Edwards-Trinity aquifer system showing CAMELS catchments coloured according to BFI_{90} and two example catchments. **(b)** Scatter plot showing BFI_{90} vs. Edwards-Trinity fraction (the green area in **(a)**; $\rho_s = 0.74$) with dots coloured according to the runoff ratio Q/P. Hydrographs of the two example catchments with estimated baseflow components for **(c)** Frio River at Concan (HU 8195000) and **(d)** Onion Creek near Driftwood (HU 8158700). Note that the y-axis is capped. Perceptual models for **(e)** Edwards Plateau catchments and **(f)** Texas Hill Country catchments. The width of the arrows indicates the amount of water relative to a normalised precipitation input.

estimates as they do not conclusively agree with information on losing and gaining catchments we found in the literature (e.g. Kleeschulte, 2000; Mugel et al., 2009, for the Ozarks), likely due to uncertainty in all water balance components (see e.g. Khan et al., 2018, for actual evapotranspiration). Instead we note that it will be important to obtain reliable estimates of regional groundwater flow to better understand baseflow signatures.

4.3.3.2 Lakes and wetlands: the Prairie Pothole Region and Florida

Lakes and wetlands are important functional units of hydrological systems. There is currently no dataset that explicitly describes surface water bodies in CAMELS (there is only a soil attribute named "water fraction"). If baseflow originates from surface water bodies, subsurface characteristics alone cannot explain the baseflow response. We explore two regions, the Prairie Pothole Region and the state of Florida, both shaped by their surface water bodies yet located in different climate zones. Both regions show a similar and distinct combination of baseflow signatures which reflects wetland connectivity.

The Prairie Pothole Region was formed by the last glaciation and the region (shown in Figure 4.1) aligns well with the boundaries of the Wisconsin glaciation (shown in Figure 4.4). Potholes provide storage that buffers against floods and provides baseflow, usually in connection with the shallow groundwater system (Winter, 1999; McLaughlin et al., 2014; Cohen et al., 2016; Ameli and Creed, 2017; Neff and Rosenberry, 2018). Fast surface connections occur only during large events and originate from wetlands near the stream. Slow subsurface connections originate from wetlands throughout the catchment, including geographically isolated ones (McLaughlin et al., 2014; Ameli and Creed, 2017). A perceptual model depicting the hydrology of the Prairie Pothole Region is shown in Figure 4.9c. The corresponding hydrograph shown in Figure 4.9a lacks a very fast response, illustrating the flood buffering effect of potholes. Baseflow is substantial but intermittent, which is indicated by a moderate BFI_5 and very low BFI_{90}. Recession exponents β_m close to 1 – the lowest of all CAMELS catchments – indicate fast late recessions, reaffirming the intermittent nature of baseflow in this region. Wetland connectivity decreases during drying (both due to evapotranspiration and discharge), as deeper layers tend to be less permeable (Cohen et al., 2016), and hence the flow ceases once the water levels have dropped below permeable layers (fill and spill; Cohen et al., 2016).

Florida is underlain by the Floridan aquifer system, a carbonate rock aquifer system that is confined by a clay rich layer in most places (Schiffer, 1998). This confining layer is overlain by unconsolidated sediments which make up the surficial aquifer system. Many lakes have developed from sinkholes, which mostly occur in places where thin or discontinuous sediment and clay layers expose the underlying carbonate rock. If the confining clay layer is intact, the Floridan aquifer system has limited influence on streams. This is the case for most of the CAMELS catchments in Florida, which lie almost exclusively in areas with thick sediment cover. In these catchments, hydrological connectivity is closely linked to the shallow aquifer system and depends

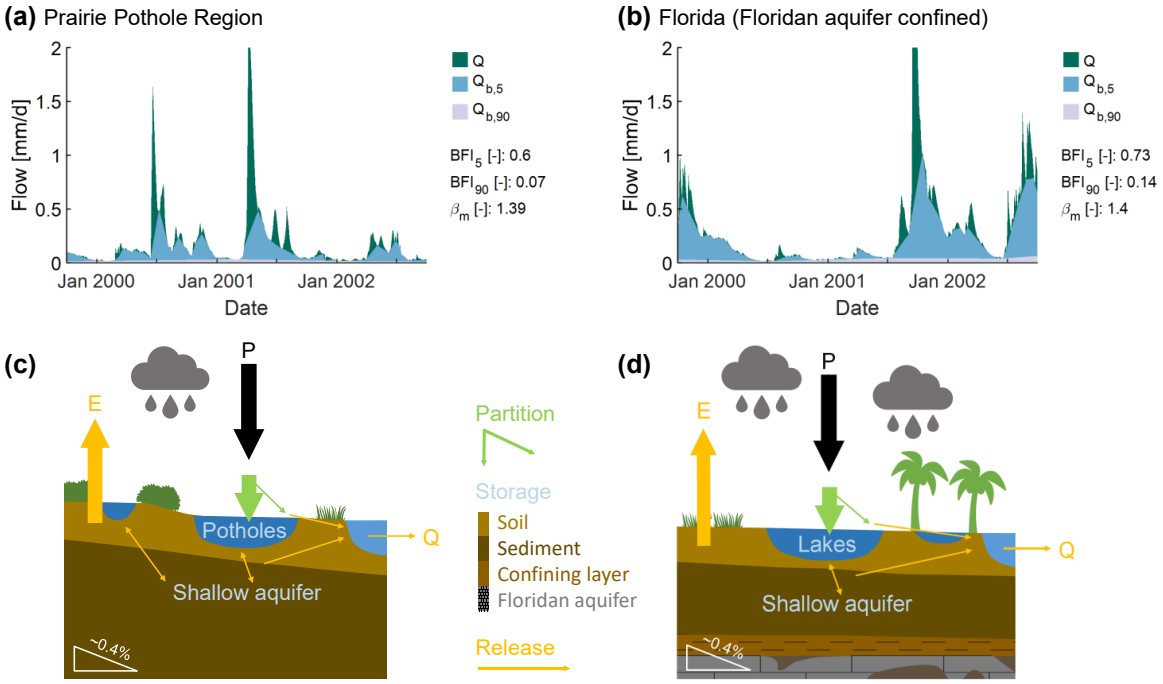

Figure 4.9: Hydrographs with estimated baseflow components for **(a)** Sheyenne River near Cooperstown, North Dakota (HU 5057000), and **(b)** Blackwater Creek near Cassia, Florida (HU 2235200). Note that the y-axis is capped. Perceptual models for **(c)** Prairie Pothole catchments and **(d)** catchments in Florida. The width of the arrows indicates the amount of water relative to a normalised precipitation input.

on the thickness and hydraulic properties of soils and sediments (Schiffer, 1998; Winter, 1999). A perceptual model of such a catchment is shown in Figure 4.9d. Similar to the Prairie Pothole Regions, the corresponding hydrograph (Figure 4.9b) lacks a very fast response and baseflow is substantial but intermittent.

As lakes can have a strong impact on the hydrological response of a catchment, we need to include information on surface water bodies in large sample datasets (see also Beck et al., 2013). In the next Section 4.3.3.3, we show that the fraction covered by surface water bodies (derived from the National Wetlands Inventory; U.S. Fish and Wildlife Service, 2020) can be used to distinguish between hydrologically different catchment groups (e.g. surface water dominated). But it is likely that more detailed information about wetland type and wetland geographic distribution will help to better understand baseflow signatures in catchments influenced by surface water bodies.

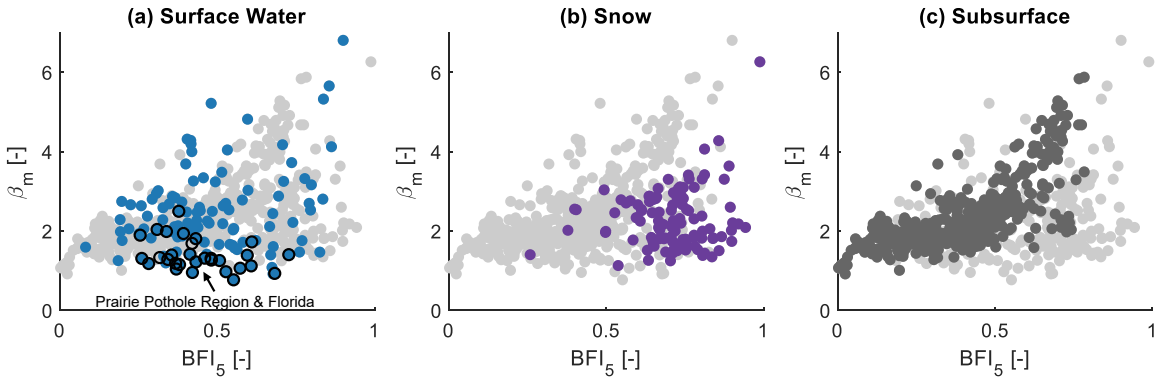

Figure 4.10: Scatter plots of median recession exponent β_m vs. BFI_5 ($\rho_s = 0.42$ for all catchments). Subplots show catchments where water is primarily stored in **(a)** in surface water bodies (>1% of area classified as lake or wetland delineated from the National Wetlands Inventory; $\rho_s = 0.15$ for the subgroup); **(b)** as snow (>30% precipitation falling as snow; $\rho_s = 0.07$); and **(c)** in the subsurface ($\rho_s = 0.72$). Note that each catchment only belongs to one class, with surface water bodies being the first criterion and snow being the second criterion. Note that the y-axis is capped. Similar plots for other signature combinations are shown in Appendix C.

4.3.3.3 Release characteristics of different baseflow sources: surface water bodies, snow, and the subsurface

Baseflow can originate from different sources, but a single signature such as BFI_5 cannot distinguish between these different sources. For example, substantial amounts of baseflow indicated by a moderate BFI_5 can be found in many regions (e.g. Oregon Cascades, Edwards Plateau, Prairie Pothole Region, Florida). But a moderate BFI_5 in conjunction with fast release dynamics indicated by a very low β_m is very typical for the surface water dominated catchments of the Prairie Pothole Region and Florida (see Section 4.3.3.2). If a catchment attribute (e.g. rock type) is important for one but unimportant for another baseflow source (e.g. groundwater storage and wetland storage), it might be difficult to link that attribute to a single signature such as BFI_5. We therefore explored the relationship between two signatures, BFI_5 and β_m, for different baseflow sources. We can divide the CAMELS catchments into three groups (McDonnell and Woods, 2004); catchments where water is primarily stored (a) in surface water bodies, (b) as snow, and (c) in the subsurface. To visualise how baseflow release dynamics are related to the amount of baseflow released, we plot the median recession exponent β_m against BFI_5, shown in Figure 4.10.

While many catchments in the Prairie Pothole Region and Florida show a similar combination of BFI_5 and β_m, there is no clear pattern for surface water dominated catchments in general (Figure 4.10a). The fact that BFI_5 and β_m form an uncorrelated point cloud shows that similar amounts of baseflow can be associated with very different baseflow dynamics and hence with different hydrological processes. Lakes and wetlands interact with local groundwater systems and are strongly influenced by seasonal climate and vegetation dynamics (Winter, 1999). Therefore,

we will need to better understand these complex, typically regional processes to understand the relationship between surface water bodies and baseflow beyond the case studies shown here.

Snow dominated catchments (Figure 4.10b) form a relatively distinct point cloud with high BFI_5 values and comparatively low β_m values. This is probably a consequence of the seasonal nature of snowmelt, which only provides baseflow for a few months in spring and summer. For example, catchments in the High Cascades (Figure 4.6d) show lower β_m values than catchments in regions with similarly significant subsurface storage such as the Ozarks (Figure 4.7d). As the partitioning of snowmelt will also depend on the subsurface, understanding baseflow processes in snow dominated regions requires the inclusion of both snow and groundwater processes (e.g. Tague and Grant, 2004; Safeeq et al., 2013).

In catchments where water is primarily stored in the subsurface, BFI_5 and β_m are strongly correlated (Figure 4.10c). High baseflow amounts (high BFI_5) are mostly associated with slow late recessions (high β_m), that is, stable low flows. This can be seen in many of our case studies, such as the Great Lakes Region (Figure 4.4d), the Appalachian Mountains (Figure 4.5c), or the Ozarks (Figure 4.7d). The remaining variability indicates that also for this subgroup, similar amounts of baseflow can be associated with different baseflow release dynamics, possibly related to different geological settings.

4.4 Discussion

4.4.1 Region-specific knowledge is underutilised in large sample studies

Large scale catchment attributes often do not reflect region-specific hydro(geo-)logical knowledge. But a wealth of – currently underutilised – region-specific qualitative and quantitative information exists and it can help us to better understand the link between catchment attributes and baseflow processes. The case studies shown here are not limited to single catchments, but often describe states or larger regions. This suggests that a better characterisation of both surface and subsurface properties will also improve our understanding at continental and global scales. Finding this information requires a creative and open search, including journal articles from related fields (e.g. geomorphology), articles from regional journals, grey literature such as technical reports from agencies (e.g. USGS), as well as communication with experts. While these additional information sources come with limitations such as a lack of external review, they proved very useful and – based on our judgement – are often of similar quality as externally reviewed academic literature. Synthesising and sharing this information requires a systematic approach, and here we have proposed and applied a framework based on standardised perceptual models.

Standardised perceptual models offer a means to formalise the relationship between catchment attributes and hydrological signatures. They have the advantage that they allow us to share qualitative or place-specific information in a systematic way (see Wagener et al., 2020).

We can use perceptual models to state explicitly how we think a system works, and this can then be developed into a testable hypothesis (c.f. Winter, 2001). If a postulated relationship between a hydrological signature and a catchment attribute is not supported by data, we can either reject (or revise) our perceptual model, or try to find other, more relevant data or updated, potentially improved datasets (see Figure 4.2). Of course, perceptual models are (by definition) subjective and some disagreement will be inevitable. But disagreement can be a useful starting point for progress, and the continuous refinement (or rejection) of these models should be seen as a learning process about processes and places (c.f. Beven, 2007).

4.4.2 Multiple baseflow signatures are needed to distinguish between different baseflow sources

Baseflow is typically defined as the portion of streamflow that is derived from groundwater and other delayed sources (Hall, 1968; Smakhtin, 2001). But baseflow signatures such as the BFI are often used without explicitly linking them to different baseflow sources. This is problematic as transferring information in both space and time requires knowledge about the processes that generate baseflow. For example, if we want to assess the impact of warmer temperatures on baseflow, we need to understand how that affects both snow and groundwater processes (e.g. Safeeq et al., 2013). Figure 4.10 shows how different sources of baseflow can lead to very different dynamics, even if the estimated amount of baseflow (quantified by BFI_5) is the same. In many catchments, the stable baseflow component BFI_{90} shows a much clearer link to geological characteristics than BFI_5 (e.g. in the Oregon Cascades, see Figure 4.6). The combination of different signatures as well as meaningful subgroups can help us to explicitly link baseflow signatures to hydrological processes. This might also help us to identify relationships between baseflow signatures and geology that are otherwise hidden.

4.4.3 Limitations: data uncertainty and hydrological signature selection

An advantage of large sample hydrology is that regional patterns make it less likely to draw wrong conclusions based on a few anomalous catchments (Gupta et al., 2014). At the same time, data errors can hide patterns if a hydrological signature is sensitive to these errors (Westerberg and McMillan, 2015). This applies both to catchment attributes (Addor et al., 2018, 2020) and hydro-meteorological data (Westerberg and McMillan, 2015). For example, regional groundwater flow can affect hydrological signatures (e.g. Figure 4.8b). But uncertainty in all hydro-meteorological data, particularly in actual evapotranspiration, makes it very difficult to quantify this effect. This substantiates the need for uncertainty estimates which large sample datasets often lack (c.f. Addor et al., 2020).

We have limited our analysis to three signatures: BFI_5, BFI_{90} and β_m. This is just one possible set of signatures and they will not capture the whole range of baseflow processes. For example, a wider range of BFI values as suggested by Stoelzle et al. (2020) might lead to a more refined

characterisation of the slow response of different catchments. Furthermore, analysing seasonal differences in both baseflow and recession behaviour might reveal more about the influence of climatic and topographic boundary conditions on the storage-discharge relationship (e.g. Zimmer and Gannon, 2018; Tashie et al., 2019). The baseflow estimation and the recession analysis are also associated with methodological uncertainty (e.g. Stoelzle et al., 2013; Dralle et al., 2017). We did not perform an extensive comparison of different signature calculation methods, but we compared the signature calculation methods used here with a few alternative methods (Lyne and Hollick, 1979; Brutsaert and Nieber, 1977); details can be found in Appendix C.

4.4.4 Next steps

4.4.4.1 Viewing catchments as systems with a history

We have seen many examples where the geomorphological history of a region does not just give us a glimpse into why a place is like it is, but also provides useful information that is hard to observe directly. The volcanic Cascades evolve from being almost entirely groundwater dominated towards having an efficient surface drainage network (Jefferson et al., 2010). The carbonatic Ozarks evolve in the other direction, as the self-perpetuating dissolution of carbonate rock leads to an increasingly efficient subsurface drainage network (Adamski et al., 1995; Hartmann et al., 2014). The Edwards Plateau might be placed somewhere in between. There is an extensive karst network below the ground, yet at the same time surface erosion has carved an extensive surface drainage network into the landscape (Woodruff and Abbott, 1979). In glacial areas, we can see the imprint of the glacial history in form of sediment composition, but also in form of fluvial erosion induced by glacial meltwater (e.g. Upper Mississippi). The hydrology of the Appalachian Mountains can be better understood by understanding the evolution and thus the architecture of their critical zone (Zimmer and Gannon, 2018). Whether these results are transferable remains to be explored. But we renew the argument that by viewing catchments as systems with a history we might be able to learn more about their present state, and perhaps about how they will evolve in the future (Harman and Troch, 2014; Troch et al., 2015). This does not necessarily imply a long history of co-evolution, as the history of a catchment can be shaped by events (faulting, glaciation; see e.g. Beven, 2015) and more recently increasingly by humans (Wagener et al., 2010)

4.4.4.2 Challenges for a geological classification at the continental scale

We have shown examples where a better characterisation of geological characteristics allows us to better explain the hydrological response at the regional scale. When extending this approach to larger scales, we will face several challenges. First, we need to merge the diverse regional classifications into a coherent framework that reflects this diversity while being general enough to be useful. Second, we need to translate qualitative information such as rock type into quantitative hydrological properties or indices. Third, we need to account more explicitly for different climatic

conditions as both long-term and short-term climatic conditions vary. For example, seasonal variability can affect baseflow (Zimmer and Gannon, 2018) and recessions (Tashie et al., 2019), and thus complicate the linkage between static catchment attributes and hydrological signatures. Similarly, differences in topography can affect recharge and hydraulic gradients, and this can alter the hydrological response even if the hydraulic properties of the subsurface stay the same (Carlier et al., 2019). At the same time, topography is related to hydrologically relevant properties of the subsurface itself (e.g. fractures; St. Clair et al., 2015; Prancevic and Kirchner, 2019). Disentangling these different, potentially co-varying processes is challenging (Price, 2011), but we will have to explicitly address them if we aim at a geological classification at the continental scale.

4.4.4.3 How much regional information do we need to predict baseflow response at the continental scale?

Our results suggest that the amount of regional information required to arrive at acceptable continental scale predictions depends both on the spatial scale and on the regions covered. We started by delineating different regions which typically covered large fractions of a state and sometimes multiple states ($\approx 10^4$–10^5 km^2). In some regions, a single attribute that characterises the subsurface can explain most of the variability in baseflow response (e.g. sinkhole density in the Ozarks, see Figure 4.7b). In other regions, more information is required, especially if baseflow originates from multiple sources (e.g. wetlands and groundwater, see Section 4.3.3.3). Continental scale predictions will require attributes that characterise all sub-regions (even though some of the attributes might only be used for some regions).

One way to approximately specify the necessary level of detail for each region would be a simple classification of the main components of our hydrological system, that is, an initial perceptual model. We might start with the three groups presented in Section 4.3.3.3 and distinguish between water that is stored in surface water bodies, as snow, and in the subsurface (McDonnell and Woods, 2004). If water is primarily stored in the subsurface, we might then further distinguish between storage in soils, sediment layers, weathered bedrock, and so forth. Such a classification could be informed by using previous glacial extents (see Section 4.3.1.1) or by a geomorphological classification (e.g. an upland vs. lowland classification, see Pelletier et al., 2016).

4.4.4.4 How can our results help to understand and predict change?

In this chapter we have focused on understanding current baseflow response in mostly natural catchments. This is a crucial first step, but ultimately we are also interested in understanding and predicting the hydrological response under change. If we better understand the drivers of baseflow generation, we can use this understanding to assess how these individual drivers and the corresponding attributes respond to change, for example, when forced by a different climate. Some attributes will be directly impacted by change (e.g. wetland extent, snow cover). Other

attributes are mostly static themselves (e.g. geological attributes), but their interaction with climatic forcing controls key hydrological processes (e.g. groundwater storage). Human impacts can be an additional driver of baseflow response and might be assessed by including attributes that characterise human interventions (e.g. land use changes; Zhang and Schilling, 2006).

Models that credibly predict change need to adequately represent the dominant hydrological processes and ideally both model structure and model parameters should be informed by process understanding rather than calibration (Sivapalan, 2005; Kirchner, 2006; Clark et al., 2017). By linking baseflow response to catchment attributes via perceptual models, our results could provide guidance on model building and a means to appraise model realism (c.f. Fenicia et al., 2014). By showing that CAMELS catchment attributes do not contain all hydrologically relevant information, we also show that we need better attributes if we want to identify model structures or parameter values based on catchment attributes. This is reinforced by a recent model intercomparison study using the same dataset which did not find a relation between model structures and static catchment attributes (Knoben et al., 2020).

4.5 Concluding remarks

In the introduction, we asked why non-climatic catchment attributes have shown limited explanatory power in recent large sample studies. We hypothesised that this is due to limitations in (a) the input data we use to inform our analyses, and (b) the hydrological signatures we use to describe the hydrological response. So what have we learned?

(a) We have found that region-specific knowledge is underutilised in large sample studies. There are many sources of information that can help us to better understand regional hydrological processes, and a key challenge will be to synthesise this information in a useful way. We suggest that this is best done through a common framework underpinned by perceptual models (i.e. "perceptual models of everywhere", cf. Beven, 2007).

(b) It is important to pay attention to the hydrological signatures we use, and we should try to explicitly link them to hydrological processes. We have shown that the use of multiple baseflow signatures – instead of a single BFI – and meaningful catchment subgroups allows us to better distinguish between different baseflow sources. A thoughtful choice of signatures will be crucial to meaningfully assess whether a catchment attribute is hydrologically relevant.

We conclude that we will be able to better link hydrological signatures to catchment attributes if we aim at a more systematic and hydrologically motivated selection of catchment attributes and hydrological signatures.

CHAPTER 5

SUMMARY AND OUTLOOK

Section 5.1 of this chapter is adapted from the abstracts of the three published works:

Gnann, S. J., Woods, R. A. and Howden, N. J. K. (2019). Is there a baseflow Budyko curve? Water Resources Research, 55(4):2838–2855. https://doi.org/10.1029/2018WR024464
Gnann, S. J., Howden, N. J. K. and Woods, R. A. (2020). Hydrological signatures describing the translation of climate seasonality into streamflow seasonality. Hydrology and Earth System Sciences, 24(2):561–580. https://doi.org/10.5194/hess-24-561-2020
Gnann, S. J., McMillan, H. K., Woods, R. A. and Howden, N. J. K. (2020). Including regional knowledge improves baseflow signature predictions in large sample hydrology. Water Resources Research, 56:e2020WR028354. https://doi.org/10.1029/2020WR028354

5.1 Summary

Baseflow is an important water resource as it provides water between precipitation events. Baseflow is also an important aspect of catchment functioning and thus a key element of hydrological theory. Yet we still lack a thorough understanding of how climate and landscape properties control baseflow generation. This thesis investigated baseflow generation across different climate zones and landscapes through a comparative hydrology approach. Specifically, we were motivated by the following three inter-related questions:

(1) How can we quantify baseflow in a meaningful way?

(2) Which climate and catchment attributes influence baseflow generation?

(3) How can we model baseflow adequately for large samples of catchments?

This thesis advances our knowledge in relation to all three questions and thereby improves our understanding of baseflow generation at the catchment scale. To quantify baseflow, we used several hydrological signatures that focus on different aspects of baseflow response. We explored which climate and catchment attributes control these signatures, and we explored whether some simple models can predict these signatures. The key findings of each chapter are summarised below, followed by an overarching discussion.

Baseflow fraction is controlled by climate aridity and catchment storage capacity

Climatic forcing determines the water and energy available to a catchment and thus is a strong control on all water balance components. We compared catchments in the US and the UK to show that the aridity index (the ratio between mean potential evapotranspiration and mean precipitation), despite being a strong control, cannot explain all the observed variability in baseflow fraction (the ratio between mean baseflow and mean precipitation). We used a simple annual water balance model (the so called Ponce-Shetty model) to explain how precipitation is partitioned into baseflow and other water balance components. In humid catchments, baseflow fraction is highly variable due to variations in a catchment's wetting potential, a parameter describing catchment storage capacity. Once a catchment's storage is "full", excess precipitation will result in fast flow. In arid catchments, most of the incoming precipitation evaporates, which leads to lower variability in baseflow fraction. While the wetting potentials and other parameters were obtained by fitting the Ponce-Shetty model to annual catchment data, their links to physical properties remain to be explored. This currently limits the model's applicability to gauged catchments with sufficiently long records.

Seasonal hydrological signatures help to better understand catchment functioning and to evaluate model performance

In many places, climatic forcing varies systematically with the seasons, and so does streamflow. Commonly used seasonal signatures consider climate or streamflow seasonality separately, but not how climate seasonality translates into streamflow seasonality. Yet this translation is insightful as it tells us more about catchment functioning, in particular about slower processes related to baseflow generation. To analyse the translation of seasonal climate input (precipitation minus potential evapotranspiration) into seasonal catchment output (streamflow), we represented the two time series by their annual Fourier mode, that is by sine waves. A catchment alters the input sine wave by reducing its amplitude and by shifting its phase. We used these quantities, the amplitude ratio and the phase shift, as seasonal hydrological signatures and calculated them for catchments in the UK and the US. Wet, rather impermeable catchments hardly attenuate the seasonal climate input. Drier catchments, especially if underlain by a productive aquifer, strongly attenuate the input sine wave leading to phase shifts up to several months. While these patterns are very pronounced in the UK, they are less clear in the US. This is in part because we used potential evapotranspiration, which restricts the applicability of the signatures to energy-limited catchments. We then used the seasonal signatures together with the baseflow index to evaluate two common hydrological models (IHACRES, GR4J) for a subset of UK catchments. We showed that GR4J cannot reproduce the signatures and the baseflow index simultaneously, pointing at structural deficiencies of the model. The signatures can therefore both be used to describe an aspect of catchment functioning and as a diagnostic tool in model building and evaluation.

Regional hydro(-geo)logical knowledge highlights the importance of catchment form on baseflow response

A catchment's hydrological response is controlled by climatic forcing and by the landscape through which water moves. It has, however, been proven difficult to link catchment characteristics (e.g. geology) to catchment response (e.g. baseflow index) in recent large sample studies. We posed two hypotheses why this might be the case: (a) the catchment attributes we use do not contain the hydrologically relevant information; and (b) the signatures we use cannot sufficiently discriminate between hydrologically different catchments. To test these hypotheses, we reviewed literature from various regions in the US and compared our findings to information contained in a large US dataset. We used standardised perceptual models to organise the findings from diverse regions and to allow for a more targeted search for catchment attributes. We found a lot of regional knowledge that is currently not represented in large scale datasets, but that is useful to better understand baseflow response. We also showed that multiple baseflow signatures are needed to better distinguish between different baseflow sources, such as the subsurface, surface water bodies, and snow. This highlights the need for a more systematic and hydrologically motivated selection of catchment attributes and hydrological signatures.

5.2 Overarching remarks

In the introduction (Chapter 1), we discussed the rather loose definition of baseflow. Now, at the end, we might ask whether we know more precisely what baseflow is, and where to go from here. We have seen – both in this thesis and in many references cited herein – that baseflow is not a crisp hydrological process, but a concept that is often used in different ways and sometimes confusingly so. This might in part be explained by the historical development of hydrology, which was often motivated by operational needs rather than scientific insights (e.g. Klemeš, 1986). Consequently, some (terminological) confusion remains to this day, and will remain if we do not clearly define what we mean by baseflow and discuss the inherent limitations of such a definition (Hewlett and Hibbert, 1967; Hall, 1968).

Many baseflow signatures are based on a separation of streamflow into baseflow and fast flow, and hence aim at quantifying the partitioning of water into different components (e.g. baseflow fraction, baseflow index). Other signatures focus more on dynamics (e.g. the phase shift of the annual cycle, recession characteristics). None of them are perfect, and in particular baseflow separation methods cannot be evaluated as baseflow (defined as the slowly varying component of streamflow) is not directly measurable (cf. Chapter 1). We have, however, seen that different baseflow signatures are very useful to describe catchment response, and that these signatures are often strongly related to landscape properties and hydrological processes. So, baseflow is still a useful umbrella term for "flow that comes from delayed sources", particularly in comparative studies. But we should move the emphasis towards understanding hydrological processes and patterns, and try to explain how they relate to climate and landscape characteristics.

We have seen and used a wide range of hydrological signatures focusing on baseflow. While multiple signatures are helpful to get a more nuanced picture of baseflow response, it certainly cannot be the aim to create an ever increasing list of signatures. We should therefore aim at identifying the most robust and the most informative signatures. Ideally, they should relate to catchment functioning in a rather explicit way, making it more likely that they can be linked to catchment form. In Chapter 2, we related mean baseflow to mean precipitation (the baseflow fraction: $\bar{Q}_b/\bar{P}$) instead of mean streamflow (the BFI: $\bar{Q}_b/\bar{Q}$), which allowed us to investigate baseflow within a catchment water balance context. In Chapter 3, we introduced seasonal signatures, in particular the phase shift, to quantify seasonal input-output dynamics. These signatures were based on theoretical considerations and thus we already had a framework at hand that we could use to interpret the observed signatures. Lastly, in Chapter 4, we looked at sub-regions with a similar climate. Having controlled for climate, two kinds of BFI – focusing on seasonal and long-term baseflow, respectively – turned out to be strongly linked to catchment form. Yet to identify different baseflow sources, we had to look at multiple signatures simultaneously. In particular, combining the BFI (as a proxy for baseflow volume) with recession characteristics (as a proxy for baseflow dynamics) allowed us to distinguish between different baseflow sources, such as wetlands, snow, or groundwater. There will not be one perfect set of baseflow signatures, as

their applicability depends on the climate, the catchment, and the problem at hand. But thinking more carefully about signature selection will help us to better link hydrological signatures to hydrological processes.

Data underpin our understanding, our theory, and our models. Calls for more – or rather more informative – data have been widespread (e.g. Kirchner, 2006; Beven, 2019). We still know very little about the subsurface as we can neither observe the medium nor the flow through it at the scales of interest. It is therefore very likely that some advancements will only be made if we get more data. But there is also a wealth of information that is available but not used (see Chapter 4). Some of that information relates rather explicitly to properties that we are interested in (e.g. hydrogeological classifications). Some of it, however, is only an indirect indicator of these properties, such as sinkholes (see Chapter 4), dynamic stream networks (Godsey and Kirchner, 2014; Prancevic and Kirchner, 2019), or vegetation (Dralle et al., 2020). Such information requires some theory to infer what we need to know. Independent of its type, the information we have is a rather fragmented collection, often of regional scope and qualitative nature. This calls for more research to systematise and synthesise this wealth of information, and ultimately combine it into a more widely applicable and quantitative classification. We made some progress towards utilising such information in Chapter 4 and found that this clarified the relation between baseflow response and catchment form.

Modelling might be seen as the last step in a scientific investigation, which serves as a test for our understanding and puts the learned into practice. We have used a simple annual water balance model in Chapter 2 and we have evaluated two commonly used rainfall-runoff models (IHACRES, GR4J) in Chapter 3. But overall we have paid rather little attention to models in this thesis. This is not accidental. For how can we build models that realistically represent processes, if many of the processes remain poorly understood? Clearly, we need to use models (especially for practical purposes), even if the models we have are imperfect. Yet it might be worth to take a step back and to ask more fundamental (and perhaps different) questions about hydrology. Evaluating the realism (or fidelity) of any model – from bucket-type hydrological models to machine learning methods – requires process understanding and good diagnostic signatures. And predictions based on (widely available) catchment attributes, and hence predictions in ungauged catchments, require informative catchment attributes and ways to make use of them. We spent our effort on investigating these basic aspects needed for model development and understanding, rather than on the modelling itself.

A main aim of (scientific) hydrology is to "simply" observe the landscape and to predict how water will be partitioned, stored, and released. Linking hydrological signatures (that meaningfully describe catchment response) to climatic and non-climatic catchment attributes – that is, linking function to forcing and form – would enable us to map these signatures across the landscape (and perhaps into the future). This would help us to identify hydrologically similar catchments and provide a solid basis for catchment classification. It would also help us to build models that

realistically represent the dominant hydrological processes without relying heavily on calibration. Lastly, it would both present a demonstration of our hydrological understanding and a means to make robust predictions in times of change. This thesis contributes to that aim, and while it cannot reach it, provides answers that will hopefully bring us a few steps closer to it.

5.3 Directions for future research

There are several research ideas that either expand or build on this thesis. We have touched only briefly on baseflow coming from snowmelt and surface water bodies. We have focused on natural catchments and neglected human impacts and the effects of (often human-induced) land use or climatic change on baseflow. These are certainly topics of high scientific and societal relevance and thus a starting point for future research. However, the topics suggested below remain focused on natural systems, as they are mostly based on the preceding research. As discussed above, baseflow generation in natural catchments is still poorly understood. Hence, it remains necessary and insightful to study baseflow generation in natural, mostly unchanging systems.

An uncalibrated annual water balance model

The parameter values of the Ponce-Shetty annual water balance model used in Chapter 2 were obtained by fitting the model to annual catchment data. A next step would therefore be to quantitatively link these parameters to catchment attributes. The wetting potential, which turned out to be the key parameter, is notionally similar to catchment storage capacity. It can be qualitatively linked to some landscape features (e.g. Chalk in the UK). But linking it quantitatively to physical properties will (probably) not be straightforward. First, as Chapter 4 has shown, many catchment attributes available at large scales lack hydrologically relevant information. It is likely that catchment attributes that fail to explain regional differences in BFI will also fail to explain parameters such as the wetting potential. Second, the wetting potential (and other parameters) might be partly influenced by climate characteristics, and hence might differ between catchments with similar subsurface characteristics. For example, higher storm intensities and more pronounced seasonality will likely lead to more fast flow, even if the storage capacity is the same (see e.g. Milly, 1994). Third, we neglected carryover storage between years (and between catchments), which might invalidate the model (or at least introduce error) in many catchments. The vaporisation potential should be similar to potential evapotranspiration, but they only show a medium correlation (Spearman rank correlation of 0.55). Also, potential evapotranspiration is itself uncertain (e.g. because we use a reference crop to estimate it) and that might further complicate such a linkage. More generally, such simple models which lump together various processes might not allow for a one to one mapping between catchment attributes and parameter values and it is unclear if such abstract parameters can be related to physical catchment characteristics at all.

Seasonal signatures in water-limited or snowy climates

The seasonal signatures introduced in Chapter 3 use potential evapotranspiration as a proxy for actual evapotranspiration. They therefore can only be applied to energy-limited catchments (with a seasonal cycle that can be reasonably described by a sine wave). We tested a simple actual evapotranspiration estimate using the Budyko framework (Budyko, 1974), which mostly affected the results in the US (see Appendix B). But this approach is rather crude and other options might be more suitable. Ideally, we would like to use (mostly) measured and independently derived actual evapotranspiration data (e.g. Martens et al., 2017), but data quality and availability of long enough daily time series might not allow that. Instead, we could use a simple water balance model to estimate actual evapotranspiration (see e.g. Pfister et al., 2017). Alternatively, there might be analytical solutions for certain simplified cases (e.g. forcing a linear reservoir with constant PET). Snowmelt inputs might also be modelled in a simple way to account for precipitation falling as snow (see e.g. Woods, 2009). This would enable us to also use theses signatures in water-limited or snowy climates.

Evaluating signature spaces of different hydrological models

In Chapter 3 we adapted a model evaluation approach following Vogel and Sankarasubramanian (2003), which could be the starting point of a more extensive model comparison experiment (e.g. using the MARRMoT toolbox; Knoben et al., 2019a). The approach works as follows. First, choose a representative set of forcing data, that is, forcing data from catchments with different climate characteristics. Second, run the chosen model with a sufficient amount of parameter sets aimed at representing the whole (realistic) parameter space. Third, calculate the chosen hydrological signatures for each model run. They form the signature space of a model, given the forcing space and the parameter space. We can then explore which ranges of signatures a model can simulate and whether that range covers the observed signatures – without calibrating it to streamflow data. This could be particularly insightful if we use multiple signatures. As shown in Chapter 3, GR4J could mostly reproduce the range of observed BFIs, but it could barely reproduce any of the observed combinations of phase shift and BFI. By comparing multiple model structures, we could explore whether there are typical model structures that are good at modelling certain (combinations of) signatures. We could also investigate more systematically how signature spaces vary with forcing. This would allow us to test whether modelled signatures are controlled more by forcing or by model structure (see Figure 5.1, which shows that for IHACRES, rising limb density is strongly controlled by forcing, whereas BFI is not). There are some technical challenges that need to be addressed, such as adequately sampling the parameter space and deciding whether an observed signature space (in multiple dimensions) is inside the modelled signature space. This might be challenging, particularly for models with many parameters, but we could draw on extensive knowledge on parameter sampling, sensitivity analysis and model optimisation to tackle these issues more efficiently.

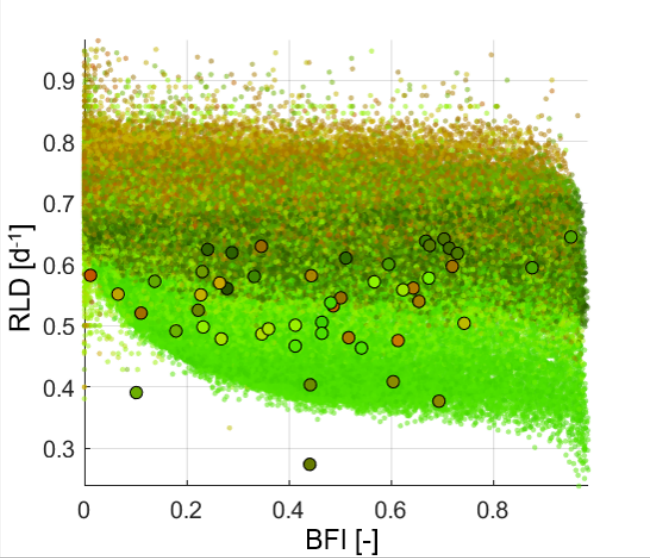

Figure 5.1: Two-dimensional signature space consisting of BFI and rising limb density (RLD). The circled dots represent observed signatures of 50 catchments. The small dots represent IHACRES model runs using the forcing of the same 50 catchments. The colours indicate forcing characteristics using the RGB colour scheme suggested by Knoben et al. (2018). More red indicates more arid catchments, more green indicates more seasonal catchments. (More blue would indicate more snowy catchments, but the catchments here are snow-free.)

From hydraulic response times to transit times

In this thesis, we only used streamflow data to characterise catchment response. While there are possibilities to infer baseflow sources if we consider information about climate (e.g. snow fraction) or landscape characteristics (e.g. wetlands, productive aquifers), this is rather indirect and often vague. Tracer data offer more explicit insights into water sources (Sprenger et al., 2019). They also add another dimension to catchment functioning. Namely, how water flows through a catchment, rather than how perturbations are propagated through a catchment (McDonnell and Beven, 2014). Since most readily available large sample datasets do not (yet) contain tracer data, we could start with collecting data from diverse regions which could then be used in a comparative study (see e.g. McNamara et al., 2011; Staudinger et al., 2017; Cooke and Buttle, 2020). Such a study might tell us more about the sources of baseflow, about mechanisms of (base-)flow generation, and it might support or challenge perceptual models based on streamflow data alone, especially our conceptions of storage. Streamflow-based storage estimates typically relate to dynamic storage (storage that is hydrologically active, i.e. filled up and released) and not total storage (which can still affect water mixing and solute transport; McNamara et al., 2011; Buttle, 2016; Staudinger et al., 2017; Dralle et al., 2018); note that the definitions relating to storage types are partly inconsistent. But also dynamic storage can consist of storage that directly drives discharge generation and storage that, while not driving discharge generation, supplies

water for evapontranspiration (Dralle et al., 2018). Similarly, many studies showed that plant water and stream water typically differ in their isotopic composition, challenging both traditional storage and flow conceptualisations (see e.g. Sprenger and Allen, 2020, and references therein). Especially during extreme conditions such as droughts, distinguishing between different types of storage might become essential to understand catchment behaviour. For example, storage that does not drive streamflow can be further depleted than it is visible from the streamflow record, and hence introduce a "memory" of past drought events that affects drought recovery (e.g. Fowler et al., 2020). While hydrometric analyses might shed some light on such processes (e.g Dralle et al., 2018; Fowler et al., 2020), tracer (and other) data can help to independently test such hypotheses, and yield crucial additional insights that advance our knowledge of catchment functioning.

A geological classification for catchment hydrology

Geology can be a strong control on catchment response. For example, the different aquifer units in the Edwards Plateau or the Ozarks can explain most of the regional variability in baseflow response (see Chapter 4). Similarly, Bloomfield et al. (2009) showed how lithology controls BFI in the Thames basin in the UK. These and many other examples use regional hydrogeological classifications. Yet we do not possess a universal catchment classification that groups catchments into geological classes that are hydrologically meaningful. This problem will not just go away by using more detailed geological maps. We also need to know how geology controls hydrology. For example, hydrologically young carbonate rock will be different to older, more karstified carbonate rock. Such and other interactions between rock type (typically contained in geological maps) and climate, topography, geomorphology, and tectonics, make such a classification a non-trivial task. But mapping geological features into metrics that are relevant to hydrology (e.g. time scales, storages) would be a huge advancement that could inform catchment classification, hydrological model building, and predictions in ungauged catchments. As a first step, we could explore semi-empirically how landscape metrics (e.g. rock type, landforms) control catchment response, and whether we can find catchment groups that are both hydrologically and geologically similar. Instead of a purely data-based approach, we could start with delineating catchment groups based on a priori (qualitative) knowledge. For example, we know that tectonically active uplands, where regolith is close to the surface, have a different subsurface architecture than lowlands covered by thick sediment layers. Such landscapes are likely to be hydrologically different and thus require different descriptors (e.g. primary hydraulic conductivity might be meaningless if most flow is through fractures). Eventually, we would like – if they exist – simple mechanistic theories that incorporate the relevant landscape attributes and their heterogeneity at the scale of interest. Yet these might be different for different landscapes, so a semi-empirical grouping might be a good starting point to delineate regions in which similar processes are at work.

Perceptual models and subjectivity in science

Perceptual models can help us to discuss hydrological systems more systematically and to share the often subjective and quantitative experience of hydrologists (e.g. Wagener et al., 2020). Since perceptual models are by definition subjective and influenced by the background of a hydrologist, would other hydrologists come up with different perceptual models as we did in Chapter 4? Probably, but how different? Would we agree on the key processes (relevant for the purpose at hand)? And if not, what conclusions would we draw from that? How would we evaluate process realism (or fidelity) in models if we cannot agree on what a realistic process representation is? Subjectivity is often thought (or perhaps hoped) to be outside the realm of science. However, we make (and cannot avoid) decisions that are somewhat subjective (e.g. we stick to the models we have experience with, even if they are not the most adequate; see Addor and Melsen, 2019). Especially since Kuhn (1962), philosophers of science have started to pay more attention to the social character of science and the role consensus plays in establishing scientific knowledge (e.g. Oreskes, 2019). That is not to say that most science (or hydrology) is subjective and that we should not strive for evidence-based knowledge. Yet it can only benefit us to be open about (inevitable) subjective decisions, and about things we do not know, to avoid "appearance of scientific knowledge where there is ignorance" (Klemeš, 1986). One aspect of that would be to share and discuss our perceptual models. To address this question, we have started to work on a "perceptual model survey" in which we ask hydrologists to draw perceptual models of the same catchments. This survey aims at getting a snapshot impression of how our community thinks about hydrological systems and at fostering a debate about perceptual models in general.

APPENDIX A

SUPPORTING INFORMATION FOR CHAPTER 2

A.1 Derivation of a theoretical equation for BFI

To obtain an equation for the BFI we make use of another catchment index presented in Sivapalan et al. (2011), the runoff ratio K_R:

$$K_R = \frac{\bar{Q}_f + \bar{Q}_b}{\bar{P} - \lambda_W V_p} \tag{A.1}$$

K_R can be approximated theoretically by:

$$K_R = \frac{\tilde{P}(1+\tilde{V}_p)}{\tilde{P} + \tilde{V}_p + \tilde{V}_p\tilde{P}} \tag{A.2}$$

We can write the BFI using K_B and K_R:

$$\text{BFI} = \frac{\bar{Q}_b}{\bar{Q}_f + \bar{Q}_b} = \frac{K_B}{K_R} \tag{A.3}$$

$$\text{BFI} = \frac{\tilde{P}(1+\tilde{P})^{-1}}{\tilde{P} + \tilde{V}_p + \tilde{V}_p\tilde{P}} \left(\frac{\tilde{P}(1+\tilde{V}_p)}{\tilde{P} + \tilde{V}_p + \tilde{V}_p\tilde{P}} \right)^{-1} \tag{A.4}$$

$$\text{BFI} = \frac{1}{(1+\tilde{P})(1+\tilde{V}_p)} \tag{A.5}$$

A.2 Data and code availability

The CAMELS dataset is available from `https://ral.ucar.edu/solutions/products/camels`. Information about the UK Benchmark Network can be obtained from `https://nrfa.ceh.ac.uk/benchmark-network`. Streamflow data and catchment attributes are available from `https://nrfa.ceh.ac.uk`. CEH-GEAR precipitation data are available from `https://doi.org/10.5285/33604ea0-c238-4488-813d-0ad9ab7c51ca`. CHESS-PE potential evapotranspiration data are

available from https://doi.org/10.5285/8baf805d-39ce-4dac-b224-c926ada353b7.
Map colours are based on www.ColorBrewer.org by Cynthia A. Brewer, Penn State.

APPENDIX B

SUPPORTING INFORMATION FOR CHAPTER 3

B.1 The response of linear reservoirs to periodic forcing

The derivations in B.1.1–B.1.3 and the resulting equations are not novel. We think, however, that presenting them altogether might be a useful overview for the interested reader. General overviews on linear systems theory are given for example by Dooge (1973) for hydrology or by Smith III (2007) for signal processing. The response of linear reservoirs to periodic forcing was for example described by Eriksson (1971) or Peters et al. (2003).

B.1.1 Single linear reservoir

The outflow Q from a linear reservoir is described by Eq. (B.1):

$$Q = \frac{S}{\tau} \tag{B.1}$$

Conservation of mass requires that the rate of change of storage S equals the inflow Q_{in} minus the outflow Q from the reservoir:

$$\frac{dS}{dt} = Q_{\text{in}} - Q \tag{B.2}$$

For simplicity, we consider a simple sinusoidal input signal with unit amplitude, with zero phase (i.e. aligned with the cycle of interest) and with zero mean. We also replace the period T by the angular frequency $\omega = \frac{2\pi}{T}$.

$$Q_{\text{in}}(t) = \sin(\omega t) \tag{B.3}$$

Combining Eq. (B.1), Eq. (B.2), and Eq. (B.3) yields:

$$\frac{dQ}{dt} + \frac{Q}{\tau} = \frac{\sin(\omega t)}{\tau} \tag{B.4}$$

which is a first-order ordinary differential equation (ODE) that can be solved with the help of an integrating factor $\exp(t/\tau)$ and by using the product rule:

$$\exp\left(\frac{t}{\tau}\right)\left(\frac{dQ}{dt}+\frac{Q}{\tau}\right)=\exp\left(\frac{t}{\tau}\right)\frac{\sin(\omega t)}{\tau} \tag{B.5}$$

$$\frac{d}{dt}\left(\exp\left(\frac{t}{\tau}\right)Q\right)=\exp\left(\frac{t}{\tau}\right)\frac{\sin(\omega t)}{\tau} \tag{B.6}$$

$$\left[\exp\left(\frac{t}{\tau}\right)Q\right]_0^t=\frac{1}{\tau}\int_0^t\exp\left(\frac{t}{\tau}\right)\sin(\omega t)dt \tag{B.7}$$

$$\exp\left(\frac{t}{\tau}\right)Q(t)-Q(0)=\frac{1}{\tau}\int_0^t\exp\left(\frac{t}{\tau}\right)\sin(\omega t)dt \tag{B.8}$$

Since we are interested in the steady-state periodic response of our system, we set $Q(0)=0$. We can solve the integral on the right hand side using an integration rule (see e.g. Spiegel, 1968):

$$\exp\left(\frac{t}{\tau}\right)Q(t)=\frac{1}{\tau}\frac{\exp\left(\frac{t}{\tau}\right)}{\sqrt{(\frac{1}{\tau})^2+\omega^2}}\sin\left(\omega t+\arccos\left(\frac{1}{\tau}\frac{1}{\sqrt{(\frac{1}{\tau})^2+\omega^2}}\right)\right) \tag{B.9}$$

$$Q(t)=\frac{1}{\sqrt{1+(\omega\tau)^2}}\sin\left(\omega t+\arccos\left(\frac{1}{\sqrt{1+(\omega\tau)^2}}\right)\right) \tag{B.10}$$

If we rewrite $A=\frac{1}{\sqrt{1+(\omega\tau)^2}}$ (Eq. (3.9)) and $\phi=\arccos(A)$ (Eq. (3.10)), we obtain:

$$Q(t)=A\sin(\omega t+\phi) \tag{B.11}$$

Since the system is linear, other inflow amplitudes can be accounted for by scaling (multiplication) and a non-zero mean by addition. We could also superimpose other inputs of different periods. Note that we can obtain the same result using the transfer function approach in linear systems theory, as for example described in Dooge (1973) or Smith III (2007).

B.1.2 Linear reservoirs in series

Linear reservoirs in series can be conceptualised as follows. The outflow from the first reservoir is the inflow to the second reservoir, the outflow from the second reservoir is the inflow to the third reservoir, and so forth. Let's denote the outflow from the first reservoir by Q_1:

$$Q_1(t)=A_1\sin(\omega t+\phi_1) \tag{B.12}$$

If we use Eq. (B.12) as inflow to the second reservoir (which is also the total outflow), we obtain:

$$\frac{dQ_2}{dt}+\frac{Q_2}{\tau_2}=\frac{A_1\sin(\omega t+\phi_1)}{\tau_2} \tag{B.13}$$

This can be solved in a similar fashion as before (Eq. (B.4)) and we get:

$$Q_2(t)=A_1A_2\sin(\omega t+\phi_1+\phi_2) \tag{B.14}$$

If we continue to do this for n reservoirs, we get:

$$Q_n(t) = A_1 A_2 \dots A_n \sin(\omega t + \phi_1 + \phi_2 + \dots + \phi_n) \tag{B.15}$$

$$Q_n(t) = \prod_{i=1}^{n} A_i \sin\left(\omega t + \sum_{i=1}^{n} \phi_i\right) \tag{B.16}$$

The total amplitude ratio is thus obtained by multiplication of all individual amplitude ratios and the total phase shift by addition of all individual phase shifts. If all the reservoirs have the same time constant ($\tau_1 = \tau_2 = \dots = \tau_n$), we obtain the so called Nash cascade (Nash, 1957).

B.1.3 Linear reservoirs in parallel

Linear reservoirs in parallel are the weighted sum of the outflow from each reservoir. The resulting flow is hence a sum of sine waves of the same angular frequency, weighted by the fraction p_i going into each reservoir. For two reservoirs we can write:

$$Q_{12}(t) = Q_1(t) + Q_2(t) \tag{B.17}$$

where Q_{12} is the combined outflow from the two reservoirs Q_1 and Q_2. We can use Eq. (B.11) to get the outflow from each of the two reservoirs:

$$Q_{12}(t) = p_1 A_1 \sin(\omega t + \phi_1) + p_2 A_2 \sin(\omega t + \phi_2) \tag{B.18}$$

The sum of two sine waves (Eq. (B.18)) can be rewritten to obtain only one sine wave:

$$Q_{12}(t) = p_1 A_1 \sin(\omega t + \phi_1) + p_2 A_2 \sin(\omega t + \phi_2) \tag{B.19}$$

$$Q_{12}(t) = A_{12} \sin(\omega t + \phi_{12}) \tag{B.20}$$

where the total amplitude ratio A_{12} (Eq. (3.14)) and the total phase shift ϕ_{12} (Eq. (3.15)) are given by (see e.g. Smith III, 2007):

$$A_{12} = \sqrt{\left[p_1 A_1 \cos\phi_1 + p_2 A_2 \cos\phi_2\right]^2 + \left[p_1 A_1 \sin\phi_1 + p_2 A_2 \sin\phi_2\right]^2} \tag{B.21}$$

$$\phi_{12} = \arctan\left(\frac{p_1 A_1 \sin\phi_1 + p_2 A_2 \sin\phi_2}{p_1 A_1 \cos\phi_1 + p_2 A_2 \cos\phi_2}\right) \tag{B.22}$$

We could do the same for more (n) reservoirs by stepwise adding the resulting sine wave to the next sine wave (e.g. $Q_{12} + Q_3$).

B.1.4 Non-linear reservoirs

In the following we investigate how non-linear reservoirs respond to periodic forcing. A non-linear reservoir can be described by (Kirchner, 2009):

$$Q = Q_{\text{ref}} \left(\frac{S - S_0}{m}\right)^n \tag{B.23}$$

where Q_{ref} is an arbitrary reference discharge, S_0 is a reference storage, m is a scaling coefficient (it has the units of storage), and n is a non-linearity parameter ($n = 1$ results in the linear reservoir). For non-linear reservoirs, there are no general analytical solutions available. Non-linear reservoirs cannot be characterised by an invariant time constant, as their outflow rate depends on the storage, as it can be seen from Eq. (B.23). Hence, we model the response of a single non-linear reservoir numerically.

The outflow from a non-linear reservoir forced by a sinusoidal input is shown in Figure B.1. The outflow is still periodic, but the "sine curve" is somewhat squeezed. The maxima and minima do not have the same distance and hence there is no unique phase shift. For the example shown in Figure B.1, the phase shift between the maxima of Q_{in} and Q_{out} is 68 days, and the phase shift between the minima is 70 days. The difference between the maximum amplitude and the mean amplitude (= 0.49 mm) is 0.20 mm, and the difference between the minimum amplitude and the mean amplitude is 0.18 mm. This asymmetry is partly due to numerical inaccuracies, however, probably mostly due to the non-linearity of the reservoir. Because non-linear reservoirs with $n > 1$ drain more slowly as they empty, their minima are closer to the mean and the minima have a larger phase shift (and vice versa for $n < 1$).

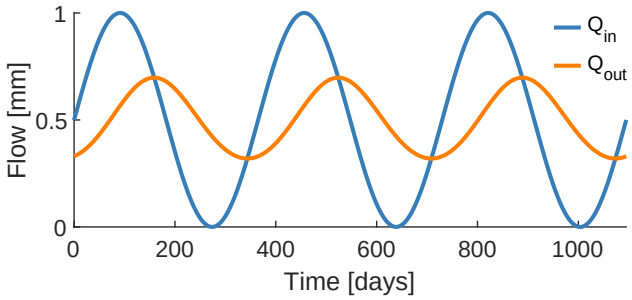

Figure B.1: Numerical approximation of steady-state (sinusoidal) inflow to and outflow from a non-linear reservoir ($Q_{\text{ref}} = 1$ mm, $S_0 = 0$ mm, $m = 200$ mm, $n = 2$). The starting time is chosen arbitrarily.

We can plot amplitude ratios and phase shifts for non-linear reservoirs with different parameter values. We therefore do not need to specify a characteristic time constant. This is shown in Figure B.2. Even though the outflow is not exactly a sine wave, it is possible to define a mean phase shift and amplitude ratio, and in practice this might be a reasonably good approximation and hardly distinguishable from an actual sine wave. So, even if the reservoir is non-linear, its steady state behaviour (or the response to seasonal inputs) might be reasonably well approximated by a linear reservoir with a time constant that reflects the outflow characteristics at a characteristic storage level (e.g. mean storage).

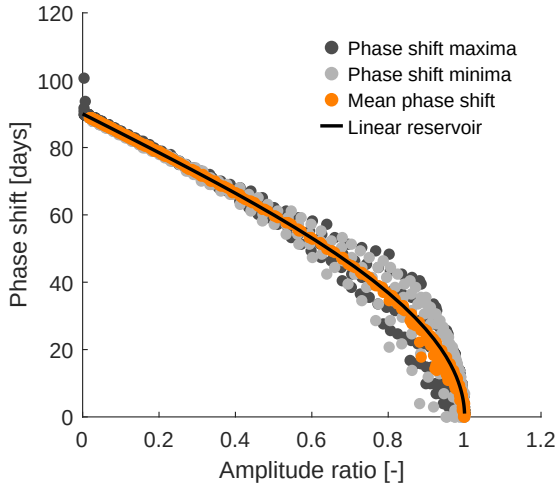

Figure B.2: Amplitude ratio and phase shift for a single non-linear reservoir for varying parameter values. The phase shifts of the maxima, the minima, and the mean phase shifts are indicated by coloured dots. The non-linearity parameter n ranges from 0.5 to 10.

B.2 Additional analyses

B.2.1 Extracting seasonal components from time series

We have tested two methods to extract seasonal components from time series. The first method is a multiple linear regression. The second method makes use of the cross-covariance of two time series, both are described below.

B.2.1.1 Multiple linear regression

A basic sine wave is given by:

$$x(t) = \delta_x \bar{x} \sin(2\pi \frac{t}{T} - \phi_x) + \bar{x} \tag{B.24}$$

We can rewrite Eq. (B.24) as follows (see e.g. Kirchner, 2016):

$$x(t) = \alpha \cos(2\pi \frac{t}{T}) + \beta \sin(2\pi \frac{t}{T}) + \bar{x} \tag{B.25}$$

We can rewrite Eq. (B.25) in vector form:

$$\begin{bmatrix} \cos(2\pi \frac{t_1}{T}) & \sin(2\pi \frac{t_1}{T}) & 1 \\ \cos(2\pi \frac{t_2}{T}) & \sin(2\pi \frac{t_2}{T}) & 1 \\ \vdots & \vdots & \vdots \\ \cos(2\pi \frac{t_n}{T}) & \sin(2\pi \frac{t_n}{T}) & 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta \\ \bar{x} \end{bmatrix} = \begin{bmatrix} x(t_1) \\ x(t_2) \\ \vdots \\ x(t_n) \end{bmatrix} \tag{B.26}$$

We can solve for α, β and $\bar{x}$ in Eq. (B.26) by means of multiple linear regression (e.g. by using Matlab's *mldivide* function). We can then solve for δ_x and ϕ_x via the identities:

$$\delta_x \bar{x} = \sqrt{\alpha^2 + \beta^2} \tag{B.27}$$

$$\phi_x = \arctan\left(\frac{\beta}{\alpha}\right) \tag{B.28}$$

Note that the *atan2* function is required to obtain an unambiguous phase shift.

B.2.1.2 Cross-covariance method

The unbiased estimate of the cross-covariance of two sine waves x and y is given by:

$$\gamma_{xy}(k) = \frac{1}{2}\delta_x \bar{x} \delta_y \bar{y} \cos\left((\phi_x - \phi_y) - 2\pi \frac{t}{T} k\right) \tag{B.29}$$

where k is the lag between the sine waves, $\delta_x \bar{x}$ and $\delta_y \bar{y}$ are their amplitudes, and ϕ_x and ϕ_y are their phase shifts. If we define x to be the signal of interest (e.g. streamflow) and y to be a dummy cycle of unit amplitude and zero initial phase shift ($\delta_y \bar{y} = 1$ and $\phi_y = 0$), Eq. (B.29) simplifies to:

$$\gamma_{xy}(k) = \frac{1}{2}\delta_x \bar{x} \cos\left(\phi_x - 2\pi \frac{t}{T} k\right) \tag{B.30}$$

We can calculate the empirical cross-correlation between the signal of interest and the dummy cycle and fit a sine curve to it (via nonlinear least squares). The parameters of that sine curve can then be used to find the parameters of our cycle of interest ($\delta_x \bar{x}$ and ϕ_x) via Eq. (B.30).

The fitting methods (multiple linear regression and cross-covariance method) show an almost perfect match (Figure B.3), which indicates that the extraction of the seasonal component is not sensitive to the method. While this means that we can reliably extract the sinusoidal component of the period of interest (1 year), that is, the annual Fourier mode, it does not mean that we perfectly extracted the "(annual) seasonal component" of the variable of interest. A sine wave is just a parsimonious approximation of the seasonal behaviour and the choice of a sine wave to model seasonality is also associated with uncertainty.

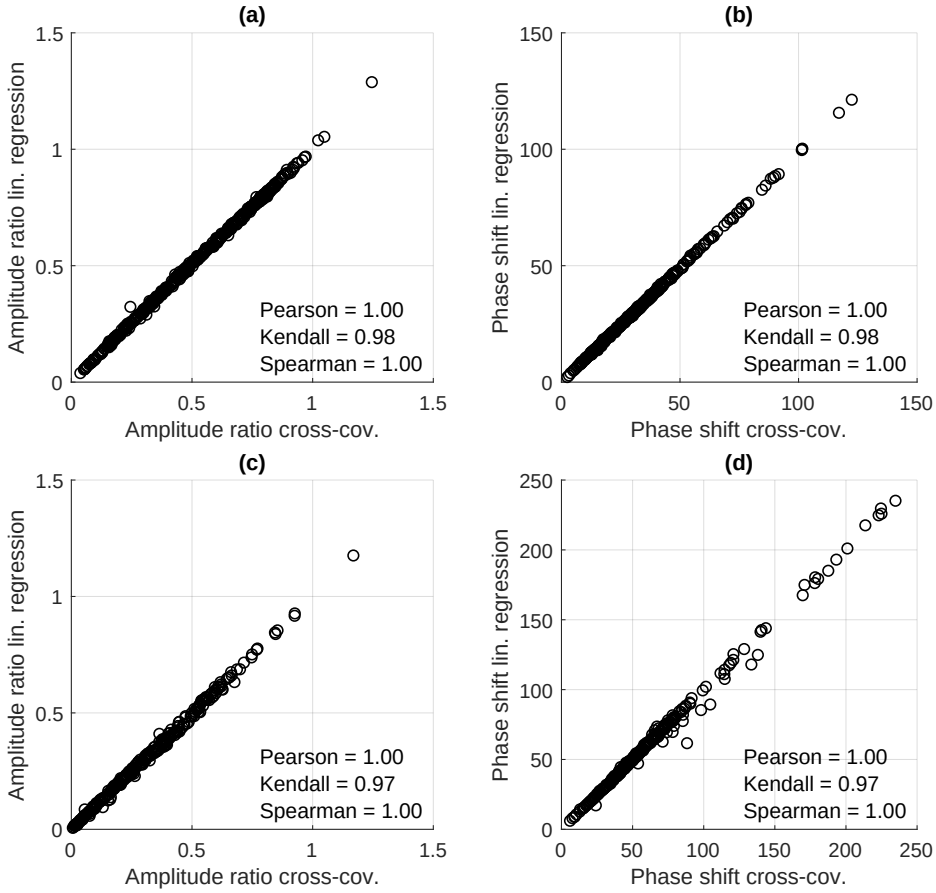

Figure B.3: Comparison of amplitude ratio and phase shift using the different sine fitting methods for UK catchments from 1989 to 2009. Panel **(a)** and **(b)** show UK catchments. Panel **(c)** and **(d)** show CAMELS catchments. Note that both axes are limited.

B.2.1.3 Robustness of seasonal signatures

To check whether the seasonal signatures are robust, we calculate the signatures for two different time periods: from 1989 to 1999, and from 1999 to 2009.

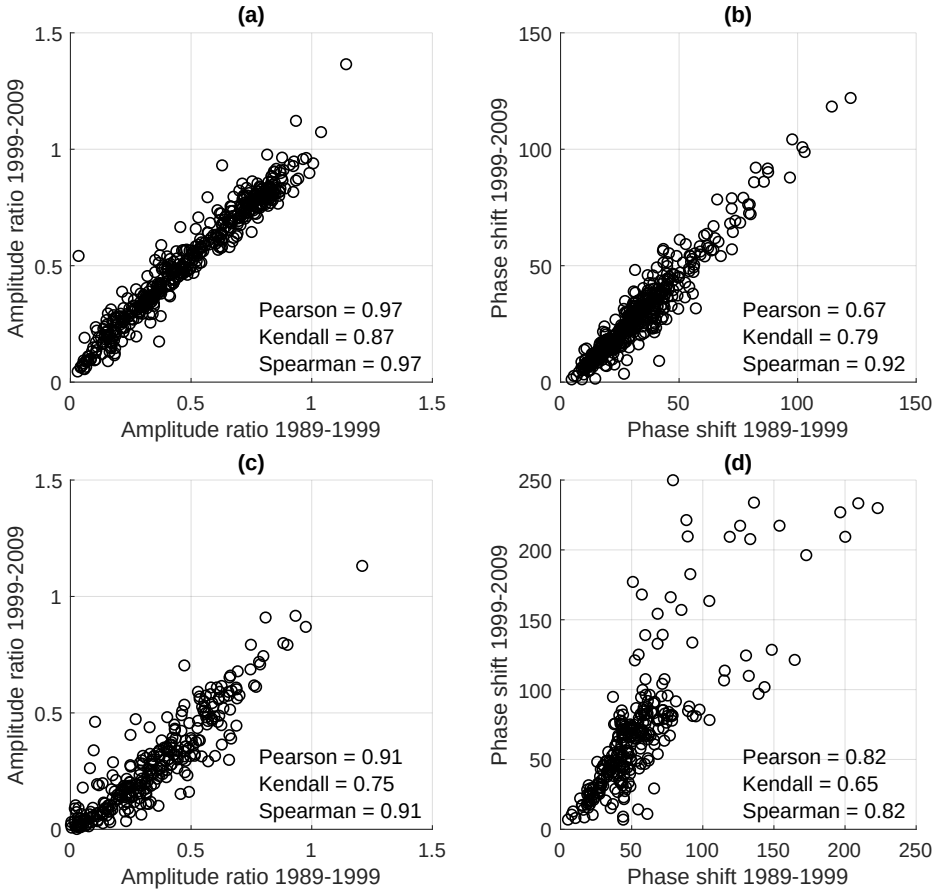

Figure B.4: Comparison of amplitude ratio and phase shift using different time periods. Panel **(a)** and **(b)** show UK catchments. Panel **(c)** and **(d)** show CAMELS catchments. Note that both axes are limited.

Figure B.4 shows that the amplitude ratio and the phase shift show good agreement for the different time periods analysed. This means that the period from 1989 to 1999 does not exhibit a fundamentally different behaviour than the period from 1999 to 2009, that is, the signatures are robust. Some variability can be expected as the forcing varies from year to year, due to human influences (UK), and because the signatures can be unreliable particularly in arid climates (US).

B.2.2 Fourier spectra of forcing and streamflow

To check whether the annual periodic component (annual Fourier mode) is the strongest periodic component of our time series, we can calculate one-sided power spectra for all catchments. For almost every catchment investigated here (> 99%) the strongest forcing Fourier mode is the annual mode. For a few catchments in the US a 0.5 year mode is the strongest mode, yet there is also an annual mode present. Some of the streamflow data have strongest modes different from one year, yet again there is also an annual mode present. Figure B.5 shows one sided power spectra for two catchments (the same catchments as shown in Figure 3.4). We can see that in both

catchments the annual mode is the strongest mode. Figure B.5b also shows a strong multi-annual mode (of about 7 years, see also Rust et al., 2019, for more information on multi-annual modes in the UK).

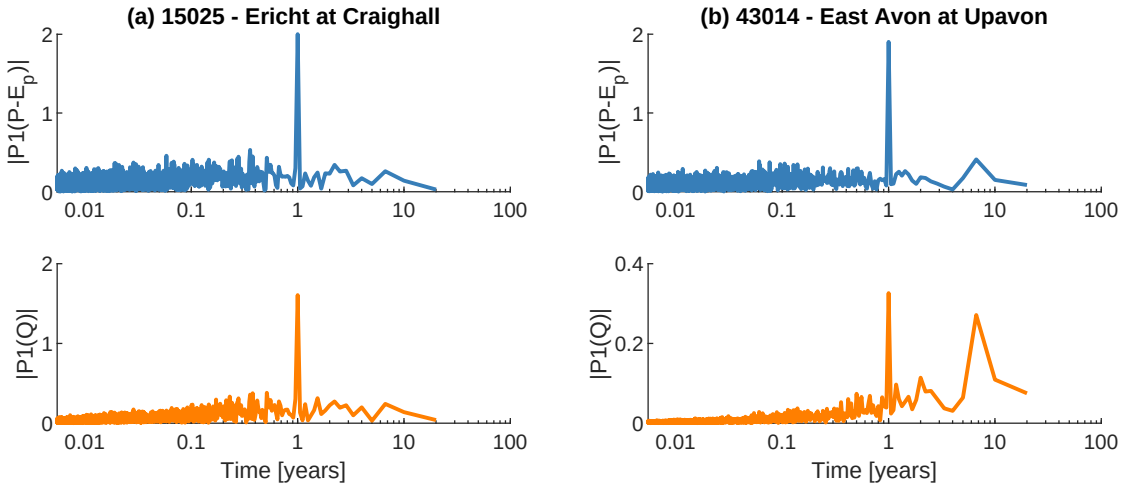

Figure B.5: One sided power spectra of climate input ($P - E_p$; blue) and catchment output (Q; orange) for two catchments in the UK, and their respective seasonal components. The Ericht is a rather responsive catchment (BFI = 0.47), while the East Avon has a large baseflow component (BFI = 0.89).

B.2.3 Catchments with precipitation falling as snow

Snow presents a different storage process that is not considered in the current approach. We therefore remove snowy catchments, defined as catchments with a snow fraction f_s (Knoben et al., 2018) larger than 0.001, from the analysis. Figure B.6 shows the snowy catchments and the corresponding snow fraction.

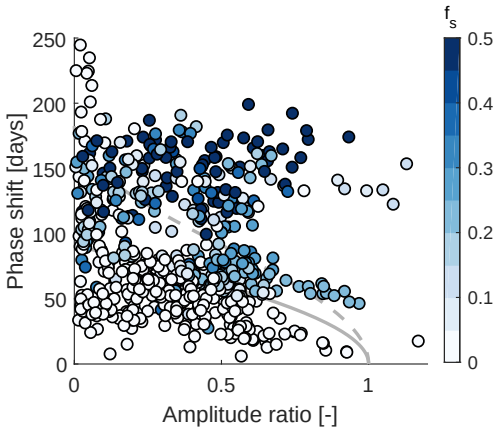

Figure B.6: Amplitude ratio against phase shift for CAMELS catchments. Colour indicates the fraction of precipitation falling as snow. Note that both axes are limited.

B.2.4 On the use of potential evapotranspiration as input signal

As mentioned in Chapter 3, we use precipitation P minus potential evapotranspiration E_p as a proxy for the input signal to a catchment (the forcing F). To test the validity of the assumption that $E_a = E_p$ in a rather straightforward manner, we adjust the seasonal component of E_p to obtain E_a by means of the Budyko framework (Budyko, 1974). We reduce the peak of the seasonal component of E_p so that it equals the mean of E_a, which we estimate using the following equation (Budyko, 1974):

$$\frac{\bar{E}_a}{\bar{P}} = \sqrt{\frac{\bar{E}_p}{\bar{P}} \tanh\left(\frac{\bar{P}}{\bar{E}_p}\right)\left(1 - \exp\left(-\frac{\bar{E}_p}{\bar{P}}\right)\right)} \tag{B.31}$$

We therefore obtain a new sine curve for the seasonal component of E_a, which has a reduced amplitude ($\delta_{E_a}\bar{E}_a$) and a reduced mean ($\bar{E}_a$), but the same phase as the sine curve for the seasonal component of E_p. This will increase the amplitude ratio and might change the phase shift whenever $\bar{E}_a < \bar{E}_p$.

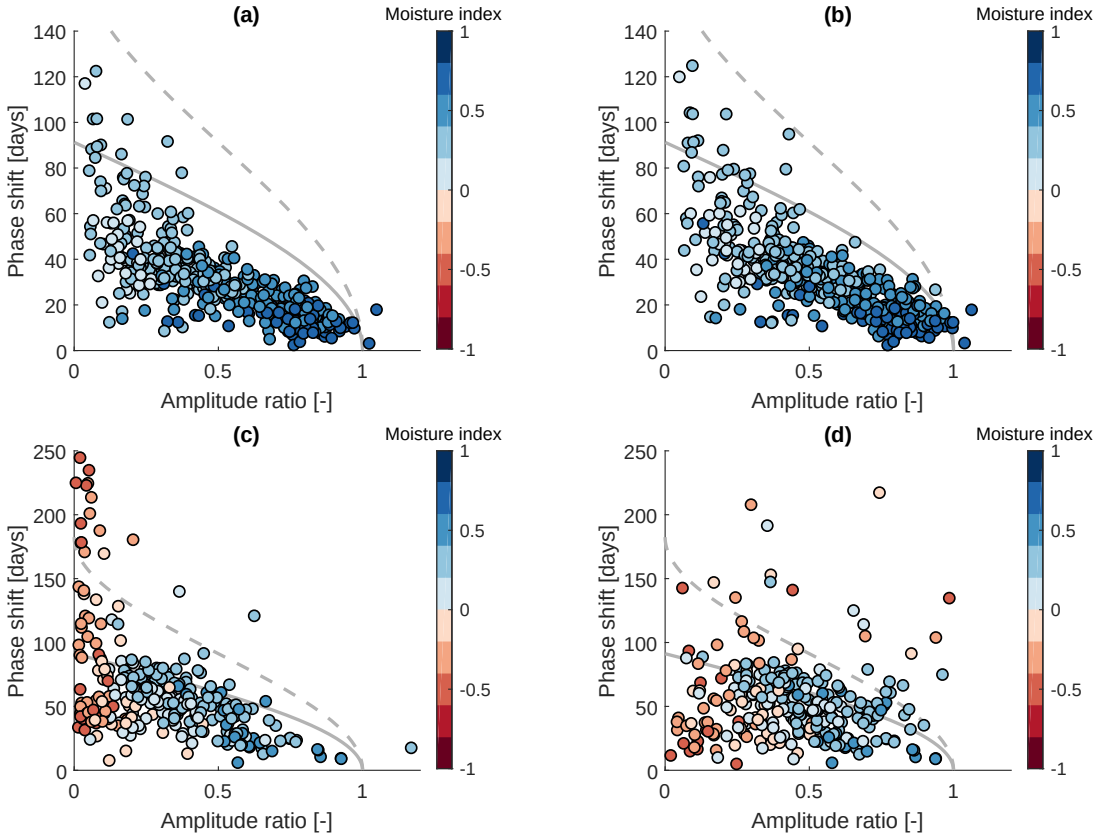

Figure B.7: Amplitude ratio against phase shift for **(a)** UK catchments and **(c)** CAMELS catchments using potential evapotranspiration in the forcing ($F = P - E_p$) and for **(b)** UK catchments and **(d)** CAMELS catchments using estimated actual evapotranspiration in the forcing ($F = P - E_a$). Grey solid line indicates a single linear reservoir, grey dashed line indicates the outer envelope for two reservoirs in parallel. Colours indicate the moisture index. Note that both axes are limited and that the y-axes differ in their range.

Figure B.7 shows the resulting amplitude ratios and phase shifts for the UK and the US together with the ones using E_p as input. Energy-limited catchments (high I_m), in particular the UK catchments, show a similar pattern for E_p and estimated E_a. The amplitude ratios are slightly higher, that is, the whole point cloud is slightly shifted to the right. In water-limited catchments in the US, where the annual water balance already implies that $E_a < E_p$, the amplitude ratios are often higher and the phase shifts smaller. Many of the dark red dots that are close to zero in Figure B.7b now plot much further to the right (Figure B.7d). Some of the large phase shifts in these arid catchments might still be a consequence of a poorly estimated input signal. The approach based on the Budyko framework (Eq. (B.31)) only reduces the amplitude of the evapotranspiration component, it does not change the timing of E_a compared to E_p. In reality, however, it is possible (and likely) that E_a also has a different phase compared to E_p. Especially in catchments where P and E_p are out of phase, the soil moisture reservoir will fill up during the wet months (peak rainfall) and dry out during the dry months (peak evapotranspiration). If

$E_p \gg P$, the soil will likely dry out before we reach peak E_p, and hence we reach peak E_a before peak E_p. The difference in the P peak and the E_p peak (e.g. the phase shift we observe in Figure B.7) would therefore be larger than the difference between the P peak and the E_a peak (i.e. the actual phase shift caused by the catchment). To overcome that, we would either need modelled E_a, for example, from a (simple) hydrological model, or measured E_a. Modelling E_a comes at the cost of introducing more modelling steps and therefore more complexity. Measurements of E_a are typically not available at a daily time scale. We therefore leave this for future work.

B.3 Modelling experiment

We use two common rainfall-runoff models from the MARRMoT rainfall-runoff modelling toolbox (Knoben et al., 2019a): IHACRES and GR4J. The parameter ranges are specified in Tables B.1 and B.2. The parameter ranges for IHACRES mostly follow the MARRMoT default values. The fast flow routing delay τ_q is set to range from 0.0001 to 10 days to represent fast flow (the default range is from 1 to 700 days). The flow delay τ_d (a pure delay function) is set to 0 (inactive) which makes the model conceptually equal to the version used by Croke and Jakeman (2004). The parameter ranges for GR4J are equal to the ranges used by Smith et al. (2019).

We determine the initial storages by repeatedly simulating the first water year, using the storages at the end of the year as new initial storages, until we reach an equilibrium (< 1% change) or 20 iterations (in case of non-convergence).

Table B.1: Parameter ranges for IHACRES. *This parameter is inactive.

Parameter	Unit	Description	Min	Max
lp	mm	Wilting point	0	2000
d	mm	Threshold for flow generation	0	2000
p	-	Flow response non-linearity	0	10
α	-	Fast/slow flow division	0	1
τ_q	d	Fast flow routing delay	0.0001	10
τ_s	d	Slow flow routing delay	1	700
(τ_d*	d	Flow delay	0	0)

Table B.2: Parameter ranges for GR4J.

Parameter	Unit	Description	Min	Max
x_1	mm	Maximum soil moisture storage	0.0001	3000
x_2	mm d^{-1}	Subsurface water exchange	-20	20
x_3	-	Routing store depth	0.0001	2000
x_4	mm	Unit hydrograph time base	0.5	15

B.3.1 Subset of catchments for modelling experiment

The subset of catchments is chosen as follows. We only use benchmark catchments (Harrigan et al., 2017) and remove the catchments with a low flow score of 0 and with any missing values between 1989 and 2009. The remaining set of catchments is manually thinned out to evenly occupy the signature space shown in Figure 3.5. Figure B.8 shows amplitude ratios, phase shifts, and BFIs of the subset.

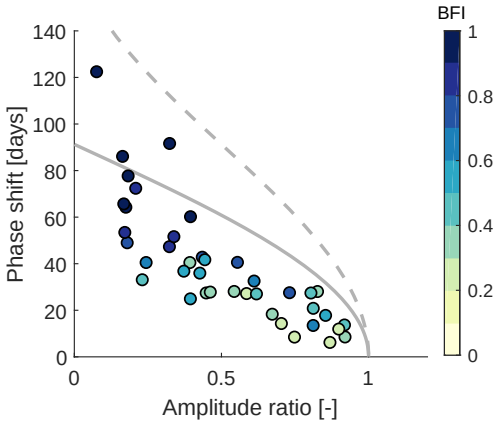

Figure B.8: Amplitude ratio against phase shift for the subset of UK catchments used in the modelling experiment. Grey solid line indicates a single linear reservoir, grey dashed line indicates the outer envelope for two reservoirs in parallel. Colours indicate the BFI.

B.3.2 Robustness of parameter sampling

To test whether the sample size used in the modelling experiment is large enough, we run the models with 2000, 5000, 10000, and 20000 parameter sets, respectively, all generated using Latin Hypercube sampling. The results from all 40 catchments (see Figure B.8) are summarised using box plots. The results are shown in Figure B.9 for IHACRES, and in Figure B.10 for GR4J.

While small differences are visible, the overall pattern is stable. While sampling more parameter sets is unlikely to change the results (except perhaps for some "outliers"), sampling with a different sampling scheme might influence the shapes of the resulting probability distributions. Furthermore, using different parameter ranges might affect the results.

B.3.3 Code availability

A repository with Matlab code used for the analysis and the resulting data is available from `https://github.com/SebastianGnann/Seasonal_signatures_paper_public`. Colours are based on `www.ColorBrewer.org`, by Cynthia A. Brewer, Penn State.

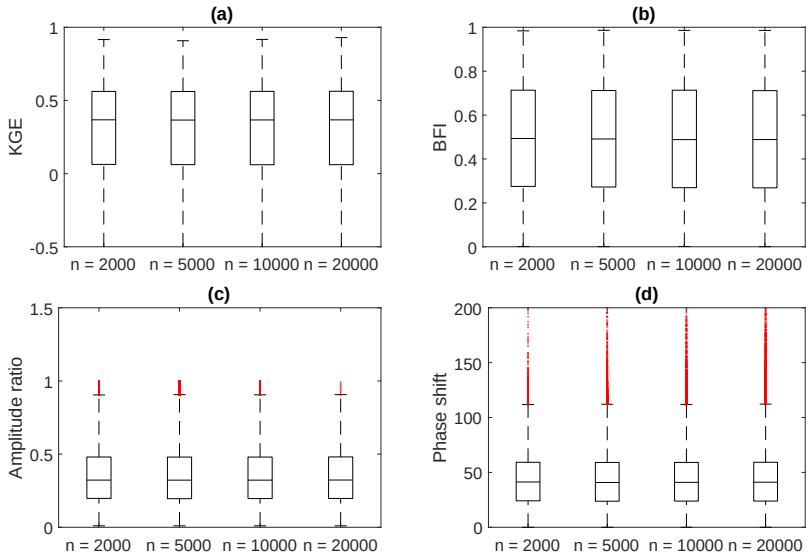

Figure B.9: Box plots summarising model performances for different parameter sample sizes for 40 UK catchments for IHACRES. Boxes indicate the median (middle), the 25th (bottom) and 75th (top) percentiles, respectively. Whiskers show the range of data points not considered outliers. Outliers are shown in red. The metrics used are **(a)** KGE, **(b)** BFI, **(c)** amplitude ratio, and **(d)** phase shift. Note that the y-axes are limited.

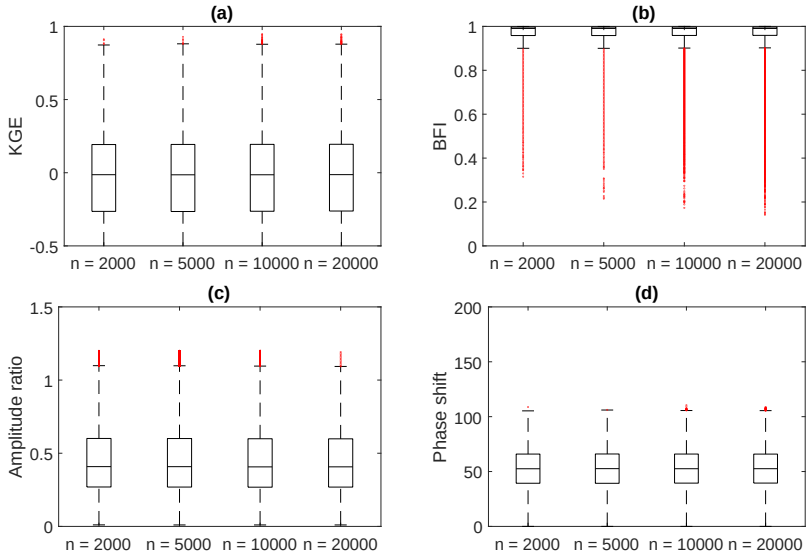

Figure B.10: Box plots summarising model performances for different parameter sample sizes for 40 UK catchments for GR4J. Boxes indicate the median (middle), the 25th (bottom) and 75th (top) percentiles, respectively. Whiskers show the range of data points not considered outliers. Outliers are shown in red. The metrics used are **(a)** KGE, **(b)** BFI, **(c)** amplitude ratio, and **(d)** phase shift. Note that the y-axes are limited.

B.4 Data and code availability

The CAMELS dataset is available from https://ral.ucar.edu/solutions/products/camels. Information about the UK Benchmark Network can be obtained from https://nrfa.ceh.ac.uk/benchmark-network. Streamflow data and catchment attributes are available from https://nrfa.ceh.ac.uk. CEH-GEAR precipitation data are available from https://doi.org/10.5285/33604ea0-c238-4488-813d-0ad9ab7c51ca. CHESS-PE potential evapotranspiration data are available from https://doi.org/10.5285/8baf805d-39ce-4dac-b224-c926ada353b7. A repository with Matlab code used for the analysis and the resulting data are available from https://github.com/SebastianGnann/Seasonal_signatures_paper_public. Map colours are based on www.ColorBrewer.org by Cynthia A. Brewer, Penn State.

APPENDIX C

SUPPORTING INFORMATION FOR CHAPTER 4

C.1 Calculation of catchment attributes

C.1.1 ArcGIS raster preparation

We project every dataset to a WGS84 Web Mercator projection (*Project*). If a dataset consists of multiple sub-datasets (e.g. states), we merge all sub-datasets (*MosaicToNewRaster, merge*). We transform the datasets (polygons, rasters of different resolution) to a raster dataset of approximately 1 km resolution covering all CAMELS shapefiles (*PolygonToRaster, Resample*). We use the Hydrosheds DEM as reference dataset, and snap all other rasters to it. Each cell is assigned the value of the maximum combined area. If resampling is necessary, we choose bilinear interpolation. We then transform the raster to an ASCII file (*RasterToASCII*). This process is mostly done manually, as most datasets differ in size, format, projection, and so forth.

C.1.2 Catchment attribute calculation with Matlab

We use Matlab to calculate catchment averages based on the CAMELS shapefiles and the ASCII datasets created with ArcGIS. We check which grid cells are inside a catchment shapefile and use them to calculate a catchment attribute, for example, the arithmetic mean. While this process can lead to some uncertainties for small catchments and for datasets with fine resolution, we deem it an acceptable compromise between computational effort and precision. As a simple check, we calculated two attributes that are also contained in CAMELS (subsurface porosity and subsurface permeability from GLHYMPS; Gleeson et al., 2014). The resulting catchment averages show excellent agreement (Spearman rank correlation $\rho_s = 0.99$ for both attributes).

C.1.3 Details for different datasets

Glacial areas We edited the Generalized Glacial Limit Lines by hand to obtain two polygons for the extent of the Wisconsin glacial area and the Pre-Wisconsin glacial area.

Geological age The State Geologic Map Compilation contains minimum and maximum values for geological age. We calculated the arithmetic mean for each geological unit and used that for subsequent analyses.

Lakes and wetlands The National Wetlands Inventory comes with one or several files for each state, and thus needs to be merged first. The Python code used to do that with ArcGIS is contained in the corresponding GitHub repository. We grouped the 8 groups contained in the National Wetlands Inventory (see `https://www.fws.gov/wetlands/documents/Wetlands-Mapper-Documentation-Manual-May-2019.pdf` (last access: 14/10/2020)) into four categories: estuarine fraction, freshwater fraction, lake fraction, and other. For Chapter 4, we classified every catchment as surface water dominated which has a combined freshwater fraction and lake fraction of more than 1%.

C.2 Calculation of baseflow signatures

C.2.1 Baseflow estimation

In Chapter 4, we use a baseflow estimate obtained with the smoothed minima method (UKIH method; Institute of Hydrology, 1980) using two different time window values: the "default" 5 days and 90 days. We compared the default method using a 5 day window with another common filter, the Lyne-Hollick filter (Lyne and Hollick, 1979, using a parameter value of 0.925 and three passes). The resulting BFI values are very similar ($\rho_s = 0.98$). Note that BFI_{90} is strongly correlated with the normalised 5% flow quantile $Q_5/\bar{Q}$ ($\rho_s = 0.95$).

C.2.2 Recession analysis

We calculate the median recession exponent β_m following Roques et al. (2017). The "classic" Brutsaert and Nieber (1977) method applied to individual recessions leads to very similar results for the median recession exponent ($\rho_s = 0.99$).

C.2.3 Correlation between baseflow signatures

The three baseflow signatures used in Chapter 4 are correlated and thus not completely independent (see Figure C.1). In particular, BFI_5 and BFI_{90} are strongly correlated as BFI_5 "contains" BFI_{90} (this can be seen in the hydrographs shown in Chapter 4). However, Figure C.1 also shows

that there is considerable variability left to explain. For the same BFI_5, BFI_{90} can vary by up to 0.5 and thus may provide very different information for some regions.

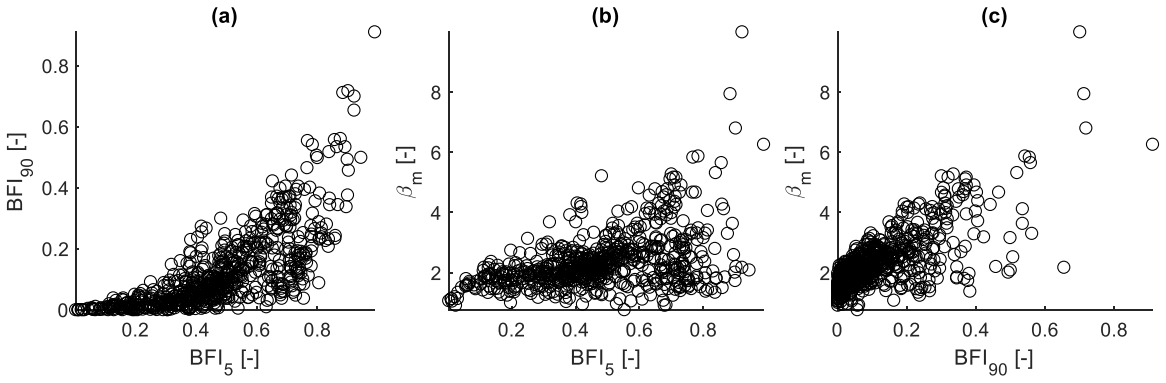

Figure C.1: Scatter plots showing correlations between **(a)** BFI_5 and BFI_{90} ($\rho_s = 0.81$), **(b)** BFI_5 and β_m ($\rho_s = 0.42$), and **(c)** BFI_{90} and β_m ($\rho_s = 0.68$).

C.3 Baseflow signature relationships for catchment subgroups

Figures C.2 and C.3 complement Figure 4.10 which shows scatter plots of median recession exponent β_m vs. BFI_5 for three catchment subgroups.

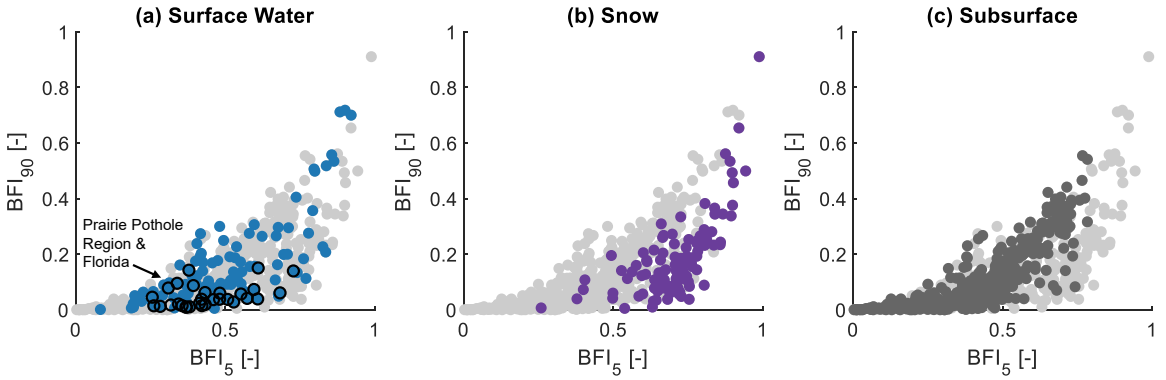

Figure C.2: Scatter plots of median recession exponent β_m vs. BFI_{90} ($\rho_s = 0.68$ for all catchments). Subplots show catchments where water is primarily stored in **(a)** in surface water bodies (>1% of area classified as lake or wetland delineated from the National Wetlands Inventory; $\rho_s = 0.70$ for the subgroup); **(b)** as snow (>30% precipitation falling as snow; $\rho_s = 0.26$); and **(c)** in the subsurface ($\rho_s = 0.82$). Note that each catchment only belongs to one class, with surface water bodies being the first criterion and snow being the second criterion. Note that the y-axis is capped.

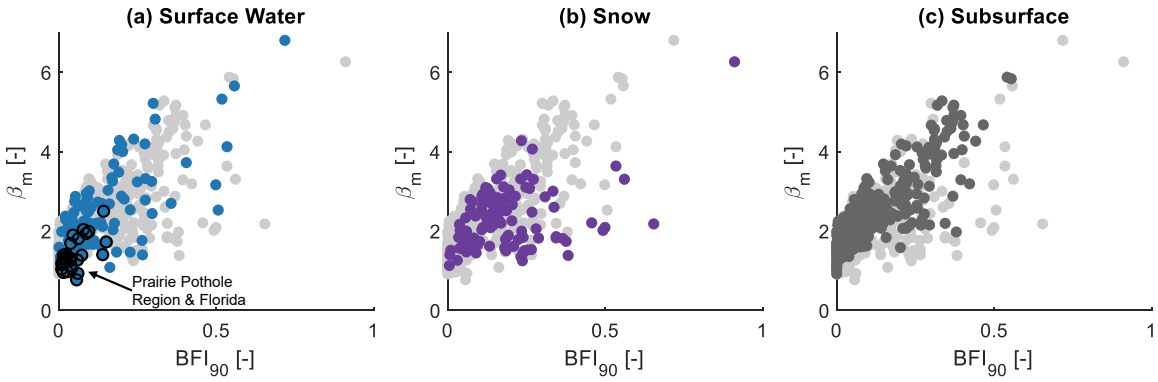

Figure C.3: Scatter plots of median recession exponent BFI_5 vs. BFI_{90} ($\rho_s = 0.81$ for all catchments). Subplots show catchments where water is primarily stored in **(a)** in surface water bodies (>1% of area classified as lake or wetland delineated from the National Wetlands Inventory; $\rho_s = 0.66$ for the subgroup); **(b)** as snow (>30% precipitation falling as snow; $\rho_s = 0.66$); and **(c)** in the subsurface ($\rho_s = 0.87$). Note that each catchment only belongs to one class, with surface water bodies being the first criterion and snow being the second criterion. Note that the y-axis is capped.

C.4 Losing and gaining catchments

To explore how regional groundwater flow Q_{gw} influences baseflow signatures, we followed the idea of Schaller and Fan (2009). We use actual evapotranspiration estimates E_a to calculate Q_{gw} via the water balance:

$$\bar{Q}_{gw} = \bar{P} - \bar{E}_a - \bar{Q} \tag{C.1}$$

where $\bar{Q}_{gw}$ is mean regional groundwater flow, $\bar{P}$ is mean precipitation, $\bar{E}_a$ is mean actual evapotranspiration, and $\bar{Q}$ is mean streamflow. We use Q and P from CAMELS and two different E_a products: MODIS (Mu et al., 2011) and GLEAM (Miralles et al., 2011; Martens et al., 2017). MODIS provides average E_a from 2000 to 2013. GLEAM is available as a NetCDF file with yearly data, which we average from 2000 to 2013 in ArcGIS. We then convert the GIS raster files to ASCII files and extract the catchment averages as described above for the other catchment attributes. We average the other water balance components using Matlab and calculate $\bar{Q}_{gw}$ using Eq. (C.1). Figure C.4 shows maps of estimated mean regional groundwater flow normalised by mean annual precipitation.

In agreement with Schaller and Fan (2009), the results suggest many losing and many gaining catchments all over the US (see Figure C.4). The results did, however, not conclusively agree with information we found in the literature (e.g. Kleeschulte, 2000; Mugel et al., 2009, for the Ozarks). Since all of the water balance components are uncertain, the obtained values for Q_{gw} might rather reflect the uncertainty of the water balance components. The uncertainties for E_a typically exceed 10% (Mu et al., 2011; Khan et al., 2018), which is in the order of magnitude of streamflow in arid catchments (e.g. the Edwards Plateau). Therefore, we decided against the use

of such an estimate at the continental scale, noting that it will be important to obtain reliable estimates of regional groundwater flow to better understand baseflow signatures.

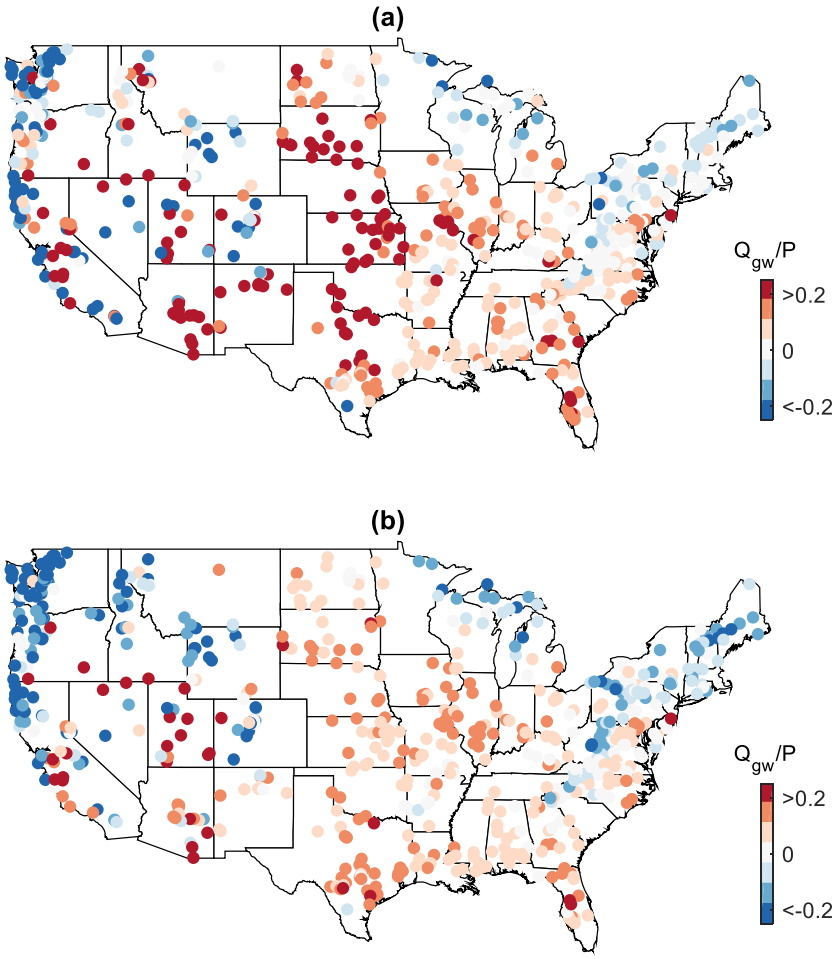

Figure C.4: Estimated mean regional groundwater flow normalised by mean annual precipitation for **(a)** MODIS and **(b)** GLEAM evapotranspiration estimates. Red colours indicate losing catchments, blue colours indicate gaining catchments.

C.5 Data and code availability

New catchment attributes and baseflow signatures are available from `https://doi.org/10.5281/zenodo.4071983`. Matlab and Python code used for this study are available from `https://github.com/SebastianGnann/Baseflow_signatures`. Links to the source datasets can be found in Table C.1. Colours are based on `www.ColorBrewer.org` by Cynthia A. Brewer, Penn State.

Table C.1: Dataset URLs.

Dataset name	URL
CAMELS	https://ral.ucar.edu/solutions/products/camels (last access: 14/10/2020)
Datasets in CAMELS	
STATSGO	http://www.soilinfo.psu.edu/index.cgi?soil_data&index.html (last access: 14/10/2020)
GLiM	https://www.geo.uni-hamburg.de/en/geologie/forschung/geochemie/glim.html (last access: 14/10/2020)
GLHYMPS	https://dataverse.scholarsportal.info/dataset.xhtml?persistentId=doi:10.5683/SP2/DLGXYO (last access: 14/10/2020)
Additional datasets	
HydroSHEDS	https://hydrosheds.org/downloads (last access: 14/10/2020)
Generalized Glacial Limit Lines	https://purl.stanford.edu/vz874sc7648 (last access: 14/10/2020)
Physiographic Divisions of the US	https://water.usgs.gov/GIS/metadata/usgswrd/XML/physio.xml#Metadata_Reference_Information (last access: 14/10/2020)
USGS Geological Map	https://pubs.er.usgs.gov/publication/ds1052 (last access: 14/10/2020)
Principal Aquifers of the US	https://water.usgs.gov/ogw/aquifer/map.html (last access: 14/10/2020)
MGS Sinkhole Points	https://apps5.mo.gov/geostrat/ (last access: 14/10/2020)
TWDB Major Aquifers	http://www.twdb.texas.gov/mapping/gisdata.asp (last access: 14/10/2020)
National Wetlands Inventory	https://www.fws.gov/wetlands/index.html (last access: 14/10/2020)
MODIS	http://www.ntsg.umt.edu/project/modis/mod16.php (last access: 14/10/2020)
GLEAM	https://www.gleam.eu/ (last access: 14/10/2020)

APPENDIX D

TOSSH: A TOOLBOX FOR STREAMFLOW SIGNATURES IN HYDROLOGY

This appendix contains information on TOSSH: A toolbox for streamflow signatures in hydrology, an open source Matlab toolbox that provides accessible, standardised signature calculations. The paper describing the toolbox has been published as a research article in Environmental Modelling & Software.

TOSSH has been developed to approximately equal parts by HKM and SJG, with input from GC, RAW, and NJKH. It is therefore listed in the appendix and not as research chapter in the main body of the thesis.

Citation: Gnann S. J., Coxon G., Woods R. A., Howden N. J. K., and McMillan H. K. (2021). TOSSH: A toolbox for streamflow signatures in hydrology. Environmental Modelling & Software, 138:104983. https://doi.org/10.1016/j.envsoft.2021.104983

Abstract

We present a Matlab toolbox to calculate hydrologic signatures, which are metrics that quantify streamflow dynamics. Signatures are widely used for catchment characterisation, hydrologic model evaluation, and assessment of instream habitat, but standardisation across applications and advice on signature selection is lacking. The toolbox provides accessible, standardised signature calculations, with clear information on methodological decisions and recommended parameter values. The toolbox implements three categories of signatures: basic signatures that describe the five components of a natural streamflow regime, signatures from benchmark papers, and an extended set of process-based signatures. The toolbox is designed for ease of use, including

documentation, workflow scripts and example data to demonstrate implementation procedures, and visualisation options. We demonstrate the accuracy and robustness of the signature calculations by applying reproducible workflows to large streamflow datasets. The modular design of the toolbox allows for flexibility and easy future expansion. The toolbox is available from `https://github.com/TOSSHtoolbox/TOSSH` (`https://doi.org/10.5281/zenodo.4313275`).

APPENDIX E

UNANSWERED QUESTIONS ON THE BUDYKO FRAMEWORK

This appendix contains information on an opinion paper entitled "Unanswered questions on the Budyko framework" that has been published in Hydrological Processes. Slight modifications have been made to better fit the general layout of this thesis.

WRB conceived the idea and prepared the manuscript with contributions from SJG and RAW. It is therefore listed in the appendix and not as research chapter in the main body of the thesis.

Citation: Berghuijs, W. R., Gnann, S. J., and Woods, R. A. (2020). Unanswered questions on the Budyko framework. Hydrological Processes, 34:5699—5703. https://doi.org/10.1002/hyp.13958

Abstract

Landscapes and their hydrology are complex and sui generis. As a result, few theories exist that (without calibration) usefully describe or predict catchment-scale hydrological behaviour (Beven, 2000; Sivapalan, 2005). The Budyko hypothesis (Budyko, 1951, 1974) is a rare exception: its simple parameterisation of how aridity (the ratio of long-term mean precipitation to long-term mean potential evapotranspiration) controls the long-term mean partitioning of precipitation into streamflow and evapotranspiration captures the behaviour of many catchments around the world. In recent years, the Budyko framework has increasingly been used to interpret and predict (often non-stationary) water balances. While uses of the framework have become diverse and widespread, they are typically founded in common principles that rely on largely untested assumptions and strongly relate to questions for which no clear answers exist. Therefore, we

believe that answering several basic questions around the Budyko framework can strengthen (or invalidate) many old, recent, and future applications. We realise that similar questions have been contemplated by others, but we hope that presenting them in the following manner may prove useful.

APPENDIX F

CURRICULUM VITAE

Sebastian Johannes Gnann
sebastian.gnann@bristol.ac.uk

Higher Education

01/2017–present	**PhD Water Informatics Science & Engineering (WISE CDT)** University of Bristol, Bristol, UK Thesis: Baseflow generation at the catchment scale — An investigation using comparative hydrology Supervisors: Dr. Nicholas J. K. Howden, Dr. Ross A. Woods
01/2020—03/2020	**Research Visit** San Diego State University, San Diego, USA Supervisor: Dr. Hilary K. McMillan
10/2014—10/2016	**M.Sc. Applied & Environmental Geoscience** University of Tübingen, Tübingen, Germany Thesis: Joint spatial interpolation of groundwater quality parameters Supervisors: Prof. Dr-Ing. Olaf A. Cirpka, Dr-Ing. Claus P. Haslauer
10/2010—09/2013	**B.Sc. Umweltnaturwissenschaften (Environmental Sciences)** University of Tübingen, Tübingen, Germany Thesis: Redox reactions between iron-bearing clays (SWy-2) and model quinones (AQDS) Supervisors: Prof. Dr. Stefan B. Haderlein, Dr. Silvia Orsetti

Work and Teaching Experience

10/2018—11/2020	Teaching assistant: Water Resources Project, Numerical Analysis Using Matlab (University of Bristol)
04/2012-–02/2014	Lab work in the work group Environmental Mineralogy & Chemistry (University of Tübingen)
04/2012—07/2013	Tutorial in Biogeochemistry (University of Tübingen)
10/2011—02/2012	Tutorial in Geomicrobiology (University of Tübingen)

Civilian Service

08/2009—04/2010	**Alternative civilian service (Zivildienst)** Parksanatorium Aulendorf, Germany

Skills and Service

Languages	German (native language), English (fluent)
Software	Matlab, Python, ArcGIS, LaTeX
Events	Organiser of group seminars, journal club, diversity & inclusion workshops (University of Bristol), session convener at EGU General Assembly
Reviews	Reviewer for Water Resources Research, Journal of Hydroinformatics, Computers and Geoscience, Environmental Science & Technology, Journal of Hydrology, Earth System Science Data

Conferences

04/2018	Integrated Hydrosystem Modelling 2018 Conference (Tübingen, Germany)
04/2018	04/2018 EGU General Assembly 2018 (Vienna, Austria)
04/2019	EGU General Assembly 2019 (Vienna, Austria)
07/2019	Catchment Science: Interactions of Hydrology, Biology and Geochemistry (Gordon Research Conference in Andover, NH, USA)
05/2020	EGU General Assembly 2020 (Online conference)

Publications

Gnann, S. J., Coxon, G., Woods, R. A., Howden, N. J. K., and McMillan H. K. (2021). TOSSH: A toolbox for streamflow signatures in hydrology. Environmental Modelling & Software, 138:104983.

Gnann, S. J., McMillan, H. K., Woods, R. A. and Howden, N. J. K. (2020). Including regional knowledge improves baseflow signature predictions in large sample hydrology. Water Resources Research, 56:e2020WR028354.

Berghuijs, W. R., **Gnann, S. J.**, and Woods, R. A. (2020). Unanswered questions on the Budyko framework. Hydrological Processes, 34:5699—5703.

Gnann, S. J., Howden, N. J. K. and Woods, R. A. (2020). Hydrological signatures describing the translation of climate seasonality into streamflow seasonality. Hydrology and Earth System Sciences, 24(2):561–580.

Gnann, S. J., Woods, R. A. and Howden, N. J. K. (2019). Is there a baseflow Budyko curve? Water Resources Research, 55(4):2838–2855.

Gnann, S. J., Allmendinger, M. C., Haslauer, C. P. and Bárdossy, A. (2018). Improving copula-based spatial interpolation with secondary data. Spatial Statistics, 28:105–127.

References

Adamski, J. C., Petersen, J. C., Freiwald, D. A., and Davis, J. V. (1995). Environmental and hydrologic setting of the Ozark Plateaus study unit, Arkansas, Kansas, Missouri, and Oklahoma. *U.S. Geological Survey Water-Resources Investigations Report*, 94-4022.

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K. J. A., and Mendoza, P. A. (2020). Large-sample hydrology: Recent progress, guidelines for new datasets and grand challenges. *Hydrological Sciences Journal*, 65(5):712–725.

Addor, N. and Melsen, L. A. (2019). Legacy, rather than adequacy, drives the selection of hydrological models. *Water Resources Research*, 55(1):378–390.

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., and Clark, M. P. (2018). A ranking of hydrological signatures based on their predictability in space. *Water Resources Research*, 54(11):8792–8812.

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

Ameli, A. A. and Creed, I. F. (2017). Quantifying hydrologic connectivity of wetlands to surface water systems. *Hydrology and Earth System Sciences*, 21(3):1791–1808.

Andréassian, V. and Perrin, C. (2012). On the ambiguous interpretation of the Turc-Budyko nondimensional graph. *Water Resources Research*, 48(10):1–5.

Barker, R. A. and Ardis, A. F. (1996). Hydrogeologic framework of the Edwards-Trinity aquifer system, west-central Texas. *U.S. Geological Survey Professional Paper*, 1421 B.

Beck, H. E., de Roo, A., and van Dijk, A. I. J. M. (2015). Global maps of streamflow characteristics based on observations from several thousand catchments. *Journal of Hydrometeorology*, 16(4):1478–1501.

Beck, H. E., van Dijk, A. I. J. M., Miralles, D. G., De Jeu, R. A. M., Bruijnzeel, L. A., McVicar, T. R., and Schellekens, J. (2013). Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. *Water Resources Research*, 49(12):7843–7863.

Berghuijs, W. R., Larsen, J. R., van Emmerik, T. H., and Woods, R. A. (2017). A global assessment of runoff sensitivity to changes in precipitation, potential evaporation, and other factors. *Water Resources Research*, 53(10):8475–8486.

Berghuijs, W. R., Sivapalan, M., Woods, R. A., and Savenije, H. H. G. (2014). 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.

Berghuijs, W. R. and Woods, R. A. (2016). Correspondence: Space-time asymmetry undermines water yield assessment. *Nature Communications*, 7:1–2.

Bettis, E. A., Benn, D. W., and Hajic, E. R. (2008). Landscape evolution, alluvial architecture, environmental history, and the archaeological record of the Upper Mississippi River Valley. *Geomorphology*, 101(1-2):362–377.

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

Beven, K. (2002). Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrological Processes*, 16(2):189–206.

Beven, K. (2007). Towards integrated environmental models of everywhere: Uncertainty, data and modelling as a learning process. *Hydrology and Earth System Sciences*, 11(1):460–467.

Beven, K. (2012). *Rainfall-Runoff Modelling: The Primer, Second Ed.* John Wiley & Sons.

Beven, K. (2015). What we see now: Event-persistence and the predictability of hydro-eco-geomorphological systems. *Ecological Modelling*, 298:4–15.

Beven, K. (2019). How to make advances in hydrological modelling. *Hydrology Research*, 50(6):1481–1494.

Beven, K. and Germann, P. (1982). Macropores and water flow in soils. *Water Resources Research*, 18(5):1311–1325.

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

Black, P. E. (1997). Watershed functions. *Journal of the American Water Resources Association*, 33(1):1–11.

Bloomfield, J. P., Allen, D. J., and Griffiths, K. J. (2009). Examining geological controls on baseflow index (BFI) using regression analysis: An illustration from the Thames Basin, UK. *Journal of Hydrology*, 373(1-2):164–176.

Boorman, D. B., Hollis, J. M., and Lilly, A. (1995). Hydrology of soil types: A hydrologically-based classification of the soils of United Kingdom. *Institute of Hydrology Report*, 126.

Boughton, W. (1993). A hydrograph-based model for estimating the water yield of ungauged catchments. In *Hydrology and Water Resources Symposium, Newcastle, IEAust, 1993*.

Brutsaert, W. and Nieber, J. L. (1977). Regionalized drought flow hydrographs from a mature glaciated plateau. *Water Resources Research*, 13(3):637–643.

Budyko, M. I. (1951). On climatic factors of runoff (in Russian). *Problemy Fiz Geografii*, 16.

Budyko, M. I. (1974). *Climate and Life: English Ed. edited by David H. Miller*. New York: Academic Press.

Buttle, J. (2016). Dynamic storage: A potential metric of inter-basin differences in storage properties. *Hydrological Processes*, 30(24):4644–4653.

Buttle, J. (2018). Mediating stream baseflow response to climate change: The role of basin storage. *Hydrological Processes*, 32(3):363–378.

Carlier, C., Wirth, S. B., Cochand, F., Hunkeler, D., and Brunner, P. (2018). Geology controls streamflow dynamics. *Journal of Hydrology*, 566(July):756–769.

Carlier, C., Wirth, S. B., Cochand, F., Hunkeler, D., and Brunner, P. (2019). Exploring geological and topographical controls on low flows with hydrogeological models. *Groundwater*, 57(1):48–62.

Cartwright, I., Gilfedder, B., and Hofmann, H. (2014). Contrasts between estimates of baseflow help discern multiple sources of water contributing to rivers. *Hydrology and Earth System Sciences*, 18(1):15.

Cayan, D. R., Riddle, L. G., and Aguado, E. (1993). The influence of precipitation and temperature on seasonal streamflow in California. *Water Resources Research*, 29(4):1127–1140.

Chapman, T. G. (1991). Comment on "Evaluation of automated techniques for base flow and recession analyses" by R. J. Nathan and T. A. McMahon. *Water Resources Research*, 27(7):1783–1784.

Cheng, J. and Druzdzel, M. J. (2000). Latin hypercube sampling in Bayesian networks. In *FLAIRS Conference*, pages 287–292.

Clark, M. P., Bierkens, M. F., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V. R., Cai, X., Wood, A. W., and Peters-Lidard, C. D. (2017). The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. *Hydrology and Earth System Sciences*, 21(7):3427–3440.

Clark, M. P., McMillan, H. K., Collins, D. B., Kavetski, D., and Woods, R. A. (2011). Hydrological field data from a modeller's perspective: Part 2: Process-based evaluation of model hypotheses. *Hydrological Processes*, 25(4):523–543.

Clark, M. P., Rupp, D. E., Woods, R. A., Tromp-van Meerveld, H. J., Peters, N. E., and Freer, J. E. (2009). Consistency between hydrological models and field observations: Linking processes at the hillslope scale to hydrological responses at the watershed scale. *Hydrological Processes*, 23(2):311–319.

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. (2016). Improving the theoretical underpinnings of process-based hydrologic models. *Water Resources Research*, 52(3):2350–2365.

Clausen, B. and Biggs, B. (2000). Flow variables for ecological studies in temperate streams: Groupings based on covariance. *Journal of Hydrology*, 237(3-4):184–197.

Cohen, M. J., Creed, I. F., Alexander, L., Basu, N. B., Calhoun, A. J., Craft, C., D'Amico, E., DeKeyser, E., Fowler, L., Golden, H. E., Jawitz, J. W., Kalla, P., Kirkman, L. K., Lane, C. R., Lang, M., Leibowitz, S. G., Lewis, D. B., Marton, J., McLaughlin, D. L., Mushet, D. M., Raanan-Kiperwas, H., Rains, M. C., Smith, L., and Walls, S. C. (2016). Do geographically isolated wetlands influence landscape functions? *Proceedings of the National Academy of Sciences of the United States of America*, 113(8):1978–1986.

Colwell, R. K. (1974). Predictability, constancy, and contingency of periodic phenomena. *Ecology*, 55(5):1148–1153.

Cooke, C. D. and Buttle, J. M. (2020). Assessing basin storage: Comparison of hydrometric-and tracer-based indices of dynamic and total storage. *Hydrological Processes*, 34(9):2012–2031.

Costelloe, J. F., Peterson, T. J., Halbert, K., Western, A., and McDonnell, J. J. (2015). Groundwater surface mapping informs sources of catchment baseflow. *Hydrology and Earth System Sciences*, 19:1599–1613.

Court, A. (1962). Measures of streamflow timing. *Journal of Geophysical Research*, 67(11):4335–4339.

Coxon, G., Freer, J. E., Lane, R. A., Dunne, T., Knoben, W. J. M., Howden, N. J. K., Quinn, N., Wagener, T., and Woods, R. A. (2019). DECIPHeR v1: Dynamic fluxEs and ConnectIvity for Predictions of HydRology. *Geoscientific Model Development*, 12(6):2285–2306.

Croke, B. F. and Jakeman, A. J. (2004). A catchment moisture deficit module for the IHACRES rainfall-runoff model. *Environmental Modelling and Software*, 19(1):1–5.

DeWalle, D. R., Edwards, P. J., Swistock, B. R., Aravena, R., and Drimmie, R. J. (1997). Seasonal isotope hydrology of three Appalachian forest catchments. *Hydrological Processes*, 11(15):1895–1906.

Dooge, J. C. I. (1973). *Linear Theory of Hydrologic Systems (Technical Bulletin No. 1468)*. Agricultural Research Service, U.S. Department of Agriculture.

Dooge, J. C. I. (1986). Looking for hydrologic laws. *Water Resources Research*, 22(9 S):46S–58S.

Dralle, D. N., Hahm, W. J., Rempe, D. M., Karst, N., Anderegg, L. D., Thompson, S. E., Dawson, T. E., and Dietrich, W. E. (2020). Plants as sensors: Vegetation response to rainfall predicts root-zone water storage capacity in Mediterranean-type climates. *Environmental Research Letters*, 15(10):104074.

Dralle, D. N., Hahm, W. J., Rempe, D. M., Karst, N. J., Thompson, S. E., and Dietrich, W. E. (2018). Quantification of the seasonal hillslope water storage that does not drive streamflow. *Hydrological Processes*, 32(13):1978–1992.

Dralle, D. N., Karst, N. J., Charalampous, K., Veenstra, A., and Thompson, S. E. (2017). Event-scale power law recession analysis: Quantifying methodological uncertainty. *Hydrology and Earth System Sciences*, 21(1):65–81.

Dralle, D. N., Karst, N. J., and Thompson, S. E. (2015). a, b careful! Consequences of scale invariance in power-law models of the streamflow recession. *Geophysical Research Letters*, 42:9285–9293.

Duan, Q., Schaake, J., Andréassian, V., Franks, S. W., Goteti, G., Gupta, H. V., Gusev, Y., Habets, F., Hall, A., Hay, L. E., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E. F. (2006). Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, 320(1-2):3–17.

Eckhardt, K. (2005). How to construct recursive digital filters for baseflow separation. *Hydrological Processes*, 19(2):507–515.

Ehret, U., Gupta, H. V., Sivapalan, M., Weijs, S. V., Schymanski, S. J., Blöschl, G., Gelfan, A. N., Harman, C. J., Kleidon, A., Bogaard, T. A., Wang, D., Wagener, T., Scherer, U., Zehe, E., Bierkens, M. F., Di Baldassarre, G., Parajka, J., van Beek, L. P. H., Van Griensven, A., Westhoff, M. C., and Winsemius, H. C. (2014). Advancing catchment hydrology to deal with predictions under change. *Hydrology and Earth System Sciences*, 18(2):649–671.

Eriksson, E. (1971). Compartment models and reservoir theory. *Annual Review of Ecology and Systematics*, 2(1):67–84.

Erskine, A. and Papaioannou, A. (1997). The use of aquifer response rate in the assessment of groundwater resources. *Journal of Hydrology*, 202(1-4):373–391.

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

Falkenmark, M. and Chapman, T. (1989). *Comparative Hydrology: An Ecological Approach to Land and Water Resources*. The Unesco Press.

Fan, Y. (2019). Are catchments leaky? *WIREs Water*, 6(6):1–25.

Farmer, D., Sivapalan, M., and Jothityangkoon, C. (2003). Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: Downward approach to water balance analysis. *Water Resources Research*, 39(2):1–21.

Farvolden, R. (1963). Geologic controls on ground-water storage and base flow. *Journal of Hydrology*, 1(3):219–249.

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M. H., Ghezzehei, T. A., Hengl, T., Kollet, S., Agam, N., and Avissar, R. (2020). Soil structure is an important omission in Earth System Models. *Nature Communications*, 11(1).

Fenicia, F., Kavetski, D., and Savenije, H. H. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, 47(11).

Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., and Freer, J. E. (2014). Catchment properties, function, and conceptual model representation: Is there a correspondence? *Hydrological Processes*, 28(4):2451–2467.

Fenicia, F., Savenije, H., Matgen, P., and Pfister, L. (2006). Is the groundwater reservoir linear? Learning from data in hydrological modelling. *Hydrology and Earth System Sciences*, 10:139–150.

Fenneman, N. and Johnson, D. (1946). Physiographic divisions of the conterminous U.S., available at `https://water.usgs.gov/GIS/metadata/usgswrd/XML/physio.xml` (last access: 14 October 2020).

Ficklin, D. L., Robeson, S. M., and Knouft, J. H. (2016). Impacts of recent climate change on trends in baseflow and stormflow in United States watersheds. *Geophysical Research Letters*, 43(10):5079–5088.

Fowler, K. J. A., Coxon, G., Freer, J. E., Peel, M. C., Wagener, T., Western, A., Woods, R. A., and Zhang, L. (2018). Simulating runoff under changing climatic conditions: A framework for model improvement. *Water Resources Research*, 54(12):9812–9832.

Fowler, K. J. A., Knoben, W., Peel, M. C., Peterson, T., Ryu, D., Saft, M., Seo, K.-W., and Western, A. (2020). Many commonly used rainfall-runoff models lack long, slow dynamics: Implications for runoff projections. *Water Resources Research*, 56(5):e2019WR025286.

Freeze, R. A. (1972). Role of subsurface flow in generating surface runoff: 1. Base flow contributions to channel flow. *Water Resources Research*, 8(3):609–623.

Gleeson, T., Moosdorf, N., Hartmann, J., and Van Beek, L. P. (2014). A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. *Geophysical Research Letters*, 41(11):3891–3898.

Godsey, S. and Kirchner, J. W. (2014). Dynamic, discontinuous stream networks: Hydrologically driven variations in active drainage density, flowing channels and stream order. *Hydrological Processes*, 28(23):5791–5803.

Gonzales, A. L., Nonner, J., Heijkers, J., and Uhlenbrook, S. (2009). Comparison of different base flow separation methods in a lowland catchment. *Hydrology and Earth System Sciences*, 13(11):2055–2068.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2):80–91.

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

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

Hall, F. R. (1968). Base-flow recessions – A review. *Water Resources Research*, 4(5):973–983.

Harman, C. J. (2019). Age-ranked storage-discharge relations: A unified description of spatially-lumped flow and water age in hydrologic systems. *Water Resources Research*, pages 1–23.

Harman, C. J. and Sivapalan, M. (2009). Effects of hydraulic conductivity variability on hillslope-scale shallow subsurface flow response and storage-discharge relations. *Water Resources Research*, 45(1).

Harman, C. J., Sivapalan, M., and Kumar, P. (2009). Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics. *Water Resources Research*, 45(9):1–13.

Harman, C. J. and Troch, P. A. (2014). What makes Darwinian hydrology "Darwinian"? Asking a different kind of question about landscapes. *Hydrology and Earth System Sciences*, 18(2):417–433.

Harman, C. J., Troch, P. A., and Sivapalan, M. (2011). Functional model of water balance variability at the catchment scale: 2. Elasticity of fast and slow runoff components to precipitation change in the continental United States. *Water Resources Research*, 47(2):1–12.

Harrigan, S., Hannaford, J., Muchan, K., and Marsh, T. J. (2017). Designation and trend analysis of the updated UK Benchmark Network of river flow stations: the UKBN2 dataset. *Hydrology Research*, 49(2):552–567.

Harrigan, S., Prudhomme, C., Parry, S., Smith, K., and Tanguy, M. (2018). Benchmarking ensemble streamflow prediction skill in the UK. *Hydrology and Earth System Sciences*, 22(3):2023–2039.

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., and Weiler, M. (2014). Karst water resources in a changing world: Review of hydrological modeling approaches. *Reviews of Geophysics*, 52(3):218–242.

Hartmann, J. and Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, 13(12):1–37.

Harvey, E. J. (1981). Ground water in the Springfield-Salem Plateaus of southern Missouri and northern Arkansas. *U.S. Geological Survey Water-Resources Investigations Report*, 80-101.

Hays, P. D., Knierim, K. J., Breaker, B., Westerman, D. A., and Clark, B. R. (2016). Hydrogeology and hydrologic conditions of the Ozark Plateaus aquifer system. *U.S. Geological Survey Scientific Investigations Report*, 2016-5137.

He, Y., Bárdossy, A., and Zehe, E. (2011). A review of regionalisation for continuous streamflow simulation. *Hydrology and Earth System Sciences*, 15(11):3539.

Hewlett, J. D. and Hibbert, A. R. (1967). Factors affecting the response of small watersheds to precipitation in humid areas. *Forest Hydrology*, 1:275–290.

Horton, J., San Juan, C., and Stoeser, D. (2017). The State Geologic Map Compilation (SGMC) geodatabase of the conterminous United States (ver. 1.1, August 2017), available at `https://pubs.er.usgs.gov/publication/ds1052` (last access: 14 October 2020).

Horton, R. E. (1933). The role of infiltration in the hydrologic cycle. *Eos, Transactions American Geophysical Union*, 14(1):446–460.

Hrachowitz, M., Benettin, P., Van Breukelen, B. M., Fovet, O., Howden, N. J. K., Ruiz, L., Van Der Velde, Y., and Wade, A. J. (2016). Transit times – the link between hydrology and water quality at the catchment scale. *WIREs Water*, 3(5):629–657.

Hrachowitz, M. and Clark, M. P. (2017). HESS Opinions: The complementary merits of competing modelling philosophies in hydrology. *Hydrology and Earth System Sciences*, 21(8):3953–3973.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J. E., Savenije, H. H. G., and Gascuel-Odoux, C. (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resources Research*, 50(9):7445–7469.

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

Huyck, A. A., Pauwels, V. R., and Verhoest, N. E. (2005). A base flow separation algorithm based on the linearized Boussinesq equation for complex hillslopes. *Water Resources Research*, 41(8).

Institute of Hydrology (1980). *Low Flow Studies Report No. 1: Research Report*. Institute of Hydrology.

Jachens, E. R., Rupp, D. E., Roques, C., and Selker, J. S. (2020). Recession analysis revisited: Impacts of climate on parameter estimation. *Hydrology and Earth System Sciences*, 24(3):1159–1170.

Jasechko, S., Birks, S. J., Gleeson, T., Wada, Y., Fawcett, P. J., Sharp, Z. D., McDonnell, J. J., and Welker, J. M. (2014). The pronounced seasonality of global groundwater recharge. *Water Resources Research*, 50(11):8845–8867.

Jefferson, A., Grant, G. E., Lewis, S. L., and Lancaster, S. T. (2010). Coevolution of hydrology and topography on a basalt landscape in the Oregon Cascade Range, USA. *Earth Surface Processes and Landforms*, 35(7):803–816.

Jehn, F. U., Bestian, K., Breuer, L., Kraft, P., and Houska, T. (2020). Using hydrological and climatic catchment clusters to explore drivers of catchment behavior. *Hydrology and Earth System Sciences*, 24(3):1081–1100.

Jothityangkoon, C., Sivapalan, M., and Farmer, D. (2001). Process controls of water balance variability in a large semi-arid catchment: Downward approach to hydrological model development. *Journal of Hydrology*, 254(1-4):174–198.

Käser, D. and Hunkeler, D. (2016). Contribution of alluvial groundwater to the outflow of mountainous catchments. *Water Resources Research*, 52(2):680–697.

Khan, M. S., Liaqat, U. W., Baik, J., and Choi, M. (2018). Stand-alone uncertainty characterization of GLEAM, GLDAS and MOD16 evapotranspiration products using an extended triple collocation approach. *Agricultural and Forest Meteorology*, 252:256–268.

Kirchner, J. W. (2003). A double paradox in catchment hydrology and geochemistry. *Hydrological Processes*, 17(4):871–874.

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, 42(3):1–5.

Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research*, 45(2):1–34.

Kirchner, J. W. (2016). Aggregation in environmental systems – Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments. *Hydrology and Earth System Sciences*, 20(1):279–297.

Klaus, J. and McDonnell, J. J. (2013). Hydrograph separation using stable isotopes: Review and evaluation. *Journal of Hydrology*, 505:47–64.

Kleeschulte, M. J. (2000). Ground-and surface-water relations in the Eleven Point and Current River Basins, south-central Missouri. *U.S. Geological Survey Fact Sheet*, 032-00.

Klemeš, V. (1986). Dilettantism in hydrology: Transition or destiny? *Water Resources Research*, 22(9S):177S–188S.

Knoben, W. J. M., Freer, J. E., Fowler, K. J. A., Peel, M. C., and Woods, R. A. (2019a). 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.

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., and Woods, R. (2020). A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. *Water Resources Research*, 56(9):e2019WR025975.

Knoben, W. J. M., Woods, R. A., and Freer, J. E. (2018). A quantitative hydrological climate classification evaluated with independent streamflow data. *Water Resources Research*, 54(7):5088–5109.

Knoben, W. J. M., Woods, R. A., and Freer, J. E. (2019b). Global bimodal precipitation seasonality: A systematic overview. *International Journal of Climatology*, 39(1):558–567.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, 55(12):11344–11354.

Kuentz, A., Arheimer, B., Hundecha, Y., and Wagener, T. (2017). Understanding hydrologic variability across Europe through catchment classification. *Hydrology and Earth System Sciences*, 21(6):2863–2879.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.

Laaha, G. and Blöschl, G. (2006). Seasonality indices for regionalizing low flows. *Hydrological Processes*, 20(18):3851–3878.

Laaha, G., Demuth, S., Hisdal, H., Kroll, C. N., van Lanen, H. A. J., Nester, T., Rogger, M., Sauquet, E., Tallaksen, L. M., Woods, R. A., and Young, A. (2013). Prediction of low flows in ungauged basins. In *Runoff Prediction in Ungauged Basins*, pages 163–188. Cambridge University Press.

Lacey, G. C. and Grayson, R. B. (1998). Relating baseflow to catchment properties in south-eastern Australia. *Journal of Hydrology*, 204(1-4):231–250.

Lane, R. A., Coxon, G., Freer, J. E., Wagener, T., Johnes, P. J., Bloomfield, J. P., Greene, S., Macleod, C. J. A., and Reaney, S. M. (2019). Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain. *Hydrology and Earth System Sciences*, 23(10):4011–4032.

Larson, G. and Schaetzl, R. (2001). Origin and evolution of the Great Lakes. *Journal of Great Lakes Research*, 27(4):518–546.

Lehner, B., Verdin, K., and Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, 89(10):93–94.

Longobardi, A. and Villani, P. (2008). Baseflow index regionalization analysis in a Mediterranean area and data scarcity context: Role of the catchment permeability index. *Journal of Hydrology*, 355(1-4):63–75.

L'vovich, M. I. (1979). *World Water Resources and Their Future, Translated Ed.* American Geophysical Union.

Lyne, V. and Hollick, M. (1979). Stochastic time-variable rainfall-runoff modelling. In *Institute of Engineers Australia National Conference*, pages 89–93. Institute of Engineers Australia Barton, Australia.

Martens, B., Miralles, D. G., Lievens, H., Van Der Schalie, R., De Jeu, R. A., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, 10(5):1903–1925.

McDonnell, J. J. and Beven, K. (2014). Debates – The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. *Water Resources Research*, 50(6):5342–5350.

McDonnell, J. J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R. P., Kirchner, J. W., Roderick, M. L., Selker, J. S., and Weiler, M. (2007). Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research*, 43(7):1–6.

McDonnell, J. J. and Woods, R. A. (2004). On the need for catchment classification. *Journal of Hydrology*, 299(1-2):2–3.

McGlynn, B. L., McDonnell, J. J., and Brammer, D. D. (2002). A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchments, New Zealand. *Journal of Hydrology*, 257(1-4):1–26.

McGuire, K. J. and McDonnell, J. J. (2006). A review and evaluation of catchment transit time modeling. *Journal of Hydrology*, 330(3-4):543–563.

McLaughlin, D. L., Kaplan, D. A., and Cohen, M. J. (2014). A significant nexus: Geographically isolated wetlands influence landscape hydrology. *Water Resources Research*, 50(9):7153–7166.

McMillan, H. K. (2020). Linking hydrologic signatures to hydrologic processes: A review. *Hydrological Processes*, 34(6):1393–1409.

McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M., and Woods, R. A. (2011). Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure. *Hydrological Processes*, 25(4):511–522.

McMillan, H. K., Gueguen, M., Grimon, E., Woods, R. A., Clark, M. P., and Rupp, D. E. (2014). Spatial variability of hydrological processes and model structure diagnostics in a 50km^2 catchment. *Hydrological Processes*, 28(18):4896–4913.

McMillan, H. K., Krueger, T., and Freer, J. E. (2012). Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrological Processes*, 26(26):4078–4111.

McMillan, H. K., Westerberg, I. K., and Branger, F. (2017). Five guidelines for selecting hydrological signatures. *Hydrological Processes*, 31(26):4757–4761.

McNamara, J. P., Tetzlaff, D., Bishop, K., Soulsby, C., Seyfried, M., Peters, N. E., Aulenbach, B. T., and Hooper, R. (2011). Storage as a metric of catchment comparison. *Hydrological Processes*, 25(21):3364–3371.

Miller, D. A. and White, R. A. (1998). A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. *Earth Interactions*, 2(2):1–26.

Milly, P. C. D. (1994). Climate, soil water storage, and the average annual water balance. *Water Resources Research*, 30(7):2143–2156.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J. (2008). Stationarity is dead: Whither water management? *Science*, 319(5863):573–574.

Miralles, D. G., Holmes, T. R., De Jeu, R. A., Gash, J. H., Meesters, A. G., and Dolman, A. J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, 15(2):453–469.

Missouri Geological Survey (2018). Missouri Geological Survey Geosciences Technical Resource Assessment Tool (GeoSTRAT), available at `https://apps5.mo.gov/geostrat/` (last access: 14 October 2020).

Montanari, A. and Toth, E. (2007). Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins? *Water Resources Research*, 43(5):1–10.

Moore, R. J. (2007). The PDM rainfall-runoff model. *Hydrology and Earth System Sciences*, 11(1):483–499.

Mu, Q., Zhao, M., and Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, 115(8):1781–1800.

Mugel, D. N., Richards, J. M., and Schumacher, J. G. (2009). Geohydrologic investigations and landscape characteristics of areas contributing water to springs, the Current River, and Jacks Fork, Ozark National Scenic Riverways, Missouri. *U.S. Geological Survey Scientific Investigations Report*, 2009–5138.

Nash, J. E. (1957). The form of the instantaneous unit hydrograph. *International Association of Scientific Hydrology, Publ*, 3:114–121.

Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I – A discussion of principles. *Journal of Hydrology*, 10(3):282–290.

National Atlas of the United States (2005). Generalized Glacial Limit Lines: Geology of the conterminous United States, available at `https://purl.stanford.edu/vz874sc7648` (last access: 14 October 2020).

National River Flow Archive (2019). NERC CEH, Wallingford, available at https://nrfa.ceh.ac.uk (last access: 31 January 2020).

Naylor, S., Letsinger, S. L., Ficklin, D. L., Ellett, K. M., and Olyphant, G. A. (2016). A hydropedological approach to quantifying groundwater recharge in various glacial settings of the mid-continental USA. *Hydrological Processes*, 30(10):1594–1608.

Neff, B. P., Day, S., Piggott, A., and Fuller, L. (2005). Base flow in the Great Lakes Basin. *U.S. Geological Survey Scientific Investigations Report*, 2005-5217.

Neff, B. P. and Rosenberry, D. O. (2018). Groundwater connectivity of upland-embedded wetlands in the Prairie Pothole Region. *Wetlands*, 38(1):51–63.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, 19(1):209–223.

NRCS (2004). National Engineering Handbook: Part 630 – Hydrology. *USDA Soil Conservation Service: Washington, DC, USA*.

Obergfell, C., Bakker, M., and Maas, K. (2019). Estimation of average diffuse aquifer recharge using time series modeling of groundwater heads. *Water Resources Research*, 55(3):2194–2210.

Olden, J. D. and Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, 19(2):101–121.

Oreskes, N. (2019). *Why Trust Science?* Princeton University Press.

Oudin, L., Salavati, B., Furusho-Percot, C., Ribstein, P., and Saadi, M. (2018). Hydrological impacts of urbanization at the catchment scale. *Journal of Hydrology*, 559:774–786.

Parry, S., Wilby, L. R., Prudhomme, C., and Wood, J. P. (2016). A systematic assessment of drought termination in the United Kingdom. *Hydrology and Earth System Sciences*, 20(10):4265–4281.

Peel, M. C. and Blöschl, G. (2011). Hydrological modelling in a changing world. *Progress in Physical Geography*, 35(2):249–261.

Pelletier, A. and Andréassian, V. (2020). Hydrograph separation: An impartial parametrisation for an imperfect method. *Hydrology and Earth System Sciences*, 24(3):1171–1187.

Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G., Williams, Z., Brunke, M. A., and Gochis, D. (2016). A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *Journal of Advances in Modeling Earth Systems*, 8(1):41–65.

Peña-Arancibia, J. L., Van Dijk, A. I. J. M., Mulligan, M., and Bruijnzeel, L. A. (2010). The role of climatic and terrain attributes in estimating baseflow recession in tropical catchments. *Hydrology and Earth System Sciences*, 14(11):2193.

Perrin, C., Michel, C., and Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4):275–289.

Peters, E., Torfs, P. J. J. F., van Lanen, H. A. J., and Bier, G. (2003). Propagation of drought through groundwater – A new approach using linear reservoir theory. *Hydrological Processes*, 17(15):3023–3040.

Peters, E. and van Lanen, H. A. J. (2005). Separation of base flow from streamflow using groundwater levels – Illustrated for the Pang catchment (UK). *Hydrological Processes*, 19(4):921–936.

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

Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., Sparks, R. E., and Stromberg, J. C. (1997). The natural flow regime. *BioScience*, 47(11):769–784.

Poff, N. L. and Zimmerman, J. K. H. (2010). Ecological responses to altered flow regimes: A literature review to inform the science and management of environmental flows. *Freshwater Biology*, 55(1):194–205.

Ponce, V. M. and Shetty, A. V. (1995a). A conceptual model of catchment water balance: 1. Formulation and calibration. *Journal of Hydrology*, 173(1-4):27–40.

Ponce, V. M. and Shetty, A. V. (1995b). A conceptual model of catchment water balance: 2. Application to runoff and baseflow modeling. *Journal of Hydrology*, 173(1-4):41–50.

Post, D. A. and Jakeman, A. J. (1999). Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. *Ecological Modelling*, 123(2-3):91–104.

Prancevic, J. P. and Kirchner, J. W. (2019). Topographic controls on the extension and retraction of flowing streams. *Geophysical Research Letters*, 46(4):2084–2092.

Price, K. (2011). Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review. *Progress in Physical Geography*, 35(4):465–492.

Price, K., Jackson, C. R., Parker, A. J., Reitan, T., Dowd, J., and Cyterski, M. (2011). Effects of watershed land use and geomorphology on stream low flows during severe drought conditions in the southern Blue Ridge Mountains, Georgia and North Carolina, United States. *Water Resources Research*, 47(2).

Regonda, S. K., Rajagopalan, B., Clark, M., and Pitlick, J. (2005). Seasonal cycle shifts in hydroclimatology over the western United States. *Journal of Climate*, 18(2):372–384.

Rempe, D. M. and Dietrich, W. E. (2018). Direct observations of rock moisture, a hidden component of the hydrologic cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11):2664–2669.

Richter, B. D., Baumgartner, J. V., Powell, J., and Braun, D. P. (1996). A method for assessing hydrologic alteration within ecosystems. *Conservation Biology*, 10(4):1163–1174.

Robinson, E., Blyth, E., Clark, D., Comyn-Platt, E., Finch, J., and Rudd, A. (2016). Climate hydrology and ecology research support system potential evapotranspiration dataset for Great Britain (1961-2015) [CHESS-PE]. *NERC Environmental Information Data Centre*.

Rodell, M., Beaudoing, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., Adler, R., Bosilovich, M. G., Clayson, C. A., Chambers, D., Clark, E., Fetzer, E. J., Gao, X., Gu, G., Hilburn, K., Huffman, G. J., Lettenmaier, D. P., Liu, W. T., Robertson, F. R., Schlosser, C. A., Sheffield, J., and Wood, E. F. (2015). The observed state of the water cycle in the early twenty-first century. *Journal of Climate*, 28(21):8289–8318.

Roderick, M. L. and Farquhar, G. D. (2011). A simple framework for relating variations in runoff to variations in climatic conditions and catchment properties. *Water Resources Research*, 47(12):1–11.

Roques, C., Rupp, D. E., and Selker, J. S. (2017). Improved streamflow recession parameter estimation with attention to calculation of -dQ/dt. *Advances in Water Resources*, 108:29–43.

Rouholahnejad Freund, E. and Kirchner, J. W. (2017). A Budyko framework for estimating how spatial heterogeneity and lateral moisture redistribution affect average evapotranspiration rates as seen from the atmosphere. *Hydrology and Earth System Sciences*, 21(1):217–233.

Rust, W., Holman, I., Bloomfield, J. P., Cuthbert, M., and Corstanje, R. (2019). Understanding the potential of climate teleconnections to project future groundwater drought. *Hydrology and Earth System Sciences*, 23(8):3233–3233.

Safeeq, M., Grant, G. E., Lewis, S. L., and Tague, C. L. (2013). Coupling snowpack and groundwater dynamics to interpret historical streamflow trends in the western United States. *Hydrological Processes*, 27(5):655–668.

Santhi, C., Allen, P. M., Muttiah, R. S., Arnold, J. G., and Tuppad, P. (2008). Regional estimation of base flow for the conterminous United States by hydrologic landscape regions. *Journal of Hydrology*, 351(1-2):139–153.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G. (2011). Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences*, 15(9):2895–2911.

Sayama, T., McDonnell, J. J., Dhakal, A., and Sullivan, K. (2011). How much water can a watershed store? *Hydrological Processes*, 25(25):3899–3908.

Schaefli, B. (2016). Snow hydrology signatures for model identification within a limits-of-acceptability approach. *Hydrological Processes*, 30(22):4019–4035.

Schaller, M. F. and Fan, Y. (2009). River basins as groundwater exporters and importers: Implications for water cycle and climate modeling. *Journal of Geophysical Research: Atmospheres*, 114(4):D04103.

Schiffer, D. M. (1998). Hydrology of central Florida lakes – A primer. *U.S. Geological Survey Circular*, 1137.

Schneider, M. K., Brunner, F., Hollis, J. M., and Stamm, C. (2007). Towards a hydrological classification of European soils: Preliminary test of its predictive power for the base flow index using river discharge data. *Hydrology and Earth System Sciences*, 11(4):1501–1513.

Schwartz, F. W., Liu, G., Aggarwal, P., and Schwartz, C. M. (2017). Naïve simplicity: The overlooked piece of the complexity-simplicity paradigm. *Groundwater*, 55(5):703–711.

Sefton, C. and Howarth, S. (1998). Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales. *Journal of Hydrology*, 211(1-4):1–16.

Shafii, M., Basu, N., Craig, J. R., Schiff, S. L., and Van Cappellen, P. (2017). A diagnostic approach to constraining flow partitioning in hydrologic models using a multiobjective optimization framework. *Water Resources Research*, 53(4):3279–3301.

Shafii, M. and Tolson, B. A. (2015). Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resources Research*, 51(5):3796–3814.

Shanley, J. B., Sebestyen, S. D., McDonnell, J. J., McGlynn, B. L., and Dunne, T. (2015). Water's Way at Sleepers River watershed – Revisiting flow generation in a post-glacial landscape, Vermont USA. *Hydrological Processes*, 29(16):3447–3459.

Shi, X., Wood, A. W., and Lettenmaier, D. P. (2008). How essential is hydrologic model calibration to seasonal streamflow forecasting? *Journal of Hydrometeorology*, 9(6):1350–1363.

Sivapalan, M. (2003). Prediction in ungauged basins: A grand challenge for theoretical hydrology. *Hydrological Processes*, 17(15):3163–3170.

Sivapalan, M. (2005). Pattern, process and function: Elements of a unified theory of hydrology at the catchment scale. In *Encyclopedia of Hydrological Sciences*. John Wiley & Sons.

Sivapalan, M., Yaeger, M. A., Harman, C. J., Xu, X., and Troch, P. A. (2011). Functional model of water balance variability at the catchment scale: 1. Evidence of hydrologic similarity and space-time symmetry. *Water Resources Research*, 47(2):1–18.

Sloto, R. A. and Crouse, M. Y. (1996). HYSEP: A computer program for streamflow hydrograph separation and analysis. *U.S. Geological Survey Water-Resources Investigations Report*, 96-4040.

Smakhtin, V. (2001). Low flow hydrology: A review. *Journal of Hydrology*, 240(3-4):147–186.

Smith, K. A., Barker, L. J., Tanguy, M., Parry, S., Harrigan, S., Legg, T. P., Prudhomme, C., and Hannaford, J. (2019). A multi-objective ensemble approach to hydrological modelling in the UK: An application to historic drought reconstruction. *Hydrology and Earth System Sciences*, 23(8):3247–3268.

Smith III, J. O. (2007). Introduction to Digital Filters: With Audio Applications, available at `https://ccrma.stanford.edu/~jos/filters/` (last access: 28 January 2021).

Spiegel, M. R. (1968). *Mathematical Handbook of Formulas and Tables*. McGraw-Hill.

Sprenger, M. and Allen, S. T. (2020). What ecohydrologic separation is and where we can go with it. *Water Resources Research*, 56(7):e2020WR027238.

Sprenger, M., Stumpp, C., Weiler, M., Aeschbach, W., Allen, S. T., Benettin, P., Dubbert, M., Hartmann, A., Hrachowitz, M., Kirchner, J. W., et al. (2019). The demographics of water: A review of water ages in the critical zone. *Reviews of Geophysics*, 57(3):800–834.

St. Clair, J., Moon, S., Holbrook, W. S., Perron, J. T., Riebe, C. S., Martel, S. J., Carr, B., Harman, C. J., Singha, K., and De Richter, D. B. (2015). Geophysical imaging reveals topographic stress control of bedrock weathering. *Science*, 350(6260):534–538.

Staudinger, M., Stoelzle, M., Seeger, S., Seibert, J., Weiler, M., and Stahl, K. (2017). Catchment water storage variation with elevation. *Hydrological Processes*, 31(11):2000–2015.

Stoelzle, M., Schuetz, T., Weiler, M., Stahl, K., and Tallaksen, L. M. (2020). Beyond binary baseflow separation: A delayed-flow index for multiple streamflow contributions. *Hydrology and Earth System Sciences*, 24(2):849–867.

Stoelzle, M., Stahl, K., Morhard, A., and Weiler, M. (2014). Streamflow sensitivity to drought scenarios in catchments with different geology. *Geophysical Research Letters*, 41(17):6174–6183.

Stoelzle, M., Stahl, K., and Weiler, M. (2013). Are streamflow recession characteristics really characteristic? *Hydrology and Earth System Sciences*, 17(2):817–828.

Stoelzle, M., Weiler, M., Stahl, K., Morhard, A., and Schuetz, T. (2015). Is there a superior conceptual groundwater model structure for baseflow simulation? *Hydrological Processes*, 29(6):1301–1313.

Svensson, C. (2016). Seasonal river flow forecasts for the United Kingdom using persistence and historical analogues. *Hydrological Sciences Journal*, 61(1):19–35.

Tague, C. L. and Grant, G. E. (2004). A geological framework for interpreting the low-flow regimes of Cascade streams, Willamette River Basin, Oregon. *Water Resources Research*, 40(4):1–9.

Tague, C. L., Grant, G. E., Farrell, M., Choate, J., and Jefferson, A. (2008). Deep groundwater mediates streamflow response to climate warming in the Oregon Cascades. *Climatic Change*, 86(1-2):189–210.

Tallaksen, L. (1995). A review of baseflow recession analysis. *Journal of hydrology*, 165(1-4):349–370.

Tanguy, M., Dixon, H., Prosdocimi, I., Morris, D. G., and Keller, V. D. J. (2016). Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890-2015) [CEH-GEAR]. *NERC Environmental Information Data Centre*.

Tashie, A., Pavelsky, T., and Emanuel, R. E. (2020). Spatial and temporal patterns in baseflow recession in the continental United States. *Water Resources Research*, 56(3):1–18.

Tashie, A., Scaife, C. I., and Band, L. E. (2019). Transpiration and subsurface controls of streamflow recession characteristics. *Hydrological Processes*, 33(19):2561–2575.

Texas Water Development Board (2020). Major aquifers of Texas, available at `http://www.twdb.texas.gov/mapping/gisdata.asp` (last access: 14 October 2020).

Townley, L. R. (1995). The response of aquifers to periodic forcing. *Advances in Water Resources*, 18(3):125–146.

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

Troch, P. A., Berne, A., Bogaart, P., Harman, C. J., Hilberts, A. G., Lyon, S. W., Paniconi, C., Pauwels, V. R., Rupp, D. E., Selker, J. S., Teuling, J. A., Uijlenhoet, R., and Verhoest, N. E. C. (2013). The importance of hydraulic groundwater theory in catchment hydrology: The legacy of Wilfried Brutsaert and Jean-Yves Parlange. *Water Resources Research*, 49(9):5099–5116.

Troch, P. A., Lahmers, T., Meira, A., Mukherjee, R., Pedersen, J. W., Roy, T., and Valdés-Pineda, R. (2015). Catchment coevolution: A useful framework for improving predictions of hydrological change? *Water Resources Research*, 51(7):4903–4922.

Troch, P. A., Martinez, G. F., Pauwels, V. R. N., Durcik, M., Sivapalan, M., Harman, C. J., Brooks, P. D., Gupta, H. V., and Huxman, T. (2009). Climate and vegetation water use efficiency at catchment scales. *Hydrological Processes*, 23(16):2409–2414.

Turner, M. G. (1989). Landscape ecology: The effect of pattern on process. *Annual Review of Ecology and Systematics*, 20(1):171–197.

U.S. Fish and Wildlife Service (2020). National Wetlands Inventory, available at `https://www.fws.gov/wetlands/index.html` (last access: 14 October 2020).

U.S. Geological Survey (2003). Principal aquifers of the 48 conterminous United States, Hawaii, Puerto Rico, and the U.S. Virgin Islands, available at `https://water.usgs.gov/ogw/aquifer/map.html` (last access: 14 October 2020).

Van Dijk, A. I. J. M. (2010). Climate and terrain factors explaining streamflow response and recession in Australian catchments. *Hydrology and Earth System Sciences*, 14(1):159–169.

Vega, M., Pardo, R., Barrado, E., and Debán, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, 32(12):3581–3592.

Vogel, R. M. and Sankarasubramanian, A. (2003). Validation of a watershed model without calibration. *Water Resources Research*, 39(10):1–9.

Wagener, T., Gleeson, T., Coxon, G., Hartmann, A., Howden, N. J. K., Pianosi, F., Rahman, M., Rosolem, R., Stein, L., and Woods, R. (2020). On doing large-scale hydrology with lions: Realising the value of perceptual models and knowledge accumulation.

Wagener, T. and Montanari, A. (2011). Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. *Water Resources Research*, 47(6).

Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., Kumar, P., Rao, P. S. C., Basu, N., and Wilson, J. S. (2010). The future of hydrology: An evolving science for a changing world. *Water Resources Research*, 46(5):1–10.

Wagener, T., Sivapalan, M., Troch, P. A., and Woods, R. A. (2007). Catchment classification and hydrologic similarity. *Geography Compass*, 1(4):901–931.

Wang, C., Wang, S., Fu, B., and Zhang, L. (2016). Advances in hydrological modelling with the Budyko framework: A review. *Progress in Physical Geography*, 40(3):409–430.

Wang, D. and Tang, Y. (2014). A one-parameter Budyko model for water balance captures emergent behavior in Darwinian hydrologic models. *Geophysical Research Letters*, 41(13):4569–4577.

Wang, D. and Wu, L. (2013). Similarity of climate control on base flow and perennial stream density in the Budyko framework. *Hydrology and Earth System Sciences*, 17(1):315–324.

Wang, D., Zhao, J., Tang, Y., and Sivapalan, M. (2015). A thermodynamic interpretation of Budyko and L'vovich formulations of annual water balance: Proportionality hypothesis and maximum entropy production. *Water Resources Research*, 51(4):3007–3016.

Weingartner, R., Blöschl, G., Hannah, D. M., Marks, D. G., Parajka, J., Pearson, C. S., Rogger, M., Salinas, J. L., Sauquet, E., Srikanthan, R., Thompson, S. E., and Viglione, A. (2013). Prediction of seasonal runoff in ungauged basins. In *Runoff Prediction in Ungauged Basins*, pages 102–134. Cambridge University Press.

Westerberg, I. K. and McMillan, H. K. (2015). Uncertainty in hydrological signatures. *Hydrology and Earth System Sciences*, 19(9):3951–3968.

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., and Freer, J. E. (2016). Uncertainty in hydrological signatures for gauged and ungauged catchments. *Water Resources Research*, 52(3):1847–1865.

Wilcox, B. P., Wilding, L. P., and Woodruff, J. M. (2007). Soil and topographic controls on runoff generation from stepped landforms in the Edwards Plateau of Central Texas. *Geophysical Research Letters*, 34(24):1–6.

Winter, T. C. (1999). Relation of streams, lakes, and wetlands to groundwater flow systems. *Hydrogeology Journal*, 7(1):28–45.

Winter, T. C. (2001). The concept of hydrologic landscapes. *Journal of the American Water Resources Association*, 37(2):335–349.

Winter, T. C. (2007). The role of ground water in generating streamflow in headwater areas and in maintaining base flow. *Journal of the American Water Resources Association*, 43(1):15–25.

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

Woodruff, B. M. and Abbott, P. L. (1979). Drainage-basin evolution and aquifer development in a karstic limestone terrain south-central Texas, USA. *Earth Surface Processes*, 4(4):319–334.

Woodruff, C. M. and Wilding, L. P. (2008). Bedrock, soils, and hillslope hydrology in the Central Texas Hill Country, USA: Implications on environmental management in a carbonate-rock terrain. *Environmental Geology*, 55(3):605–618.

Woods, R. A. (2009). Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks. *Advances in water resources*, 32(10):1465–1481.

Yadav, M., Wagener, T., and Gupta, H. V. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, 30(8):1756–1774.

Yokoo, Y. and Sivapalan, M. (2011). Towards reconstruction of the flow duration curve: Development of a conceptual framework with a physical basis. *Hydrology and Earth System Sciences*, 15(9):2805–2819.

Young, P. C. (1998). Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling and Software*, 13(2):105–122.

Zhang, Y., Ahiablame, L., Engel, B., and Liu, J. (2013). Regression modeling of baseflow and baseflow index for Michigan USA. *Water*, 5(4):1797–1815.

Zhang, Y. K. and Schilling, K. E. (2006). Increasing streamflow and baseflow in Mississippi River since the 1940s: Effect of land use change. *Journal of Hydrology*, 324(1-4):412–422.

Zhao, J., Wang, D., Yang, H., and Sivapalan, M. (2016). Unifying catchment water balance models for different time scales through the maximum entropy production principle. *Water Resources Research*, 52(9):7503–7512.

Zimmer, M. A. and Gannon, J. P. (2018). Run-off processes from mountains to foothills: The role of soil stratigraphy and structure in influencing run-off characteristics across high to low relief landscapes. *Hydrological Processes*, 32(11):1546–1560.