



# Unravelling abiotic and biotic controls on the seasonal water balance using data-driven dimensionless diagnostics

Simon Paul Seibert<sup>1</sup>, Conrad Jackisch<sup>1</sup>, Uwe Ehret<sup>1</sup>, Laurent Pfister<sup>2</sup>, and Erwin Zehe<sup>1</sup>

<sup>1</sup>Karlsruhe Institute of Technology (KIT), Institute for Water and River Basin Management, Chair of Hydrology, Kaiserstrasse 12, 76131 Karlsruhe, Germany

<sup>2</sup>Luxembourg Institute of Science and Technology, Department Environmental Research and Innovation, Catchment and Eco-hydrology research group, 5 avenue des Hauts-Fourneaux, 4362 Esch/Alzette, Luxembourg

Correspondence to: Simon Paul Seibert (simon.seibert@tum.de)

Received: 4 March 2016 – Revised: 9 March 2016 – Accepted: 3 April 2017 – Published: 9 June 2017

**Abstract.** The baffling diversity of runoff generation processes, alongside our sketchy understanding of how physiographic characteristics control fundamental hydrological functions of water collection, storage, and release, continue to pose major research challenges in catchment hydrology. Here, we propose innovative data-driven diagnostic signatures for overcoming the prevailing status quo in catchment inter-comparison. More specifically, we present dimensionless double mass curves (dDMC) which allow inference of information on runoff generation and the water balance at the seasonal and annual timescales. By separating the vegetation and winter periods, dDMC furthermore provide information on the role of biotic and abiotic controls in seasonal runoff formation.

A key aspect we address in this paper is the derivation of dimensionless expressions of fluxes which ensure the comparability of the signatures in space and time. We achieve this by using the limiting factors of a hydrological process as a scaling reference. We show that different references result in different diagnostics. As such we define two kinds of dDMC which allow us to derive seasonal runoff coefficients and to characterize dimensionless streamflow release as a function of the potential renewal rate of the soil storage. We expect these signatures for storage controlled seasonal runoff formation to remain invariant, as long as the ratios of release over supply and supply over storage capacity develop similarly in different catchments.

We test the proposed methods by applying them to an operational data set comprising 22 catchments (12–166 km<sup>2</sup>) from different environments in southern Germany and hydrometeorological data from 4 hydrological years. The diag-

nostics are used to compare the sites and to reveal the dominant controls on runoff formation.

The key findings are that dDMC are meaningful signatures for catchment runoff formation at the seasonal to annual scale and that the type of scaling strongly influences the diagnostic potential of the dDMC. Adding discrimination between growing season and winter period was of fundamental importance and easy to implement by means of a temperature-index model. More specifically, temperature aggregates explain over 70 % of the variability of the seasonal summer runoff coefficients. The results also show that the soil topographic index, i.e. the product of topographic gradient and saturated hydraulic conductivity, is significantly correlated with winter runoff coefficients, whereas the topographic gradient and the hydraulic conductivity alone are not. We conclude that proxies for gradients and resistances should be interpreted as a pair. Lastly, the dDMC concept reveals memory effects between summer and winter runoff regimes that are not relevant in spring between the transition from winter to summer.

## 1 Introduction

Understanding catchment-scale runoff formation and the underlying controls is the key for building hydrological models, which work for the right reasons (Kirchner, 2006), as well as for hydrological similarity assessment (Larsen et al., 1994; Barthold and Woods, 2015) in general.

The latter includes what we call “forward” and “backward” classification approaches: forward in the sense of pos-

tulating similarity of runoff generation based on similarity of catchment structural attributes; backward as a search for data-driven measures to detect similarity in hydrological response variables based either on suitable diagnostic fingerprints or dimensionless similarity indicators (Sivapalan et al., 1987; Wood et al., 1990). Prominent forward approaches include for instance hydrological response units (HRUs) (Leavesley, 1973), the concept of hydrology of soil types (HOST) (Boorman et al., 1995), or the topographic index (Kirby, 1975) which describes similarity of points within a catchment with respect to event-scale runoff formation (Beven and Kirkby, 1979). The common ground of forward approaches is that they are often used in connection with hydrological models to subdivide the catchment or hillslope into control volumes of similar runoff generation, which is represented by similar parameter sets or closure relations. Up to now, a large set of HRU types and separation methods has been suggested (Wood et al., 1990; Peschke et al., 1999; Scherrer and Naef, 2003; Schmocker-Fackel et al., 2007; Pelletier and Rasmussen, 2009; Santhi et al., 2008). Also, the representative elementary watersheds (REW) concept (Reggiani et al., 1998, 2000; Varado et al., 2006; Zhang et al., 2006) can be seen as a mathematically rigorous and thermodynamically consistent interpretation of the HRU idea. More recently a hierarchy of more specific functional units, defined on the basis of similarity of terrestrial and atmospheric controls on driving gradients and resistance terms, was proposed by Zehe et al. (2014) as a refinement of the HRU idea. The forward classification schemes are strongly model-dependent, which leaves us with questions about conclusiveness and transferability to other landscapes and how to define similarity in catchment response behaviour independently of models.

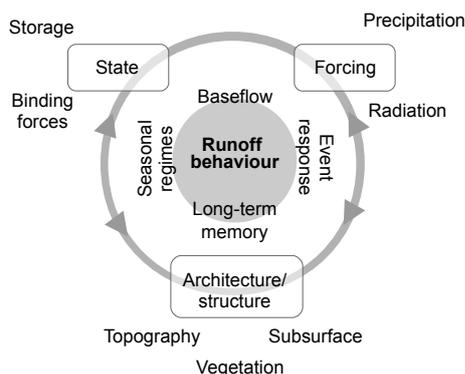
These questions are the scope of the backward approaches, which aim at detecting similarity in hydrological response variables based on suitable diagnostic fingerprints in a data-driven way. The use of backward approaches has also been promoted by the Prediction in Ungauged Basins (PUB) initiative (Sivapalan, 2003; Hrachowitz et al., 2013), which suggested diagnosing catchment functioning through runoff “signatures”. The underlying assumption is that “[...] runoff variability can be broken up into several components, each of them a manifestation of catchment functioning, albeit at different time scales, [...]” (Blöschl et al., 2013, p. 7). Today, runoff signatures are commonly accepted and usually defined as specific characteristics of the hydrograph such as autocorrelation, slope of or bias in the flow duration curve or different segments thereof, rising limb density, peak distribution, and/or as flow statistics such as mean, variance, skewness, or the coefficient of variation (Pokhrel and Yilmaz, 2012; Casper et al., 2012; Pfannerstill et al., 2014; Euser et al., 2013, 2015). Runoff signatures are widely used for similarity assessment but also for model evaluation. In the former, similarity in the signature values is interpreted as hydrological similarity among catchments, which is also the foundation of

many catchment inter-comparison studies (Merz et al., 2006; Oudin et al., 2008; Sawicz et al., 2011; Wang and Wu, 2013; Viglione et al., 2013). In model evaluation, the signature values obtained from the observation are compared to the signature values obtained from the simulation (Vrugt and Sadegh, 2013).

Ideally, we expect that similarity of data-driven diagnostics can be explained by similarity in the architecture of the catchment and hence that forward and backward approaches will yield consistent results. However, such comparisons often fail to be conclusive. One possible explanation is that similarity in hydrograph-based signatures may indeed be caused by several reasons. For instance, characteristics of the flow duration curve such as low slope values of the 33rd and 66th streamflow percentiles, implying damped runoff response, can arise from persistent year-round rainfall regimes or from the dominance of groundwater contribution to streamflow. In consequence groupings of catchments based on either signatures or physiographic properties can be rather inconsistent, as highlighted in the catchment inter-comparison study of Ali et al. (2012) where “[...] catchment groupings obtained using physical properties only did not match those obtained using flow indices, mean transit times or storage estimates”. Similar findings are reported by Oudin et al. (2010) and Ley et al. (2011). Although the potential of signatures for similarity assessment and diagnosing functional similarity is beyond question, the existing approaches to define and to detect similarity differ considerably with respect to the underlying assumptions, methods, and proposed measures. As a commonly agreed upon understanding of “signatures” and “hydrological, functional or behavioural similarity” is still missing, we can not yet distil a convergence of approaches.

In the present study we add bits and pieces to this puzzle by proposing and testing data-driven dimensionless backward signatures to diagnose and characterize seasonal runoff formation. As depicted in Fig. 1, these signatures shall unravel the influences (i) of the hydrometeorological forcing, i.e. radiation and precipitation, (ii) of key catchment structural attributes, specifically of biotic (functional vegetation) and abiotic controls (e.g. topographic gradient, subsurface hydraulic conductivity), and (iii) of the state of the catchment. The relative importance of these concurring influences on runoff generation depends on the scale. Intuitively one might expect the wetness state of the catchment to be of the highest importance for runoff generation during events, while vegetation surely has the strongest impact on seasonal runoff production.

So what could suitable data-driven diagnostic signatures look like? Inspired by the diagnostic approach to model evaluation (Gupta et al., 2008), McMillan et al. (2011) proposed combining recession analysis, soil moisture dynamic analyses, and event-scale water balance investigations to obtain insights into different runoff response timescales and to assess the impact of pre-storm wetness conditions on catchment



**Figure 1.** Conceptual sketch of the major temporal scales of runoff formation, i.e. long-term behaviour, seasonal regimes, generation of baseflow and event runoff production, and corresponding first-order controls.

runoff dynamics. Hrachowitz et al. (2011) suggested combining event runoff coefficients and different tracers to infer knowledge of hydrological processes in small catchments. Tracer data are also widely used to estimate transit time characteristics which are considered a useful similarity index for process-based catchment classification (Soulsby et al., 2010; Capell et al., 2012). Therefore, breakthrough or flushing of either contaminants (Gassmann et al., 2013), artificial tracers (Wienhöfer et al., 2009), sediments (Martínez-Carreras et al., 2010), or even diatoms as smart tracers (Martínez-Carreras et al., 2015; Klaus et al., 2015) are used. More recently, McMillan et al. (2014) proposed a “targeted analyses of catchment response data” and combined hydrograph-based signatures with signatures that evaluate characteristics of the water balance, recession analysis, and hydrological thresholds to examine the extent to which hydrological behaviour varies within a 50 km<sup>2</sup> catchment. The common ground of the mentioned approaches is that they combine multiple sources of information from rather different observations.

This is however only possible in well-instrumented research catchments. As our focus is on providing diagnostic fingerprints for comparative hydrology (Sivapalan et al., 2003), we constrain our work to operational hydrometeorological data sets. These commonly consist of meteorological variables like precipitation, air temperature, and humidity, and discharge data. Information on the catchment “state” in terms of groundwater level or soil moisture data is monitored only occasionally and often at a coarse spatial resolution. The advantage of operational data sets is that they allow us to include sufficient catchments with rather different physiographic and climate characteristics (end-members) for our analysis. A drawback of such data sets is that they mostly exist only for scales of 50 km<sup>2</sup> and larger and that specific data such as tracers, piezometric heads, or soil hydraulic properties are usually not available.

In this study we propose and test dimensionless diagnostic signatures to characterize seasonal runoff formation based on rainfall–runoff data, which shall ultimately separate the terrestrial controls on runoff formation from the meteorological forcing. The proposed concept relates to the idea of Wagener et al. (2007) that catchments have three main functions, i.e. to partition, store, and release water. In line with Black (1997) we treat catchments as lumped terrestrial filters and following Kirchner (2009) we assume that annual runoff formation is essentially a monotonously increasing function of water storage in the catchment. Accordingly, accumulated runoff is expected to be limited by accumulated water supply (rainfall minus evaporation). Water storage is controlled and limited by water supply due to rainfall, the available subsurface storage volume, as well as by its recharge and drainage properties. Consequently, we evaluate whether a diagram that plots accumulated water release scaled with total annual supply against accumulated water supply scaled by the available storage capacity provides the means to detect similarity in catchment seasonal runoff formation. Such a signature for storage controlled seasonal runoff formation is deemed to remain invariant during simple scaling operations, as further explained below.

In line with these thoughts we hypothesize that dimensionless and season-specific double mass curves (dDMC), which relate dimensionless accumulated runoff to dimensionless accumulated precipitation, are suitable for discriminating differences in how a catchment releases water as a function of the accumulated water supply. We test this hypothesis by addressing the following research questions within a catchment inter-comparison.

- Q1:** How can we obtain proper dimensionless double mass curves using site-specific characteristics to ensure comparability among catchments based on operational data sets?
- Q2:** Are temperature indices from vegetation ecology superior compared to calendrical definitions for separating summer and winter regimes?
- Q3:** Can we identify physiographic and ecological properties which explain differences in seasonal runoff behaviour revealed by differences in the dDMC?

Additionally we evaluate whether flow duration curves (FDC), which are widely used as diagnostic signatures (Ye et al., 2012; Euser et al., 2013; Viglione et al., 2013), might provide similar insights into seasonal runoff behaviour to dDMC. The paper is structured as follows: in the method section in chap. 2 we introduce the data set, the concept of dDMC, the separation of seasons, and the underlying statistical methods we apply to address Q3. After presenting the results in Sect. 3 we close with a discussion and conclusions in Sect. 4.

## 2 Concept, methods, and study area

### 2.1 Study area

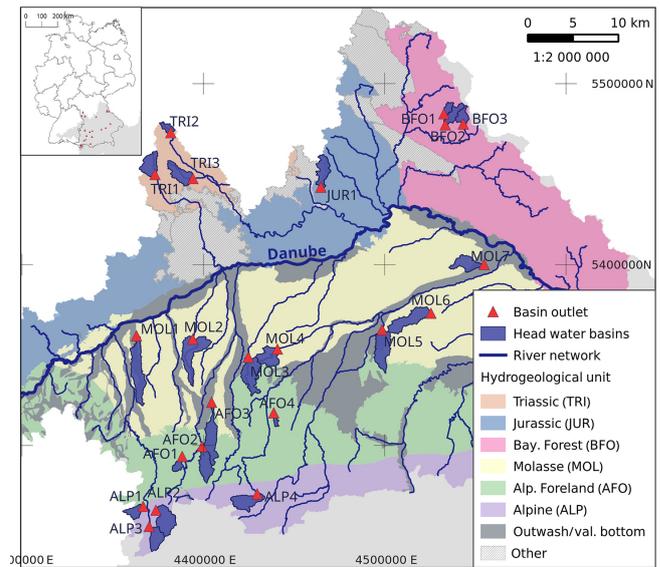
We propose and explore the potential of signatures to characterize seasonal runoff generation based on an operational data set of sub-basins from the Danube in Bavaria. In this chapter we introduce the data set and detail the differences in the climate and physiographic setting of our test catchments. Before that, we will briefly discuss the quality of the database, which in fact was in most catchments so poor that the majority of the sites had to be excluded from the analysis.

#### 2.1.1 Data quality and selection of headwater catchments

In our analysis we focus on lower mesoscale headwater catchments ( $< 200 \text{ km}^2$ ) as routing effects are still small (Robinson et al., 1995) and because the geological and pedological set-up is often still fairly homogeneous at this scale. In addition, non-convective storms often cover the majority of the catchment area and we may hence assume that the hydrometeorological forcing is well observed by the given station and that it is fairly uniform within the entire headwater. We hence select all gauged headwaters of this size within the Bavarian part of the Danube basin. For this, hourly hydrometeorological time series from the period 1 November 1999 until 31 October 2004 are available. The database of 130 potential reference catchments is analysed according to a set of quality criteria: we only include catchments where (i) at least one meteorological station was closer than 20 km (which is less than in the Mopex data set, Schaake et al., 2000; Duan et al., 2006), (ii) the total accumulated water balance error of the entire period of 4 hydrological years, considering simulated evaporation, was smaller than 5 % of total precipitation, (iii) the amount of missing and/or implausible meteorological data was  $< 5 \%$ , and (iv) the streams are not subject to any severe regulation. This screening resulted in 22 catchments being classified as suitable for the analysis. The sites represent very different hydrological regions (Fig. 2) and were roughly classified based upon geology, climate, and elevation. Moving from the north-west to the south we differentiate Triassic (TRI), Jurassic (JUR), Bavarian Forest (BFO), Faulted Molasse (MOL), Alpine Foreland (AFO), and Alpine (ALP) landscapes. The catchment identifiers reflect these units and include an index of the sample size, e.g. MOL1, ..., MOL7. Catchment locations, real gauges, and corresponding stream names are provided in Table C1 in the Appendix.

#### 2.1.2 Data and physiographic site properties

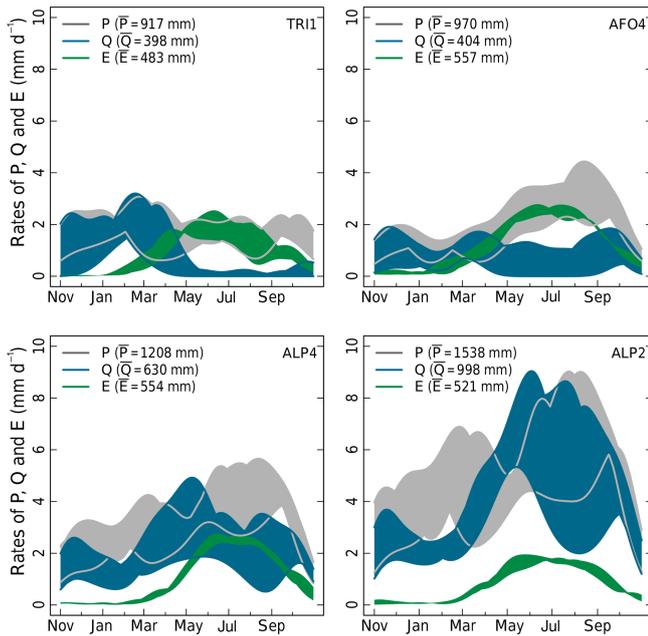
The Bavarian part of the Danube basin exhibits considerable differences in the hydrometeorological regimes. To illustrate them we plot regime curves for four different catchments in different environmental settings in Fig. 3. The catchment



**Figure 2.** Upper Danube catchment in southern Germany with selected headwater basins (blue polygons), corresponding gauges (red triangles), and major river networks (blue lines). The site identifiers (IDs) refer to the corresponding (hydro)geological unit (colour coded map in the background, adapted from BGR and SGD, 2015) and a single Arabic numeral. Moving from the north-west to the south, we differentiate Triassic (TRI), Jurassic (JUR), Bavarian Forest (BFO), Faulted Molasse (MOL), Alpine Foreland (AFO), and Alpine (ALP) locations. Table C1 provides links between the site IDs and the real gauge names. The inset in the upper left corner shows Germany's federal state boundaries, the individual headwater outlets, and the basin of the Danube (grey area). The grid coordinates refer to the Gauss–Krüger zone 4 projection (CRS identifier EPSG:31468).

TRI1 (Fig. 3, top left) receives a fairly constant input in precipitation ( $P$ ) throughout the year, but releases discharge ( $Q$ ) with a strong seasonality and pronounced minimum during summer. Compared to the other sites the inter-annual variation in evapotranspiration ( $E$ ) is rather large. The catchment AFO4 (Fig. 3, top right) in contrast shows seasonality in  $P$  but a fairly constant output in  $Q$ . ALP4 (Fig. 3, bottom left) and ALP2 (Fig. 3, bottom right), which are both Alpine sites, show a pronounced minimum in  $Q$  during February due to snow storage. ALP2 however shows a very large range in both  $P$  and  $Q$  during summer, which suggests little buffering and a high reactivity. In contrast, ALP4 has a much more damped response to  $P$  during summer and a more pronounced seasonality in  $E$ .

To characterize the structural set-up of the different catchments, we employ various landscape analysis techniques. Using these methods we derive nearly 30 different physiographic characteristics for each of the catchments. They are summarized in Tables A1 and A2 in Appendix A. Topographic information was extracted from a digital elevation model with a resolution of 25 m. We use it to calculate the



**Figure 3.** Regime curves of observed areal precipitation ( $P$ ) (grey), discharge ( $Q$ ) (blue), and calculated areal mean evapotranspiration ( $E$ , Penman–Monteith) (green) of four selected headwater catchments. The width of the individual bands illustrates the inter-annual variation during the 4-year long period. The curves are based on kernel density estimates using identical kernels and bandwidths for variables of the same type.  $\bar{P}$ ,  $\bar{Q}$ , and  $\bar{E}$  provide the 4-year mean annual average of the different variables, respectively.

median topographic gradient ( $\phi$ ) (–) of each catchment according to McGuire et al. (2005) as the flowpath length from each pixel to the stream divided by the corresponding difference in height using the Whitebox geographic analysis toolbox (Lindsay, 2014). Areal shares of various surface cover and soil properties like average sand, silt, or skeleton contents, root-zone effective field (eFC) (mm), and air capacities (AC) (mm) among others were derived from CORINE land use data as of 2006 and the digital soil map of Germany (scale: 1 : 1 000 000) (Hartwich et al., 1995). The average saturated hydraulic conductivity ( $K_s$ ) ( $\text{m s}^{-1}$ ) is estimated for each catchment based on available grain size distributions using *Rosetta's* pedo-transfer functions (Schaap et al., 2001). Due to Darcy's law we also included a soil topographic index ( $\tau$ ) ( $\text{m s}^{-1}$ ), i.e. the product of the topographic gradient and saturated hydraulic conductivity, in the analysis, as we expect them to act in a group:

$$\tau = \phi \cdot K_s. \quad (1)$$

Last but not least, we employed the conceptual hydrological Large Area Runoff Simulation Model (LARSIM) (Ludwig and Bremicker, 2006) to simulate  $E$  and snow water equivalent ( $W$ ) and to obtain areal estimates of  $E$ ,  $P$ , and  $W$  for each catchment. For this purpose a gridded version of LARSIM (spatial resolution: 1  $\text{km}^2$ ; time step: 1 h), which is

calibrated and in operational use at the Bavarian flood forecasting agency, was made available by the LfU. To obtain areal estimates we forced the model using observed hydrometeorological station data like temperature ( $T$ ), wind speed, humidity, radiation ( $R$ ), and  $P$ , among others. LARSIM simulates  $E$  using the Penman–Monteith equation. Interpolation is done using the grid point method (NOAA, 1972). It is important to note that we use LARSIM exclusively as a uniform interpolation tool for  $P$  and  $E$  and that we do not use any other model output. Note that Table A1 also provides hydrometeorological catchment characteristics like 4-year mean annual precipitation ( $\bar{P}$ ), discharge ( $\bar{Q}$ ), runoff coefficient ( $\bar{CR}$ ), and streamflow coefficient of variation ( $\nu_{\bar{Q}}$ ). These quantities are all calculated based upon the available observables, except for  $E$ . Additional information on the Bavarian part of the Danube basin, the hydrometeorological data, and the LARSIM model can be obtained from Seibert et al. (2014).

## 2.2 Dimensionless double mass curves to characterize runoff formation and the water balance

Essentially, we suggest that dDMC, describing dimensionless accumulated release as a function of dimensionless accumulated rainfall supply, are feasible for characterizing storage controlled runoff formation in a scale-invariant way. In this chapter we provide the corresponding principles (Sect. 2.2.1) and introduce two different approaches to derive dDMC (Sect. 2.2.2) which allow us to address Q1. To address research questions Q2 and Q3 we employ a temperature-index method to obtain season-specific dDMC and relate their properties to the available catchment descriptors. The corresponding methods are described in chaps. 2.3 and 2.4, respectively.

### 2.2.1 Theory

Pfister et al. (2002) introduced double mass curves as a straightforward way of comparing seasonal runoff formation and runoff coefficients among catchments situated in different geologies in the Alzette basin. The double mass curve quantifies cumulated runoff as a function of cumulated precipitation within the hydrological year. These diagrams are also helpful for learning about the inter-annual variability of the interplay of streamflow release and rainfall supply (Jackisch, 2015) and as a diagnostic signature to improve the development of perceptual models (Wrede et al., 2015). A straightforward means of ensuring the comparability of double mass curves within the same clean geological and climate setting is to relate fluxes instead of the flows. This implies independence of a simple scaling of the respective flows with catchment area. Jackisch (2015) and Pfister et al. (2002) showed that the inter-annual variation of double mass curves and their shape is rather invariant within a given geological setting but reveals distinct differences among different ge-

ologies. The key question is, however, whether these differences indeed reflect non-trivial differences in terrestrial controls of runoff formation or whether these are (partly) caused by volumetric “scale effects”, i.e. differences in storage volume, which arise from different vertical extents and not from differences in specific soil and bedrock porosity.

A nice example that underlines that a dimensionless formulation removes such simple scale effects is the dimensionless breakthrough curve (BTC). BTCs are used in soil physics to reveal (dis)similarities in transport and adsorption properties of a soil “filter” independently of the spatial extent of the probe (Hillel, 2004). Conceptually, BTCs are plots of scaled cumulated solute outflow against cumulated infiltration scaled by the pore volume of the sample. The BTC is hence invariant for a given solute as long as the ratio of accumulated water supply and storage capacity remains the same. The key to scale independence is to use the ratio of cumulated irrigation and storage volumes instead of using the cumulated irrigation alone.

The essential difference between a breakthrough experiment and the catchment water balance is, however, that the catchment releases water vapour to the atmosphere and liquid water to the stream at two different interfaces and that the subsurface extent of the storage volume and the flow path length to the riparian zone are largely unknown. Within our analogy, we regard liquid water release as the analogue to solute breakthrough, because both processes are controlled by gravity-driven water fluxes through the subsurface. In contrast,  $E$  feeds mostly from soil water which is stored against gravity. This part of soil water is largely immobile similar to adsorbed solutes (Zehe and Jackisch, 2016). The second difference is that the underlying controls on  $Q$  and particular on  $E$  do not only depend on properties of the soil but are also a function of time-dependent system states and boundary conditions like the meteorological forcing or the pressure difference between the groundwater system and the stream. Whereas the driving force on  $Q$ , i.e. gravity, is constant, the radiative forcing and the impact of vegetation on  $E$  follow a pronounced seasonal cycle – at least in temperate environments.

### 2.2.2 Derivation of dimensionless double mass curves

The key to obtaining scale-invariant dimensionless quantities is to divide a state variable of interest – for instance a force, velocity, or length, by a characteristic quantity of the system (Blöschl and Sivapalan, 1995). A popular example is the Reynolds number in fluid mechanics, which relates inertial forces to viscous forces. It is defined as flow velocity times a characteristic length divided by the dynamic viscosity of the fluid, which is suited to comparing turbulence in open channel flow independent of the channel width. The best known example of a dimensionless response measure in hydrology is the runoff coefficient (CR) (–) defined as specific discharge over specific rainfall, both with units of volume per time per

unit area. The latter is often used as a “diagnostic” variable to detect scale-invariant differences in generation of runoff volume (Merz et al., 2006; Graeff et al., 2012).

As the double mass curve will relate the accumulated dimensionless water release to the dimensionless accumulated water supply of the catchment, the dependent and independent variables need to be scaled properly. We define the dimensionless accumulated water release ( $Q^*$ ) (–) of the catchment based on the ratio of accumulated specific discharge (cum. $Q$ ) (mm) and annual totals of  $P$  ( $\sum P_{yr}$ ) (mm) of the corresponding hydrological year:

$$Q^* = \frac{\text{cum. } Q}{\sum P_{yr}}. \quad (2)$$

We are aware that this definition is, while being close to the definition of a runoff coefficient, not exactly the desired ratio of accumulated release to total annual supply. The latter, i.e. infiltration, needs to be determined based on the difference of accumulated  $P$  and accumulated  $E$ . As evaporation is however not easy to observe, Eq. (2) might still be a suitable data-driven estimate.

Removing the dimension of the abscissa of the double mass curve, i.e. cum. $P$  (mm), is less trivial. An easy way is to use  $\sum P_{yr}$  (mm) again to scale accumulated rainfall supply (Eq. 3). This nicely scales both axes to the range of [0...1] and allows us to compare multiple sites and years.

$$P^* = \frac{\text{cum. } P}{\sum P_{yr}} \quad (3)$$

An advantage of plotting  $Q^*$  against  $P^*$  is that it yields curves that are fairly intuitive to interpret, as this scaling preserves the shape of the original double mass curve.

- In snow-free periods with negligible  $E$  catchment “filter” properties naturally are expected to cause dDMC with overall slopes smaller than unity. Step changes and deviations from the overall pattern towards smaller slopes suggest either activation of additional storage or that water leaves the catchment in the form of evaporation, or both. Differentiation among these options is unfortunately not straightforward. A possible remedy is to focus on summer and winter periods individually, because snow is not relevant during summer and because  $E$  is deemed to be small during winter (see also chap. 2.3). To indicate the activation of storage in the snowpack, we highlight dDMC periods with temperatures  $< 0^\circ\text{C}$ .
- Segments with slopes of 1 indicate that release equals supply and that water is neither stored in the system nor that significant amounts of water leave the catchment in the vapour phase. Segments with a slope of one or close to it are expected to occur during intensity controlled runoff formation, i.e. Hortonian overland flow

(Horton, 1933) or preferential flow (Beven and Germann, 2013). These mechanisms are not controlled by storage capacity but by infiltrability and conductivity of the subsurface (Struthers and Sivapalan, 2007). dDMC segments with slopes close to one may however also indicate fast responding catchments e.g. due to dominance of poorly developed shallow soils, impermeable substratum or strong topographic gradients.

- Segments of the dDMC that expose slopes higher than 1 indicate that release is larger than supply. Provided that routing effect and runoff concentration times are negligible, these phenomena should be restricted to periods of significant snowmelt. They should hence only occasionally occur in the winter and coincide with periods of snow accumulation that are followed by warming periods.

A considerable disadvantage of scaling the abscissa using  $\sum P_{\text{yr}}$  (mm), named type I dDMC hereafter, is however that the same ordinate values in different hydrological years do not correspond to the same mass input. A comparison of type I dDMC among different sites and years might thus suggest greater similarity than there actually is.

Motivated by the dimensionless breakthrough curve, we propose an alternative approach to scale the accumulated water supply, named type II dDMC hereafter. It can be derived from the widely used HBV beta store (Bergstroem, 1976). The latter conceptualizes storage controlled formation of direct runoff ( $Q_d$ ) as a function of the relative saturation of the soil storage ( $S_m(t)$ ), actual precipitation  $P(t)$ , and a form parameter ( $\beta$ ) (Eq. 4).

$$Q_d(t) = P \left( \frac{S_m(t)}{S_{\text{max}}} \right)^\beta \quad (4)$$

Considering the mass balance and  $Q_d$  as a surrogate for rainfall-driven streamflow generation, it can be shown from Eq. (4) that accumulated  $Q_d$  within the interval  $t_0$  and  $t_e$  scales with the ratio of accumulated precipitation amount (cum.  $P$ ) ( $t_e$ ) and total storage ( $S_{\text{max}}$ ) as one key factor as shown in Eq. (5). The second factor that impacts the integral of  $Q_d$  is the integral of the relative storage change scaled with the relative saturation of the soil storage power of  $\beta - 1$ . Whereas both factors are strongly dependent on the unknown form parameter  $\beta$ , the second factor implicitly also depends on  $E$ . In Appendix B we derive Eq. (5) from the beta store concept, Eq. (4), and illustrate the role of the different terms using the case  $\beta = 1$  (compare Eq. B7).

$$\int_{t_0}^{t_e} Q_d dt = \frac{\text{cum. } P(t_e)}{S_{\text{max}}} \left( \frac{S_m(t_e)^\beta}{(S_{\text{max}})^{\beta-1}} \right) - \int_{t_0}^{t_e} \left( \frac{\text{cum. } P(t)}{S_{\text{max}}} \beta \left( \frac{S_m(t)}{S_{\text{max}}} \right)^{\beta-1} \frac{dS}{dt} \right) dt \quad (5)$$

In accordance with Eq. (5) we define the type II dDMC as accumulated streamflow release scaled by the annual precipitation amount, plotted against cum.  $P(t)$  scaled by  $S_{\text{max}}$ . We expect this curve to be sensitive to differences among catchments with respect to the ratio of rainfall supply and storage volume as well as to differences that arise from the second factor. Differences in the latter might arise from differences in how contributing areas grow with relative saturation of the soil storage. We expect the latter to be identifiable during the wet period. When increasing  $\beta$  to values larger than 1, the second factor becomes non-linear, which yields an increasingly convex function, which implies increasingly smaller winter slopes in the type II dDMC. When decreasing  $\beta$  to values smaller than unity, the second factor becomes non-linear and increasingly concave, which implies increasingly larger type II dDMC winter slopes. Differences in the change in water storage  $dS/dt$  due to differences in  $E$  (compare Eq. B6) are expected to reflect the inter-annual variations at a given site (warm and cold summer) or climate differences among different sites. Those differences should be identifiable during the summer period.

Deriving type II dDMC requires however an estimate of  $S_{\text{max}}$ , i.e. information about the characteristic vertical extent of the storage volume, which is largely not observable. Standard data for soil storage that are available in Germany (Hartwich et al., 1995) include estimates of specific effective field capacity (eFC) and air capacity (AC) (both in mm/dm) which we multiplied by root zone depth to obtain a suitable unit (mm). Though these data are uncertain, particularly at the catchment scale, we use them to scale cum  $P$ .

$$P^* = \frac{\text{cum. } P}{\text{eFC} + \text{AC}} \quad (6)$$

### 2.3 Season-specific evaluation of the double mass curve

During the growing period evapotranspiration often makes up the largest portion of the water balance. A possible way to isolate abiotic controls on runoff formation from the impact of the vegetation is to focus on data from frost-free periods of the “dormant” fall and winter season. Accordingly we derive season-specific double mass curves. This however requires a meaningful separation of the dormant, i.e. the “winter period” and the vegetation dominated “summer period”. Such a separation of seasons is not straightforward – particular for inter-comparison studies. Conventional definitions like hydrological years or the astronomical equinoxes (varying between 19 and 21 March and 22 and 24 September in central Europe) are easy to implement and thus frequently used. However, they often introduce uncertainty as they fail in predicting the onset of vegetation and thus the end of the dormant period. Furthermore, the period of active vegetation does not only depend on the location of a catchment with respect to latitude but also with respect to altitude.

As an alternative we test temperature aggregates which are used as predictors of bulk ecological activity in plant phys-

iology (Mäkelä et al., 2004). Specifically we use a temperature index model (Menzel et al., 2003) to define the onset and the duration of the period of vegetation. In this approach the growing season is defined as the period between the last spring frost and the first fall frost in which the average daily temperature constantly equals or exceeds  $5^{\circ}\text{C}$ . We evaluate this method visually by plotting the results into the double mass curves and by comparing them to the Gregorian definition of the beginning of spring which is 20 March in our climate. This way we address the second research question (Q2).

The separation of seasons allows us to derive average summer and winter runoff coefficients,  $\text{CR}_S$  (–) and  $\text{CR}_W$  (–), respectively, for the type I dDMC, by calculating the slope of linear regression lines which we fit to the summer and winter segments of the type I dDMC. For the type II dDMC we evaluate only the winter regimes, as we expect them to be primarily sensitive to the shape parameter  $\beta$ , and not to  $E$ , as suggested by the proof in Appendix B. Analogously to the type I dDMC, we fit linear regression to the winter segments of the type II dDMC. Here, the regression slope ( $m_W$ ) (–) characterizes how fast relative release increases with the potential renewal rate of the soil stock.

#### 2.4 Catchment inter-comparison and statistical evaluation of the double mass curves

To explain differences in seasonal runoff formation and hence to address Q3 we compare dDMC properties of the 22 catchment based on their average  $m_W$  (–),  $\text{CR}_S$  (–),  $\text{CR}_W$  (–) and  $\text{CR}_{\text{YR}}$  (–) values and corresponding inter-annual variations. Specifically we compare the spatial patterns of the different quantities and test if they coincide with the groupings of our catchments which is based on hydrological regions as outlined in chap. 2.1.2. In a statistical evaluation we attribute differences in  $m_W$  (–),  $\text{CR}_S$  (–) and  $\text{CR}_W$  (–) to the structural and climatic properties of the catchments listed in Tables A1 and A2 to identify the underlying controls but also to evaluate the different scaling schemes of the dependent variable against the dimensionless formulation of accumulated supply. For this we use simple and multiple linear regression analyses and corresponding tests of significance which are implemented in the statistical computing platform R (R Core Team, 2015). In the multiple regression analysis we kept all variables that increased the  $r^2$  by at least 5 % and which had  $p$ -values  $\leq 0.05$ . Assuming that linear relationships would suffice to identify important explanatory variables, we did not test non-linear relationships as clear hypotheses on the kind of these relationships were missing.

Last but not least we compare what we learn from the dDMC against what is to be learned from the well-known and widely used FDC. FDC describe the distribution of stream-flow magnitudes with respect to exceedance probability. According to Sawicz et al. (2011) the slope of the flow duration curve (sFDC (–)) can be considered as a proxy for av-

erage flow variability. Specifically we compare  $\text{CR}_S$  (–) and  $\text{CR}_W$  (–) against sFDC between the 33rd and 66th percentiles within a linear regression analysis. Since exceedance probabilities are not sensitive to seasonality, we also derive season-specific FDC based on the temperature-indexed derived seasons. Using the seasonal FDC we calculate inter-percentile slopes of the dormant period (sFDC<sub>W</sub> (–)) and that of the period of vegetation (sFDC<sub>S</sub> (–)) and include them in the analysis.

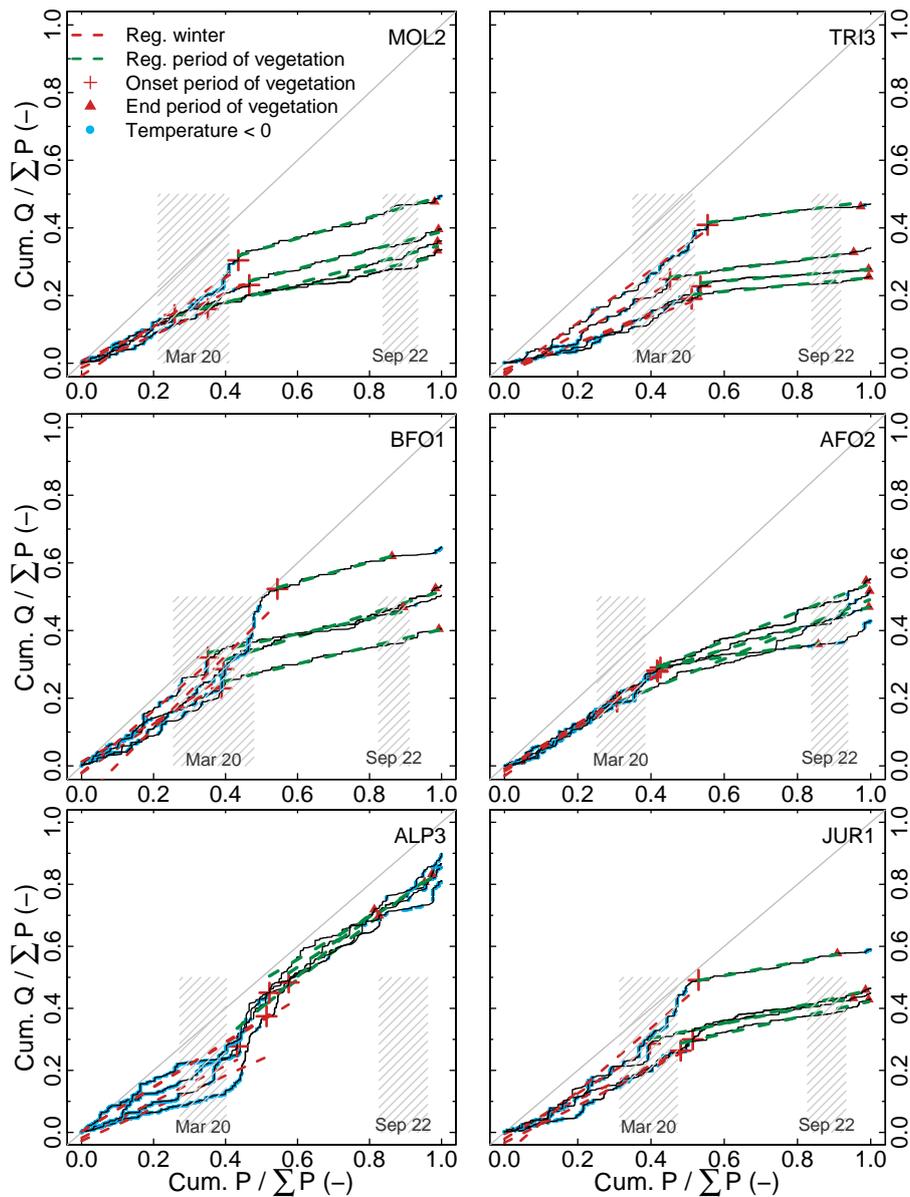
### 3 Results

#### 3.1 Dimensionless double mass curves

##### 3.1.1 Type I dDMC and temperature-index-based season separation

The season-specific evaluation of the type I dDMC revealed a surprisingly consistent pattern with similar mass curves in all catchments in terms of a fairly steep increase during winter and a clear regime shift towards much flatter, partly zero regression slopes in the vegetation period. In fact the slopes of the dDMC for different years of a catchment are almost constant, just shifted in parallel, during the period of vegetation at many sites and for many years as depicted in Fig. 4. As indicated by Table 1 the seasonal winter runoff coefficients ( $\text{CR}_W = 0.67$ ) exceed the average summer runoff coefficients ( $\text{CR}_S = 0.32$ ) by a factor of 2 on average which is in line with the findings of (Hellebrand et al., 2008) – with exception of the two Alpine sites ALP2 and ALP3. The inter-annual variation of the seasonal runoff coefficients are twice as large during winter ( $\sigma_{\text{CR}_W} = 0.1$ ) as during summer ( $\sigma_{\text{CR}_S} = 0.06$ ), indicating a rather strong homogeneity during summer.

As depicted in Fig. 4 the temperature-based estimates for the beginning and end of winter and summer period (Menzel et al., 2003) nicely coincide with the points where the slope of the dDMC changes (in terms of  $\text{cum. } P / \sum P$ ). This applies for different years and the entire range of physiographic conditions of our catchments, except for the Alpine region which is shown in the bottom left panel of Fig. 4. In the Alpine settings the accumulation and melting of snow has a stronger impact on the seasonal water balance than the period of vegetation. The spring and fall equinoxes (drawn as hatched areas in Fig. 4) and the temperature-based estimates of the tipping points deviate by  $-38 \pm 17$  and  $-26 \pm 15$  days, respectively. This confirms that the use of conventional definitions for the onset and end of the period of vegetation may yield significantly different results compared to temperature-based estimates. We conclude that temperature aggregates can be included easily in hydrological analyses and suggest that they may have a higher explanatory power than static conventional definitions for the beginning and end of the period of vegetation such as the equinoxes.

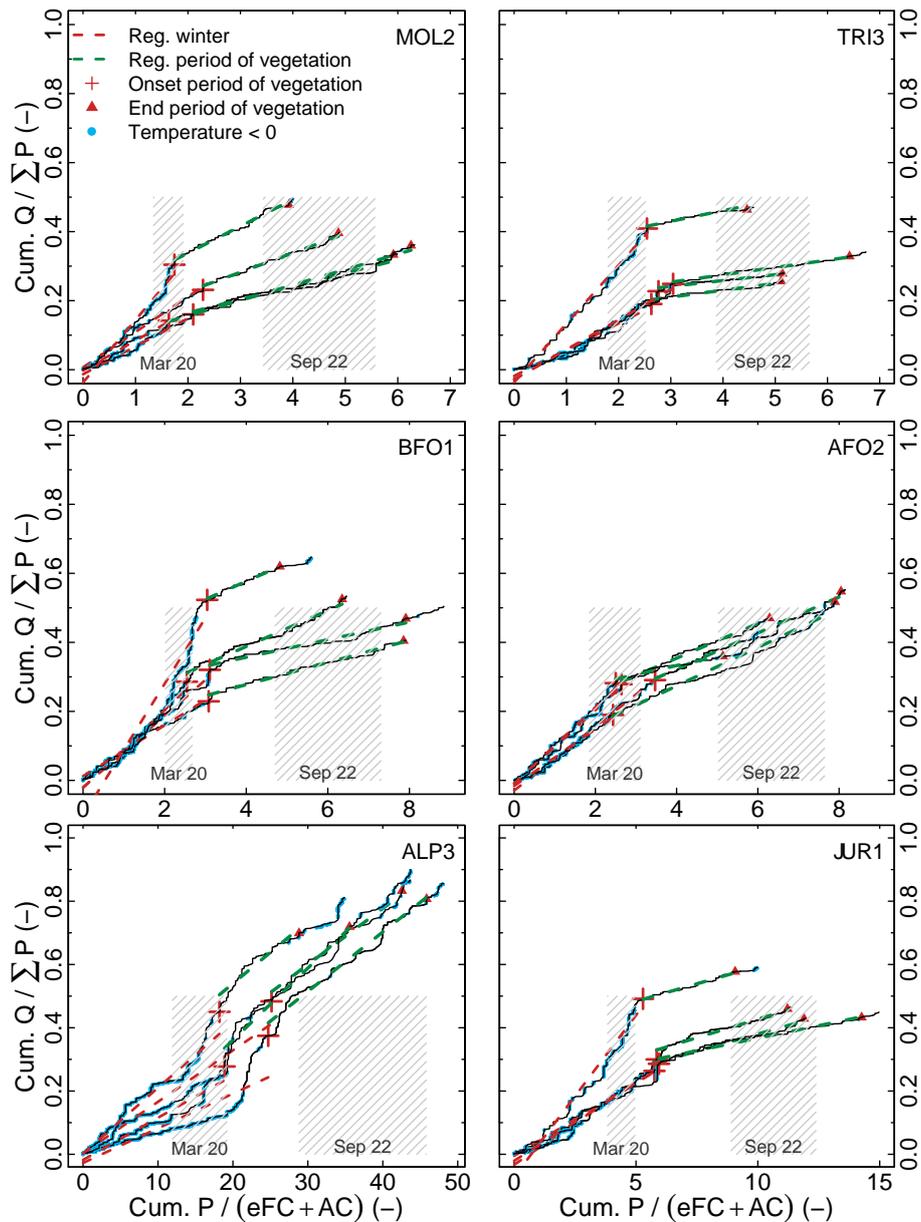


**Figure 4.** Type I dimensionless double mass curves for catchments of different geological units for the hydrological years 1999–2003. Onset and end of the period of vegetation are determined using a temperature-index model. Regression lines are fitted to both periods (dotted lines in red/green); their slopes are interpreted as seasonal runoff coefficients. Periods with temperatures  $< 0^{\circ}\text{C}$  are highlighted in blue. Gregorian definitions for the first start of spring (20 March) and the last start of autumn (22 September) are added to the cum.  $P/\sum P$  plane as hatched rectangles (to improve readability) to highlight their differences to temperature-based estimates at the onset and end of the period of vegetation. Statistical properties of the dDMC are summarized in Table 1.

### 3.1.2 Season-specific evaluation of type II double mass curves

The type II dDMC, which scale accumulated supply by the specific storage volume of the soil in Fig. 5 reveal a similar shape as the type I dDMC in Fig. 4. They exhibit an obvious regime shift between steep winter and flat summer regression slopes but expose a larger spread and a less lin-

ear progression. More importantly, these curves reveal differences among the different hydrological years. Dry and wet years are nicely separated, as indicated by the end points of the individual curves. The type II dDMC is also much more suited to revealing differences among catchments, in terms of how “fast” the catchments release a fraction of rainfall input in terms of how often the annual rainfall supply could potentially renew the soil water stock. This is particularly



**Figure 5.** Type II dimensionless double mass curves for six catchments of different geological units. The dDMC depicted here are the same as those depicted in Fig. 4 but with the difference that the abscissa is scaled using the sum of average effective field capacity (eFC) and air capacity (AC). All other properties are analogous to Fig. 4.

visible when comparing the ALP3 catchment in the lower left panel of Fig. 5, with a potential renewal rate of 30 to 40 times the soil pore volume, against the catchments from the other settings with potential renewal rates almost 1 order of magnitude lower. Another advantage is the visibility of differences in winter slopes, which are more uniform in the non-Alpine catchments compared to type I dDMC. Average  $m_W$  values and corresponding inter-annual variations ( $\sigma_{m_W}$ ) are provided in Table 1.

With respect to the shape of the winter segment of the type II dDMC, we find that it is fairly linear for catchments

located in the Faulted Molasse, Triassic, and pre-Alpine regions (MOL2, TRI3, and AFO2) (Fig. 5, red lines). Step changes towards dDMC segments with smaller slopes indicate activation of additional storage and thus a slower growth of contributing areas with catchment relative storage. Relating to the beta store concept, this corresponds to a larger  $\beta$  parameter. This is particularly visible in the winter periods of the Alpine catchments (ALP3), but also in the Bavarian Forest (BFO1) due to storage in the snowpack. This is also suggested by temperature data as these segments coincide with frost periods and are followed by segments of warm-

**Table 1.** Mean seasonal winter ( $CR_W$ ), summer ( $CR_S$ ), and annual runoff coefficients ( $CR_{YR}$ ) as indicated by the slope of linear regression lines fitted to the different segments of the type I dDMC. The analogue for the type II dDMC is  $m_W$ , which describes how fast dimensionless relative release grows with the potential renewal rate of the soil stock during winter. The inter-annual variation of these quantities within the hydrological years (1999–2003) is quantified using the mean absolute deviation and is given by  $\sigma_{CR_W}$ ,  $\sigma_{CR_S}$ ,  $\sigma_{CR_{YR}}$ , and  $\sigma_{m_W}$ , respectively.  $sFDC_W$ ,  $sFDC_S$ , and  $sFDC$  characterize the slope of seasonal flow duration curves between the 33rd and 66th percentiles of the dormant period, the period of vegetation, and all data, respectively. The separation of seasons was done using a temperature-index model (Menzel et al., 2003). All quantities are dimensionless.

Site	$CR_W$	$CR_S$	$CR_{YR}$	$m_W$	$\sigma_{CR_W}$	$\sigma_{CR_S}$	$\sigma_{CR_{YR}}$	$\sigma_{m_W}$	$sFDC_W$	$sFDC_S$	$sFDC$
TRI1	0.72	0.12	0.43	0.098	0.058	0.030	0.031	0.026	-1.15	-0.88	-1.46
TRI2	0.70	0.07	0.37	0.111	0.198	0.027	0.105	0.033	-1.24	-1.04	-1.94
TRI3	0.55	0.12	0.34	0.105	0.133	0.024	0.069	0.040	-0.93	-0.65	-0.99
JUR1	0.73	0.25	0.48	0.063	0.150	0.017	0.054	0.039	-0.88	-0.49	-0.78
BFO1	0.82	0.28	0.52	0.120	0.169	0.037	0.068	0.035	-0.58	-0.48	-0.61
BFO2	0.85	0.30	0.54	0.114	0.143	0.020	0.067	0.034	-0.69	-0.39	-0.64
BFO3	0.93	0.29	0.57	0.126	0.173	0.034	0.089	0.035	-0.82	-0.46	-0.76
MOL1	0.60	0.24	0.38	0.121	0.103	0.040	0.021	0.009	-0.60	-0.54	-0.63
MOL2	0.56	0.27	0.40	0.113	0.080	0.022	0.049	0.028	-0.38	-0.22	-0.36
MOL3	0.62	0.34	0.46	0.121	0.069	0.019	0.045	0.010	-0.32	-0.24	-0.27
MOL4	0.56	0.23	0.37	0.101	0.084	0.021	0.051	0.020	-0.53	-0.31	-0.50
MOL5	0.69	0.22	0.41	0.134	0.055	0.028	0.026	0.027	-0.56	-0.36	-0.57
MOL6	0.60	0.20	0.36	0.152	0.066	0.018	0.025	0.005	-0.61	-0.42	-0.58
MOL7	0.35	0.27	0.31	0.110	0.139	0.089	0.031	0.003	-0.55	-0.99	-0.71
AFO1	0.72	0.35	0.51	0.135	0.031	0.120	0.042	0.027	-1.23	-1.25	-1.32
AFO2	0.68	0.34	0.49	0.098	0.036	0.099	0.043	0.003	-0.77	-0.58	-0.76
AFO3	0.56	0.24	0.38	0.115	0.130	0.068	0.031	0.016	-0.72	-0.84	-0.75
AFO4	0.66	0.22	0.39	0.136	0.050	0.090	0.030	0.035	-0.73	-1.42	-1.03
ALP1	0.89	0.50	0.75	0.045	0.098	0.088	0.031	0.043	-1.21	-0.93	-1.14
ALP2	0.71	0.82	0.83	0.017	0.067	0.161	0.047	0.037	-0.65	-0.76	-0.83
ALP3	0.64	0.84	0.86	0.015	0.082	0.109	0.025	0.013	-1.01	-0.66	-1.07
ALP4	0.66	0.53	0.64	0.024	0.064	0.066	0.032	0.059	-0.52	-0.30	-0.49
Mean	0.67	0.32	0.49	0.099	0.10	0.06	0.046	0.026	-0.76	-0.66	-0.83

ing periods with much steeper slopes. Clear step changes towards steeper slopes in the frost-free period are visible in the Alpine settings (ALP3) and to a smaller degree in AFO2 and BFO1, suggesting activation of rapid flowpaths connected to the gauge. At the Jurassic site (JUR1), where the subsurface is karstified, the likely rapid runoff formation can not be distilled by either the type II or the type I dDMC. A possible explanation is that intensity controlled processes such as the formation of Hortonian overland flow and rapid subsurface flow take place at a temporal sub-scale of the dDMC which is tailored for the seasonal and annual timescales.

### 3.1.3 Spatial patterns and inter-comparison of seasonal runoff formation

With respect to the different physiographic settings, the type I dDMC reveal distinct seasonal and spatial patterns. As indicated by the values in Table 1 and Fig. 7, the highest average winter runoff coefficients ( $CR_W = 0.8$ – $0.9$ ) occur in the north-eastern catchments (BFO1, BFO2, BFO3) which are rather densely forested, but also in the Alpine ALP1 catchment. ALP4, ALP3, and ALP2, which are also Alpine catch-

ments and located at similar altitudes, show much smaller  $CR_W$  values of 0.64 to 0.71 on average, most likely due to storage in the snowpack. The smallest winter runoff coefficients (0.35 and 0.55) occur in MOL7 and TRI3, respectively. With respect to the inter-annual winter variance, we encounter small mean absolute deviations  $\leq 0.05$  in lowland sites of the Molasse and glacial drift areas, e.g. MOL5, AFO4, AFO2, and AFO1. High mean absolute deviations  $\geq 0.15$  occur in different geologies, including the TRI2, BFO1, JUR1, and BFO3 sites. Please also note that  $CR_{YR} > CR_W$  in a few cases where snow exhibits a strong control on winter runoff regimes (e.g. ALP3 or ALP2; see Table 1). Here, fitting linear regressions to the double mass curves is not suitable for estimating seasonal winter runoff coefficients.

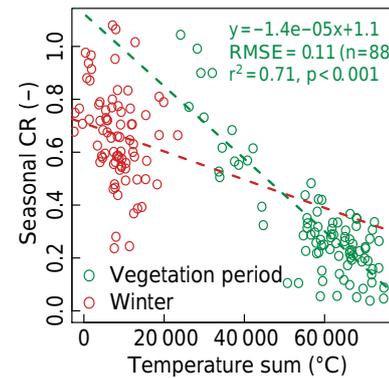
The summer season is characterized by an opposite spatial pattern.  $CR_S \geq 0.8$  occur in the snow-dominated Alpine catchments of ALP3 and ALP2. The smallest  $CR_S$  with values between 0.07 and 0.12 are encountered at the Triassic sites (TRI3, TRI2, TRI1). In contrast to these end-members, the variations within and among the Bavarian Forest, Molasse, and Alpine Foreland regions differ between 0.20 and

0.35, indicating similar seasonal regimes at many sites. This implies that the large-scale geological setting, which we used for the naming of the sites (chap. 2.1), does not coincide with the spatial patterns in  $CR_S$  and  $CR_W$ . It is also important to note that for several low-land sites the  $CR_S$  shows very little inter-annual variance, as indicated by mean absolute deviations  $\leq 0.03$  (e.g. MOL5, TRI3, TRI2, MOL6, MOL2, MOL4, JUR1, MOL3, and others) (Table 1). In these catchments the slopes of the dDMC are fairly constant throughout different hydrological years, indicating a very strong control of evapotranspiration on the water balance during summer. At these sites the curves of the dDMC in summer have nearly identical slopes, which is due to their state at the onset of vegetation activity in spring.

### 3.1.4 Regression analysis of dDMC characteristics with catchment characteristics

To better understand the physiographic controls on the seasonal runoff regimes, we regress summer and winter runoff coefficients against the 24 different site-specific variables provided in Tables A1 and A2. A key finding in this evaluation is that temperature sums, which can be considered as bulk parameters for ecological activity, explain 71 % of the variance of the  $CR_S$  with respect to the entire range of the physiographic setting in our data set. During winter, temperature aggregates are not significant and without predictive power (Fig. 6). We thus state that the period of the vegetation exerts first-order controls on runoff formation as it dominates the seasonal interplay in storage and release in all of our physiographic and climatic settings except for the snow-dominated Alpine region.

With respect to  $CR_W$ , which correspond to the winter slopes of the type I dDMC, we did not find any single variable or combination of variables which explained more than 30 % of the variance in the  $CR_W$ . The best significant predictors for the latter, all with  $p$ -values  $\leq 0.001$ , are sand content ( $r^2 = 0.29$ ), the soil topographic index ( $r^2 = 0.22$ ), and silt content ( $r^2 = 0.22$ ), which are followed by forest coverage ( $r^2 = 0.16$ ), skeleton content ( $r^2 = 0.15$ ), number of frost days ( $r^2 = 0.14$ ), effective field capacity ( $r^2 = 0.13$ ), and absolute sum of negative temperatures ( $r^2 = 0.12$ ). All other variables provided in Tables A1 and A2 either have coefficients of determination  $r^2 \leq 0.10$  or the relation was not significant. In several multiple linear regressions based on the variables in Tables A1 and A2, the best result is obtained for a combination of the soil topographic index, forest cover, and absolute sum of negative temperatures (multiple  $r^2 = 0.30$ ,  $p$ -value  $< 0.001$ ). Summer temperature sums, length, or end of the period of vegetation from the previous hydrological year do not improve the prediction of the actual  $CR_W$ . The key finding of this analysis is that  $\tau$  yields  $r^2 = 0.22$ , whereas  $\phi$  and  $K_s$  alone only explain 0.02 % ( $p$ -value = 0.18) and 0.08 % ( $p$ -value = 0.009) of the variance in the  $CR_W$ , respectively. This confirms that surrogates for



**Figure 6.** Hourly summer and winter temperature sums, calculated for the dormant period and the period of vegetation, plotted against seasonal summer (green) and winter (red) runoff coefficients for all sites ( $n = 22$ ) and years ( $n = 4$ ). The dotted lines are linear regressions. Statistical information on the summer model is plotted in green. During winter we did not find a significant statistical relation ( $r^2 = 0.04$ ,  $p = 0.062$ ). Statistical properties of all dDMC are summarized in Table 1.

gradients and resistances act as a team, as their product jointly controls the water flux in Darcy's law. The finding hence supports the concept of functional units described by Zehe et al. (2014). With respect to  $CR_W$  their joint impact is, despite the coarse spatial resolution of the underlying soil and topographic map, even detectable at the lower mesoscale. We hence suggest that physically meaningful combinations of catchment descriptors should gain more attention in catchment inter-comparison studies.

To further explore runoff formation during winter, we evaluate average  $m_W$  values of the type II double mass curves within a similar regression exercise. The latter characterizes how fast accumulated dimensionless water release grows with the potential renewal rate of the soil storage. The most explanatory variables for  $m_W$ , again all with  $p$ -values  $\leq 0.001$ , are  $\phi$  ( $r^2 = 0.44$ ), relative fraction of rock outcrops ( $r^2 = 0.34$ ), length of the previous period of vegetation and corresponding temperature sums ( $r^2 = 0.34$ ,  $r^2 = 0.33$ ), total modelled snow water equivalent ( $r^2 = 0.30$ ), skeleton content ( $r^2 = 0.22$ ), and number of frost days ( $r^2 = 0.19$ ). All correlations are negative except for the two variables that relate to the previous period of vegetation. Average root depth, top soil air capacity (AC), and effective field capacity (eFC) also showed high and significant coefficients of determination ( $r^2 = 0.39$ , 0.35, and 0.26, respectively), but non-trivial spurious fractions of correlation (Pearson, 1987; Kenney, 1982) make these variables difficult to interpret and, hence, of little use. With respect to all sites and the 4 hydrological years, potential renewal rates for winter differ by more than 1 order of magnitude (minimum = 1.1, maximum = 19.2). Average and standard deviations of all sites are 3.8 and 4.0, respectively.

There are two main findings in this evaluation. The first is that properties of the preceding summer period control the speed of relative water release in winter. This implies a long-term catchment memory, which is most likely caused by the baseflow regimes, at some sites. In contrast, this is not the case for the period of vegetation, when antecedent conditions from the winter period seem to have little or no impact on the growth of dimensionless water release to the stream. The second important finding is that the topographic gradient is the most explanatory variable for  $m_W$ , while the soil topographic index is of little use here. This might suggest that capacity controlled runoff formation in the riparian zones is dominant in most of the areas and that relative water release to the stream grows with accumulated water supply into these wetlands. However, the correlation of  $m_W$  and topographic gradient, fraction of rock outcrops, modelled snow water equivalent, skeleton content, and number of frost days may also be a signal for the impact of physiographic and hydrometeorological conditions. In our data set, all of these variables increase with altitude and have their maximum in the Alpine region.

In summary, we state that dimensionless and season-specific double mass curves provide manifold insights into the interplay of accumulated rainfall forcing and release among mesoscale catchments. As expected, the type I dDMC, which is based on a trivial scaling of the abscissa, revealed a much more homogeneous picture of seasonal runoff formation than the type II dDMC, which uses proper scaling by considering the capacity of the soil storage as a characteristic quantity for storage-driven runoff formation. We hence state that the type II dDMC is the signature of choice, but we require more information.

### 3.2 Comparison of type I and type II dDMC and (seasonal) flow duration curves

Season-specific evaluations reveal that seasonality in the rainfall–runoff formation is only weakly reflected by the slope of the flow duration curve. As shown in Table 1, the slopes of the flow duration curve between the 33rd and 66th percentiles are smaller during the growing period than during the dormant period. On average, absolute differences between  $sFDC_W$  and  $sFDC_S$  amounted to 0.24. Correlation analysis between  $sFDC_W$  and  $sFDC_S$  yields a coefficient of correlation of 0.64 ( $r^2 = 0.40$ ,  $p = 0.0014$ ). The slopes of the “all-data” flow duration curves  $sFDC$  and  $sFDC_W$  are more narrowly correlated ( $r^2 = 0.82$ ,  $p < 0.001$ ) than  $sFDC$  and  $sFDC_S$  ( $r^2 = 0.52$ ,  $p < 0.001$ ), implying that  $sFDC$  rather reflects winter conditions than summer conditions. Comparing  $sFDC$  to seasonal coefficients of summer and winter flow variation yields coefficients of determination of  $r^2 = 0.58$  and  $r^2 = 0.47$ , respectively. In contrast, we find only weak and insignificant correlations (all  $r < 0.38$ ) among the different (seasonal) runoff coefficients and the slopes of the (seasonal) flow duration curves. This shows that season-

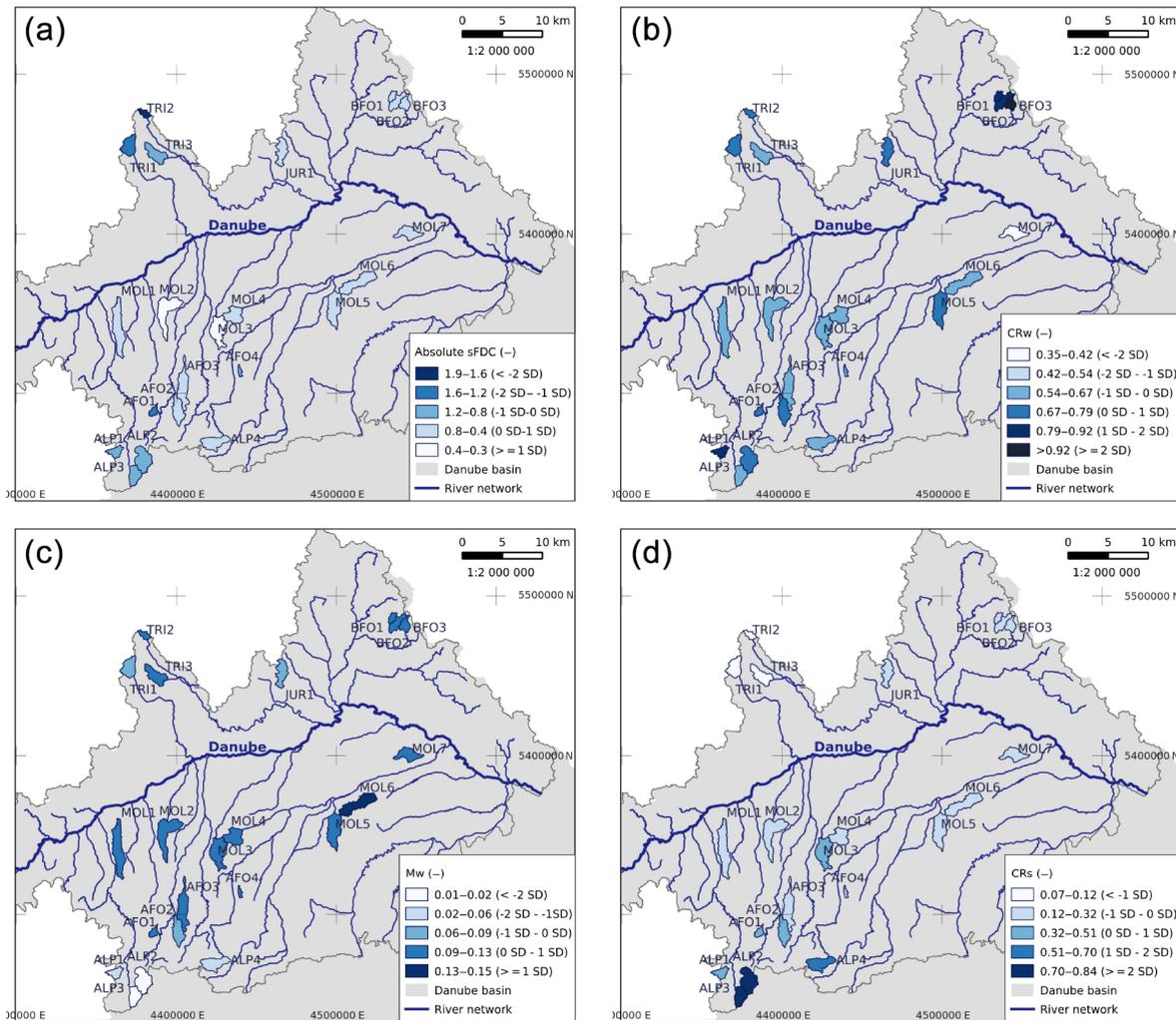
ality in the rainfall–runoff regimes is poorly reflected by FDC regardless of whether they are derived season-specific or for longer periods of time.

Comparing  $sFDC$ ,  $CR_W$ ,  $CR_S$ ,  $m_W$ , and the corresponding spatial patterns in Fig. 7, we find that the  $sFDC$  values (top left) characterize the Triassic flow regimes as the most variable ones, followed by Alpine sites and catchments from the Alpine Foreland. The smallest  $sFDC$  values, implying the most damped response, occur in the Molasse area (MOL2 and MOL3). The remaining catchments from the Molasse area, the catchments located in the Bavarian Forest, and two sites from the Alpine Foreland are uniformly classified. The rate of dimensionless release (panel in the bottom left) illustrates a different pattern. The  $m_W$  values from the Alpine region are 1 order of magnitude smaller than those in the Molasse area. However, the Triassic region is classified similarly to the Bavarian Forest and similarly to most of the catchments in the Molasse area. The spatial patterns of the  $CR_W$  and  $CR_S$  values are again different.  $CR_W$  (top right panel) suggests that there is little differentiation among most catchments except for the sites located in the Alpine region and the Bavarian Forest (BFO). In contrast,  $CR_S$  (bottom right panel) end-members are the Alpine sites and the catchments located in the Triassic region. The differences among the other sites are small.

In line with the spatial pattern analyses, we conclude that geology alone is not a good predictor of the hydrological behaviour of the selected catchments and that the different approaches reveal fairly different information. Specifically, dDMC provide means to characterize seasonal rainfall–runoff regimes that can not be extracted from other signatures such as FDC which are frequently used to characterize hydrological similarity. This holds true even if FDC are based on seasonal data.

## 4 Discussion and conclusions

The rationale of this study was to derive data-driven dimensionless diagnostics for seasonal runoff formation and the water balance. Motivated by the notion of the three main hydrological functions of catchments proposed by Wagener et al. (2007), we tested the feasibility of different types of dimensionless double mass curves for this purpose. Conceptually we designed these curves to characterize annual patterns of dimensionless water release to the stream as a function of dimensionless water supply to the catchment storage. Their feasibility was tested using operational data from 22 lower mesoscale catchments. Within this test we compared two different approaches to obtain dDMC and put special emphasis on detecting and explaining (a) seasonal differences in annual runoff formation and (b) differences among the different catchments, which we a priori classified into five major geological and climatological settings. Our results provide evidence that dDMC are straightforward to implement and



**Figure 7.** The panels show maps of (a) absolute values of the slope of the flow duration curves (sFDC) (–), (b) winter runoff coefficients ( $CR_w$ ) (–) obtained from the type I dDMC, (c) the winter rate of how fast dimensionless relative release grows with the potential renewal rate of the soil stock ( $m_w$ ), and (d) summer runoff coefficients ( $CR_s$ ) (–). The applied colour code is derived from the standard deviation of each data set.

well suited for characterizing the seasonal pattern of dimensionless water supply and release to streamflow as discussed here below.

#### 4.1 Potential and limitations of dimensionless double mass curves

We demonstrated the potential for dimensionless formulation and season-specific evaluation of the traditional double mass curves to separate influences that relate to simple scaling of total supply from those that relate to catchment-specific contributions controlled by saturated area dynamics. This is a step ahead compared to the findings of Pfister et al. (2002) and Jackisch (2015). In line with their findings, we also found characteristic regime shifts in dDMC between “steep” winter and “flat” summer conditions; however, our test catchments are 1 to 2 orders of magnitude larger than those used in

their studies. Our results revealed also that neither seasonal FDC nor usual FDC, which are frequently used to characterize runoff regimes in catchment inter-comparison studies (Oudin et al., 2010; Sawicz et al., 2011; Casper et al., 2012; Viglione et al., 2013), yielded similar information. We hence conclude that season-specific dDMC are a well-suited fingerprint for characterizing seasonal runoff formation in mesoscale catchments of temperate environments and that dDMC are suited for inter-comparison studies.

A drawback of dDMC is, however, that accumulated dimensionless water supply is at best an estimator for total relative storage in the catchment. We neither have information on where the water is stored nor on whether it is subject to strong, weak, or no capillary or osmotic forces. Even more importantly, dDMC do not characterize the entire water balance. We cannot infer whether the mass balance residual of

a hydrological year is partly stored within the catchment or whether it leaves the system as evaporation. The real object of desire is, hence, a dimensionless triple mass curve which adds dimensionless  $E$  as a third dimension to the double mass curve. As suggested by Seibert et al. (2016b), we expect that the resulting 3-D curves could also serve as an exhaustive diagnostic fingerprint for similarity in the annual and seasonal water balance. In combination with an analysis of the autocorrelation of the mass balance residuals, triple mass curves could serve as an indicator of long-term memory effects. The drawback is here, however, that triple mass curves are not entirely based on observables, as measurements of  $E$  are commonly not available. Data-driven estimates of  $E$  can at best be provided in terms of maximum potential evaporation, e.g. by dividing net radiation with the latent heat of vaporization.

Last but not least we suggest that dDMC can be used for model evaluation. Specifically, we propose that the model-based reproduction of the tipping point between winter and summer (regime shift) can serve as a powerful benchmark for the accuracy of a model with respect to timing, which is important but rarely done in hydrology (Seibert et al., 2016a). The reproduction of this signature is particularly relevant for models which are used for climate change studies where coping with the most likely non-stationary role of biotic controls (Milly et al., 2008) is of utmost importance.

#### 4.2 Appropriate scaling of catchment input and output

Proper scaling means to divide a state variable of a system or a flux leaving the system by a time-invariant characteristic quantity (often a length or time) which limits the processes, i.e. the system dynamic of interest. The Reynolds number scales driving inertial forces, which scale with velocity and a characteristic length of the system, with counteracting viscous forces to characterize turbulence in a scale-independent manner (Reynolds, 1883). The Damköhler number relates residence time in a reactor to the first-order reaction constant, to characterize its efficiency in a scale-independent manner (Fogler, 1999; Oldham et al., 2013). We may also formulate cumulated input–output relations in a scale-independent form, e.g. the dimensionless breakthrough curve characterizing transport and adsorption properties of a soil probe in a scale-invariant manner (Hillel, 2004). In analogy to this curve, we designed the double mass curve to characterize accumulated dimensionless release of water to the stream as a function of dimensionless water supply. Our choice of this input–output relation, and in particular our approaches to scale input and output fluxes, are based on implicit assumptions about the dominant runoff generation processes.

Our scaling of the accumulated output by the annual rainfall depth essentially assumes that primarily the amount of rainfall limits the annual amount of streamflow generation and not so much the temporal pattern within the year or rainfall totals at the event scale. Scaled cum.  $P$  can however

also be interpreted as the characteristic time for accumulated direct runoff formation, and time stands still in case of no precipitation. Here we essentially assume that it is only the precipitation amount that counts, but not its intensity. This assumption might become inappropriate in areas where Hortonian overland flow dominates runoff generation (Mualem et al., 1990; Cerdan et al., 2002) because the intensity of the precipitation input crucially determines whether runoff leaves the system or not (Reaney et al., 2014). In such areas the same annual rainfall total can hence cause different amounts of annual release, depending on the subscale rainfall intensity pattern.

In line with this, we assume water release to the stream to increase monotonically with relative water storage. If this was true, accumulated release should scale with the ratio of cum.  $P$  over  $S_{\max}$  as shown in Eq. (B5). By estimating the potential renewal rate of the soil storage, we furthermore assume that transport through the subsurface is not limiting. Due to the absence of detailed information about active subsurface storage volumes in the target area, we scale cum.  $P$  with the specific storage volume of the soil ( $AC + eFC$ ) in the root zone to obtain what we call the type II dDMC. We are aware that this is not the best choice as estimates of  $AC$  and  $eFC$  are uncertain (particularly at the catchment scale) and refer to the root zone instead of to active storage – but this is what is available from operational data sets in southern Germany. In any case we suggest testing additional scaling approaches by replacing the denominator in Eq. (6) with a more meaningful surrogate that relates to active storage. Alternatively, we scaled the annual accumulated rainfall with the annual rainfall depth of the respective hydrological year (Seibert et al., 2016b). The type I dDMC preserve the shape of the non-scaled double mass curve and surrogate similarity among the different hydrological years in the same catchment and among different catchments as well. This is because both axes of the type I dDMC are scaled to unity. However, as the same step on the abscissa does not imply the same amount of mass input into the system, this similarity is in fact pseudo-similarity; dry and wet years appear as the same. In line with Blöschl and Sivapalan (1995) we hence conclude that proper scaling requires consideration of a suitable characteristic length and that the annual rainfall amount cannot be regarded as a characteristic length scale for supply in this respect.

In contrast, we found that the type II dDMC are well suited to distinguishing inter-annual differences between dry and wet years within the same catchment. They furthermore nicely visualize that the growth of relative water release with the potential renewal rate of the storage was rather different between the different geological settings, while being typical within those settings. We hence conclude that the type II dDMC is well suited to discriminating seasonal patterns of runoff formation among different landscape settings and that its shape allows evaluation of whether the above-mentioned assumptions are met or not. A linear growth of dimensionless

water release with the ratio of cum.  $P$  over  $S_{\max}$  indicates that runoff generation grows rather linearly with relative saturation, i.e. if the form parameter in the beta store is close to unity. This can be for instance due to the dominance of saturation excess runoff formation in the riparian zone and is consistent with our finding that the topographic gradient was the best predictor for the winter slopes of the type II dDMC within the catchment inter-comparison. Deviations towards steeper slopes indicate that runoff generation is to a lesser extent growing with potential renewal of the storage. This might for instance reflect the importance of intensity controlled runoff formation such as Hortonian overland flow or rapid preferential flow. In summary, we state the type I dDMC focus on the “plain water balance”. Their strength is in characterizing seasonal regimes, whereas type II dDMC relate to storage referenced states.

It is also worth briefly reflecting on meaningful separations and definitions of seasons and years. We argue that concepts like the hydrological year or conventional definitions like the equinoxes which are often used as a proxy for the beginning and end of the period of vegetation or the filling of the catchment storage are fairly limited. This is in line with Hellebrand et al. (2008), who state that a meaningful separation of seasons is not straightforward to determine as hydrologically meaningful definitions of these seasons beyond calendaric ones are not available. The same applies for instance for dynamic storage (Sayama et al., 2011). A comparison of dynamic storage time series would require that integration starts at the same relative storage amount in all catchments – and not at the same date. Significant dry or wet periods when subsurface wetness can be deemed as being either near saturation or near the minimum could for instance help to identify these points in time. These considerations differ from the dimensionless time used for instance by Woods (2003), who defined dimensionless time relative to the length of a seasonal cycle, which was 1 year in his study for sites outside the tropics.

### 4.3 Physiographic and ecological controls on seasonal runoff formation

The analysis of physiographic and ecological controls revealed several important findings. The first is that temperature data prove to be good predictors for summer runoff coefficients, independent of the physiographic and climatological conditions represented by our data set. This is in line with approaches used in plant and tree physiology where air temperature aggregates are widely used for instance to estimate photosynthetic capacity (Mäkelä et al., 2004) or as harvest and hence biomass predictors (Perry et al., 1993; Rawson and Gomez-Macpherson, 2000). The applied temperature sum definitions however differ with respect to the process of interest. In Finland effective temperature sums are for instance defined as the sum of positive differences between the diurnal mean temperature and 5 °C (Solantie,

2004). Holdridge (1967) includes accumulated mean daily temperature (in °C) above zero, starting on 1 January among other variables to measure vegetative development in the tropics. In our analysis we employ hourly temperature sums for the dormant period and the period of vegetation. It remains open for future research to find out whether different definitions of temperature aggregates can further increase the prediction of summer runoff coefficients. In this context we also raise the question of whether characteristic ecological timescales are adequately represented in hydrological models. The parametrization of variables such as albedo, leaf-area index, stomata, and cuticle resistance which govern plant and canopy roughness still prevails in the form of time-invariant annual look-up tables which have typically been derived within a specific hydro-climatic setting – albeit that they exhibit pronounced seasonal and spatial dynamics (Reichenau et al., 2016). This even applies for so-called “physical” models such as Catflow (Maurer, 1997; Zehe et al., 2001) or WaSiM-ETH (Schulla, 1997), and it is in particular relevant for climate change studies (Milly et al., 2008). For further reading please refer to Loritz et al. (2017), who conduct additional studies on this topic and, among others, confront modelled  $E$  estimates with sap-flow data and both Gregorian and temperature-based estimates for the beginning of the vegetation period.

The second important finding is that the soil topographic index is an important predictor of the average winter runoff coefficients, while the topographic gradient and the saturated hydraulic conductivity alone are not. Because gradients and resistances exert first-order controls, it is not surprising that both variables are often interpreted as independent predictors (McGuire et al., 2005; Santhi et al., 2008; Sayama et al., 2011; Li et al., 2012), while they in fact act as a group. The latter is essentially also supported by the gradient–flux relationship: picturing a synthetic hillslope, we intuitively expect that a doubled topographic gradient and a half  $K_s$  would yield the same Darcy flux. We hence suggest that physically meaningful combinations of catchment descriptors should gain more attention in catchment inter-comparison studies. In this context it is also important to note that seasonal runoff coefficients are independent of the dDMC. They only depend on the separation of seasons. Please note also that the areal share of an impermeable substratum within the catchment is considered a good predictor of  $CR_W$  (Hellebrand et al., 2008). We did not test this variable as quantitative estimates were operationally not available.

The last important aspect is that type II dDMC provide a means to analyse how fast relative release increases with the potential renewal rate of the soil stock. dDMC with a linear “first-order type” shape suggest that streamflow release from a catchment works similarly to how saturation excess systems operate and that the form parameter in the beta store is close to unity. This is in line with the concept proposed by Kirchner (2009). Step changes and deviations of the dDMC towards “flatter” and “steeper” segments may indicate the ac-

tivation of additional storage, e.g. in the snowpack in winter or snowmelt, and the relevance of intensity controlled processes in summer. Care must be taken if snow dominates parts of the dDMC, as the interpretation of these periods is ambiguous. We also want to highlight that type II dDMC are able to reveal memory effects between winter runoff conditions and the previous period of vegetation that can not be extracted from type I dDMC. This confirms that scaling the abscissa using a storage proxy is more meaningful than the use of total annual rainfall. To increase confidence in the proposed signatures, we suggest applying them (i) to a larger number of catchments and (ii) to a (nested) set of small and densely instrumented catchments with more homogeneous physiognomic predictors.

## 5 Conclusions

In summary, we conclude that data-driven dDMC turn out to be an easy-to-compute yet very powerful diagnostic signature to characterize seasonal runoff formation and the water balance – also with respect to comparative analyses. They allow separation of terrestrial controls on runoff formation from that of the meteorological forcing, proper scaling and a season-specific evaluation provided. This also holds true as a comparison of dDMC and (seasonal) flow duration curves shows that the former provide information that can not be extracted from the latter.

The use of cumulated precipitation to estimate supply can also be interpreted as the characteristic time for accumulated direct runoff formation, and time stands still in case of no precipitation. Here we essentially assume that it is only the precipitation amount which counts, but not its intensity. Overall we conclude that dimensionless double mass curves, in combination with an ecological temperature index, are well suited to unravelling biotic and abiotic controls on seasonal runoff formation – as long as runoff formation monotonously increases with catchment storage. They are particularly suited to depicting shifts between the winter and summer regimes, and where either runoff generation or evapotranspiration dominates water release.

Nevertheless, it is clear that seasonal scale is only one perspective of runoff formation and that understanding functional similarity, i.e. the behaviour of catchments (Schaeffli et al., 2011), requires consideration of not only diagnostics for different temporal scales of runoff formation, but also diagnostics for threshold processes.

*Data availability.* The data and models used in this study were provided within the scope of a research project by the flood forecasting unit of the Bavarian Environmental Agency (LfU). It can be obtained upon request from the flood forecasting agency ([www.hnd.bayern.de](http://www.hnd.bayern.de), Bayerisches Landesamt für Umwelt, 2017).

## Appendix A: Soil and physiographic catchment properties

**Table A1.** Physiographic catchment properties in terms of topography, land use, and hydro-meteorology. The columns contain the site identifier (ID), catchment size ( $A$ ), mean catchment elevation above sea level (elev), median topographic gradient ( $\phi$ ), soil topographic index ( $\tau$ ), relative land coverage ratios for infrastructure (infr), arable land (arab), pasture (past), forest (frst), wetlands (wet) and rock outcrops (rock), the 30-year mean annual precipitation (MAP), 4-year mean annual precipitation ( $\bar{P}$ ), discharge ( $\bar{Q}$ ), runoff coefficient ( $\bar{CR}$ ), and streamflow coefficient of variation ( $v_{\bar{Q}}$ ).

ID	topography				% land use coverage						hydrometeorological characteristics				
	A (km <sup>2</sup> )	elev (m)	$\phi$ (-)	$\tau$ (m s <sup>-1</sup> )	infr	arab	past	frst	wet	rock	MAP (mm)	$\bar{P}$ (mm)	$\bar{Q}$ (mm)	$\bar{CR}$ (-)	$v_{\bar{Q}}$ (-)
TRI1	88	481	2.8e-02	1.0e-06	0.04	0.54	0.20	0.21	0	0	802	919	0.045	0.43	2.00
TRI2	26	460	2.5e-02	9.8e-07	0	0.60	0.12	0.29	0	0	707	801	0.033	0.36	2.20
TRI3	93	468	3.8e-02	8.3e-07	0.02	0.62	0.07	0.30	0	0	738	829	0.032	0.33	1.80
JUR1	90	518	7.2e-02	2.0e-06	0.01	0.59	0.15	0.26	0	0	833	839	0.046	0.48	0.83
BFO1	25	620	6.9e-02	2.5e-06	0	0.35	0.06	0.60	0	0	889	933	0.054	0.51	0.89
BFO2	64	635	6.1e-02	1.3e-06	0.02	0.39	0.12	0.47	0.01	0	893	920	0.055	0.53	0.91
BFO3	58	624	7.9e-02	1.2e-06	0.01	0.24	0.21	0.55	0	0	908	825	0.052	0.56	1.10
MOL1	166	543	1.1e-02	1.6e-08	0.07	0.42	0.29	0.23	0	0	889	973	0.042	0.38	0.83
MOL2	163	515	4.0e-02	6.9e-08	0.03	0.28	0.37	0.32	0	0	901	1010	0.045	0.39	0.90
MOL3	163	558	1.4e-02	2.0e-08	0.05	0.69	0.10	0.15	0	0	933	1100	0.057	0.45	0.64
MOL4	97	517	2.5e-02	3.9e-08	0.04	0.77	0.03	0.15	0	0	888	1016	0.042	0.36	0.93
MOL5	133	473	2.0e-02	4.4e-08	0.05	0.81	0.05	0.09	0	0	883	1016	0.047	0.40	1.00
MOL6	146	484	4.2e-02	7.8e-08	0.02	0.79	0.04	0.15	0	0	856	721	0.029	0.35	1.60
MOL7	87	379	2.4e-02	3.4e-08	0.02	0.73	0.01	0.24	0	0	744	733	0.026	0.31	1.10
AFO1	45	840	3.4e-02	3.6e-08	0.01	0.11	0.27	0.62	0	0	1388	1243	0.073	0.51	1.90
AFO2	95	777	4.0e-02	9.1e-08	0.01	0.11	0.55	0.31	0.01	0	1292	1466	0.083	0.50	1.30
AFO3	136	751	2.2e-02	5.2e-08	0.05	0.12	0.62	0.22	0	0	1198	1015	0.045	0.38	0.72
AFO4	12	688	2.9e-02	3.2e-08	0.07	0.32	0.31	0.29	0	0	1114	1024	0.047	0.40	1.20
ALP1	47	1279	3.3e-01	1.8e-06	0.00	0.05	0.50	0.45	0	0	2212	2662	0.230	0.75	1.40
ALP2	127	1433	4.0e-01	7.7e-07	0.01	0.02	0.55	0.28	0	0.15	2315	2526	0.240	0.83	1.10
ALP3	76	1539	5.1e-01	7.6e-07	0.01	0.01	0.59	0.18	0	0.21	2438	2181	0.210	0.86	0.89
ALP4	114	1270	4.2e-01	5.1e-07	0.01	0.01	0.25	0.61	0.02	0.09	1826	1684	0.120	0.64	1.00

**Table A2.** Soil properties and climatological characteristics of the test catchments. Next to the site identifier (ID) the columns contain average clay, silt, sand, and skeleton contents, root zone depth ( $h_{RZ}$ ), effective field (eFC) and air capacity (AC) therein, and the corresponding average saturated hydraulic conductivity (Ks). All soil properties are provided by the national soil map of Germany (Hartwich et al., 1995) except for Ks, which was estimated based on pedo-transfer functions (Schaap et al., 2001). The climatological variables include 4-year mean winter and summer temperature sums ( $\sum \overline{T_W}$  and  $\sum \overline{T_S}$ ), the length of the period of vegetation ( $\overline{t_{veg}}$ ), and the number of frost days ( $\overline{n_{FD}}$ ).

ID	soil properties								climatological characteristics			
	clay (%)	silt (%)	sand (%)	skeleton (%)	$h_{RZ}$ (m)	eFC (mm)	AC (mm)	Ks ( $m s^{-1}$ )	$\sum \overline{T_W}$ ( $^{\circ}C$ )	$\sum \overline{T_S}$ ( $^{\circ}C$ )	$\overline{t_{veg}}$ (d)	$\overline{n_{FD}}$ (-)
TRI1	42.7	9.3	48.6	1.8	0.68	73.9	47.1	3.6e-05	15 941	65 040	174	45
TRI2	39.9	9.0	51.7	1.9	0.67	71.0	49.8	3.9e-05	10 601	69 717	191	45
TRI3	23.3	18.0	58.4	2.5	0.76	94.2	59.2	2.2e-05	9788	67 822	188	49
JUR1	39.2	16.8	43.5	3.7	0.26	32.5	37.0	2.8e-05	8209	61 808	178	61
BFO1	7.5	9.5	83.2	4.0	0.60	74.0	55.5	3.6e-05	3960	59 262	176	79
BFO2	12.4	17.0	71.1	3.6	0.55	74.5	40.8	2.1e-05	3345	58 475	176	82
BFO3	14.3	18.7	67.6	3.8	0.58	67.8	36.4	1.5e-05	3685	58 760	176	81
MOL1	26.4	58.8	14.4	1.1	0.85	144.8	44.6	1.5e-06	10 227	67 692	192	48
MOL2	22.3	56.5	21.8	1.4	0.79	146.1	44.3	1.7e-06	9567	67 994	192	50
MOL3	22.0	48.5	28.9	1.9	0.87	151.3	55.4	1.5e-06	8485	67 990	190	56
MOL4	22.7	57.2	21.0	1.2	0.89	132.5	46.2	1.6e-06	8805	69 246	192	53
MOL5	19.0	46.5	34.8	1.9	0.88	138.3	54.3	2.2e-06	8783	68 689	190	53
MOL6	21.0	54.0	24.7	1.3	0.88	129.5	47.7	1.9e-06	7423	68 895	190	58
MOL7	23.0	66.8	11.2	1.0	0.90	167.1	47.5	1.4e-06	7571	70 519	192	55
AFO1	21.0	39.0	40.0	2.0	1.0	153.0	76.0	1.1e-06	8088	56 910	176	65
AFO2	19.8	39.5	40.8	2.2	0.74	149.6	57.0	2.3e-06	11 966	57 158	187	50
AFO3	19.3	33.7	47.1	2.5	0.79	138.4	61.8	2.4e-06	14 974	54 775	180	50
AFO4	20.8	38.0	41.2	1.8	0.80	142.2	59.5	1.1e-06	10 437	61 874	182	58
ALP1	15.8	23.7	60.5	2.5	0.56	90.2	38.5	5.5e-06	243	39 834	141	109
ALP2	28.2	28.8	44.5	4.2	0.28	42.2	18.8	1.9e-06	-4490	31 694	123	129
ALP3	29.4	28.7	42.4	4.4	0.24	35.2	16.0	1.5e-06	-7934	30 122	118	141
ALP4	37.5	24.4	39.3	4.3	0.29	41.1	18.7	1.2e-06	-220	37 509	135	114

## Appendix B: Derivation of the type II dDMC

In this proof we show that accumulated direct runoff ( $Q_d$ ) scales with the ratio of cumulate precipitation (cum.  $P$ ) over storage capacity of the soil ( $S_{\max}$ ). It is based on the integral of Eq. (4) from  $t_0$  to  $t_e$  which equals accumulated  $Q_d$  in a hydrological year – Eq. (B1):

$$\int_{t_0}^{t_e} Q_d dt = \int_{t_0}^{t_e} P(t) \left( \frac{S_m(t)}{S_{\max}} \right)^\beta dt. \quad (\text{B1})$$

As a product, Eq. (B1) needs to be evaluated using the formula for partial integration (which can be found in any textbook).

$$\int_{t_0}^{t_e} u'(t) v dt = [u(t)v(t)]_{t_0}^{t_e} - \int_{t_0}^{t_e} u(t)v'(t) dt \quad (\text{B2})$$

$u'(t)$  denotes the temporal derivative of  $u(t)$ . Here we set  $P(t) = u'(t)$ , implying that  $u(t)$  corresponds to the integral of  $P(t)$  over time + a constant  $u(t) = \int P(t)dt + c$ . According to the chain rule,  $v(t) = \left( \frac{S_m(t)}{S_{\max}} \right)^\beta$  and  $v'(t) = \beta \left( \frac{S_m(t)}{S_{\max}} \right)^{\beta-1} \frac{1}{S_{\max}} \frac{dS}{dt}$ . Since  $P(t)$  is no analytical function of time, there is no analytical anti-derivative, and we define the integral of  $P(t)$  in the Riemann sense as the area under the curve. Inserting these terms yields

$$\int_{t_0}^{t_e} Q_d dt = \left[ \int P(t)dt \left( \frac{S_m(t)}{S_{\max}} \right)^\beta \right]_{t_0}^{t_e} - \int_{t_0}^{t_e} \left( \int_{t_0}^{t_e} P(t)dt \cdot \beta \left( \frac{S_m(t)}{S_{\max}} \right)^{\beta-1} \frac{1}{S_{\max}} \frac{dS}{dt} \right) dt. \quad (\text{B3})$$

Setting  $\int_{t_0}^{t_e} P(t)dt = \text{cum. } P$  and given the fact that  $\text{cum. } P(t_0) = 0$ , we obtain

$$\int_{t_0}^{t_e} Q_d dt = \text{cum. } P(t_e) \left( \frac{S_m(t_e)}{S_{\max}} \right)^\beta - \int_{t_0}^{t_e} \left( \text{cum. } P(t) \cdot \beta \left( \frac{S_m(t)}{S_{\max}} \right)^{\beta-1} \frac{1}{S_{\max}} \frac{dS}{dt} \right) dt. \quad (\text{B4})$$

Factoring out  $1/S_{\max}$  in the first summand and rearranging the terms in the second summand, we finally yield

$$\int_{t_0}^{t_e} Q_d dt = \frac{\text{cum. } P(t_e)}{S_{\max}} \left( \frac{S_m(t_e)^\beta}{(S_{\max})^{\beta-1}} \right) - \int_{t_0}^{t_e} \left( \frac{\text{cum. } P(t)}{S_{\max}} \beta \left( \frac{S_m(t)}{S_{\max}} \right)^{\beta-1} \frac{dS}{dt} \right) dt. \quad (\text{B5})$$

Equation (B5) illustrates that accumulated rainfall-driven water release is a monotonically increasing function of

cum.  $P(t_e)/S_{\max}$ . It hence justifies the scaling of the abscissa of the type II dDMC with  $S_{\max}$ . Specifically, Eq. (B5) shows that accumulated rainfall-driven water release at the end of the hydrological year is equal to the product of cum.  $P(t)$  with two summands. These are (a) the relative saturation at the end of the hydrological year to the power of  $\beta$  and (b) the integral of the relative storage change scaled with the relative saturation of the storage to the power of  $\beta - 1$ . Note also that the second summand is, depending on the form parameter  $\beta$  a strongly non-linear function and depends implicitly also on  $E(t)$  as the latter affects  $dS/dt$ . This is obvious if we insert the HBV equation (Eq. 4) into the water balance equation  $\frac{dS}{dt} = P(t) - Q_d(t) - E(t)$ :

$$\frac{dS}{dt} = \left[ P(t) \left( 1 - \left( \frac{S_m(t)}{S_{\max}} \right)^\beta \right) - E(t) \right]. \quad (\text{B6})$$

Inserting Eq. (B6) into the second summand of Eq. (B5), we obtain, for a linear store, i.e.  $\beta = 1$ , two linear terms:

$$\int_{t_0}^{t_e} Q_d dt = \frac{\text{cum. } P(t_e)}{S_{\max}} S_m(t_e) - \int_{t_0}^{t_e} \frac{\text{cum. } P(t)}{S_{\max}} \left[ P(t) \left( 1 - \frac{S_m(t)}{S_{\max}} \right) - E(t) \right] dt. \quad (\text{B7})$$

The first term depends on storage and scales with cum.  $P/S_{\max}$ . The second equals the integral of cum.  $P/S_{\max}$  times infiltration minus evapotranspiration.

## Appendix C: Link table

**Table C1.** Link table that relates the site identifiers (ID) introduced in Sect. 2.1 to the corresponding gauge and stream names. Gauge locations are provided in Gauß–Krüger zone 4 coordinates (GKR and GKH, CRS identifier EPSG:31468).

ID	Gauge	Stream	GKR	GKH
TRI1	Reichenbach (REIB)	Wörnitz	4 373 327	5 449 863
TRI2	Binzwangen (BINZ)	Altmühl	4 381 996	5 473 002
TRI3	Bechhofen (BECH)	Wieseth	4 394 270	5 447 640
JUR1	Holnstein (HOLN)	Unterbürger Laber	4 464 800	5 442 860
BFO1	Gartenried (GART)	Murach	4 532 661	5 483 477
BFO2	Untereppenried (UEPR)	Ascha	4 533 425	5 477 338
BFO3	Tiefenbach (TIEF)	Bayerische Schwarzach	4 543 360	5 477 800
MOL1	Roth (ROTR)	Roth	4 363 140	5 360 723
MOL2	Fleinhausen (FLEI)	Zusam	4 394 141	5 358 887
MOL3	Mering (MERI)	Paar	4 424 840	5 348 870
MOL4	Odelzhausen (ODZH)	Glonn	4 440 860	5 353 360
MOL5	Appolding (APPO)	Strogen	4 498 575	5 364 071
MOL6	Dietelskirchen (DIKI)	Kleine Vils	4 525 540	5 373 175
MOL7	Wallersdorf (WALR)	Reißingerbach	4 554 850	5 400 160
AFO1	Unterthingau (alt) (UTHI)	Kirnach	4 388 313	5 294 058
AFO2	Hörmanshofen (HOER)	Geltnach	4 399 272	5 299 593
AFO3	Buchloe (BUCH)	Gennach	4 404 574	5 323 974
AFO4	Herrssigmang (HERR)	Kienbach	4 438 860	5 318 140
ALP1	Gunzesried (GZRI)	Gunzesrieder Ach	4 366 798	5 266 382
ALP2	Reckenberg (RECK)	Ostrach	4 373 822	5 264 305
ALP3	Oberstdorf (OBTR)	Trettach	4 370 128	5 255 320
ALP4	Oberammergau (OAMM)	Ammer	4 429 723	5 273 332

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* We gratefully acknowledge the support of the editor and the three anonymous reviewers whose comments helped us to improve the quality of the manuscript. Furthermore, we acknowledge data provision by the Bavarian Environmental Agency (LfU) and the German Weather Service (DWD). This study is part of the DFG funded CAOS project “From Catchments as Organised Systems to Models based on Dynamic Functional Units” (FOR 1598). Additional funding was provided by the LfU, which is gratefully acknowledged as well. Lastly we acknowledge support by the Deutsche Forschungsgemeinschaft and the Open Access Publishing Fund of the Karlsruhe Institute of Technology.

The article processing charges for this open-access publication were covered by a Research Centre of the Helmholtz Association.

Edited by: F. Fenicia

Reviewed by: three anonymous referees

## References

- Ali, G., Tetzlaff, D., Soulsby, C., McDonnell, J. J., and Capell, R.: A comparison of similarity indices for catchment classification using a cross-regional dataset, *Adv. Water Resour.*, 40, 11–22, <https://doi.org/10.1016/j.advwatres.2012.01.008>, 2012.
- Barthold, F. K. and Woods, R. A.: Stormflow generation: A meta-analysis of field evidence from small, forested catchments, *Water Resour. Res.*, 51, 3730–3753, <https://doi.org/10.1002/2014WR016221>, 2015.
- Bayerisches Landesamt für Umwelt: Hochwassernachrichtendienst Bayern, available at: <http://www.hnd.bayern.de/>, 2017.
- Bergstroem, S.: Development and application of a conceptual runoff model for Scandinavian catchments, Tech. rep., Swedish Meteorological and Hydrological Institute (SMHI), Report RHO 7, Norrköping, 1976.
- Beven, K. and Germann, P.: Macropores and water flow in soils revisited, *Water Resour. Res.*, 49, 3071–3092, <https://doi.org/10.1002/wrcr.20156>, 2013.
- Beven, K. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, *Hydrological Sciences Bulletin*, 24, 43–69, 1979.
- BGR and SGD: Hydrogeological spatial structure of Germany (HYRAUM). Digital map data v3.2. German Federal States Geological Surveys (SGD) and Federal Institute for Geosciences and Natural Resources (BGR), 2015.
- Black, P. E.: Watershed Functions, *J. Am. Water Resour. As.*, 33, 1–11, <https://doi.org/10.1111/j.1752-1688.1997.tb04077.x>, 1997.
- Blöschl, G. and Sivapalan, M.: Scale issues in hydrological modelling: A review, *Hydrol. Process.*, 9, 251–290, <https://doi.org/10.1002/hyp.3360090305>, 1995.
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H. H. G.: *Runoff prediction in ungauged basins: synthesis across processes, places and scales*, Cambridge University Press, Cambridge, 2013.
- Boorman, D. B., Hollis, J. M., and Lilly, A.: *Hydrology of soil types: a hydrologically-based classification of the soils of United Kingdom.*, Tech. Rep. 126, Institut of Hydrology, Wallingford, 1995.
- Capell, R., Tetzlaff, D., Hartley, A. J., and Soulsby, C.: Linking metrics of hydrological function and transit times to landscape controls in a heterogeneous mesoscale catchment, *Hydrol. Process.*, 26, 405–420, <https://doi.org/10.1002/hyp.8139>, 2012.
- Casper, M. C., Grigoryan, G., Gronz, O., Gutjahr, O., Heinemann, G., Ley, R., and Rock, A.: Analysis of projected hydrological behavior of catchments based on signature indices, *Hydrol. Earth Syst. Sci.*, 16, 409–421, <https://doi.org/10.5194/hess-16-409-2012>, 2012.
- Cerdan, O., Souchère, V., Lecomte, V., Couturier, A., and Le Bissonnais, Y.: Incorporating soil surface crusting processes in an expert-based runoff model: Sealing and Transfer by Runoff and Erosion related to Agricultural Management, *Catena*, 46, 189–205, [https://doi.org/10.1016/S0341-8162\(01\)00166-7](https://doi.org/10.1016/S0341-8162(01)00166-7), 2002.
- Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y. M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O. N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E. F.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, *J. Hydrol.*, 320, 3–17, <https://doi.org/10.1016/j.jhydrol.2005.07.031>, 2006.
- Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, *Hydrol. Earth Syst. Sci.*, 17, 1893–1912, <https://doi.org/10.5194/hess-17-1893-2013>, 2013.
- Euser, T., Hrachowitz, M., Winsemius, H. C., and Savenije, H. H.: The effect of forcing and landscape distribution on performance and consistency of model structures, *Hydrol. Process.*, 29, 3727–3743, <https://doi.org/10.1002/hyp.10445>, 2015.
- Fogler, S. H.: *Elements of Chemical Reaction Engineering*, Prentice Hall, New Jersey, 1999.
- Gassmann, M., Stamm, C., Olsson, O., Lange, J., Kümmerer, K., and Weiler, M.: Model-based estimation of pesticides and transformation products and their export pathways in a headwater catchment, *Hydrol. Earth Syst. Sci.*, 17, 5213–5228, <https://doi.org/10.5194/hess-17-5213-2013>, 2013.
- Graeff, T., Zehe, E., Blume, T., Francke, T., and Schröder, B.: Predicting event response in a nested catchment with generalized linear models and a distributed watershed model, *Hydrol. Process.*, 26, 3749–3769, <https://doi.org/10.1002/hyp.8463>, 2012.
- Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, *Hydrol. Process.*, 22, 3802–3813, <https://doi.org/10.1002/hyp.6989>, 2008.
- Hartwich, R., Behrens, J., Eckelmann, W., Haase, G., Richter, A., Roeschmann, G., and Schmidt, R.: *Bodenübersichtskarte der Bundesrepublik Deutschland 1 : 1 000 000, Karte mit Erläuterungen, Textlegende und Leitprofilen*, 1995.
- Hellebrand, H., Van Den Bos, R., Hoffmann, L., Juilleret, J., and Pfister, L.: The potential of winter stormflow coefficients for hydrological regionalization purposes in poorly gauged basins of the middle Rhine region, *Hydrolog. Sci. J.*, 53, 773–788, <https://doi.org/10.1623/hysj.53.4.773>, 2008.

- Hillel, D.: Introduction to Environmental Soil Physics, Academic Press, Elsevier, San Diego, California, 2004.
- Holdridge, L. R.: Life zone ecology, Tropical Science Center, San Jose, Costa Rica, 1967.
- Horton, R. E.: The role of infiltration in the hydrologic cycle, Transactions of the American Geophysical Union, 14, 446–460, 1933.
- Hrachowitz, M., Bohte, R., Mul, M. L., Bogaard, T. A., Savenije, H. H. G., and Uhlenbrook, S.: On the value of combined event runoff and tracer analysis to improve understanding of catchment functioning in a data-scarce semi-arid area, Hydrol. Earth Syst. Sci., 15, 2007–2024, <https://doi.org/10.5194/hess-15-2007-2011>, 2011.
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fencica, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB) a review, Hydrolog. Sci. J., 58, 1198–1255, <https://doi.org/10.1080/02626667.2013.803183>, 2013.
- Jackisch, C.: Linking structure and functioning of hydrological systems. How to achieve necessary experimental and model complexity with adequate effort., PhD thesis, Institut für Wasser und Gewässerentwicklung, Bereich Hydrologie, Karlsruher Institut für Technologie (KIT), <https://doi.org/10.5445/IR/1000051494>, 2015.
- Kenney, B. C.: Beware of spurious self-correlations!, Water Resour. Res., 18, 1041–1048, <https://doi.org/10.1029/WR018i004p01041>, 1982.
- Kirby, M.: Hydrograph modelling strategies, in: Progress in Physical and Human Geograph, edited by: Peel, R., 69–90, Heinemann, London, 1975.
- Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42, 5 pp., <https://doi.org/10.1029/2005WR004362>, 2006.
- Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, Water Resour. Res., 45, W02429, <https://doi.org/10.1029/2008WR006912>, 2009.
- Klaus, J., Wetzel, C. E., Martínez-Carreras, N., Ector, L., and Pfister, L.: A tracer to bridge the scales: On the value of diatoms for tracing fast flow path connectivity from headwaters to meso-scale catchments, Hydrol. Process., 29, 5275–5289, <https://doi.org/10.1002/hyp.10628>, 2015.
- Larsen, J. E., Sivapalan, M., Coles, N. A., and Linnet, P. E.: Similarity analysis of runoff generation processes in real-world catchments, Water Resour. Res., 30, 1641–1652, <https://doi.org/10.1029/94WR00555>, 1994.
- Leavesley, G. H.: A mountain watershed simulation model, PhD thesis, Colorado State University, Fort Collins, 1973.
- Ley, R., Casper, M. C., Hellebrand, H., and Merz, R.: Catchment classification by runoff behaviour with self-organizing maps (SOM), Hydrol. Earth Syst. Sci., 15, 2947–2962, <https://doi.org/10.5194/hess-15-2947-2011>, 2011.
- Li, H., Sivapalan, M., and Tian, F.: Comparative diagnostic analysis of runoff generation processes in Oklahoma DMIP2 basins: The Blue River and the Illinois River, J. Hydrol., 418–419, 90–109, <https://doi.org/10.1016/j.jhydrol.2010.08.005>, 2012.
- Lindsay, J. B.: The Whitebox Geospatial Analysis Tools project and open-access GIS, in: Proceedings of the GIS Research UK 22nd Annual Conference, p. 8, <https://doi.org/10.13140/RG.2.1.1010.8962>, 2014.
- Loritz, R., Hassler, S. K., Jackisch, C., Allroggen, N., van Schaik, L., Wienhöfer, J., and Zehe, E.: Picturing and modeling catchments by representative hillslopes, Hydrol. Earth Syst. Sci., 21, 1225–1249, <https://doi.org/10.5194/hess-21-1225-2017>, 2017.
- Ludwig, K. and Bremicker, M.: The Water Balance Model LAR-SIM, Tech. Rep. 22, Institut für Hydrologie der Universität Freiburg i.Br., 2006.
- Mäkelä, A., Hari, P., Berninger, F., Hänninen, H., and Nikinmaa, E.: Acclimation of photosynthetic capacity in Scots pine to the annual cycle of temperature., Tree physiology, 24, 369–76, <https://doi.org/10.1093/TREEPHYS/24.4.369>, 2004.
- Martínez-Carreras, N., Krein, A., Gallart, F., Iffly, J. F., Pfister, L., Hoffmann, L., and Owens, P. N.: Assessment of different colour parameters for discriminating potential suspended sediment sources and provenance: A multi-scale study in Luxembourg, Geomorphology, 118, 118–129, <https://doi.org/10.1016/j.geomorph.2009.12.013>, 2010.
- Martínez-Carreras, N., Wetzel, C. E., Frenness, J., Ector, L., McDonnell, J. J., Hoffmann, L., and Pfister, L.: Hydrological connectivity inferred from diatom transport through the riparian-stream system, Hydrol. Earth Syst. Sci., 19, 3133–3151, <https://doi.org/10.5194/hess-19-3133-2015>, 2015.
- Maurer, T.: Physikalisch begründete zeitkontinuierliche Modellierung des Wassertransports in kleinen ländlichen Einzugsgebieten, Phd thesis, Mitteilungen des Instituts für Hydrologie und Wasserwirtschaft, Bd. 61, Universität Fridericiana zu Karlsruhe (TH), 1997.
- McGuire, K. J., McDonnell, J. J., Weiler, M., Kendall, C., McGlynn, B. L., Welker, J. M., and Seibert, J.: The role of topography on catchment-scale water residence time, Water Resour. Res., 41, W05002, <https://doi.org/10.1029/2004WR003657>, 2005.
- McMillan, H., Gueguen, M., Grimon, E., Woods, R., Clark, M., and Rupp, D. E.: Spatial variability of hydrological processes and model structure diagnostics in a 50 km<sup>2</sup> catchment, Hydrol. Process., 28, 4896–4913, <https://doi.org/10.1002/hyp.9988>, 2014.
- McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M., and Woods, R. A.: Hydrological field data from a modeller’s perspective: Part 1. Diagnostic tests for model structure, Hydrol. Process., 25, 511–522, <https://doi.org/10.1002/hyp.7841>, 2011.
- Menzel, A., Jakobi, G., Ahas, R., Scheifinger, H., and Estrella, N.: Variations of the climatological growing season (1951–2000) in Germany compared with other countries, Int. J. Climatol., 23, 793–812, <https://doi.org/10.1002/joc.915>, 2003.
- Merz, R., Blöschl, G., and Parajka, J.: Spatio-temporal variability of event runoff coefficients, J. Hydrol., 331, 591–604, <https://doi.org/10.1016/j.jhydrol.2006.06.008>, 2006.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Stationarity Is Dead: Whither Water Management?, Science, 319, 573–574, <https://doi.org/10.1126/science.1151915>, 2008.
- Mualem, Y., Assouline, S., and Rohdenburg, H.: Rainfall induced soil seal (A) – A critical review of observations and

- models, *CATENA*, 17, 185–203, [https://doi.org/10.1016/0341-8162\(90\)90008-2](https://doi.org/10.1016/0341-8162(90)90008-2), 1990.
- NOAA: National Weather Service River Forecast System Forecast Procedures, Tech. rep., NWS-Hydro-14, NOAA Technical Memorandum, US Department of Commerce, Washington, DC, 1972.
- Oldham, C. E., Farrow, D. E., and Peiffer, S.: A generalized Damköhler number for classifying material processing in hydrological systems, *Hydrol. Earth Syst. Sci.*, 17, 1133–1148, <https://doi.org/10.5194/hess-17-1133-2013>, 2013.
- Oudin, L., Andreassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, 44, W03413, <https://doi.org/10.1029/2007WR006240>, 2008.
- Oudin, L., Kay, A., Andreassian, V., and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, *Water Resour. Res.*, 46, W11558, <https://doi.org/10.1029/2009wr008887>, 2010.
- Pearson, K.: *Mathematical Contributions to the Theory of Evolution. – On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs*, *P. R. Soc. London*, 60, 489–498, <https://doi.org/10.1098/rsp1.1896.0076>, 1987.
- Pelletier, J. D. and Rasmussen, C.: Geomorphically based predictive mapping of soil thickness in upland watersheds, *Water Resour. Res.*, 45, W09417, <https://doi.org/10.1029/2008WR007319>, 2009.
- Perry, K., Sanders, D., Granberry, D., Thomasgarrett, J., Decoteau, D., Nagata, R., Dufault, R., Deanbatal, K., and McLaurin, W.: Heat units, solar radiation and daylength as pepper harvest predictors, *Agr. Forest Meteorol.*, 65, 197–205, [https://doi.org/10.1016/0168-1923\(93\)90004-2](https://doi.org/10.1016/0168-1923(93)90004-2), 1993.
- Peschke, G., Etzenberg, C., Müller, G., Töpfer, J., and Zimmermann, S.: Das wissenschaftsbasierte System FLAB – ein Instrument zur rechnergestützten Bestimmung von Landschaftseinheiten mit gleicher Abflussbildung, Tech. rep., Internationales Hochschulinstitut Zittau, IHI-Schriften, BD. 10, Zittau, 1999.
- Pfannerstill, M., Guse, B., and Fohrer, N.: Smart low flow signature metrics for an improved overall performance evaluation of hydrological models, *J. Hydrol.*, 510, 447–458, <https://doi.org/10.1016/j.jhydrol.2013.12.044>, 2014.
- Pfister, L., Iffly, J. F., and Hoffmann, L.: Use of regionalized stormflow coefficients with a view to hydroclimatological hazard mapping, *Hydrolog. Sci. J.*, 47, 479–491, <https://doi.org/10.1080/02626660209492948>, 2002.
- Pokhrel, P. and Yilmaz, K. K.: Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures, *J. Hydrol.*, 418, 49–60, <https://doi.org/10.1016/j.jhydrol.2008.12.004>, 2012.
- R Core Team: *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, 2015.
- Rawson, H. M. and Gomez-Macpherson, H.: *Irrigated wheat – managing your crop*, Food and Agriculture Organization of the United Nations (FAO), Rome, 2000.
- Reaney, S. M., Bracken, L. J., and Kirkby, M. J.: The importance of surface controls on overland flow connectivity in semi-arid environments: results from a numerical experimental approach, *Hydrol. Process.*, 28, 2116–2128, <https://doi.org/10.1002/hyp.9769>, 2014.
- Reggiani, P., Sivapalan, M., and Majid Hassanizadeh, S.: A unifying framework for watershed thermodynamics: balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics, *Adv. Water Resour.*, 22, 367–398, [https://doi.org/10.1016/S0309-1708\(98\)00012-8](https://doi.org/10.1016/S0309-1708(98)00012-8), 1998.
- Reggiani, P., Sivapalan, M., and Hassanizadeh, S. M.: Conservation equations governing hillslope responses: Exploring the physical basis of water balance, *Water Resour. Res.*, 36, 1845, <https://doi.org/10.1029/2000WR900066>, 2000.
- Reichenau, T. G., Korres, W., Montzka, C., Fiener, P., Wilken, F., Stadler, A., Waldhoff, G., and Schneider, K.: Spatial heterogeneity of Leaf Area Index (LAI) and its temporal course on arable land: Combining field measurements, remote sensing and simulation in a Comprehensive Data Analysis Approach (CDAA), *PLoS ONE*, 11, 1–24, <https://doi.org/10.1371/journal.pone.0158451>, 2016.
- Reynolds, O.: An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels, *Philos. T. R. Soc. Lond.*, 174, 935–982, <https://doi.org/10.1098/rstl.1883.0029>, 1883.
- Robinson, J. S., Sivapalan, M., and Snell, J. D.: On the relative roles of hillslope processes, channel routing, and network geomorphology in the hydrologic response of natural catchments, *Water Resour. Res.*, 31, 3089–3101, <https://doi.org/10.1029/95WR01948>, 1995.
- Santhi, C., Allen, P. M., Mutiah, R. S., Arnold, J. G., and Tuppad, P.: Regional estimation of base flow for the conterminous United States by hydrologic landscape regions, *J. Hydrol.*, 351, 139–153, <https://doi.org/10.1016/j.jhydrol.2007.12.018>, 2008.
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, *Hydrol. Earth Syst. Sci.*, 15, 2895–2911, <https://doi.org/10.5194/hess-15-2895-2011>, 2011.
- Sayama, T., McDonnell, J. J., Dhakal, A., and Sullivan, K.: How much water can a watershed store?, *Hydrol. Process.*, 25, 3899–3908, <https://doi.org/10.1002/hyp.8288>, 2011.
- Schaake, J., Duan, Q., Smith, M., and Koren, V.: *Criteria to select basins for hydrologic model development and testing*, 15th Conference on Hydrology (P1.8), AMS, January 9–14, Long Beach, CA, 2000.
- Schaap, M. G., Leij, F. J., and van Genuchten, M. T.: Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions, *J. Hydrol.*, 251, 163–176, [https://doi.org/10.1016/S0022-1694\(01\)00466-8](https://doi.org/10.1016/S0022-1694(01)00466-8), 2001.
- Schaefli, B., Harman, C. J., Sivapalan, M., and Schymanski, S. J.: HESS Opinions: Hydrologic predictions in a changing environment: behavioral modeling, *Hydrol. Earth Syst. Sci.*, 15, 635–646, <https://doi.org/10.5194/hess-15-635-2011>, 2011.
- Scherrer, S. and Naef, F.: A decision scheme to indicate dominant hydrological flow processes on temperate grassland, *Hydrol. Process.*, 17, 391–401, <https://doi.org/10.1002/hyp.1131>, 2003.
- Schmocker-Fackel, P., Naef, F., and Scherrer, S.: Identifying runoff processes on the plot and catchment scale, *Hydrol. Earth Syst. Sci.*, 11, 891–906, <https://doi.org/10.5194/hess-11-891-2007>, 2007.
- Schulla, J.: *Hydrologische Modellierung von Flussgebieten zur Abschätzung der Folgen von Klimaänderungen*, Phd the-

- sis, *Zürcher Geographische Schriften*, Bd. 69, ETH Zürich, <https://doi.org/10.3929/ethz-a-001763261>, 1997.
- Seibert, S. P., Skublics, D., and Ehret, U.: The potential of coordinated reservoir operation for flood mitigation in large basins – A case study on the Bavarian Danube using coupled hydrological-hydrodynamic models, *J. Hydrol.*, 517, 1128–1144, <https://doi.org/10.1016/j.jhydrol.2014.06.048>, 2014.
- Seibert, S. P., Ehret, U., and Zehe, E.: Disentangling timing and amplitude errors in streamflow simulations, *Hydrol. Earth Syst. Sci.*, 20, 3745–3763, <https://doi.org/10.5194/hess-20-3745-2016>, 2016a.
- Seibert, S. P., Jackisch, C., Pfister, L., Ehret, U., and Zehe, E.: Exploring the interplay between state, structure and runoff behaviour of lower mesoscale catchments, *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2016-109>, in review, 2016b.
- Sivapalan, M.: Prediction in ungauged basins: a grand challenge for theoretical hydrology, *Hydrol. Process.*, 17, 3163–3170, <https://doi.org/10.1002/hyp.5155>, 2003.
- Sivapalan, M., Beven, K., and Wood, E. F.: On hydrologic similarity: 2. A scaled model of storm runoff production, *Water Resour. Res.*, 23, 2266–2278, 1987.
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V., Karambiri, H., Lakshmi, W., Liang, X., McDonnell, J., Mendionde, E. M., O’Connell, P., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, 48, 857–880, 2003.
- Solantie, R.: Daytime temperature sum - a new thermal variable describing growing season characteristics and explaining evapotranspiration, *Boreal Environ. Res.*, 9, 319–333, 2004.
- Soulsby, C., Tetzlaff, D., and Hrachowitz, M.: Are transit times useful process-based tools for flow prediction and classification in ungauged basins in montane regions?, *Hydrol. Process.*, 24, 1685–1696, <https://doi.org/10.1002/hyp.7578>, 2010.
- Struthers, I. and Sivapalan, M.: A conceptual investigation of process controls upon flood frequency: role of thresholds, *Hydrol. Earth Syst. Sci.*, 11, 1405–1416, <https://doi.org/10.5194/hess-11-1405-2007>, 2007.
- Varado, N., Braud, I., Galle, S., Le Lay, M., Séguis, L., Kamagate, B., and Depraetere, C.: Multi-criteria assessment of the Representative Elementary Watershed approach on the Donga catchment (Benin) using a downward approach of model complexity, *Hydrol. Earth Syst. Sci.*, 10, 427–442, <https://doi.org/10.5194/hess-10-427-2006>, 2006.
- Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria, *Hydrol. Earth Syst. Sci.*, 17, 2263–2279, <https://doi.org/10.5194/hess-17-2263-2013>, 2013.
- Vrugt, J. A. and Sadeh, M.: Toward diagnostic model calibration and evaluation: Approximate Bayesian computation, *Water Resour. Res.*, 49, 4335–4345, <https://doi.org/10.1002/wrcr.20354>, 2013.
- Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment Classification and Hydrologic Similarity, *Geography Compass*, 1, 901–931, <https://doi.org/10.1111/j.1749-8198.2007.00039.x>, 2007.
- Wang, D. and Wu, L.: Similarity of climate control on base flow and perennial stream density in the Budyko framework, *Hydrol. Earth Syst. Sci.*, 17, 315–324, <https://doi.org/10.5194/hess-17-315-2013>, 2013.
- Wienhöfer, J., Germer, K., Lindenmaier, F., Färber, A., and Zehe, E.: Applied tracers for the observation of subsurface stormflow at the hillslope scale, *Hydrol. Earth Syst. Sci.*, 13, 1145–1161, <https://doi.org/10.5194/hess-13-1145-2009>, 2009.
- Wood, E. F., Sivapalan, M., and Beven, K.: Scale and similarity in catchment storm response, *Rev. Geophys.*, 28, 1–18, 1990.
- Woods, R.: The relative roles of climate, soil, vegetation and topography in determining seasonal and long-term catchment dynamics, *Adv. Water Resour.*, 26, 295–309, [https://doi.org/10.1016/S0309-1708\(02\)00164-1](https://doi.org/10.1016/S0309-1708(02)00164-1), 2003.
- Wrede, S., Fenicia, F., Martinez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., Savenije, H. H. G., Uhlenbrook, S., Kavetski, D., and Pfister, L.: Towards more systematic perceptual model development: A case study using 3 Luxembourgish catchments, *Hydrol. Process.*, 29, 2731–2750, <https://doi.org/10.1002/hyp.10393>, 2015.
- Ye, S., Yaeger, M., Coopersmith, E., Cheng, L., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves – Part 2: Role of seasonality, the regime curve, and associated process controls, *Hydrol. Earth Syst. Sci.*, 16, 4447–4465, <https://doi.org/10.5194/hess-16-4447-2012>, 2012.
- Zehe, E. and Jackisch, C.: A Lagrangian model for soil water dynamics during rainfall-driven conditions, *Hydrol. Earth Syst. Sci.*, 20, 3511–3526, <https://doi.org/10.5194/hess-20-3511-2016>, 2016.
- Zehe, E., Maurer, T., Ihringer, J., and Plate, E.: Modeling water flow and mass transport in a loess catchment, *Phys. Chem. Earth Pt. B*, 26, 487–507, [https://doi.org/10.1016/S1464-1909\(01\)00041-7](https://doi.org/10.1016/S1464-1909(01)00041-7), 2001.
- Zehe, E., Ehret, U., Pfister, L., Blume, T., Schröder, B., Westhoff, M., Jackisch, C., Schymanski, S. J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, J., van Schaik, L., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V., and Kleidon, A.: HESS Opinions: From response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments, *Hydrol. Earth Syst. Sci.*, 18, 4635–4655, <https://doi.org/10.5194/hess-18-4635-2014>, 2014.
- Zhang, G. P., Savenije, H. H. G., Fenicia, F., and Pfister, L.: Modelling subsurface storm flow with the Representative Elementary Watershed (REW) approach: application to the Alzette River Basin, *Hydrol. Earth Syst. Sci.*, 10, 937–955, <https://doi.org/10.5194/hess-10-937-2006>, 2006.