Hydrology and
Earth System
Sciences

Open Access

Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration

S. Gharari[1,2]**, M. Hrachowitz**[1]**, F. Fenicia**[3,1]**, H. Gao**[1]**, and H. H. G. Savenije**[1]

[1]Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands
[2]Public Research Center – Gabriel Lippmann, Belvaux, Luxembourg
[3]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

Correspondence to: S. Gharari (s.gharari@tudelft.nl)

Abstract. Conceptual environmental system models, such as rainfall runoff models, generally rely on calibration for parameter identification. Increasing complexity of this type of models for better representation of hydrological process heterogeneity, typically makes parameter identification more difficult. Although various, potentially valuable, approaches for better parameter estimation have been developed, strategies to impose general conceptual understanding of how a catchment works into the process of parameter estimation has not been fully explored. In this study we assess the effects of imposing semi-quantitative, relational inequality constraints, based on expert-knowledge, for model development and parameter specification, efficiently exploiting the complexity of a semi-distributed model formulation. Making use of a topography driven rainfall-runoff modeling (FLEX-TOPO) approach, a catchment was delineated into three functional units, i.e., wetland, hillslope and plateau. Ranging from simple to complex, three model setups, FLEX^A, FLEX^B and FLEX^C were developed based on these functional units, where FLEX^A is a lumped representation of the study catchment, and the semi-distributed formulations FLEX^B and FLEX^C progressively introduce more complexity. In spite of increased complexity, FLEX^B and FLEX^C allow modelers to compare parameters, as well as states and fluxes, of their different functional units to each other, allowing the formulation of constraints that limit the feasible parameter space. We show that by allowing for more landscape-related process heterogeneity in a model, e.g., FLEX^C, the performance increases even without traditional calibration. The additional introduction of relational constraints further improved the performance of these models.

1 Introduction

Lumped conceptual and distributed physically based models are the two endpoints of the modeling spectrum, ranging from simplicity to complexity, which here is defined as the number of model parameters. Each modeling strategy is characterized by advantages and limitations. In hydrology, physically based models are typically applied under the assumptions that (a) the spatial resolution and the complexity of the model are warranted by the available data, and (b) the catchment response is a mere aggregation of small scale processes. However, these two fundamental assumptions are commonly violated. As a result, the predictive power and hydrological insight achievable via these models is limited (e.g., Beven, 1989, 2001; Grayson et al., 1992; Blöschl, 2001; Pomeroy et al., 2007; Sivapalan, 2006; McDonnell et al., 2007; Hrachowitz et al., 2013b).

In contrast, lumped conceptual models require less data for model parameters estimation. This advantage comes at the expense of considerable limitations. Representing system integrated processes, model structures and parameters are not directly linked to observable quantities. Their estimation therefore strongly relies on calibration. To limit

parameter identifiability issues arising from calibration, these models are often oversimplified abstractions of the system, and if inadequately tested they may act as *mathematical marionettes* (Kirchner, 2006). They may outperform more complex distributed models (e.g., Refsgaard and Knudsen, 1996; Ajami et al., 2004; Reed et al., 2004), but often fail to provide realistic representations of the underlying processes, leading to limited predictive power (e.g., Freer et al., 2003; Seibert, 2003; Kirchner, 2006; Beven, 2006; Kling and Gupta, 2009; Andréassian et al., 2012; Euser et al., 2013; Gharari et al., 2013).

Various strategies have been suggested to allow for increased model complexity and to thereby improve the physical realism of conceptual models. These strategies included the attempt to incorporate different data sources in the parameter estimation process, such as ground- and soil water dynamics (e.g., Seibert and McDonnell, 2002; Freer et al., 2004; Fenicia et al., 2008a; Matgen et al., 2012; Sutanudjaja et al., 2013), remotely sensed evaporation (e.g., Winsemius et al., 2008), snow dynamics (e.g., Parajka and Blöschl, 2008) or tracer data (e.g., Vaché and McDonnell, 2006; Dunn et al., 2008; Son and Sivapalan, 2007; Birkel et al., 2011; Capell et al., 2012; Hrachowitz et al., 2013a). Alternatively, one may seek to extract more information from available data, for example through the development of signatures representing different aspects of the data (e.g., Gupta et al., 1998, 2008; Boyle et al., 2000, 2001; Madsen, 2000; Fenicia et al., 2006; Rouhani et al., 2007; Khu et al., 2008; Winsemius et al., 2009; Bulygina and Gupta, 2010; McMillan et al., 2011; Clark et al., 2011; Euser et al., 2013; He et al., 2014; Hrachowitz et al., 2014).

Traditional parameter estimation of conceptual models relies on the availability of calibration data, which, however, are frequently not available for the time period or the spatiotemporal resolution of interest. A wide range of regionalization techniques for model parameters and hydrological signatures have been developed to avoid calibration in such data scarce environments (e.g., Bárdossy, 2007; Yadav et al., 2007; Perrin et al., 2008; Zhang et al., 2008; Kling and Gupta, 2009; Samaniego et al., 2010; Kumar et al., 2010; Wagener and Montanari, 2011; Kapangaziwiri et al., 2012; Viglione et al., 2013). However, it is challenging to identify suitable functional relationships between catchment characteristics and model parameters (e.g., Merz and Blöschl, 2004; Kling and Gupta, 2009), and only recently did Kumar et al. (2010, 2013a) show that multi-scale parameter regionalization (MPR) can yield global parameters that perform consistently over different catchment scales. In a further study they successfully transferred parameters obtained by the MPR technique to ungauged catchments in Germany and the USA (Kumar et al., 2013b).

Related to these difficulties with parameter identifiability, the lack of sufficient representation of processes heterogeneity (i.e., complexity) in conceptual models limits the degree of realism of these models. The concept of

hydrological response units (HRUs) can be exploited as a strategy for an efficient tradeoff between model simplicity, required for adequate parameter identifiability, and a sufficient degree of process heterogeneity. HRUs are units within a catchment, characterized by a different hydrological function and can be represented by different model structures or parameters. In most cases HRUs are defined based on soil types, land cover and other physical catchment characteristics (e.g., Knudsen et al., 1986; Flügel, 1995; Grayson and Blöschl, 2000; Krcho, 2001; Winter, 2001; Scherrer and Naef, 2003; Uhlenbrook et al., 2004; Wolock et al., 2004; Pomeroy et al., 2007; Scherrer et al., 2007; Schmocker-Fackel et al., 2007; Efstratiadis and Koutsoyiannis, 2008; Lindström et al., 2010; Nalbantis et al., 2011; Kumar et al., 2010).

A wide range of studies also points towards the potential value of using topographical indices, readily available from digital elevation models (DEMs), to account for process heterogeneity (e.g., McGlynn and McDonnell, 2003; Seibert, 2003; McGuire et al., 2005; Hrachowitz et al., 2009; Jencso et al., 2009; Detty and McGuire, 2010; Gascuel-Odoux et al., 2010). Because standard metrics of landscape organization, such as absolute elevation, slope or curvature, as used in the catena concept (Milne, 1935; Park and van de Giesen, 2004), are often not strong enough descriptors to infer hydrological function, alternative concepts have been sought. The development of derived metrics such as the *Topographic Wetness Index* (Beven and Kirkby, 1979) facilitated an important step forward, being at the core of TOPMODEL (e.g., Beven and Kirkby, 1979; Beven and Freer, 2001a), which has proven to be a valuable approach in specific environmental settings meeting the assumptions of the model. A different descriptor allowing a potentially more generally applicable and hydrologically meaningful landscape classification has recently been suggested by Rennó et al. (2008): the *Height Above the Nearest Drainage* (HAND). Nobre et al. (2011) showed the hydrological relevance of HAND by investigating long term groundwater behavior and land use.

Explicitly invoking the co-evolution of topography, vegetation and hydrology, Savenije (2010) argued that catchments, as self-organizing systems, need to fulfill the contrasting hydrological functions of efficient drainage and sufficient water storage to allow, in a feedback process, topography and vegetation to develop the way they did. These distinct hydrological functions can be associated with different landscape elements or HRUs as defined by HAND and slope, such that each HRU is represented by a model structure best representing its function in the ecosystem (cf. Savenije, 2010).

While HAND-based landscape classification can potentially show a way forward, it does not solve the problem arising when moving from lumped to HRU-guided, semi-distributed model formulations: multiple parallel model structures typically result in an increased number of parameters, which, when not adequately constrained, may increase equifinality and thereby limit predictive uncertainty (e.g.,

Gupta and Sorooshian, 1983;Beven, 2006;Gupta et al., 2008). To better satisfy the contrasting priorities of model complexity and predictive power, new strategies are sought to more efficiently utilize the modelers' understanding of the system, particularly when for constraining the feasible model- and parameter space is scarce (e.g., Gupta et al., 2008; Wagener and Montanari, 2011; Singh and Bárdossy, 2012; Andréassian et al., 2012; Gharari et al., 2013; Hrachowitz et al., 2013a; Razavi and Tolson, 2013).

In contrast to earlier attempts to constrain models using multiple evaluation criteria or a priori information on catchment properties such as land use or soil type (e.g., Koren et al., 2008), this study tests the utility of a different and so far underexploited type of constraint, based on a priori understanding of the system. The concept of topography-driven conceptual modeling involves the identification of HRUs that operate in parallel. Linked to the technique of regularization (e.g., Tikhonov, 1963; Engl et al., 1996), this opens the possibility to impose semi-quantitative, expert knowledge based, relational constraints of catchment behavior on model parameters, similar to what was suggested by Pokhrel et al. (2008) and Yilmaz et al. (2008). In contrast to those studies, the suggested concept introduces relations between parallel HRUs exclusively based on *hydrological reasoning* to ensure that similar processes between parallel model structures are represented in an internally consistent way, thereby reducing the feasible parameter space. The advantage of this method is that there is only minimal need to quantify the constraints or the prior parameter distributions (e.g., Koren et al., 2000, 2003; Kuzmin et al., 2008; Duan et al., 2006), while allowing for a meaningful and potentially more realistic representation of the system in which each model component is, within certain limits, constrained to do what it is designed to do, rather than allowing it to compensate for data and model structural errors.

The objective of this paper is to test the hypothesis that application of model constraints based on expert knowledge (regarding relations between parameters, fluxes and states) to semi-distributed conceptual models defined by a hydrologically meaningful, topography-based, landscape classification system combined can (1) increase model internal consistency and thus the level of process realism as compared to lumped model setups, (2) increase the predictive power compared to lumped model setups and (3) reduce the need for model calibration.

2 Study area and data

The outlined methodology will be illustrated and tested with a case study using data of the Wark catchment in the Grand Duchy of Luxembourg. The catchment has an area of $82 \, km^2$ with the catchment outlet located downstream of the town of Ettelbrück at the confluence with the Alzette River ($49.85°$ N, $6.10°$ E, Fig. 1). With an annual mean

precipitation of $850 \, mm \, yr^{-1}$ and an annual mean potential evaporation of $650 \, mm \, yr^{-1}$ the annual mean runoff is approximately $250 \, mm \, yr^{-1}$. The geology in the northern part is dominated by schist while the southern part of the catchment is mostly underlain by sandstone and conglomerate. Hillslopes are generally characterized by forest, while plateaus and valley bottoms are mostly used as crop land and pastures, respectively. Drogue et al. (2002) quantified land use in the catchment as 4.3 % urban areas, 52.7 % agricultural land and 42.9 % forest. In addition they reported that 61 % of catchment is covered by permeable soils while the remainder is characterized by lower permeability substrate. The elevation varies between 195 to 532 m, with a mean value of 380 m. The slope of the catchment varies between 0 and 200 %, with a mean value of 17 % (Gharari et al., 2011).

The hydrological data used in this study include discharge measured at the outlet of the Wark catchment, potential evaporation estimated by the Hamon equation (Hamon, 1961) with temperature data measured at Luxembourg airport (Fenicia et al., 2006); and precipitation measured by a tipping bucket rain gauge located at Reichlange. The temporal resolution used in this study is 3 h.

3 FLEX-TOPO framework

Realizing the potential of "reading the landscape" in a *systems approach* (cf. Sivapalan et al., 2003), Savenije (2010) argued that due to the co-evolution of topography, soil and vegetation, all of which define the hydrological function of a given location, an efficient, hydrologically meaningful descriptor of topography together with land use could be used to distinguish different HRUs. HAND, which can be loosely interpreted as the hydraulic head at a given location in a catchment, may be such a descriptor as it potentially allows for meaningful landscape classification (e.g., Rennó et al., 2008). It was argued previously (Gharari et al., 2011) that, in central European landscapes, HAND can efficiently distinguish between wetlands, hillslopes and plateaus. These are landscape elements that may also be assumed to fulfill distinct hydrological functions (HRUs) in the study catchment (Savenije, 2010). Wetlands, located at low elevations above streams, are characterized by shallow ground water tables with limited fluctuations. Due to reduced storage capacity between ground water table and soil surface, potentially exacerbated by the relative importance of the capillary fringe, wetlands tend to be saturated, and thus connected, earlier during a rainfall event than the two other landscape elements with arguably higher storage capacity, thus frequently becoming the dominant source of storm flow during comparably dry periods (e.g., Seibert, 2003; McGlynn et al., 2004; Molénat et al., 2005; Blume et al., 2008; Anderson et al., 2010; Kavetski et al., 2011). The dominant runoff process in wetlands can therefore be assumed to be saturation overland flow. In contrast, forested hillslopes,
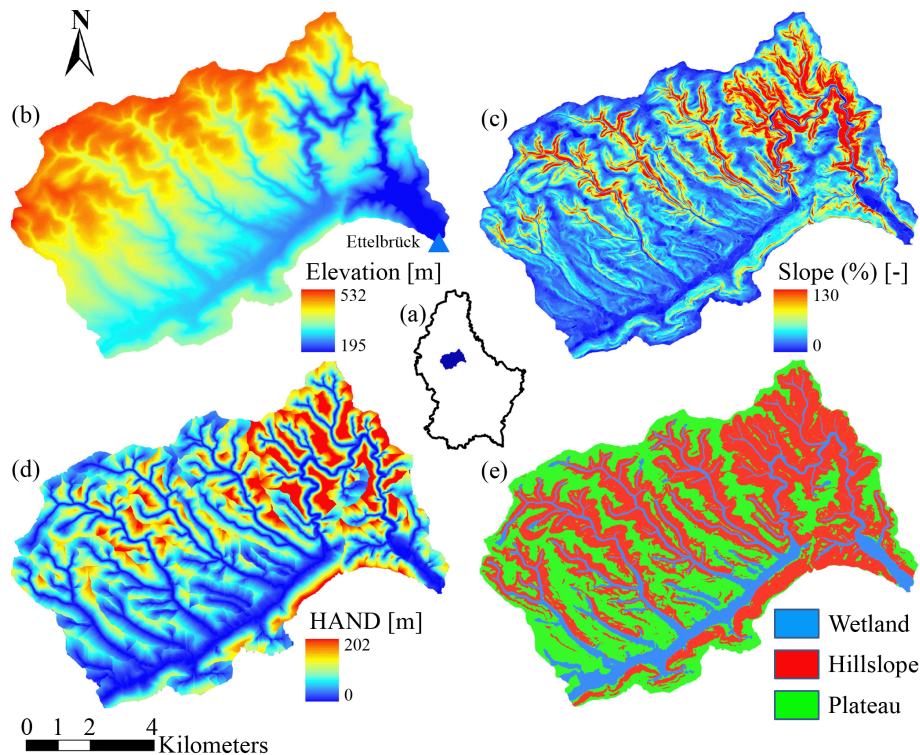
Figure 1. (a) Location of the Wark catchment in the Grand Duchy of Luxembourg, **(b)** digital elevation model (DEM) of the Wark catchment with cell size of 5 m × 5 m, **(c)** local slopes (%) in the Wark catchment derived from a DEM with resolution of 5 m × 5 m [–], **(d)** HAND of the Wark catchment derived from a DEM with resolution of 5 m × 5 m [m], **(e)** the classified landscape units wetland, hillslope and plateau using the combined HAND and slope thresholds of 5 m and 11 %, respectively (from Gharari et al., 2011).

landscape elements with steeper slopes than the wetlands or plateaus, require a balance between sufficient storage capacity and efficient drainage to develop and maintain the ecosystem (Savenije, 2010). A dual system combining sufficient water storage in the root zone and efficient lateral drainage through preferential flow networks, controlled by a suite of activation thresholds as frequently observed on hillslopes (e.g., Hewlett, 1961; Beven and Germann, 1982; Sidle et al., 2001; Freer et al., 2002; Weiler et al., 2003; McNamara et al., 2005; Tromp-van Meerveld and McDonnell, 2006a, b; Zehe and Sivapalan, 2009; Spence, 2010) can be seen as the dominant mechanism. Finally, plateaus are landforms with low to moderate slopes and comparably deep ground water tables. In absence of significant topographic gradients and due to the potentially increased unsaturated storage capacity, it can be hypothesized that the primary functions of plateaus are sub-surface storage and groundwater recharge (Savenije, 2010). Although plateaus may experience infiltration excess overland flow in specific locations, the topographical gradients may not be sufficient to generate surface runoff connected to the stream network (Liu et al., 2012). In the FLEX-TOPO approach the proportions of the hydrologically distinct landscape units, i.e., HRUs, in a given catchment need to be determined on the basis of topographical and land cover information. Subsequently suitable model structures and parameterizations (read constitutive functions) will be assigned to the different HRUs (Clark et al., 2009; Fenicia et al., 2011; Hrachowitz et al., 2014). The integrated catchment output, i.e., runoff and evaporative fluxes, can then be obtained by combining the computed proportional outputs from the individual HRUs. Note that the three landscape classes tested for suitability in this study, i.e., wetland, hillslope and plateau together with their assumed dominant runoff process are designed for the Wark catchment and are likely to be different for other environmental settings (e.g., Gao et al., 2014).

3.1 Landscape classification

As the objective of FLEX-TOPO is to efficiently extract and use hydrologically relevant information from worldwide readily available topographic data, i.e., DEMs, the Height Above the Nearest Drainage (HAND; Rennó et al., 2008; Nobre et al., 2011; Vannametee et al., 2014) is a potentially powerful metric to classify landscapes into HRUs with distinct hydrological function, as discussed above. Testing a suite of HAND-based classification methods Gharari et al. (2011) found that results best matching observed landscape types could be obtained by using HAND together with the local slope. Based on a probabilistic framework to map the desired

HRUs which were then compared with in situ observations they obtained a threshold for HAND and slope of approximately 5 m and 11 % for the Wark catchment. Following that, wetlands were defined to be areas with HAND $\leq$ 5 m. Areas with HAND > 5 m and local slopes > 11 % were classified as hillslopes, while areas with HAND > 5 m and slope $\leq$ 11 % were defined as plateaus. The HAND and slope map of the study catchment together with the classified landscape entities (wetland, hillslope and plateau) are presented in Fig. 1. The proportion of the individual HRUs wetland, hillslope, and plateau are 15 %, 45 %, and 40 %, respectively.

3.2 Model setup

In this study a lumped conceptual model of the Wark catchment, hereafter referred to as FLEX[A], is used as similar lumped conceptual models are frequently used in catchment hydrology (e.g., Merz and Blöschl, 2004; Clark et al., 2008; Perrin et al., 2008; Seibert and Beven, 2009; Fenicia et al., 2014). The above discussed concept of FLEX-TOPO (Savenije, 2010) is thereafter tested with a stepwise increased number of parallel landscape units (FLEX[B], FLEX[C]), thereby increasing the conceptualized process heterogeneity and thus the model complexity. The core of the three model setups is loosely based on the FLEX model (Fenicia et al., 2006, 2008b).

3.2.1 FLEX[A]

This model setup represents the catchment in a lumped way. The FLEX[A] model structure consists of four storage elements representing interception, unsaturated, slow (i.e., groundwater) and fast responding reservoirs (i.e., preferential flow and saturation overland flow). A schematic illustration of FLEX[A] is shown in Fig. 2a. The water balance equations and constitutive functions used are given in Table 2.

Interception reservoir (S_I)

The interception reservoir is characterized by its maximum storage capacity (I_{max} [L]). After precipitation (P [L T^{-1}]) enters this reservoir the excess precipitation, hereafter referred to as effective precipitation (P_e [L T^{-1}]), is distributed between the unsaturated (S_U), slow (S_S) and fast (S_F) reservoirs. Interception (I [L T^{-1}]) is then dependent on the potential evaporation (E_{pot} [L T^{-1}]) and the amount of water stored in the interception reservoir (S_I).

Unsaturated reservoir (S_U)

The unsaturated reservoir is characterized by a parameter that loosely reflects the maximum soil moisture capacity in the root zone ($S_{U,max}$ [L]). Part of the effective precipitation (P_e) enters the unsaturated zone according to the coefficient C_r, which here is defined by a power function with exponent β [–], reflecting the spatial heterogeneity of thresholds for activating fast lateral flows from S_F. This coefficient C_r will be 1 when soil moisture (S_U) is lower than a specific percentage of maximum soil moisture capacity ($S_{U,max}$) defined by relative soil moisture at field capacity (F_C [–]), meaning that the entire incoming effective precipitation (P_e) at a given time step is stored in the unsaturated reservoir (S_U). The soil moisture reservoir feeds the slow reservoir through matrix percolation (R_p [L T^{-1}]), expressed as a linear relation of the available moisture in the unsaturated zone (S_U) and the maximum percolation capacity (P_{Per} [L T^{-1}]). The reverse process, capillary rise (R_C), feeds the unsaturated reservoir from the saturated zone. Capillary rise (R_C [L T^{-1}]) has an inverse linear relation with the moisture content in the unsaturated zone and is characterized by the maximum capillary rise capacity (C [L T^{-1}]). Soil moisture is depleted by plant transpiration. Transpiration is assumed to be moisture constrained when the soil moisture content is lower than a fraction L_p [–] of the maximum unsaturated capacity ($S_{U,max}$). When the soil moisture content in the unsaturated reservoir is higher than this fraction (L_p) transpiration is assumed to be equal to the available potential evaporation ($E_{pot} - I$).

Splitter and transfer functions

The proportion of effective rainfall which is not stored in the unsaturated zone, i.e., $1 - C_r$, is further regulated by the partitioning coefficient (D [–]), distributing flows between preferential groundwater recharge (R_S [L T^{-1}]) to S_S and water that is routed to the stream by fast lateral processes from S_F (e.g., preferential flow or saturation overland flow, R_F). Both fluxes are lagged by rising linear lag functions with parameters N_{lagf} and N_{lags}, respectively (Fenicia et al., 2006).

Fast reservoir (S_F)

The fast reservoir is a linear reservoir characterized by reservoir coefficient S_F.

Slow reservoir (S_S)

The slow reservoir is a linear reservoir characterized by reservoir coefficient S_S.

3.2.2 FLEX[B]

As discussed above, a range of process studies suggested that wetlands can frequently exhibit storage–discharge dynamics that are decoupled from other parts of a catchment, in particular due to their typically reduced storage capacity and closeness to the stream. FLEX[B] explicitly distinguishes wetlands from the rest of the catchment, the *remainder* (i.e., hillslopes and plateaus), which is represented in a lumped way, to account for this difference. The FLEX[B] model setup therefore consists of two parallel model structures which are connected with a common groundwater reservoir (Fig. 2b), similar to
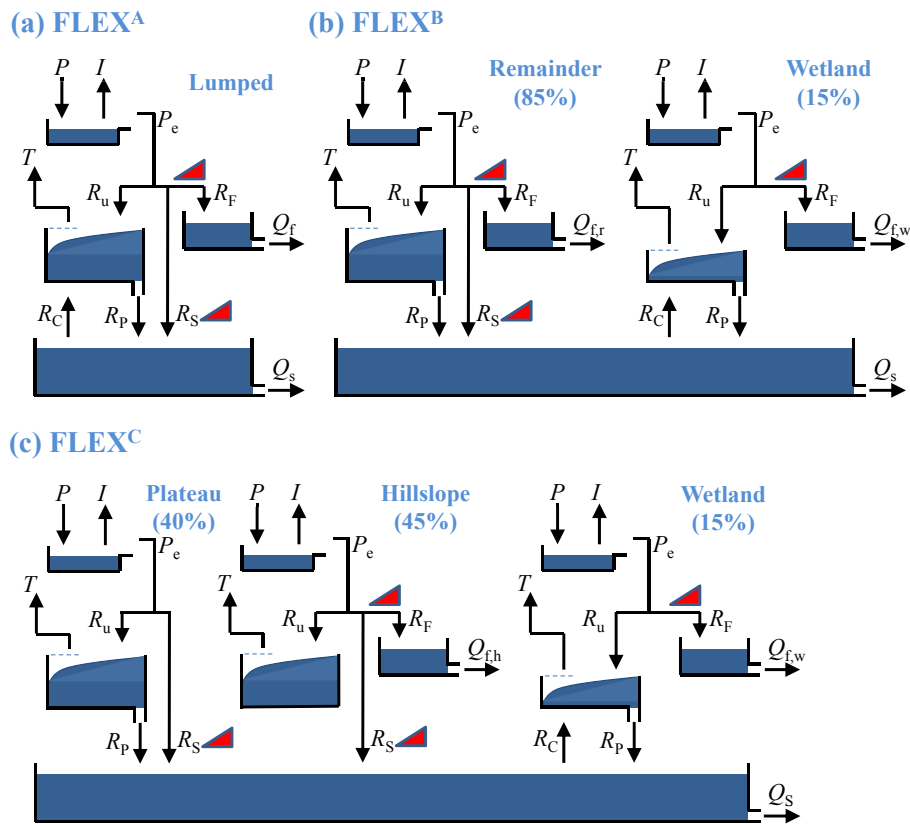
Figure 2. The model structures for **(a)** FLEX^A, **(b)** FLEX^B and **(c)** FLEX^C.

what has been suggested by Knudsen et al. (1986). One major difference between the two parallel structures is that capillary rise is assumed to be a relevant process only in the wetland, while it is considered negligible in the remainder of the catchment due to the deeper groundwater. Further, since the wetlands are predominantly ex-filtration zones of potentially low permeability, preferential recharge is considered negligible in wetlands. The areal proportions of wetland and the remainder (i.e., hillslope and plateau) of the catchment are 15 % and 85 %, respectively (Gharari et al., 2011).

3.2.3 FLEX^C

This model setup offers a complete representation of the three HRUs in the study catchment: wetland, hillslope and plateau (Fig. 2c). The formulation of the wetland module in FLEX^C is identical to the one suggested above for FLEX^B. The hillslope HRU is represented by a model structure resembling the FLEX^A setup. Plateaus are assumed to be dominated by vertical fluxes, while direct lateral movement in the form of Hortonian overland flow is considered negligible compared to those generated from hillslope and wetland HRUs. Therefore the plateau model structure does not account for these fast fluxes. Analogous to FLEX^B, the FLEX^C setup is characterized by one single groundwater

reservoir linking the three dominant HRUs in this catchment. The individual proportions of wetland, hillslope, and plateau are 15 %, 45 %, and 40 %, respectively (Gharari et al., 2011). The proportions of these HRUs are used to compute the total discharge based on the contribution of each landscape unit.

The connection between the parallel structures of FLEX^B and FLEX^C is through the surface drainage network (the stream network) and through the slow (groundwater) reservoir.

3.3 Introducing realism constraints in selecting behavioral parameter sets

With increasing process heterogeneity from FLEX^A over FLEX^B to FLEX^C, the respective model complexities and therefore the number of calibration parameters also increase. This, in the absence of sufficient suitable data to efficiently constrain a model, typically leads to a situation where parameters have increased freedom to compensate for errors in data and model structures, as recently reiterated by Gupta et al. (2008). In this study, two fundamentally different types of model constraints were applied to test their value for reducing equifinality in complex model setups, *parameter constraints* and *process constraints*.

Table 1. Uniform prior parameter distributions for the three model setups.

	Unit		FLEX^A	FLEX^B wetland	FLEX^B remainder	FLEX^C wetland	FLEX^C hillslope	FLEX^C plateau
I_{max}	mm	Interception storage for forest*			2–5			
		Interception storage for cropland and grassland*			1–3			
$S_{U,max}$	mm	Maximum unsaturated storage	0–500	0–100	0–500	0–100	0–500	0–500
β	–	Soil moisture distribution exponent	0–5	0–5	0–5	0–5	0–5	0–5
L_P	–	Transpiration coefficient	0.5	0.5	0.5	0.5	0.5	0.5
F_C	–	Relative soil moisture at field capacity	0–0.3	0	0–0.3	0	0–0.3	0–0.3
D	–	Partitioning fast and slow reservoir	0–1	0	0–1	0	0–1	1
C	mm $(3\,h)^{-1}$	Maximum capillary rise rate	0	0–0.3	–	0–0.3	–	–
P_{per}	mm $(3\,h)^{-1}$	Maximum percolation rate	0–0.5	0–0.5	0–0.5	0–0.5	0	0–0.5
N_{lagf}	3 h	Lag time for flux to fast reservoir	1–7	1–3	1–5	1–3	1–5	–
N_{lags}	3 h	Lag time for preferential recharge	1–7	–	1–7	–	1–7	1–7
K_F	$3\,h^{-1}$	Fast reservoir coefficient	0–1	0–1	0–1	0–1	0–1	–
K_S	$3\,h^{-1}$	Slow reservoir coefficient	0.005–0.05	0.005–0.05			0.005–0.05	

* Inferred from Breuer et al. (2003).

Table 2. Water balance equations and constitutive functions used in FLEX^A.

Reservoir	Water balance equations	Constitutive functions
Interception reservoir	$\frac{\Delta S_I}{\Delta t} = P - P_e - I$	$P_e = \begin{cases} 0 & S_I < I_{max} \\ (S_I - I_{max})/\Delta t & S_I = I_{max} \end{cases}$ $I = \begin{cases} E_{pot} & E_{pot}\Delta t < S_I \\ S_I/\Delta t & E_{pot}\Delta t \geq S_I \end{cases}$
Unsaturated reservoir	$\frac{\Delta S_U}{\Delta t} = R_u - T - R_P + R_C$	$R_u = C_r P_e$ $C_r = \begin{cases} 1 - \left[\frac{(S_U - S_{U,max} F_C)}{(S_{U,max} - S_{U,max} F_C)}\right]^\beta & S_U \geq S_{U,max} F_C \\ 1 & S_U < S_{U,max} F_C \end{cases}$ $T = K_T (E_{pot} - I)$ $K_T = \begin{cases} \left[\frac{S_U}{S_{U,max} L_p}\right] & S_U < S_{U,max} L_p \\ 1 & S_U \geq S_{U,max} L_p \end{cases}$ $R_p = [S_U/S_{U,max}] P_{per}$ $R_C = [1 - (S_U/S_{U,max})] C$
Fast reservoir	$\frac{\Delta S_F}{\Delta t} = R_{F,lag} - Q_f$	$R_F = (1 - D)(1 - C_r) P_e$ $R_{F,lag} = R_F^* N_{lagf}$ $Q_f = K_F S_F$
Slow reservoir	$\frac{\Delta S_S}{\Delta t} = R_{S,lag} - Q_s + R_P - R_C$	$R_S = D(1 - C_r) P_e$ $R_{S,lag} = R_S^* N_{lags}$ $Q_s = K_S S_S$

* is the convolution operator.

3.3.1 Parameter constraints

Inequality conditions between parameters of parallel model units, hereafter referred to as *parameter constraints*, were imposed *before* each model evaluation run. These a priori constraints ensure that the individual parameter values for the same process in the parallel units, reflect the modeler's perception of the system. For example, it can be argued that the maximum interception capacity (I_{max}) of a forested HRU needs to be higher than that of a non-forested one. In the absence of more detailed information this does not only allow overlapping prior distributions but it also avoids the need for quantification of the constraints themselves. In the following, a set of parameter constraints imposed on the different model structure are listed. The applicability of each

parameter constraint for every model structure is summarized in Table 3. The subscripts w, h, and p indicate parameters for wetland, hillslope, and plateau, respectively.

Interception

Different land cover proportions of individual landscape units, here wetlands, hillslopes and plateaus, can be used to define the relation between interception thresholds (I_{max}) of these individual units. The land uses are defined as two general classes for this case study: (1) forested areas, (2) cropland and grassland areas. The maximum interception capacity (I_{max}) for each landscape entity can be estimated from the proportion of land-use classes and their maximum interception capacities, selected from their respective prior

distributions as given in Table 1:

$$I_{\text{max,w}} = a_{\text{w}} I_{\text{max,forest}} + b_{\text{w}} I_{\text{max,crop-grassland}}, \qquad (1)$$

$$I_{\text{max,h}} = a_{\text{h}} I_{\text{max,forest}} + b_{\text{h}} I_{\text{max,crop-grassland}}, \qquad (2)$$

$$I_{\text{max,p}} = a_{\text{p}} I_{\text{max,forest}} + b_{\text{p}} I_{\text{max,crop-grassland}}. \qquad (3)$$

The proportions of forested area are indicated with a_{w}, a_{h} and a_{p} for wetland, hillslope, and plateau and are fixed at 42 %, 60 %, and 29 %, respectively. The proportions of cropland and grassland areas are indicated by b_{w}, b_{h}, and b_{p} for wetland, hillslope, and plateau and are fixed at 58 %, 40 %, and 71 %, respectively. Moreover the parameter sets which are selected for maximum interception capacity of forest are expected to be higher than cropland or grassland:

$$I_{\text{max,crop-grassland}} < I_{\text{max,forest}}. \qquad (4)$$

Lag functions

Preferential recharge (R_{S}) is routed to the slow reservoir by a lag function. Due to a deeper groundwater table on plateaus it can be assumed that the lag time for (R_{S}) is longer for plateaus than for hillslopes. It can also be assumed that the lag function used for fast reservoir for hillslopes is longer than for wetlands due to the on average higher distance and therefore longer travel times from hillslopes to the stream.

$$N_{\text{lags,w}} \le N_{\text{lags,h}} \le N_{\text{lags,p}} \qquad (5)$$

Soil moisture capacity

Wetlands have shallower groundwater tables than the other two landscape entities in this study. Therefore the unsaturated zone of wetland should have a lower maximum soil moisture capacity ($S_{\text{U,max}}$) than hillslopes and plateaus. Moreover, as hillslopes in the study catchment are predominantly covered with forest, it can, due to the deeper root zone of forests, be expected that the maximum unsaturated soil moisture capacity ($S_{\text{U,max}}$) in the root zone of hillslopes is deeper than the other two landscape entities.

$$S_{\text{U,max,w}} < S_{\text{U,max,p}} < S_{\text{U,max,h}} \qquad (6)$$

Reservoir coefficients

The reservoir coefficient of the wetland fast reservoir (K_{F}) is assumed to be higher than reservoir coefficient of the hillslope fast reservoir as, once connectivity is established, the flow velocities of saturation overland flow in wetlands are assumed to exceed the integrated flow velocities of preferential flow networks (cf. Anderson et al., 2009). Likewise, the reservoir coefficient of the slow reservoir should be lower than both wetland and hillslope fast reservoirs.

$$K_{\text{S}} < K_{\text{F,h}} < K_{\text{F,w}} \qquad (7)$$

The reservoir constraints can be applied to all models while the other constraints can only be applied to FLEX[B] and FLEX[C].

3.3.2 Process constraints

In contrast to the parameters constraints discussed above, which are set a priori, *process constraints* are applied a posteriori. Only parameters which generate model flux and state dynamics in agreement with the modeler's perception of these dynamics are retained as feasible. Hence, while with the use of parameter constraints there is no need to run the model for rejected parameter sets, here it is necessary to run the model to evaluate it with respect to the process constraints.

Process constraints are defined for dry and wet periods as well as for peak, high, and low flows. Here wet periods were defined to be the months from November to April, while the dry periods in the study catchment occur between May and October. The thresholds for distinguishing between high and low flow were chosen to be 0.05 and 0.2 mm (3 h)$^{-1}$ respectively for dry and wet periods. Furthermore, events during which discharge increases with a rate of more than 0.2 mm (3 h)$^{-2}$ are defined as peak flows. Note that in the following the subscripts peak, high and low indicate peak, high, and low flows. The applicability of each process constraint for every model structure is summarized in Table 3.

Transpiration

Transpiration typically exhibits a clear relationship with the normalized difference vegetation index (NDVI, Szilagyi et al., 1998). Therefore the ratios between NDVI values of different landscape units can serve as constraints on modeled transpiration obtained from the individual parallel model components. A rough estimation of the ratio between transpiration from plateau and hillslope can be derived from LANDSAT 7 images. For this ratio, seven cloud free images were selected (acquisition dates of 20 April 2000, 6 March 2000, 11 September 2000, 18 February 2001, 6 March 2001, 26 March 2001 and 29 August 2001). The ratio of transpiration between hillslope and plateau (R_{trans}) can be estimated by assuming a linear relation (Szilagyi et al., 1998) with slope of α and intercept zero between transpiration and mean NDVI for each landscape unit (μ_{NDVI}).

$$R_{\text{trans}} = \frac{\alpha \mu_{\text{NDVI,h}}}{\alpha \mu_{\text{NDVI,p}}} = \frac{\mu_{\text{NDVI,h}}}{\mu_{\text{NDVI,p}}} \qquad (8)$$

Mean (μ_{Rtrans}) and standard deviation (σ_{Rtrans}) of the transpiration ratio (R_{trans}) can be used to estimate acceptable limits of the transpiration ratios for hillslope and plateau. Therefore the annual transpiration can be confined between two values as follows:

$$\mu_{\text{Rtrans}} - \sigma_{\text{Rtrans}} < \frac{\sum T_{\text{h}} \text{d}t}{\sum T_{\text{p}} \text{d}t} < \mu_{\text{Rtrans}} + \sigma_{\text{Rtrans}}. \qquad (9)$$

Based on the mean ($\mu_{\text{Rtrans}} = 1.2$) and standard deviation ($\sigma_{\text{Rtrans}} = 0.2$) of the seven LANDSAT 7 images used, the

Table 3. The applicability of different parameter and process constraints for the three different model structures, FLEX[A], FLEX[B] and FLEX[C].

	Parameter constraints				Process constraints			
	Interception	Lag functions	Soil moisture capacity	Reservoir coefficients	Transpiration	Runoff coefficient	Preferential recharge	Fast component discharge
FLEX[A]				×		×		
FLEX[B]	×	×	×	×		×		×
FLEX[C]	×	×	×	×	×	×	×	×

following process constraint on transpiration from hillslope (T_h) and plateau (T_p) was imposed:

$$1.0 < \frac{\sum T_h dt}{\sum T_p dt} < 1.4. \tag{10}$$

Similar constraints can be imposed between transpiration fluxes from wetland, hillslope or plateau; however, the spatial resolution of LANDSAT 7 data with resolution of 30 meters is coarser than the required 20 m DEM resolution for distinguishing wetlands from other landscape entities (Gharari et al., 2011).

Runoff coefficient

The runoff coefficient is a frequently used catchment signature (e.g., Sawicz et al., 2011; Euser et al., 2013) and can be used as a behavioral constraint (e.g., Duan et al., 2006; Winsemius et al., 2009). In this study the runoff coefficients of dry and wet periods as well as the annual runoff coefficient were used. Parameters that result in modeled runoff coefficients that substantially deviate from the observed ones are therefore discarded. In case of absence of suitable runoff data, the long-term mean annual runoff coefficient can be estimated from the regional Budyko curve using for example the Turc–Pike relationship (Turc, 1954; Pike, 1964; Arora, 2002). However, in this study, the runoff coefficients of each individual year and their respective dry and wet periods were used and determined the mean and standard deviation of the runoff coefficients for these periods. Here, as a conservative assumption, the limits are set to three times the standard deviation around the mean runoff coefficient. Note that the runoff coefficient is the only process constraint that is not related to model structure in this study and can therefore also be applied to the lumped FLEX[A] setup.

$$\frac{\sum Q_m \Delta t}{\sum P \Delta t} < 0.43 \tag{11}$$

$$\frac{\sum Q_m \Delta t}{\sum P \Delta t} > 0.16 \tag{12}$$

$$\frac{\sum Q_{m,dry} \Delta t}{\sum P_{dry} \Delta t} < 0.36 \tag{13}$$

$$\frac{\sum Q_{m,dry} \Delta t}{\sum P_{dry} \Delta t} > 0 \tag{14}$$

$$\frac{\sum Q_{m,wet} \Delta t}{\sum P_{wet} \Delta t} < 0.71 \tag{15}$$

$$\frac{\sum Q_{m,wet} \Delta t}{\sum P_{wet} \Delta t} > 0.40 \tag{16}$$

Preferential recharge

The slow reservoir can be recharged by both preferential and matrix percolation from the unsaturated reservoirs. Here, hillslopes and plateaus contribute to the slow reservoir by preferential recharge. It can be assumed that in a realistic model setup the long term contribution volume of preferential recharge ratio between hillslope and plateau should not be unrealistically high or low. For example, it can be assumed unrealistic that the ratio is zero or infinity, meaning that one landscape unit is constantly feeding the slow reservoir while another one is not contributing at all. To avoid such a problem, a loose and very conservative constraint was imposed on the ratio of contribution of the two fluxes.

$$0.2 < \frac{\sum R_{S,h} \Delta t}{\sum R_{S,p} \Delta t} < 5 \tag{17}$$

Fast component discharge

During dry periods, hillslopes and plateaus can exhibit significant soil moisture deficits, limiting the amount of fast runoff generated from these landscape elements. In contrast, due to their reduced storage capacity, wetlands are likely to generate fast flows at lower moisture levels, thus dominating event response during dry periods (cf. Beven and Freer, 2001b; Seibert, 2003; Molénat et al., 2005; Anderson et al., 2010; Birkel et al., 2010). It can thus be assumed that for peak flows during dry periods, the fast component of wetlands ($Q_{f,w,dry,peaks}$) contributes more to runoff than the fast component of hillslopes ($Q_{f,h,dry,peaks}$). In contrast, high flows

during wet periods are predominantly generated by fast reaction from hillslopes ($Q_{f,h,wet}$; $Q_{f,h,wet,high}$) rather than of wetland ($Q_{f,w,wet}$; $Q_{f,w,wet,high}$). This process constraint is also applied to FLEX[B].

$$\frac{\sum Q_{f,h,dry,peaks} \Delta t}{\sum Q_{f,w,dry,peaks} \Delta t} < 1 \tag{18}$$

$$\frac{\sum Q_{f,h,wet,high} \Delta t}{\sum Q_{f,w,wet,high} \Delta t} > 1 \tag{19}$$

$$\frac{\sum Q_{f,h,wet} \Delta t}{\sum Q_{f,w,wet} \Delta t} > 1 \tag{20}$$

3.3.3 Calibration algorithm and objective functions

Based on uniform prior parameter distributions as well as on the parameter and process constraints the model was calibrated using MOSCEM-UA (Vrugt et al., 2003). However penalizing the objective function(s) based on the number of unsatisfied constraints, can lead to non-smooth objective functions that can cause instabilities in the search algorithm and the generation of invalid results. To resolve this issue, a recently developed stepwise search algorithm was used to find parameter sets that satisfy both parameter and process constraints (Gharari et al., 2014), and these parameter sets were then used as initial sampling parameter sets for MOSCEM-UA (instead of traditional Latin Hypercube sampling).

The models were evaluated on the basis of three different objective functions to emphasize different characteristics of the system response: (i) the Nash–Sutcliffe efficiency of the flows (Nash and Sutcliffe, 1970, E_{NS}), (ii) the Nash–Sutcliffe efficiency of the logarithm of the flows ($E_{NS,log}$) and (iii) the Nash–Sutcliffe efficiency of the flow duration curve ($E_{NS,FDC}$). These criteria evaluate the models' ability to simultaneously reproduce high flows, low flows and flow duration curves respectively. While the year 2001 was used as warm up period, the model setups were constrained and calibrated for the year 2002–2005 and validated for year 2006–2009 (see below) and vice versa (see Supplement).

3.4 Model validation and parameter evaluation

To assess the value of incorporating parameter and process constraints in increasingly complex models a four-step procedure was followed.

3.4.1 Evaluating models with *constrained but uncalibrated* parameter sets

First, all parameter sets that satisfy all the applied constraints were evaluated for their ability to reproduce the observed hydrograph; these parameter sets are referred to as *constrained but uncalibrated* parameter sets because they were obtained *without any calibration* to the observed hydrographs. Using these parameter sets, the mean performance of the three

constrained but uncalibrated models FLEX[A], FLEX[B] and FLEX[C], was evaluated using the three objective functions (E_{NS}, $E_{NS,log}$, $E_{NS,FDC}$). Note that FLEX[A], FLEX[B] and FLEX[C] have an increasing number of constraints and so this tests both whether the higher complexity models also result in better model performance and how the predictive uncertainty is affected by increased complexity and model realism. To investigate how well the hydrographs match the observed hydrograph, the simulated 95% uncertainty intervals were generated and uncertainty was estimated as the area contained within the 95% uncertainty intervals.

To further study the effect of constraints on the performance and uncertainty of *constrained but uncalibrated* parameter sets, three benchmark models were considered in which the aforementioned constraints were not applied. This simply means the models can produce any possible output without any restriction on parameters, fluxes and states. However the percentages of each landscape for model FLEX[B] and FLEX[C] remains intact.

3.4.2 Evaluating models with *constrained and calibrated* parameter sets

In the second step, the three models FLEX[A], FLEX[B] and FLEX[C] were calibrated while being constrained to the parameter space that satisfies all of the imposed parameter and process constraints. Calibration was performed using a multi-objective strategy (E_{NS}, $E_{NS,log}$, $E_{NS,FDC}$), and the obtained Pareto optimal model parameters are referred to as *constrained and calibrated*.

Uncertainty intervals were evaluated based on the *constrained and calibrated* Pareto members and the uncertainty was estimated on the basis of the area within the uncertainty bands.

Again, the results were compared to the calibrated but unconstrained benchmarks.

3.4.3 Comparison of model performance and uncertainty for *constrained but uncalibrated* and *constrained and calibrated* parameter sets

To assess the added value of incorporating constraints in higher complexity models, the performance and uncertainties of the three models FLEX[A], FLEX[B] and FLEX[C] were compared for both the *constrained but uncalibrated* and the *constrained and calibrated* case during calibration and validation periods.

3.4.4 Comparison of modeled hydrograph components for different model structures

One of the main reasons for imposing constraints on model parameters is to ensure realistic internal dynamics. Comparing different fluxes contributing to the modeled hydrograph can provide insights into the performance of imposed constraints on the model. The effect of

imposing behavioral constraints on fast and slow components of the three models structures, FLEX[A], FLEX[B] and FLEX[C] is compared visually. The fast component of the lumped model, FLEX[A], is compared with fast components of FLEX[B] that are wetland and remainder of catchment and fast components of FLEX[C] which are wetland and hillslope. This visual comparison is based on normalized average contribution of each component for Pareto optimal parameter sets in every time step.

4 Results and discussion

4.1 Evaluating the performance of *constrained but uncalibrated* parameter sets

The median and the 95 % uncertainty intervals of the performance of modeled hydrographs for *constrained but uncalibrated* parameter sets is presented in Table 4 for the 2002–2005 calibration and 2006–2009 validation periods together with their (*unconstrained*) benchmarks. The lumped FLEX[A] model has only one parameter and one process constraint, i.e., the reservoir coefficient and the runoff coefficient, respectively. Hence, this model is free within the limits of this relatively weak condition, resulting in a wide range of feasible parameter sets, many of which cannot adequately reproduce the system response. As a consequence, the overall performance is poor ($E_{\mathrm{NS,median}} = 0.18$, $E_{\mathrm{NS,log,median}} = 0.05$, $E_{\mathrm{NS,FDC,median}} = 0.39$) (Table 4, Fig. 3).

FLEX[B], run with the set of *constrained but uncalibrated* parameters shows a substantial improvement in overall performance ($E_{\mathrm{NS,median}} = 0.56$, $E_{\mathrm{NS,log,median}} = 0.33$, $E_{\mathrm{NS,FDC,median}} = 0.87$) compared to FLEX[A], as FLEX[B] not only allows for more process heterogeneity but, more importantly, it is conditioned with an increased number of constraints.

The additional process heterogeneity and constraints allowed by FLEX[C], results in the highest overall performance for all three objective functions ($E_{\mathrm{NS,median}} = 0.66$, $E_{\mathrm{NS,log,median}} = 0.36$, $E_{\mathrm{NS,FDC,median}} = 0.93$) (Table 4, Fig. 3).

These results clearly illustrate that the imposed relational constraints force the model and its parameters towards a more realistic behavior, which significantly improves model performance. Additionally, the comparison of result of the three models with their unconstrained benchmarks (Table 4) clearly shows that the incorporation of constraints improves the median performance and 95 % uncertainty intervals of all the models by rejecting parameter sets that violate the constraints and cannot reproduce certain aspects of the response patterns. In addition, the comparison between the unconstrained benchmark models themselves suggests that more complex model structures improve the performance, implying that model structures themselves already contain

a considerable degree of information even in absence of any constraints or calibration attempts.

The 95 % uncertainty areas mapped by simulated hydrographs indicate that FLEX[C], which might be expected to produce the highest uncertainty interval due to its complexity, is providing a lower uncertainty compared to FLEX[B]. Although FLEX[C] cannot outperform FLEX[A] in terms of a narrower uncertainty interval in the validation period, overall performance of this model is better than FLEX[A] as discussed earlier (Table 4, Fig. 3).

Flipping calibration and validation gave equivalent results, which are, for brevity, provided in Table S1; Fig. S1 in the Supplement.

4.2 Evaluating the performance of *constrained and calibrated* parameter sets

The comparison of the *constrained and calibrated* model setups shows that all three models setups can reproduce the hydrograph similarly well (Table 5, Fig. 4). FLEX[A] exhibits a slightly better calibration performance, based on $E_{\mathrm{NS,log,median}}$, compared to the other two model setups. This can partly be attributed to the lower number of parameters which leads, with the same number of samples, to a more exhaustive sampling of the parameter space and a smoother identification of Pareto optimal solutions. In addition, FLEX[A] has the lowest number of imposed constraints, i.e., only the runoff coefficient and one parameter constraints, compared to FLEX[B] and FLEX[C]. This allows the model more freedom in exploiting the parameter space to produce mathematically good fits between observed and modeled system response in the calibration period.

For the validation period, arguably more important for model assessment because it provides independent information on model consistency (cf. Klemeš, 1986; Andréassian et al., 2009; Euser et al., 2013) and predictive uncertainty, the performances of the three model setups exhibit quite different patterns (Table 5). The simplest model, the lumped FLEX[A], is characterized by performance deterioration from calibration to validation. In contrast, FLEX[B] and FLEX[C] exhibit a performance improvement in the validation period. Although the increase in performance is subjected to the nature of the forcing and observed discharge data in calibration and validation period, and formally no meaningful comparison between Nash–Sutcliffe efficiencies for different periods can be made, these results nevertheless indicate that the more complex model structure together with its constraints performs in a more stable manner outside of the calibration period. When flipping the calibration and validation periods the difference between model performance in calibration and validation is not as strong (Table S2). A possible explanation could be that the observed data quality is not informative enough for calibration (period 2002–2005). Constraints then prevent the model to over-fit and thus enable the models to maintain a more reliable performance outside the calibration

Table 4. The median model performances (in brackets their corresponding 95 % uncertainty intervals) and the area spanned by the 95 % uncertainty interval of hydrograph derived from uncalibrated parameter sets which satisfy the complete set of constraints for the three model setups FLEX[A], FLEX[B] and FLEX[C], for the three modeling objectives (E_{NS}, $E_{NS,log}$, $E_{NS,FDC}$) in the calibration (2002–2005) and validation (2006–2009) periods. The italic values indicate performance and 95 % uncertainty interval of hydrograph for the unconstrained benchmark models.

|		E_{NS}	$E_{NS,log}$	$E_{NS,FDC}$	95 % uncertainty area [mm]
FLEX[A]	Calibration	0.18 [0.09, 0.29]	0.05 [−0.40, 0.49]	0.39 [0.25, 0.69]	1325
		0.16[−0.16, 0.30]	*0.10[−1.11, 0.51]*	*0.35[−0.12, 0.67]*	*1814*
	Validation	0.23 [0.12, 0.39]	0.29 [−0.02, 0.59]	0.45 [0.28, 0.76]	1243
		0.18[−0.37, 0.39]	*0.29[−2.53, 0.56]*	*0.38[−0.35, 0.76]*	*1888*
FLEX[B]	Calibration	0.56 [0.00, 0.73]	0.33 [−1.36, 0.65]	0.87 [0.66, 0.95]	1827
		0.44[−1.03, 0.72]	*0.07[−3.06, 0.60]*	*0.77[0.05, 0.93]*	*2615*
	Validation	0.52 [−0.06, 0.77]	0.45 [−1.15, 0.73]	0.89 [0.62, 0.99]	2042
		0.45[−1.44, 0.76]	*0.30[−3.50, 0.73]*	*0.81[0.08, 0.97]*	*2993*
FLEX[C]	Calibration	0.66 [0.22, 0.75]	0.36 [−2.37, 0.70]	0.93 [0.82, 0.96]	1274
		0.54[−0.24, 0.75]	*0.34[−2.30, 0.69]*	*0.86[0.60, 0.94]*	*2015*
	Validation	0.67 [−0.06, 0.80]	0.50 [−0.33, 0.74]	0.95 [0.88, 0.99]	1294
		0.59[−0.11, 0.79]	*0.58[−2.89, 0.75]*	*0.93[0.65, 0.99]*	*2287*
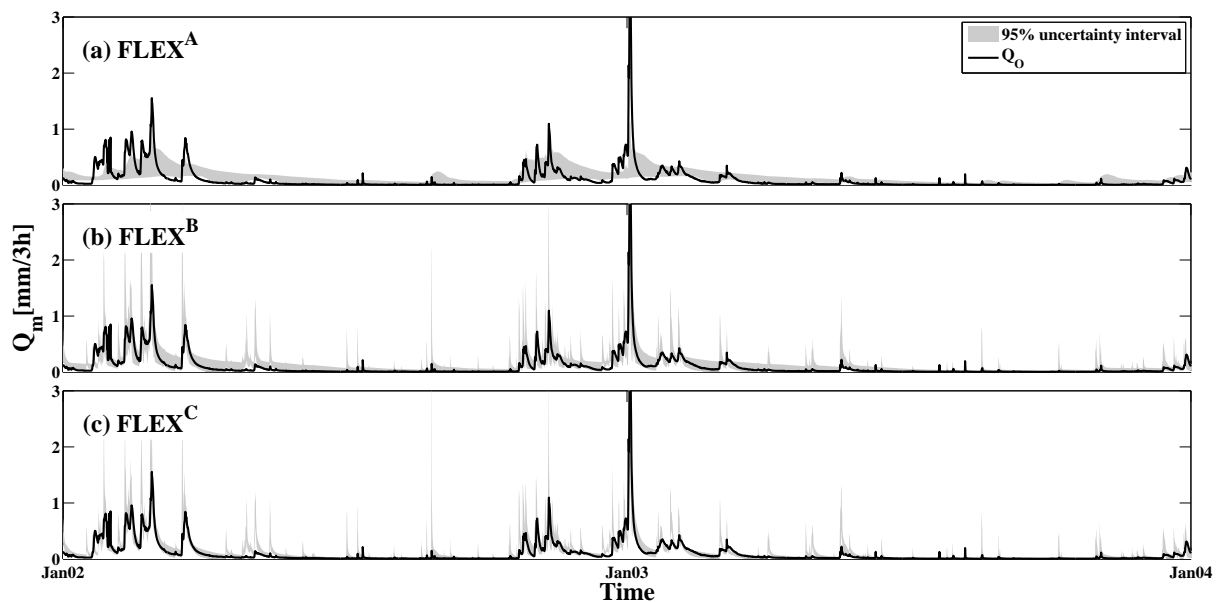
Figure 3. The observed hydrograph and the 95 % uncertainty interval of the modeled hydrograph derived from the complete set of *constrained but uncalibrated* parameter sets for the three different model setups **(a)** FLEX[A], **(b)** FLEX[B] and **(c)** FLEX[C] for 2 years (2002–2003) of calibration.

period. In contrast, if the calibration period is informative (period 2006–2009), constraints may not affect performance outside the calibration period that much. However constraints remain necessary to reduce model uncertainty both during calibration and validation. In addition to formal performance and uncertainty measures, it can be seen visually in Fig. 4 that FLEX[C] can adequately predict the high flows during a dry period, while FLEX[A] misses most of the peaks.

4.3 Comparison of *constrained but uncalibrated* and *constrained and calibrated* models

The following comparison of the performances of FLEX[A], FLEX[B] and FLEX[C] for *constrained but uncalibrated*, *constrained and calibrated* and their unconstrained benchmarks is focused on E_{NS} only, for the reason of brevity (Fig. 5, gray box plots indicate the benchmark models). In Fig. 5a and b the model performances based on the *constrained but*

Table 5. The median model performances (in brackets their corresponding Pareto uncertainty intervals) and the area spanned by the uncertainty interval of the hydrograph derived from the Pareto optimal solutions of the *constrained and calibrated* model setups FLEX[A], FLEX[B] and FLEX[C] for the three modeling objectives (E_{NS}, $E_{NS,log}$, $E_{NS,FDC}$) in the calibration (2002–2005) and validation (2006–2009) periods. The italic values indicate performance and 95 % uncertainty interval of hydrograph for the benchmark models (without any constraints).

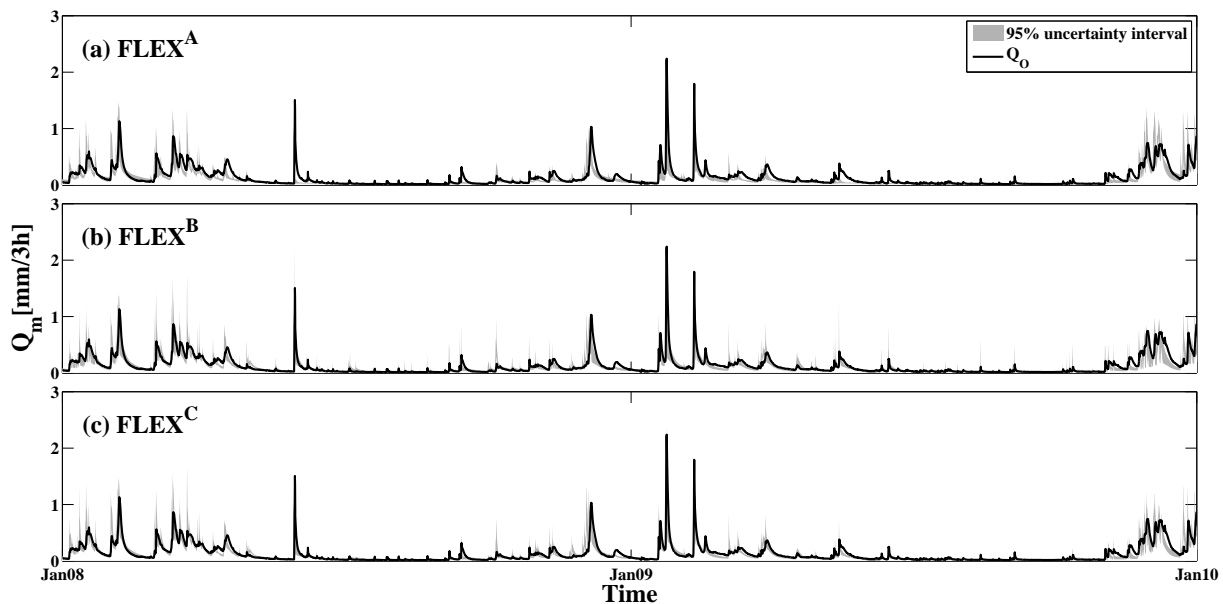|		E_{NS}	$E_{NS,log}$	$E_{NS,FDC}$	95 % uncertainty area [mm]
FLEX[A]	Calibration	0.71 [0.51, 0.83]	0.80 [0.70, 0.85]	0.97 [0.95, 0.99]	709
		0.71[0.51, 0.84]	*0.79[0.68, 0.85]*	*0.97[0.95, 0.99]*	*732*
	Validation	0.63 [0.45, 0.78]	0.73 [0.65, 0.80]	0.95 [0.93, 0.97]	844
		0.63[0.46, 0.78]	*0.73[0.63, 0.80]*	*0.95[0.93, 0.97]*	*870*
FLEX[B]	Calibration	0.75 [0.50, 0.80]	0.71 [0.40, 0.79]	0.96 [0.92, 0.98]	790
		0.74[0.51, 0.80]	*0.72[0.46, 0.82]*	*0.96[0.92, 0.98]*	*826*
	Validation	0.76 [0.32, 0.82]	0.79 [0.63, 0.85]	0.97 [0.93, 1.00]	999
		0.72[0.45, 0.82]	*0.78[0.48, 0.84]*	*0.96[0.94 0.99]*	*986*
FLEX[C]	Calibration	0.74 [0.53, 0.82]	0.72 [0.47, 0.81]	0.96 [0.92, 0.98]	763
		0.74[0.48, 0.82]	*0.71[−0.17, 0.83]*	*0.96[0.90, 0.98]*	*864*
	Validation	0.78 [0.45, 0.82]	0.83 [0.72, 0.85]	0.99 [0.98, 1.00]	927
		0.73[0.42, 0.83]	*0.78[−0.05, 0.85]*	*0.98[0.95, 0.99]*	*1047*

Figure 4. The observed hydrograph and the 95 % Pareto uncertainty interval of the modeled hydrograph for *constrained and calibrated* parameter sets for the three different model setups **(a)** FLEX[A], **(b)** FLEX[B] and **(c)** FLEX[C] for the 2 years (2008–2009) of validation period.

uncalibrated parameter sets, that satisfy the full set of constraints, are shown for the calibration and validation periods. As discussed in detail above, although uncalibrated, increasing the number of constraints from FLEX[A] to FLEX[C] increases the overall performance of the models while reducing uncertainty (Fig. 5c and d; note that these are zoom-ins). Further, comparison to the uncalibrated benchmarks, suggests that improving the model structure based on landscape units in itself substantially increases the performance of the model.

However additional constraints will eventually reduce the uncertainty and improve the performance (Fig. 5a and b).

Figure 5e compares model performance based on *constrained and calibrated* parameter sets for the calibration period. When comparing the individual model performances of the *constrained and calibrated* models during the validation period (Fig. 5f), it can be seen that FLEX[A] not only shows the strongest performance deterioration compared to the calibration period but also that FLEX[A] is also the model with the poorest performance in the validation period. This implies
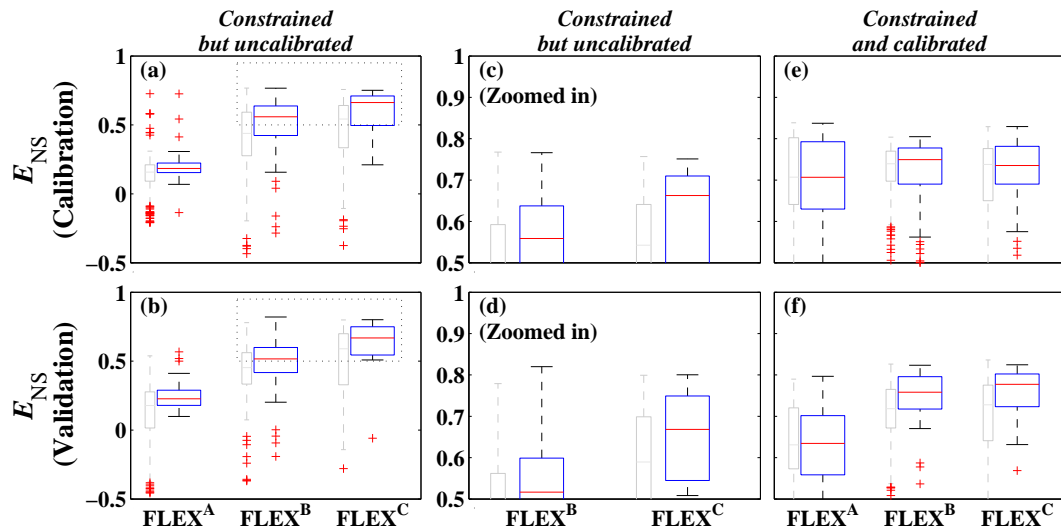
Figure 5. Model performance (E_{NS}) based on *constrained but uncalibrated* (**a–d**) and *constrained and calibrated* (**e–f**) parameter sets for calibration (2002–2005) and validation (2006–2009) periods for the three different model setups FLEX[A], FLEX[B] and FLEX[C]. Note that (**c**) and (**d**) are zoom-ins of (**a**) and (**b**) and the gray box-plots represent the unconstrained benchmark models. The box plots indicate the median value in red and 25 and 75 % quartile. Whiskers represent the 1.5 times the interquartile range (IQR) and the red crosses show outliers.

that although FLEX[C] is the most complex model, the realism constraints together with landscape related structure imposed on this model generate the most reliable outputs when used for prediction, i.e., in the validation period. When the calibration and validation periods are switched, the performance of FLEX[C] remains comparable to above, although in this case FLEX[A] performs best during validation (see Fig. S5). This strongly underlines that the widely accepted notion of complex models necessarily being subject to higher predictive uncertainty is not generally valid when the feasible parameter space can be constrained based on assumptions of realistic functionality of a catchment. As explained earlier, this also indicates that when the data of the calibration period are not sufficiently informative, imposing constraints will force the model to perform better outside the calibration period.

In addition, a second crucial aspect was revealed by comparing *constrained but uncalibrated* and *constrained and calibrated* models. It can be seen that *constrained but uncalibrated* FLEX[C], shows significant improvement in performance approaching the performance of the calibrated lumped model, FLEX[A]. Interestingly, it was found that in validation the *constrained but uncalibrated* FLEX[C] can, depending on the performance measure used and the information content of the calibration period (i.e., climatic variability and data quality), reach the performance level of the *constrained and calibrated* FLEX[A] (Figs. 5 and S5). This highlights the value of semi- and non-quantitative hydrological expert knowledge for finding suitable model parameter sets for ungauged basins.

4.4 Comparison of flow contributions from different model components

The comparison of the fluxes generated from the individual model components in the three model setups helps to assess to which degree the model internal dynamics reflect the modeler's perception of the system and thus to a certain degree the realism of the models.

Fast and slow responses of each tested model setup have been visually illustrated in Fig. 6. Predominance of slow responses of all the three models are indicated by green color; predominance of fast responses of FLEX[A], fast responses of the remainder of the catchment of FLEX[B] and fast responses of hillslope of FLEX[C] are indicated by red color; wetland fast responses of FLEX[B] and FLEX[C] are indicated by predominance of blue color.

The colors in Fig. 6 are an illustration using three colors (red, green and blue) for the models' responses based on their weight of contribution to the modeled runoff. As it can be seen in Fig. 6a the fast component of FLEX[A] is dominant just during peak flows and even the recession shortly after peak flows are accounted for mainly by ground water. Analysis of the individual model components computed by Pareto optimal parameter sets (not shown here for brevity), indicates that some Pareto optimal parameters can generate peak flows by predominant contributions from slow responses while fast reaction tends to be inactive during these events.

In accordance with the perception of the system that wetlands are predominantly responsible for peak flows during dry conditions, Fig. 6b and c show that wetland fast responses in FLEX[B] and FLEX[C] control the rapid response
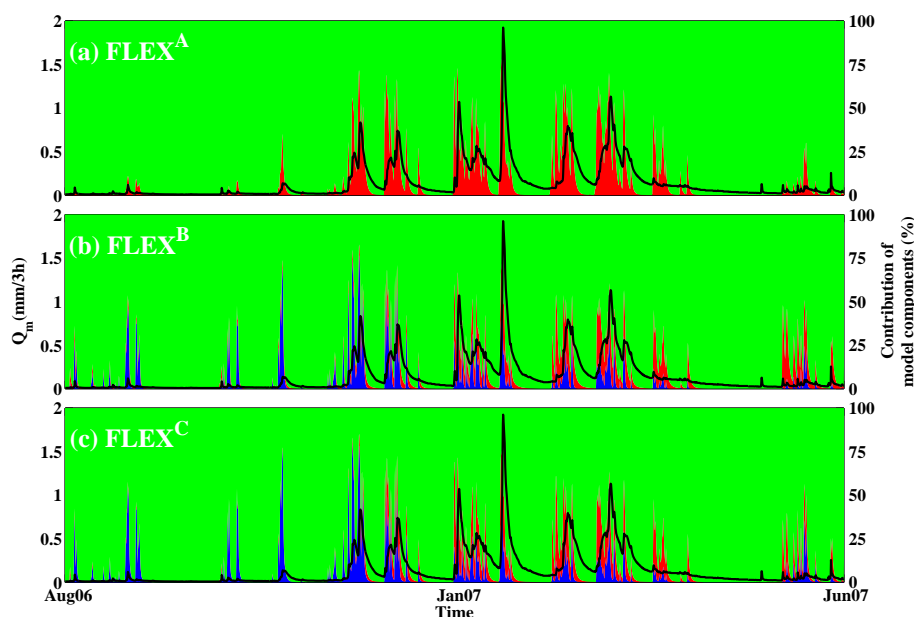
Figure 6. Comparison between mean proportions of Pareto members for model components of the three model setups in part of the validation periods (August 2006–June 2007) **(a)** FLEX[A], **(b)** FLEX[B], and **(c)** FLEX[C]. The green color indicates the relative contribution of the slow reservoir for the three different models. Red indicates relative contribution from the fast components, i.e., fast reservoir in FLEX[A], fast reservoir of the remainder of the catchment in FLEX[B] and fast reservoir of hillslope of FLEX[C]. The blue color indicates the relative contribution of fast wetland component of FLEX[B] and FLEX[C].

during wetting up periods (dry to wet transition), before hillslope fast processes become more important at higher moisture levels. When the system is saturated the hillslope contribution to modeled runoff becomes significantly higher compared to the wetland response. Note that the response of the wetland may not correspond well to individual events, as a consequence of the fact that the corresponding constraint was set for an aggregated period.

4.5 Wider implications

The results of this study quite clearly indicate that discretizing the catchment into hydrological response units (HRUs) and incorporating expert knowledge in model development and testing is a potentially powerful strategy for runoff prediction, even where insufficient data for model calibration (e.g., Koren et al., 2003; Duan et al., 2006; Winsemius et al., 2009) or only comparatively unreliable regionalization tools are available (e.g., Wagener and Wheater, 2006; Bárdossy, 2007; Parajka et al., 2007; Oudin et al., 2008; Laaha et al., 2014). It was found that the performance and the predictive power of a comparatively complex uncalibrated conceptual model, based on posterior parameter distributions obtained merely from relational, semi- and non-quantitative realism constraints inferred from expert knowledge, can approach or even be as efficient as the calibration of a lumped conceptual model (Figs. 5 and S3).

Typically it is expected that, if not warranted by data, models with higher complexity suffer from higher predictive uncertainty. As stated by Beven (2001): "More complexity means more parameters, more parameters mean more calibration problems, more calibration problems will often mean more uncertainty in the predictions, particularly outside the range of the calibration data." Thus, more parameters would allow better fits of the hydrograph but would not necessarily imply a better and more robust understanding of catchment behavior or more reliable predictions.

A complex model may include many processes, i.e., hypotheses, which can usually not be rigorously tested with the available data. However, a wide range of previous studies has demonstrated that hydrologically meaningful constraints can help to limit the increased uncertainty caused by incorporating additional processes, i.e., parameters (e.g., Yadav et al., 2007; Zhang et al., 2008; Kapangaziwiri et al., 2012). These studies generally include a large set of catchments and try to relate model parameters to catchment characteristics. Although regional constraints are important, the importance of expert knowledge on the catchment scale, which leads to better understanding of hydrological behavior is highlighted in this study.

In a similar attempt, Pokhrel et al. (2008, 2012) demonstrated use of regularization for model parameters and reduction of model parameter space dimensionality by linking model parameters using super-parameters to catchment characteristics. However, no explicit hydrological reasoning

is typically applied for such "regularization rules" (e.g., Pokhrel et al., 2012). On the other hand, Kumar et al. (2010, 2013a) parameterize and successfully regionalize their models using empirical transfer functions with global parameters, developed from extensive literature study and iterative testing in a large sample of catchments In contrast, the use of relational parameter and process constraints, as presented in this study, is based on semi-quantitative, hydrologically explicit and meaningful reasoning avoiding the need for empirical transfer functions to link catchments characteristics and model parameters.

Including prior knowledge for parameters of physically based models for estimating runoff in ungauged basins was quite successfully investigated in the past (e.g., Ott and Uhlenbrook, 2004; Vinogradov et al., 2011; Fang et al., 2013; Semenova et al., 2013). These studies specifically indicate that calibration can be replaced by prior information which is a significant contribution to Predictions in Ungauged Basins (PUB). While physically based models need detailed information of catchment behavior for model parameters, the here proposed semi-distributed conceptual modeling framework, exploiting relational constraints, can be more efficiently setup using the least prior information necessary. In this study, the performances and uncertainties of the three tested model setups for *constrained but uncalibrated* parameters indicate the potential of the presented FLEX-TOPO framework for Predictions in Ungauged Basins (PUB). Hence, this framework can efficiently use expert knowledge for improving model parameter value selection in complex conceptual hydrological models, not only to increase model performance but also to reduce model predictive uncertainty even in the absence of calibration.

It should be kept in mind that the conclusions of this study remain at this point only valid for the study catchment. To generalize the findings of this study more rigorous tests should be set up (Andréassian et al., 2009) which expand this presented concept for different time series of a catchment and also a larger set of catchments such as in recent work of Gao et al. (2014) and Hrachowitz et al. (2014). Some further challenges remain, including the need to formulate generic constraints for any catchment based on available data in an automated procedure. Likewise it will be necessary to develop a better understanding of model sensitivities to different constraints and of the effectiveness and reliability of individual constraints. It is also emphasized that the constraints introduced in this study are based on the authors' subjective understanding of catchment behavior and can and should be discussed further. However, we would like to stress the notion that reaching an agreement on the relations between parameters and fluxes in different landscape units is potentially much easier than finding the most adequate parameter values together with associated uncertainties for a conceptual model based on field observations or available data on geology or soil types.

5   Conclusions

This study has tested whether a topography-driven semi-distributed formulation of a catchment-scale conceptual model, conditioned by expert knowledge based relational parameter and process constraints, can increase the level of process realism and predictive power while reducing the need for calibration.

It was found that

1. The performance of models, *although uncalibrated*, improves by accounting for different topography-based hydrological response units, even if this introduces additional complexity.

2. Imposing relational parameter and process constraints improves the performance of *uncalibrated* models and reduces their uncertainty. This illustrates the potential value of the combined use of higher complexity models and relational constraints for prediction in ungauged basins, where no time series are available for model calibration.

3. Due to the reduced feasible parameter space, the search for behavioral parameter sets focuses on the feasible parameter space only.

4. Imposing constraints prevents the model from overfitting on calibration time series and therefore enables the model to more reliably perform outside the calibration period.

The Supplement related to this article is available online at doi:10.5194/hess-18-4839-2014-supplement.

Edited by: F. Tian

References

Ajami, N. K., Gupta, H. V., Wagener, T., and Sorooshian, S.: Calibration of a semi-distributed hydrologic model for stream-flow estimation along a river system, J. Hydrol., 298, 112–135, doi:10.1016/j.jhydrol.2004.03.033, 2004.

Anderson, A. E., Weiler, M., Alila, Y., and Hudson, R. O.: Subsurface flow velocities in a hillslope with lateral preferential flow, Water Resour. Res., 45, W11407, doi:10.1029/2008WR007121, 2009.

Anderson, A. E., Weiler, M., Alila, Y., and Hudson, R. O.: Piezometric response in zones of a watershed with lateral preferential flow as a first-order control on subsurface flow, Hydrol. Process., 24, 2237–2247, doi:10.1002/hyp.7662, 2010.

Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.-H., and Valéry, A.: HESS Opinions "Crash tests for a standardized evaluation of hydrological models", Hydrol. Earth Syst. Sci., 13, 1757–1764, doi:10.5194/hess-13-1757-2009, 2009.

Andréassian, V., Le Moine, N., Perrin, C., Ramos, M.-H., Oudin, L., Mathevet, T., Lerat, J., and Berthet, L.: All that glitters is not gold: the case of calibrating hydrological models, Hydrol. Process., 2, 2206–2210, doi:10.1002/hyp.9264, 2012.

Arora, V. K.: The use of the aridity index to assess climate change effect on annual runoff, J. Hydrol., 265, 164–177, doi:10.1016/S0022-1694(02)00101-4, 2002.

Bárdossy, A.: Calibration of hydrological model parameters for ungauged catchments, Hydrol. Earth Syst. Sci., 11, 703–710, doi:10.5194/hess-11-703-2007, 2007.

Beven, K.: INTERFLOW, Ursaturated FIow in Hydrologic Modeling Theory and Practice, Springer, 191–219, 1989.

Beven, K.: Rainfall-Runoff Modelling, The Primer, John Wiley and Sons, New York, USA, 2001.

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320, 18–36, doi:10.1016/j.jhydrol.2005.07.007, 2006.

Beven, K. and Freer, J.: A dynamic TOPMODEL, Hydrol. Process., 15, 1993–2011, doi:10.1002/hyp.252, 2001a.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol., 249, 11–29, doi:10.1016/S0022-1694(01)00421-8, 2001b.

Beven, K. and Germann, P.: Macropores and water flow in soils, Water Resour. Res., 18, 1311–1325, doi:10.1029/WR018i005p01311, 1982.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrolog. Sci. J., 24, 43–69, doi:10.1080/02626667909491834, 1979.

Birkel, C., Dunn, S. M., Tetzlaff, D., and Soulsby, C.: Assessing the value of high-resolution isotope tracer data in the stepwise development of a lumped conceptual rainfall-runoff model, Hydrol. Process., 24, 2335–2348, doi:10.1002/hyp.7763, 2010.

Birkel, C., D., T., Dunn, S., and Soulsby, C.: Using time domain and geographic source tracers to conceptualize streamflow generation processes in lumped rainfall-runoff models, Water Resour. Res., 47, W02515, doi:10.1029/2010WR009547, 2011.

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

Blume, T., Zehe, E., and Bronstert, A.: Investigation of runoff generation in a pristine, poorly gauged catchment in the Chilean Andes II: Qualitative and quantitative use of tracers at three spatial scales, Hydrol. Process., 22, 3676–3688, doi:10.1002/hyp.6970, 2008.

Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, Water Resour. Res., 36, 3663–3674, doi:10.1029/2000WR900207, 2000.

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

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

Bulygina, N. and Gupta, H.: How Bayesian data assimilation can be used to estimate the mathematical structure of a model, Stoch. Environ. Res. Risk A., 24, 925–937, doi:10.1007/s00477-010-0387-y, 2010.

Capell, R., Tetzlaff, D., and Soulsby, C.: Can time domain and source area tracers reduce uncertainty in rainfall-runoff models in larger heterogeneous catchments?, Water Resour. Res., 48, W09544, doi:10.1029/2011WR011543, 2012.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, W00B02, doi:10.1029/2007WR006735, 2008.

Clark, M. P., Rupp, D. E., Woods, R. A., Tromp-van Meerveld, H. J., Peters, N. E., and Freer, J. E.: Consistency between hydrological models and field observations: linking processes at the hillslope scale to hydrological responses at the watershed scale, Hydrol. Process., 23, 311–319, doi:10.1002/hyp.7154, 2009.

Clark, M. P., Kavetski, D., and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resour. Res., 47, W09301, doi:10.1029/2010WR009827, 2011.

Detty, J. M. and McGuire, K. J.: Topographic controls on shallow groundwater dynamics: implications of hydrologic connectivity between hillslopes and riparian zones in a till mantled catchment, Hydrol. Process., 24, 2222–2236, doi:10.1002/hyp.7656, 2010.

Drogue, G., Pfister, L., Leviandier, T., Humbert, J., Hoffmann, L., Idrissi, A. E., and Iffly, J.-F.: Using 3D dynamic cartography and hydrological modelling for linear streamflow mapping, Comput. Geosci., 28, 981–994, doi:10.1016/S0098-3004(02)00028-6, 2002.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H., Gusev, Y., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, J. Hydrol., 320, 3–17, doi:10.1016/j.jhydrol.2005.07.031, 2006.

Dunn, S. M., Bacon, J. R., Soulsby, C., Tetzlaff, D., Stutter, M. I., Waldron, S., and Malcolm, I. A.: Interpretation of homogeneity in ^{18}O signatures of stream water in a nested sub-catchment system in north-east Scotland, Hydrol. Process., 22, 4767–4782, doi:10.1002/hyp.7088, 2008.

Efstratiadis, A. and Koutsoyiannis, D.: Fitting Hydrological Models on Multiple Responses Using the Multiobjective Evolutionary Annealing-Simplex Approach, in: Practical Hydroinformatics,

www.hydrol-earth-syst-sci.net/18/4839/2015/

Hydrol. Earth Syst. Sci., 18, 4839–4859, 2015

edited by: Abrahart, R. J., See, L. M., Solomatine, D. P., and Singh, V., vol. 68 of Water Science and Technology Library, Springer, Berlin, Heidelberg, 259–273, doi:10.1007/978-3-540-79881-1_19, 2008.

Engl, H. W., Hanke, M., and Neubauer, A.: Regularization of inverse problems, vol. 375, Kluwer Academic, Dordrecht, the Netherlands, 1996.

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, doi:10.5194/hess-17-1893-2013, 2013.

Fang, X., Pomeroy, J. W., Ellis, C. R., MacDonald, M. K., DeBeer, C. M., and Brown, T.: Multi-variable evaluation of hydrological model predictions for a headwater basin in the Canadian Rocky Mountains, Hydrol. Earth Syst. Sci., 17, 1635–1659, doi:10.5194/hess-17-1635-2013, 2013.

Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: Is the groundwater reservoir linear? Learning from data in hydrological modelling, Hydrol. Earth Syst. Sci., 10, 139–150, doi:10.5194/hess-10-139-2006, 2006.

Fenicia, F., McDonnell, J., and Savenije, H. H. G.: Learning from model improvement: On the contribution of complementary data to process understanding, Water Resour. Res., 44, W06419, doi:10.1029/2007WR006386, 2008a.

Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: Understanding catchment behavior through stepwise model concept improvement, Water Resour. Res., 44, W01402, doi:10.1029/2006WR005563, 2008b.

Fenicia, F., Kavetski, D., and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47, W11510, doi:10.1029/2010WR010174, 2011.

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

Flügel, W.-A.: Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany, Hydrol. Process., 9, 423–436, doi:10.1002/hyp.3360090313, 1995.

Freer, J., McDonnell, J. J., Beven, K. J., Peters, N. E., Burns, D. A., Hooper, R. P., Aulenbach, B., and Kendall, C.: The role of bedrock topography on subsurface storm flow, Water Resour. Res., 38, 1269, doi:10.1029/2001WR000872, 2002.

Freer, J., Beven, K., and Peters, N.: Multivariate seasonal period model rejection within the generalised likelihood uncertainty estimation procedure, Water Sci. Appl., 6, 69–87, doi:10.1029/WS006p0069, 2003.

Freer, J., McMillan, H., McDonnell, J., and Beven, K.: Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures, J. Hydrol., 291, 254–277, doi:10.1016/j.jhydrol.2003.12.037, 2004.

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

Gascuel-Odoux, C., Aurousseau, P., Durand, P., Ruiz, L., and Molenat, J.: The role of climate on inter-annual variation in stream nitrate fluxes and concentrations, Sci. Total Environ., 408, 5657–5666, doi:10.1016/j.scitotenv.2009.05.003, 2010.

Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: Hydrological landscape classification: investigating the performance of HAND based landscape classifications in a central European meso-scale catchment, Hydrol. Earth Syst. Sci., 15, 3275–3291, doi:10.5194/hess-15-3275-2011, 2011.

Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: An approach to identify time consistent model parameters: sub-period calibration, Hydrol. Earth Syst. Sci., 17, 149–161, doi:10.5194/hess-17-149-2013, 2013.

Gharari, S., Shafiei, M., Hrachowitz, M., Kumar, R., Fenicia, F., Gupta, H. V., and Savenije, H. H. G.: A constraint-based search algorithm for parameter identification of environmental models, Hydrol. Earth Syst. Sci., 18, 4861–4870, doi:10.5194/hess-18-4861-2014, 2014.

Grayson, R. and Blöschl, G.: Spatial patterns in catchment hydrology: observations and modelling, in: Chapter 14 Summary of pattern comparison and concluding remarks, Cambridge University Press, Cambridge, 355–367, 2000.

Grayson, R. B., Moore, I. D., and McMahon, T. A.: Physically based hydrologic modeling: 1. A terrain-based model for investigative purposes, Water Resour. Res., 28, 2639–2658, doi:10.1029/92WR01258, 1992.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resour. Res., 34, 751–763, doi:10.1029/97WR03495, 1998.

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, doi:10.1002/hyp.6989, 2008.

Gupta, V. K. and Sorooshian, S.: Uniqueness and observability of conceptual rainfall-runoff model parameters: The percolation process examined, Water Resour. Res., 19, 269–276, doi:10.1029/WR019i001p00269, 1983.

Hamon, W. R.: Estimating potential evapotranspiration, J. Hydraul. Div., 87, 107–120, 1961.

He, Z., Tian, F., Hu, H. C., Gupta, H. V., and Hu, H. P.: Diagnostic calibration of a hydrological model in an alpine area, Hydrol. Earth Syst. Sci. Discuss., 11, 1253–1300, doi:10.5194/hessd-11-1253-2014, 2014.

Hewlett, J. D.: Soil moisture as a source of base flow from steep mountain watersheds, US Department of Agriculture, Forest Service, Southeastern Forest Experiment Station, Asheville, North Carolina, USA, 1961.

Hrachowitz, M., Soulsby, C., D., T., Dawson, J. J. C., and Malcolm, I. A.: Regionalization of Transit Time Estimates in montane catchments by integrating landscape controls, Water Resour. Res., 45, W05421, doi:10.1029/2008WR007496, 2009.

Hrachowitz, M., Savenije, H., Bogaard, T. A., Tetzlaff, D., and Soulsby, C.: What can flux tracking teach us about water age distribution patterns and their temporal dynamics?, Hydrol. Earth Syst. Sci., 17, 533–564, doi:10.5194/hess-17-533-2013, 2013a.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V.,

Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB) – a review, Hydrolog. Sci. J., 58, 1–58, doi:10.1080/02626667.2013.803183, 2013b.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G., and Gascuel-Odoux, C.: Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, Water Resour. Res., 50, 7445–7469, doi:10.1002/2014WR015484, 2014.

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E., and Marshall, L. A.: Hydrologic connectivity between landscapes and streams: transferring reach- and plot-scale understanding to the catchment scale, Water Resour. Res., 45, W04428, doi:10.1029/2008WR007225, 2009.

Kapangaziwiri, E., Hughes, D., and Wagener, T.: Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa, Hydrolog. Sci. J., 57, 1000–1019, doi:10.1080/02626667.2012.690881, 2012.

Kavetski, D., Fenicia, F., and Clark, M. P.: Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment, Water Resour. Res., 47, W05501, doi:10.1029/2010WR009525, 2011.

Khu, S. T., Madsen, H., and de Pierro, F.: Incorporating multiple observations for distributed hydrologic model calibration: An approach using a multi-objective evolutionary algorithm and clustering, Adv. Water Resour., 31, 1387–1398, doi:10.1016/j.advwatres.2008.07.011, 2008.

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, W03S04, doi:10.1029/2005WR004362, 2006.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrolog. Sci. J., 31, 13–24, doi:10.1080/02626668609491024, 1986.

Kling, H. and Gupta, H.: On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability, J. Hydrol., 373, 337–351, doi:10.1016/j.jhydrol.2009.04.031, 2009.

Knudsen, J., Thomsen, A., and Refsgaard, J. C.: WATBAL A Semi-Distributed, Physically Based Hydrological Modelling System, Nord. Hydrol., 17, 347–362, 1986.

Koren, V., Smith, M., Wang, D., and Zhang, Z.: Use of soil property data in the derivation of conceptual rainfall-runoff model parameters, in: 15th Conference on Hydrology, Paper vol. 2, American Meteorological Society, Long Beach, 2000.

Koren, V., Smith, M., and Duan, Q.: Use of a priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models, Water Sci. Appl., 6, 239–254, 2003.

Koren, V., Moreda, F., and Smith, M.: Use of soil moisture observations to improve parameter consistency in watershed calibration, Phys. Chem. Earth A/B/C, 33, 1068–1080, doi:10.1016/j.pce.2008.01.003, 2008.

Krcho, J.: Modelling of georelief and its geometrical structure using DTM: positional and numerical accuracy, Q111 Vydavatel Stvo, 2001.

Kumar, R., Samaniego, L., and Attinger, S.: The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics, J. Hydrol., 392, 54–69, doi:10.1016/j.jhydrol.2010.07.047, 2010.

Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resour. Res., 491, 360–379, doi:10.1029/2012WR012195, 2013a.

Kumar, R., Livneh, B., and Samaniego, L.: Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme, Water Resour. Res., 49, 5700–5714, doi:10.1002/wrcr.20431, 2013b.

Kuzmin, V., Seo, D.-J., and Koren, V.: Fast and efficient optimization of hydrologic model parameters using a priori estimates and stepwise line search, J. Hydrol., 353, 109–128, doi:10.1016/j.jhydrol.2008.02.001, 2008.

Laaha, G., Skøien, J., and Blöschl, G.: Spatial prediction on river networks: comparison of top-kriging with regional regression, Hydrol. Process., 28, 315–324, doi:10.1002/hyp.9578, 2014.

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., and Arheimer, B.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, Hydrol. Res., 41, 295–319, doi:10.2166/nh.2010.007, 2010.

Liu, M., Bárdossy, A., Li, J., and Jiang, Y.: Physically-based modeling of topographic effects on spatial evapotranspiration and soil moisture patterns through radiation and wind, Hydrol. Earth Syst. Sci., 16, 357–373, doi:10.5194/hess-16-357-2012, 2012.

Madsen, H.: Automatic calibration of a conceptual rainfall–runoff model using multiple objectives, J. Hydrol., 235, 276–288, doi:10.1016/S0022-1694(00)00279-1, 2000.

Matgen, P., Fenicia, F., Heitz, S., Plaza, D., de Keyser, R., Pauwels, V. R., Wagner, W., and Savenije, H.: Can ASCAT-derived soil wetness indices reduce predictive uncertainty in well-gauged areas? A comparison with in situ observed soil moisture in an assimilation application, Adv. Water Resour., 44, 49–65, doi:10.1016/j.advwatres.2012.03.022, 2012.

McDonnell, J. J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M. L., Selker, J., and Weiler, M.: Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, Water Resour. Res., 43, W07301, doi:10.1029/2006WR005467, 2007.

McGlynn, B. L. and McDonnell, J. J.: Quantifying the relative contributions of riparian and hillslopezones to catchment runoff, Water Resour. Res., 39, 1310, doi:10.1029/2003WR002091, 2003.

McGlynn, B. L., McDonnell, J. J., Seibert, J., and Kendall, C.: Scale effects on headwater catchment runoff timing, flow sources, and groundwater-streamflow relations, Water Resour. Res., 40, W07504, doi:10.1029/2003WR002494, 2004.

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, doi:10.1029/2004WR003657, 2005.

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, doi:10.1002/hyp.7841, 2011.

McNamara, J. P., Chandler, D., Seyfried, M., and Achet, S.: Soil moisture states, lateral flow, and streamflow generation in a semi-arid, snowmelt-driven catchment, Hydrol. Process., 19, 4023–4038, doi:10.1002/hyp.5869, 2005.

Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, J. Hydrol., 287, 95–123, 2004.

Milne, G.: Some suggested units of classification and mapping particularly for East African soils, Soil Res., 4, 183–198, 1935.

Molénat, J., Gascuel-Odoux, C., Davy, P., and Durand, P.: How to model shallow water-table depth variations: the case of the Kervidy–Naizin catchment, France, Hydrol. Process., 19, 901–920, doi:10.1002/hyp.5546, 2005.

Nalbantis, I., Efstratiadis, A., Rozos, E., Kopsiafti, M., and Koutsoyiannis, D.: Holistic versus monomeric strategies for hydrological modelling of human-modified hydrosystems, Hydrol. Earth Syst. Sci., 15, 743–758, doi:10.5194/hess-15-743-2011, 2011.

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

Nobre, A. D., Cuartas, L. A., Hodnett, M., Rennó, C. D., Rodrigues, G., Silveira, A., Waterloo, M., and Saleska, S.: Height above the Nearest Drainage, a hydrologically relevant new terrain model, J. Hydrol., 404, 13–29, 2011.

Ott, B. and Uhlenbrook, S.: Quantifying the impact of land-use changes at the event and seasonal time scale using a process-oriented catchment model, Hydrol. Earth Syst. Sci., 8, 62–78, doi:10.5194/hess-8-62-2004, 2004.

Oudin, L., Andréassian, 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, doi:10.1029/2007WR006240, 2008.

Parajka, J. and Blöschl, G.: Spatio-temporal combination of MODIS images – potential for snow cover mapping, Water Resour. Res., 44, W03406, doi:10.1029/2007WR006204, 2008.

Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, Hydrol. Process., 21, 435–446, doi:10.1002/hyp.6253, 2007.

Park, S. and van de Giesen, N.: Soil-landscape delineation to define spatial sampling domains for hillslope hydrology, J. Hydrol., 295, 28–46, doi:10.1016/j.jhydrol.2004.02.022, 2004.

Perrin, C., Andréassian, V., Serna, C. R., Mathevet, T., and Moine, N. L.: Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments, Water Resour. Res., 44, W08447, doi:10.1029/2007WR006579, 2008.

Pike, J.: The estimation of annual run-off from meteorological data in a tropical climate, J. Hydrol., 2, 116–123, doi:10.1016/0022-1694(64)90022-8, 1964.

Pokhrel, P., Gupta, H. V., and Wagener, T.: A spatial regularization approach to parameter estimation for a distributed watershed model, Water Resour. Res., 44, W12419, doi:10.1029/2007WR006615, 2008.

Pokhrel, P., Yilmaz, K. K., and Gupta, H. V.: Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures, J. Hydrol., 418–419, 49–60, doi:10.1016/j.jhydrol.2008.12.004, 2012.

Pomeroy, J. W., Gray, D. M., Brown, T., Hedstrom, N. R., Quinton, W. L., Granger, R. J., and Carey, S. K.: The cold regions hydrological model: a platform for basing process representation and model structure on physical evidence, Hydrol. Process., 21(, 2650–2667, doi:10.1002/hyp.6787, 2007.

Razavi, S. and Tolson, B. A.: An efficient framework for hydrologic model calibration on long data periods, Water Resour. Res., 49, 8418–8431, doi:10.1002/2012WR013442, 2013.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., and DMIP Participants: Overall distributed model intercomparison project results, J. Hydrol., 298, 27–60, doi:10.1016/j.jhydrol.2004.03.031, 2004.

Refsgaard, J. C. and Knudsen, J.: Operational Validation and Intercomparison of Different Types of Hydrological Models, Water Resour. Res., 32, 2189–2202, doi:10.1029/96WR00896, 1996.

Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., and Waterloo, M. J.: HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia, Remote Sens. Environ., 112, 3469–3481, doi:10.1016/j.rse.2008.03.018, 2008.

Rouhani, H., Willems, P., Wyseure, G., and Feyen, J.: Parameter estimation in semi-distributed hydrological catchment modelling using a multi-criteria objective function, Hydrol. Process., 21, 2998–3008, doi:10.1002/hyp.6527, 2007.

Samaniego, L., Bárdossy, A., and Kumar, R.: Streamflow prediction in ungauged catchments using copula-based dissimilarity measures, Water Resour. Res., 46, W02506, doi:10.1029/2008WR007695, 2010.

Savenije, H. H. G.: HESS Opinions "Topography driven conceptual modelling (FLEX-Topo)", Hydrol. Earth Syst. Sci., 14, 2681–2692, doi:10.5194/hess-14-2681-2010, 2010.

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, doi:10.5194/hess-15-2895-2011, 2011.

Scherrer, S. and Naef, F.: A decision scheme to indicate dominant hydrological flow processes on temperate grassland, Hydrol. Process., 17, 391–401, doi:10.1002/hyp.1131, 2003.

Scherrer, S., Naef, F., Faeh, A. O., and Cordery, I.: Formation of runoff at the hillslope scale during intense precipitation, Hydrol. Earth Syst. Sci., 11, 907–922, doi:10.5194/hess-11-907-2007, 2007.

Schmocker-Fackel, P., Naef, F., and Scherrer, S.: Identifying runoff processes on the plot and catchment scale, Hydrol. Earth Syst. Sci., 11, 891–906, doi:10.5194/hess-11-891-2007, 2007.

Seibert, J.: Reliability of model predictions outside calibration conditions, Nord. Hydrol., 34, 477–492, 2003.

Seibert, J. and Beven, K. J.: Gauging the ungauged basin: how many discharge measurements are needed?, Hydrol. Earth Syst. Sci., 13, 883–892, doi:10.5194/hess-13-883-2009, 2009.

Seibert, J. and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, Water Resour. Res., 38, 1241, doi:10.1029/2001WR000978, 2002.

Semenova, O., Lebedeva, L., and Vinogradov, Y.: Simulation of subsurface heat and water dynamics, and runoff generation in mountainous permafrost conditions, in the Upper Kolyma River basin, Russia, Hydrogeol. J., 21, 107–119, doi:10.1007/s10040-012-0936-1, 2013.

Sidle, R. C., Noguchi, S., Tsuboyama, Y., and Laursen, K.: A conceptual model of preferential flow systems in forested hillslopes: evidence of self-organization, Hydrol. Process., 15, 1675–1692, doi:10.1002/hyp.233, 2001.

Singh, S. K. and Bárdossy, A.: Calibration of hydrological models on hydrologically unusual events, Adv. Water Resour., 38, 81–91, doi:10.1016/j.advwatres.2011.12.006, 2012.

Sivapalan, M.: Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, John Wiley & Sons, Ltd, New York, USA, doi:10.1002/0470848944.hsa012, 2006.

Sivapalan, M., Blöschl, G., Zhang, L., and Vertessy, R.: Downward approach to hydrologic prediction, Hydrol. Process., 17, 2101–2111, doi:10.1002/hyp.1425, 2003.

Son, K. and Sivapalan, M.: Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, Water Resour. Res., 43, W01415, doi:10.1029/2006WR005032, 2007.

Spence, C.: A Paradigm Shift in Hydrology: Storage Thresholds Across Scales Influence Catchment Runoff Generation, Geogr. Compass, 4, 819–833, doi:10.1111/j.1749-8198.2010.00341.x, 2010.

Sutanudjaja, E., de Jong, S., van Geer, F., and Bierkens, M.: Using {ERS} spaceborne microwave soil moisture observations to predict groundwater head in space and time, Remote Sens. Environ., 138, 172–188, doi:10.1016/j.rse.2013.07.022, 2013.

Szilagyi, J., Rundquist, D. C., Gosselin, D. C., and Parlange, M. B.: NDVI relationship to monthly evaporation, Geophys. Res. Lett., 25, 1753–1756, doi:10.1029/98GL01176, 1998.

Tikhonov, A.: Solution of incorrectly formulated problems and the regularization method, in: Soviet Math. Dokl., vol. 5, Moscow, Russia, 1035–1038, 1963.

Tromp-van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis, Water Resour. Res., 42, W02411, doi:10.1029/2004WR003800, 2006a.

Tromp-van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope, Water Resour. Res., 42, W02410, doi:10.1029/2004WR003778, 2006b.

Turc, L.: Le bilan d'eau des sols. Relation entre la precipitation, l'evaporation et l'ecoulement, Ann. Agron., 5, 491–595, 1954.

Uhlenbrook, S., Roser, S., and Tilch, N.: Hydrological process representation at the meso-scale: the potential of a distributed, conceptual catchment model, J. Hydrol., 291, 278–296, doi:10.1016/j.jhydrol.2003.12.038, 2004.

Vaché, K. and McDonnell, J.: A process-based rejectionist framework for evaluating catchment runoff model structure, Water Resour. Res., 42, W02409, doi:10.1029/2005WR004247, 2006.

Vannametee, E., Babel, L., Hendriks, M., Schuur, J., de Jong, S., Bierkens, M., and Karssenberg, D.: Semi-automated mapping of landforms using multiple point geostatistics, Geomorphology, 221, 298–319, doi:10.1016/j.geomorph.2014.05.032, 2014.

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, doi:10.5194/hess-17-2263-2013, 2013.

Vinogradov, Y. B., Semenova, O. M., and Vinogradova, T. A.: An approach to the scaling problem in hydrological modelling: the deterministic modelling hydrological system, Hydrol. Process., 25, 1055–1073, doi:10.1002/hyp.7901, 2011.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resour. Res., 39, 1214, doi:10.1029/2002WR001746, 2003.

Wagener, T. and Montanari, A.: Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, Water Resour. Res., 47, W06301, doi:10.1029/2010WR009469, 2011.

Wagener, T. and Wheater, H. S.: Parameter estimation and regionalization for continuous rainfall–runoff models including uncertainty, J. Hydrol., 320, 132–154, doi:10.1016/j.jhydrol.2005.07.015, 2006.

Weiler, M., McGlynn, B., McGuire, K., and McDonnell, J.: How does rainfall become runoff? A combined tracer and runoff transfer function approach, Water Resour. Res., 39, 1315, doi:10.1029/2003WR002331, 2003.

Winsemius, H. C., Savenije, H. H. G., and Bastiaanssen, W. G. M.: Constraining model parameters on remotely sensed evaporation: justification for distribution in ungauged basins?, Hydrol. Earth Syst. Sci., 12, 1403–1413, doi:10.5194/hess-12-1403-2008, 2008.

Winsemius, H. C., Schaefli, B., Montanari, A., and Savenije, H. H. G.: On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, Water Resour. Res., 45, W12422, doi:10.1029/2009WR007706, 2009.

Winter, T. C.: The Concept OF Hydrologic Landscapes, J. Am. Water Resour. Assoc., 37, 335–349, doi:10.1111/j.1752-1688.2001.tb00973.x, 2001.

Wolock, D. M., Winter, T. C., and McMahon, G.: Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses, Environ. Manage., 34, S71–S88, doi:10.1007/s00267-003-5077-9, 2004.

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

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resour. Res., 44, W09417, doi:10.1029/2007WR006716, 2008.

Zehe, E. and Sivapalan, M.: Threshold behaviour in hydrological systems as (human) geo-ecosystems: manifestations, controls, implications, Hydrol. Earth Syst. Sci., 13, 1273–1297, doi:10.5194/hess-13-1273-2009, 2009.

Zhang, Z., Wagener, T., Reed, P., and Bhushan, R.: Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization, Water Resour. Res., 44, W00B04, doi:10.1029/2008WR006833, 2008.