Hydrol. Earth Syst. Sci., 19, 105–123, 2015
www.hydrol-earth-syst-sci.net/19/105/2015/
doi:10.5194/hess-19-105-2015
© Author(s) 2015. CC Attribution 3.0 License.

Hydrology and Earth System Sciences
Open Access

Hydrological hysteresis and its value for assessing process consistency in catchment conceptual models

O. Fovet[1,2], L. Ruiz[1,2], M. Hrachowitz[3], M. Faucheux[1,2], and C. Gascuel-Odoux[1,2]

[1]INRA, UMR1069 SAS, 65 route de Saint Brieuc, 35042 Rennes, France
[2]Agrocampus Ouest, UMR1069 SAS, 65 route de Saint Brieuc, 35042 Rennes, France
[3]Delft University of Technology, Water Resources Section, Faculty of Civil Engineering and Applied Geosciences, Stevinweg 1, 2600 GA Delft, the Netherlands

Correspondence to: O. Fovet (ophelie.fovet@rennes.inra.fr)

Received: 10 April 2014 – Published in Hydrol. Earth Syst. Sci. Discuss.: 28 May 2014
Revised: 28 November 2014 – Accepted: 2 December 2014 – Published: 7 January 2015

Abstract. While most hydrological models reproduce the general flow dynamics, they frequently fail to adequately mimic system-internal processes. In particular, the relationship between storage and discharge, which often follows annual hysteretic patterns in shallow hard-rock aquifers, is rarely considered in modelling studies. One main reason is that catchment storage is difficult to measure, and another one is that objective functions are usually based on individual variables time series (e.g. the discharge). This reduces the ability of classical procedures to assess the relevance of the conceptual hypotheses associated with models.

We analysed the annual hysteric patterns observed between stream flow and water storage both in the saturated and unsaturated zones of the hillslope and the riparian zone of a headwater catchment in French Brittany (Environmental Research Observatory ERO AgrHys (ORE AgrHys)). The saturated-zone storage was estimated using distributed shallow groundwater levels and the unsaturated-zone storage using several moisture profiles. All hysteretic loops were characterized by a hysteresis index. Four conceptual models, previously calibrated and evaluated for the same catchment, were assessed with respect to their ability to reproduce the hysteretic patterns.

The observed relationship between stream flow and saturated, and unsaturated storages led us to identify four hydrological periods and emphasized a clearly distinct behaviour between riparian and hillslope groundwaters. Although all the tested models were able to produce an annual hysteresis loop between discharge and both saturated and unsaturated storage, the integration of a riparian component led to overall improved hysteretic signatures, even if some misrepresentation remained. Such a system-like approach is likely to improve model selection.

1 Introduction

Rainfall-runoff models are tools that mimic the low-pass filter properties of catchments. Specifically, they aim at reproducing observed stream flow time series by routing time series of meteorological drivers through a sequence of mathematically formalized processes that allow a temporal dispersion of the input signals in a way that is consistent with the modeller's conception of how the system functions. The core of most models, in particular in temperate, humid climates dominated by some type of subsurface flow, is a series of storage–discharge functions that, in the most general terms, express system output (i.e. discharge and evaporation) as a function of the system state (i.e. storage), thereby generating a signal that is attenuated and lagged with respect to the input signal (i.e. precipitation).

However, modelling efforts on the catchment scale typically face the problem that, on that scale, neither integrated internal fluxes nor the integrated storage and the partitioning between different storage components at a given time can be easily observed within limited uncertainty. Indeed, indicators of catchment storage such as groundwater levels and soil water content can be highly variable in space and exhibit heterogeneous spatio-temporal dynamics. While spatial aggregation of storage estimates (e.g. catchment averages) in lumped models may lead to a loss of crucial information

and thus to overly simplistic representations of reality, allowing for the explicit incorporation of spatial storage heterogeneity in (semi-)distributed models may prove elusive in the presence of data error and the frequent absence of detailed spatial knowledge of the properties of the flow domain. A time series of groundwater table levels from a single piezometer is not representative of the behaviour of the groundwater, even on the hillslope scale; therefore, it is difficult to link it with either a reservoir volume simulated by a lumped model or an average water table level of a grid point simulated by a fully distributed model. These problems were recently addressed in some studies that intended to assess catchment storage using all available data (McNamara et al., 2011; Tetzlaff et al., 2011) and showing the importance of this storage in thresholds observed in the response of discharge to precipitation in catchments. For example, Spence (2010) argued that the observed nonlinear relationships between stream flow and catchment storage (i.e. no unique storage–discharge relations) are the manifestation of thresholds occurring in catchment runoff generation. Thus, depending on the structure of the system, storage–discharge dynamics can exhibit hysteretic patterns, i.e. the system response depends on the history and the memory of the system (e.g. Everett and Whitton, 1952; Ali et al., 2011; Gabrielli et al., 2012; Haught and van Meerveld, 2011). Andermann et al. (2012) found a hysteretic relationship between precipitation and discharge in both glaciated and unglaciated catchments in the Himalaya Mountains that was shown to be due to groundwater storage rather than to snow or glacier melt. Hrachowitz et al. (2013a), demonstrating the presence of hysteresis in the distribution of water ages, highlighted the importance of an adequate characterization of all system-relevant internal states at a given time to predict the system response within limited uncertainty as flow can be generated from different system components depending on the wetness state of the system.

In catchment-scale rainfall-runoff models, the need for calibration remains inevitable (Beven, 2001) due to the presence of data errors (e.g. Beven, 2013) and to the typically oversimplified process representations (e.g. Gupta et al., 2012). In spite of their comparatively high degrees of freedom, such models are frequently evaluated only against one single observed output variable, e.g. stream flow. Although the calibrated models may then adequately reproduce the output variable, model equifinality (e.g. Savenije, 2001) will lead to many apparently feasible solutions that do not sufficiently well reproduce system-internal dynamics as they are mere artefacts of the mathematical optimization process rather than suitable representations of reality (Gharari et al., 2013; Hrachowitz et al., 2013b; Andréassian et al., 2012; Beven, 2006; Kirchner, 2006). The realisation that there is a need for multivariable and multiobjective model evaluation strategies to identify and discard solutions that do not satisfy all evaluation criteria applied is therefore gaining ground (e.g. Freer et al., 1996; Gupta et al., 1998, 2008, Gascuel-Odoux et al., 2010) as this will eventually lead to models that are not only capable of reproducing the observed output variables (e.g. stream flow) but that also represent the system-internal dynamics in a more realistic way (Euser et al., 2013). The value of such multivariable and/or multiobjective evaluation strategies has been demonstrated in the past, for example using groundwater levels (e.g. Fenicia et al., 2008; Molénat et al., 2005, Giustolisi and Simeone, 2006; Freer et al., 2004; Seibert, 2000; Lamb et al., 1998), soil moisture (Kampf and Burges, 2007; Parajka et al., 2006), saturated-area extension (Franks et al., 1998), snow cover patterns (e.g. Nester et al., 2012), remotely sensed evaporation, (e.g. Mohamed et al., 2006; Winsemius et al., 2008), stream flow at subcatchment outlets (e.g. Moussa et al., 2007) and even water quality data such as, e.g., chloride concentrations (Hrachowitz et al., 2011), atmospheric tracers (Molénat et al., 2013) or nitrates and sulfate concentrations (Hartmann et al., 2013a) and water isotopes such as δ^{18}O (Hartmann et al., 2013b). However, most studies using multiple response variables only evaluate them individually to identify Pareto-optimal solutions. This practice may result in the loss of critical information, such as the timing between the multiple variables. In other words it is conceivable that model calibration leads to Pareto-optimal solutions with adequate model performance for all variables while at the same time misrepresenting the dynamics between these variables. Instead, using a synthetic catchment property (Sivapalan et al., 2005) or a hydrological signature (Wagener and Montanari, 2011; Yadav et al., 2007), combining different variables into one function, may potentially serve as a instructive diagnostic tool, a calibration objective or even as a metric for catchment classification (Wagener, 2007).

Hysteretic patterns between hydrological variables are potentially good candidates to build such tools. The objective of this paper is to explore (i) the potential of using annual hysteric patterns observed between stream flow and water storage both in the saturated and unsaturated zones of the hillslope and of the riparian zone for characterizing the hydrological functioning of a small headwater catchment in French Brittany (Environmental Research Observatory ERO AgrHys (ORE AgrHys)), (ii) to which degree a suite of conceptual rainfall-runoff models with increasing complexity, which were calibrated and evaluated for this catchment in previous work using a flexible modelling framework (Hrachowitz et al., 2014), can reproduce the observed storage–discharge hysteresis and (iii) whether the use of the storage–discharge hysteresis can provide additional information for model diagnostics compared to traditional model evaluation metrics.

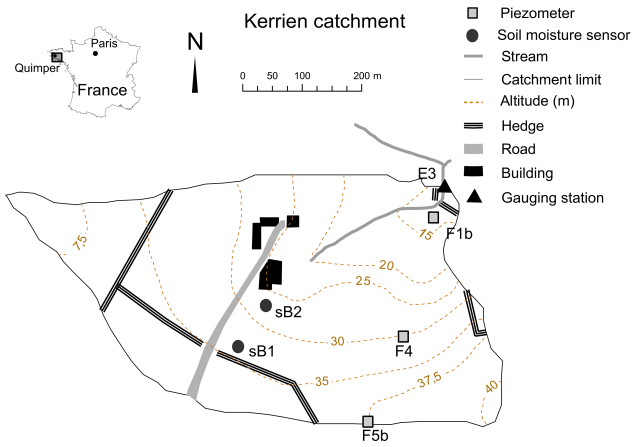


Figure 1. Study site in west Brittany (indicated by the square near Quimper) and location of the monitoring equipments. The weather station is located 500 m north of the catchment.

2 Materials and methods

2.1 Study sites

Kerrien (10.5 ha) is a headwater catchment located in southwestern Brittany (47°35′ N, 117°52′ E; see Fig. 1). Elevations range from 14 to 38 m a.s.l.; slopes are less than 8.5 %. The climate is oceanic, with a mean annual temperature of 11.9 °C with a minimum of 5.9 °C in winter and a maximum of 17.9 °C in summer. Mean annual rainfall over the period 1992–2012 is 1113 mm (±20 %) and mean annual Penman potential evapotranspiration (PET) is 700 mm (±4 %). Mean annual drainage is 360 mm (±60 %) at the outlet. There is a high water deficit in the annual budget almost every year due to underflows below the outlet (Ruiz et al., 2002). The catchment lies under granite (leucogranodiorite of Plomelin), the upper part of which is weathered from 1 to more than 20 m deep. Soils are mainly sandy loam with an upper horizon rich in organic matter; depths are between 40 and 90 cm. Soils are well drained except in the bottomlands, which represent 7 % of the total area. Agriculture dominates the land use, with 86 % of the total area covered by grassland, maize and wheat, none of them irrigated. The base flow index is about 80 to 90 %; thus, the hillslope aquifer is the main contributor to stream flow (Molénat et al., 2008; Ruiz et al., 2002). Both stream flow and shallow groundwater tables exhibit a strong annual seasonality in this catchment (Figs. 2 and 3a).

2.2 Data

Meteorological data were recorded in an automatic weather station (CIMEL, Fig. 1) which provides hourly rainfall and variables required to estimate daily Penman PET (net solar radiation, air and soil temperatures, wind speed and direction). Discharge was calculated from water level measurements at the outlet (Fig. 1) using a V-notch weir equipped with a shaft encoder with integrated data logger (OTT Thalimedes) and recorded every 10 min since 2000 (E3). Groundwater levels have been monitored every 15 min since 2001 in three piezometers – F1b, F4, and F5b (Fig. 1) – using vented pressure probe sensors (OTT Orpheus Mini).

Moisture in the unsaturated zone has been recorded every 30 min since July 2010 at seven depth (25, 55, 85, 125, 165, 215, and 265 cm) and at two locations (sB1 and sB2; Fig. 1), using capacitive probes which provide volumic humidity based on frequency domain reflectometry (EnvironScan SenteK). Due to technical problems, data are missing in December 2012 and January 2013, so only 2 complete water years were available (2010–2011 and 2011–2012). In summary, stream discharge water table levels were considered for the years 2002–2012, and soil moisture was considered for the years 2010–2012.

2.3 Catchment storage estimates

In order to obtain a proxy for the saturated-zone storage on the catchment scale, the time series of groundwater level were normalized between their minimal and maximal values over the 10 years of records so that the normalized value lies between 0 and 1. The resulting normalized variable exhibited very similar dynamics among all the piezometers (see Fig. 2a). However, the piezometer located in the riparian zone (F1b) exhibited variations at a higher frequency, especially during the winter. Therefore, in the following, we used the average of the normalized level in the two hillslope piezometers (F5b, F4) as a proxy for the hillslope groundwater storage dynamics and the normalized level in the riparian piezometer as a proxy for the riparian groundwater storage dynamics.

In order to obtain a proxy for the unsaturated-zone storage, moisture time series were also normalized using the minimal and maximal values observed in all the sensors of the two profiles over the 2 water years with complete records, setting the minimal value as 0 and the maximal value as 1. As the normalized unsaturated storage variables obtained followed very similar trends and dynamics, we used, in the following, an average of the normalized unsaturated-zone storage among all the measurement points (depths and profiles) (Fig. 2b). The two profiles are located on the upslope and downslope parts of the hillslope. Thus, we assumed that averaging their normalized values will allow us to build a proxy for the dynamics of the unsaturated-zone storage on the whole hillslope.

2.4 Hysteresis indexes

Studies on hysteretic relationships in catchments generally focus on qualitative descriptions of patterns associated with a cross-correlation analysis between the two variables (Frei et al., 2010; Hopmans and Bren, 2007; Jung et al., 2004; Salant et al., 2008; Schwientek et al., 2013; Spence et al., 2010;

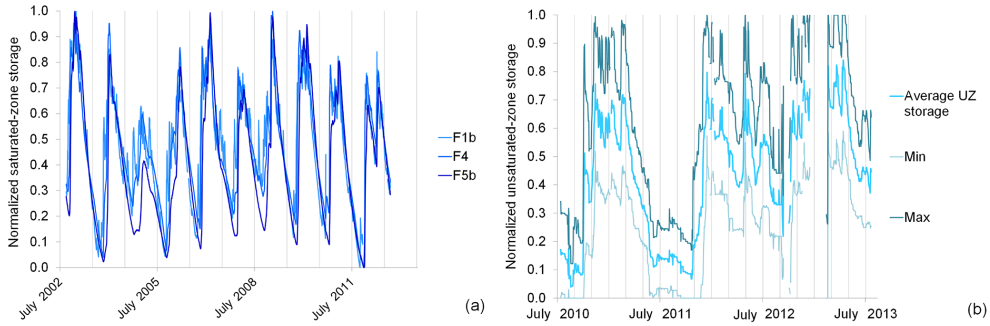


Figure 2. Normalized **(a)** groundwater levels for piezometers in the hillslope (F4 and F5b) and in the riparian zone (F1b) and **(b)** average, maximum and minimum unsaturated-zone storages for all the sensors in the two profiles in the Kerrien catchment.

Velleux et al., 2008). Some authors proposed a typology of hysteretic loops based on their rotational direction, curvature and trend to identify solute controls during storm events (Butturini et al., 2008; Evans and Davies, 1998). For storage–discharge hysteresis on the annual scale, this approach is not sufficient as the same type of hysteretic loop is likely to happen for almost all the years when a strong seasonality exists and its pattern is repeated across years. This is the case in our study, where seasonality of groundwater level and discharge showed a strong unimodal pattern for all years, except 2011–2012, which was bimodal (Figs. 2 and 3a). Moreover, a preliminary cross-correlation analysis revealed that storage and stream flow are strongly correlated, and the cross-correlation value is the greatest for a lag time of 0 days (results not shown).

Quantitative descriptions of the hysteretic loop are also found in the literature, and various ways of computing hysteresis indexes (HIs) have been proposed, for example using the relative difference between extreme concentration values (Butturini et al., 2008) or using the ratio of turbidity values in rising and falling limbs of the storm hydrograph at the midpoint discharge value (Lawler et al., 2006). The latter authors argue that computing HIs by using midpoint discharge usually allows avoiding the small convolutions which are frequently observed at both ends of the hysteretic loop.

In this paper, as the hydrological variables exhibit a strong annual unimodal cycle, we calculated the hysteresis index each year as the difference between water storages at the dates of midpoint discharge in the two phases of the hydrological year – during the recharge period (R) and the recession period (r), i.e. respectively before and after reaching the maximal discharge $Q_{\max}$ – as follows:

$$\begin{cases} \mathrm{HI} = S\left(t_{\mathrm{R,mid}}\right) - S\left(t_{\mathrm{r,mid}}\right) \\ Q\left(t_{\mathrm{R,mid}}\right) = Q_{\mathrm{mid}} \text{ and } t_{\mathrm{R,mid}} < t_{Q_{\max}} \\ Q\left(t_{\mathrm{r,mid}}\right) = Q_{\mathrm{mid}} \text{ and } t_{\mathrm{r,mid}} > t_{Q_{\max}}, \\ Q\left(t_{Q_{\max}}\right) = Q_{\max} \end{cases} \quad (1)$$

where $S(t)$ is the storage value at time t and $Q(t)$ the stream flow value at time t. The midpoint discharge Q_{mid} is defined as the mean value of discharge between Q_0, the initial value at the beginning of the hydrological year (October), and $Q_{\max}$, the maximal value reached during that year:

$$Q_{\mathrm{mid}} = \frac{Q_0 + Q_{\max}}{2}. \quad (2)$$

In order to reduce the impact of the quick variations of discharge or groundwater level due to individual storm events, we smoothed the time series using 7-day moving averages. The strong seasonal discharge cycle led us to identify two occurrences of Q_{mid} per year only – during the recharge period (t_{R}) and during the recession period (t_{r}) – while high and low stream flow values are taken several times per year as explained by Lawler et al. (2006). Computing the HI using the difference in storage was possible here because storage and stream flow values vary among years within a narrow range of magnitude, while Lawler et al. (2006) used the ratio because turbidity can differ by several orders of magnitude from one storm to the other. Computing the HI with the difference between the values of storage and not with their ratio allowed maintaining its sensitivity to the year-to-year variations of the width of the hysteretic loop. The difference in water storage dynamics in the unsaturated and saturated zones were approximated by the difference in normalized soil moisture content and by the difference in normalized groundwater level respectively.

The HI gives two types of information: (i) its sign indicates the direction of the loop (anticlockwise loop induces a negative value of the HI, whereas a clockwise loop leads to a positive value of the HI) and (ii) its absolute value is proportional to the magnitude of the hysteresis (i.e. the width of the hysteretic loop). The HI is a proxy for the importance of lag time response between variations in catchment storages (unsaturated and saturated) and stream discharge; its sign indicates whether storage reacts before or after the stream flow. Therefore, it can be used for comparing the capacity of the different models to reproduce to some extent the observed storage–discharge relationships. The normalization of the observed variables related to the storage (here either groundwater level or soil moisture) has no effect on the sign of the HI; the HI values are only being divided by the maximal amplitude ob-

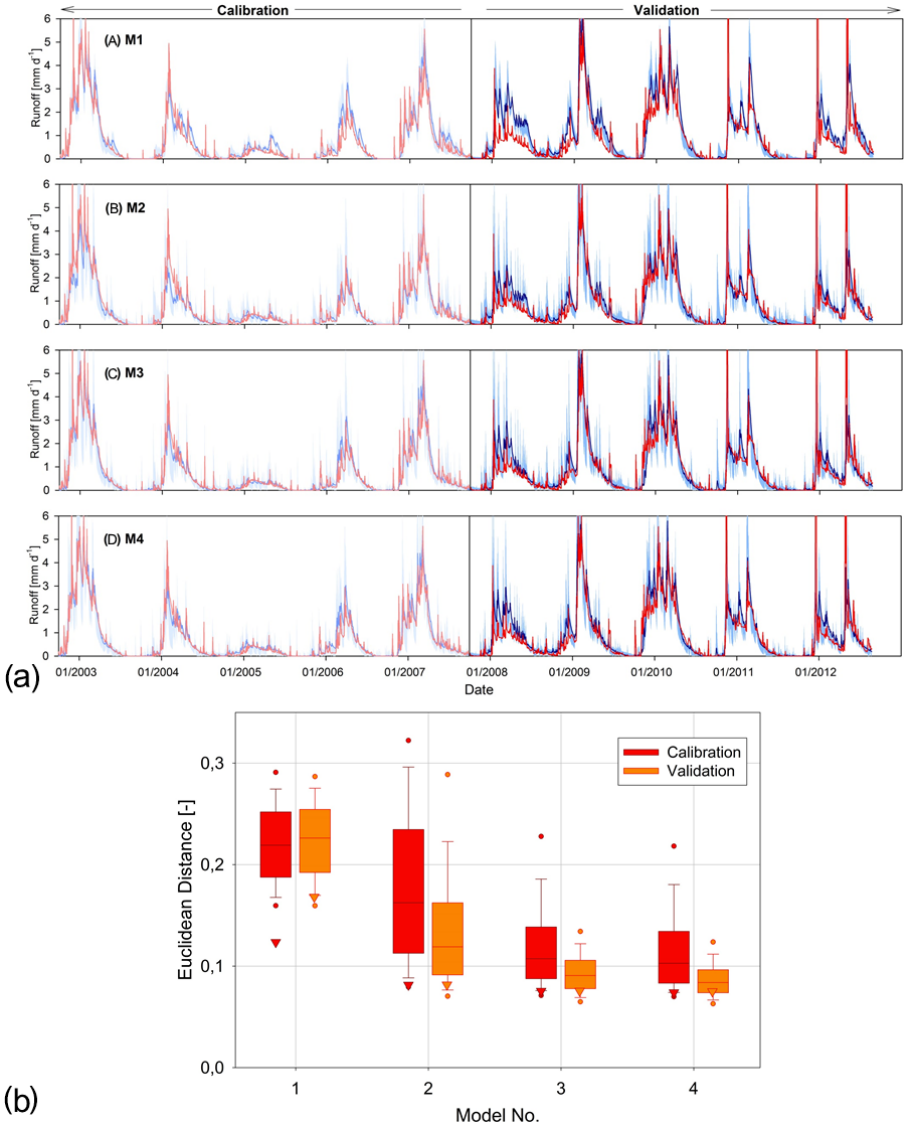


Figure 3. (a) Observed (red line) and modelled runoff for model set-ups **(A)** M1, **(B)** M2, **(C)** M3 and **(D)** M4 in calibration and independent evaluation (validation) periods. Modelled runoff shown as the most balanced solution (dark blue line) and the 5/95th uncertainty bounds (light blue shaded area). Adapted from Hrachowitz et al. (2014). **(b)** Overall model performance for all model set-ups (M1–M4) expressed as Euclidean distance from the "perfect model" computed from all calibration objectives and signatures with respect to calibration and validation periods. Triangles represent the optimal solution, i.e. the solution obtained from the parameter set with the lowest Euclidean distance during calibration. Box plots represent the Euclidean distance for the complete sets of all feasible solutions (the dots indicate 5/95th percentiles, the whiskers 10/90th percentiles and the horizontal central line the median). From Hrachowitz et al. (2014).

served in the storage during the whole period. Therefore, as long as the normalization is applied to the whole period (to all years and to both measurements and simulations), it does not affect the interpretation related to absolute values of the HI.

2.5 Models

In previous work, a range of conceptual models was calibrated and evaluated for the Kerrien catchment in a stepwise development using a flexible modelling framework (see Hrachowitz et al., 2014). This section aims at summarizing the results of this previous study as they are used as a basis for the present work. In this previous study, adopting a flexible stepwise modelling strategy, 11 models with increasing complexity, i.e. allowing for more process heterogeneity, were calibrated and evaluated for the study catchment. Four of these 11 models (hereafter referred to as M1 to M4; details given in Tables 1 and 2) were selected for the present work as they correspond to the sequence of model

Table 1. Water balance and state and flux equations of the models used.

Process	Water balance	Eq.	Models	Flux and state equations	Eq.	Models
Unsaturated zone	$\mathrm{d}S_\mathrm{U}/\mathrm{d}t = P - E_\mathrm{U} - R_\mathrm{F} - R_\mathrm{P} - R_\mathrm{S}$	(1.1)	M1, 2, 3, & 4	$E_\mathrm{U} = E_\mathrm{P}\mathrm{Min}\left(1, \frac{S_\mathrm{U}}{S_\mathrm{Umax,H}}\frac{1}{L_\mathrm{P}}\right)$	(1.2)	M1, 2, 3, & 4
				$R_\mathrm{U} = (1 - C_\mathrm{R})\,P$	(1.3)	M1, 2, 3, & 4
				$R_\mathrm{F} = C_\mathrm{R}\,(1 - C_\mathrm{P})\,P$	(1.4)	M1, 2, 3, & 4
				$R_\mathrm{S} = P_\mathrm{max}\left(\frac{S_\mathrm{U}}{S_\mathrm{Umax,H}}\right)$	(1.5)	M1, 2, 3, & 4
				$C_\mathrm{R} = \frac{1}{1+\exp\left(\frac{-S_\mathrm{U}/S_\mathrm{Umax,H}+0.5}{\beta}\right)}$	(1.6)	M1, 2, 3, & 4
Fast reservoir	$\mathrm{d}S_\mathrm{F}/\mathrm{d}t = R_\mathrm{F} - Q_\mathrm{F} - E_\mathrm{F}$	(2.1)	M1, 2, 3, & 4	$S_\mathrm{F,in} = S_\mathrm{F} + R_\mathrm{F}$	(2.2)	M1, 2, 3, & 4
				$Q_\mathrm{F} = S_\mathrm{F,in}\left(1 - e^{-k_\mathrm{F}t}\right)$	(2.3)	M1, 2, 3, & 4
				$E_\mathrm{F} = \mathrm{Min}\left(E_\mathrm{P} - E_\mathrm{U}, S_\mathrm{F,in} - Q_\mathrm{F}\right)$	(2.4)	M1, 2, 3, & 4
Slow reservoir	$\mathrm{d}S_\mathrm{S}/\mathrm{d}t = R_\mathrm{S} + R_\mathrm{P} - Q_\mathrm{S}$	(3.1)	M1	$S_\mathrm{S,in} = S_\mathrm{S} + R_\mathrm{S} + R_\mathrm{P}$	(3.2)	M1
				$Q_\mathrm{S} = S_\mathrm{S,in}\left(1 - e^{-k_\mathrm{S}t}\right)$	(3.3)	M1
	$\mathrm{d}S_\mathrm{s,a}/\mathrm{d}t = \begin{cases} S_\mathrm{s,a} - \mathrm{Max}\left(0, S_\mathrm{S,tot,out}\right), & S_\mathrm{S,tot,in} > 0 \\ 0, \quad S_\mathrm{S,tot,in} \le 0 \end{cases}$	(3.4)	M2, 3 & 4	$Q_\mathrm{S} = \mathrm{Max}\left(0, Q_\mathrm{S,tot} - Q_\mathrm{L,cst}\right)$	(3.7)	M2, 3 & 4
	$\mathrm{d}S_\mathrm{s,p}/\mathrm{d}t = \begin{cases} S_\mathrm{s,p} + \mathrm{Min}\left(0, S_\mathrm{S,tot,out}\right), & S_\mathrm{S,tot,in} > 0 \\ S_\mathrm{s,p} + S_\mathrm{S,tot,out}, \quad S_\mathrm{S,tot,in} \le 0 \end{cases}$	(3.5)	M2, 3 & 4	$S_\mathrm{S,tot,in} = S_\mathrm{s,a} + S_\mathrm{s,p} + R_\mathrm{S} + R_\mathrm{P}$	(3.8)	M2, 3 & 4
	$\mathrm{d}S_\mathrm{s}/\mathrm{d}t = \mathrm{d}S_\mathrm{s,a}/\mathrm{d}t + \mathrm{d}S_\mathrm{s,p}/\mathrm{d}t = R_\mathrm{S} + R_\mathrm{P} - Q_\mathrm{L,cst}$	(3.6)	M2, 3 & 4	$S_\mathrm{S,tot,out} = \begin{cases} S_\mathrm{S,tot,in}e^{-k_\mathrm{S}t} - \frac{Q_\mathrm{L,cst}}{k_\mathrm{S}}\left(1 - e^{-k_\mathrm{S}t}\right), & S_\mathrm{S,tot,in} > 0 \\ S_\mathrm{S,tot,in} - Q_\mathrm{L,cst}, \quad S_\mathrm{S,tot,in} \le 0 \end{cases}$	(3.9)	M2, 3 & 4
				$Q_\mathrm{L,cst} = \mathrm{constant}$	(3.10)	M2, 3 & 4
Unsaturated riparian zone	$\mathrm{d}S_\mathrm{U,R}/\mathrm{d}t = P - E_\mathrm{U,R} - R_\mathrm{R}$	(4.1)	M3 & 4	$E_\mathrm{U,R} = E_\mathrm{P}\mathrm{Min}\left(1, \frac{S_\mathrm{U,R}}{S_\mathrm{Umax,R}}\frac{1}{L_\mathrm{P}}\right)$	(4.2)	M3 & 4
				$R_\mathrm{R} = C_\mathrm{R,R}\,P$	(4.3)	M3 & 4
				$C_\mathrm{R,R} = \mathrm{Min}\left(1, \frac{S_\mathrm{U,R}}{S_\mathrm{Umax,R}}\right)$	(4.4)	M3
				$C_\mathrm{R,R} = \mathrm{Min}\left(1, \left(\frac{S_\mathrm{U,R}}{S_\mathrm{Umax,R}}\right)^{\beta_R}\right)$	(4.5)	M4
Riparian reservoir	$\mathrm{d}S_\mathrm{R}/\mathrm{d}t = R_\mathrm{R} - Q_\mathrm{R} - E_\mathrm{R}$	(5.1)	M3 & 4	$S_\mathrm{R,in} = S_\mathrm{R} + R_\mathrm{R}$	(5.2)	M3 & 4
				$Q_\mathrm{R} = S_\mathrm{R,in}\left(1 - e^{-k_R t}\right)$	(5.3)	M3 & 4
				$E_\mathrm{R} = \mathrm{Min}\left(E_\mathrm{P} - E_\mathrm{U,R}, S_\mathrm{R,in} - Q_\mathrm{R}\right)$	(5.4)	M3 & 4
Total runoff	$Q_\mathrm{T} = Q_\mathrm{F} + Q_\mathrm{S}$	(6.1)	M1 & 2			
Total evaporative fluxes	$Q_\mathrm{T} = (1 - f)\,(Q_\mathrm{F} + Q_\mathrm{S}) + f\,Q_\mathrm{R}$	(6.2)	M3 & 4			
	$E_\mathrm{A} = E_\mathrm{U} + E_\mathrm{F}$	(7.1)	M1 & 2			
	$E_\mathrm{A} = (1 - f)\,(E_\mathrm{U} + E_\mathrm{F}) + f\left(E_\mathrm{U,R} + E_\mathrm{R}\right)$	(7.2)	M3 & 4			

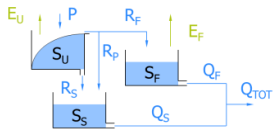
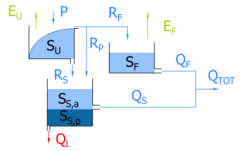
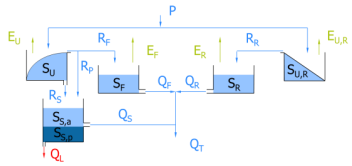
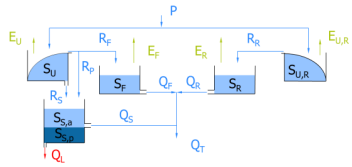
List of symbols: C_P – preferential recharge coefficient [-]; P – total precipitation [L T^{-1}]; S_F – storage in fast reservoir [L]; C_R – hillslope runoff generation coefficient [-]; E_F – transpiration fast responding reservoir [L T^{-1}]; S_R – storage in riparian reservoir [L]; $C_\mathrm{R,R}$ – riparian runoff generation coefficient [-]; E_P – potential evaporation [L T^{-1}]; S_S – storage in slow reservoir [L]; k_F – storage coefficient of fast reservoir [T^{-1}]; E_R – transpiration from riparian reservoir [L T^{-1}]; $S_\mathrm{S,a}$ – active storage in slow reservoir [L]; k_S – storage coefficient of slow reservoir [T^{-1}]; E_U – transpiration from unsaturated reservoir [L T^{-1}]; $S_\mathrm{S,p}$ – passive storage in slow reservoir [L]; k_L – storage coefficient for deep infiltration loss [T^{-1}]; $E_\mathrm{U,R}$ – transpiration unsaturated riparian reservoir [L T^{-1}]; $S_\mathrm{S,tot}$ – total storage in slow reservoir [L]; k_R – storage coefficient of riparian reservoir [T^{-1}]; Q_R – runoff from riparian reservoir [L T^{-1}]; S_U – storage in unsaturated reservoir [L]; f – proportion wetlands in the catchment [-]; Q_S – runoff from slow reservoir [L T^{-1}]; $S_\mathrm{S,tot,in}$ – total storage incoming in slow reservoir [L]; L_P – transpiration threshold [-]; Q_F – runoff from fast reservoir [L T^{-1}]; $S_\mathrm{S,tot,out}$ – total storage outcoming from slow reservoir [L]; P_max – percolation capacity [L T^{-1}]; $Q_\mathrm{L,const}$ – constant deep infiltration loss [L T^{-1}]; $S_\mathrm{Umax,H}$ – unsaturated hillslope storage capacity [L]; R_F – recharge of fast reservoir [L T^{-1}]; $S_\mathrm{Umax,R}$ – unsaturated riparian storage capacity [L]; R_P – preferential recharge of slow reservoir [L T^{-1}]; β – hillslope shape parameter for CR [–]; R_R – recharge of riparian reservoir [L T^{-1}]; β_R – riparian shape parameter for $C_\mathrm{R,R}$ [-]; R_S – recharge of slow reservoir [L T^{-1}]; R_U – infiltration into unsaturated reservoir [L T^{-1}].

architectures that provide the most significant performance improvements among the tested set-ups. As a starting point and benchmark, Model M1 with seven parameters, resembling many frequently used catchment models, such as HBV, was used (e.g. Bergström, 1995). The three boxes represent respectively an unsaturated zone, a slow-responding and a fast-responding reservoir. In Model M2, additional deep infiltration losses are integrated from the slow store to take into account the significant groundwater export to adjacent catchments in this study catchment as indicated by the observed long-term water balance (Ruiz et al., 2002). This is done by adding a second outlet together with a threshold to this storage to allow for continued groundwater export from a storage volume below the stream during zero-flow conditions, i.e. when the stream runs dry. As riparian zones frequently exhibit a distinct hydrological functioning (e.g. Molénat et al., 2005; Seibert et al., 2003), indicated in the study catchment by distinct response dynamics in the riparian piezometers (Martin et al., 2006), Models M3 and M4 additionally integrate a wetland/riparian zone component, composed of an unsaturated-zone store and a fast-responding reservoir, parallel to the other boxes. The riparian unsaturated zone generates flow using a linear function in M3 and a nonlinear function in M4. The complete set of water balance and constitutive model equations of the four models is listed in Table 1, while the model structures are schematized in Table 2.

2.6 Calibration and evaluation

This section is also a summary of the findings of Hrachowitz et al. (2014) that served as a basis for this study and does not consist of the results of the current study. The models have been calibrated for the period 1 October 2002–30 September 2007 after a 1-year warm-up period, using a multiobjective calibration strategy (e.g. Gupta et al., 1998) based on Monte Carlo sampling (10^7 realizations). The uniform prior parameter distributions used for M1–M4 are provided in Table 3. To reduce parameter and associated predictive uncertainty, the models were calibrated using a total of four calibration objective functions (see Table 4), i.e. the Nash–Sutcliffe efficiencies (Nash and Sutcliffe, 1970) for stream flow ($E_\mathrm{NS,}Q$), for the logarithm of the stream flow ($E_\mathrm{NS,log(}Q\mathrm{)}$) and for the flow duration curve ($E_\mathrm{NS,FDC}$) as well as the volumetric efficiency for stream flow ($V_\mathrm{E,}Q$; Criss and Winston, 2008). To facilitate a clearer assessment, the calibration objective functions ($n = 4$) were combined in a

Table 2. Model structures and parameters.

Model structure	Name	Parameters	Equations
	M1	k_F, k_S, P_{max}, L_P, $S_{Umax,H}$, β, C_P	(1.1) to (1.6); (2.1) to (2.4); (3.1) to (3.3); (6.1) & (7.1)
	M2	k_F, k_S, P_{max}, L_P, $S_{Umax,H}$, β, C_P, $Q_{L,cst}$	(1.1) to (1.6); (2.1) to (2.4); (3.4) to (3.10); (6.1) & (7.1)
	M3	k_F, k_S, P_{max}, L_P, $S_{Umax,H}$, $S_{Umax,R}$ β, C_P, $Q_{L,cst}$, k_R, f,	(1.1) to (1.6); (2.1) to (2.4); (3.4) to (3.10); (4.1) to (4.4); (5.1) to (5.4); (6.2) & (7.2)
	M4	k_F, k_S, P_{max}, L_P, $S_{Umax,H}$, β, C_P, $Q_{L,cst}$, f, k_R, $S_{Umax,R}$, β_R	(1.1) to (1.6); (2.1) to (2.4); (3.4) to (3.10); (4.1) to (4.3); (4.5); (5.1) to (5.4); (6.2) & (7.2)

single calibration metric: the Euclidean distance to the perfect model ($D_{E,cal}$; e.g. Hrachowitz et al., 2013a; Gascuel-Odoux et al., 2010):

$$D_E = \sqrt{\frac{(1-E_{NS,Q})^2 + (1-E_{NS,\log(Q)})^2 + (1-E_{V,Q})^2 + (1-E_{NS,FDC})^2}{n}}. \quad (3)$$

As mathematically optimal parameter sets are frequently hydrologically suboptimal, i.e. unrealistic (e.g. Beven, 2006), all parameter sets within the 4-dimensional space spanned by the calibration Pareto fronts, as approximated by the cloud of sample points, were retained as feasible.

The calibrated models were then evaluated against their respective skills to predict the system response with respect to a selection of 13 catchment signatures (described in Table 4) in a multicriteria posterior evaluation strategy. Figure 3 and Table 4 show the global performance D_E of the four models in terms of the Euclidean distance to the perfect model, constructed from all calibration objective functions and evaluation signatures. Model M1 provided good performance in calibration on the objective functions while its validation performances were considerably decreased. Its ability to reproduce the different signatures showed that it failed in particular to reproduce flow in wet periods (such as the evaluation period in Fig. 3a) and groundwater dynamics. Model M2 led to calibration performances slightly lower than model M1 but higher validation performances. The hydrological signatures simulated by M2 exhibited lower uncertainties both in validation and calibration periods because of a better simulation of low-flow conditions and groundwater dynamics. Model M3 provided similar performances to M2 for calibration and for validation but with clearly reduced uncertainty bounds. Overall signature reproduction was improved because of a clear improvement of low-flow and groundwater-related signatures even if performance in calibration objective functions remained lower than that for model M1. Model M4 exhibited similar performances to the previous models both in calibration and validation periods but a better performance for the whole set of signatures and lower uncertainties.

More details on the model calibration and evaluation with respect to hydrological signatures can be found in Hrachowitz et al. (2014; note that M1, M2 , M3 and M4 presented in this study correspond respectively to M1, M6, M8 and M11 in the original paper). Within the obtained range of parameter uncertainty, the types of simulated hysteresis patterns were not affected by the parameter values but only by the model structures. Note that we restricted the following analysis only to the optimal parameter set in each case, first for the sake of clarity and also because, at this stage, our interest was in assessing the ability of model structures to reproduce the observed general features in hysteresis pat-

terns and not in quantifying their performance in fitting the observations.

In the present work the sensitivity of the hysteresis indexes to parameter uncertainty is investigated by computing the HI values for the all sets of feasible parameters.

3 Results and discussion

3.1 Hysteretic pattern of the groundwater storage–discharge relationship

3.1.1 Observations in hillslope and riparian zones: saturated storage vs. flow

The 2-dimensional observed relationship between saturated storage in the hillslope (HSS) or in the riparian zone (RSS) and stream discharge (Q) for each year was hysteretic, highlighting the nonuniqueness of the response of discharge to storage depending on the initial conditions and a lag time between both variable dynamics, in particular during the recharge period, as illustrated in Fig. 4 for two contrasting water years.

The direction of the hysteretic loop was different depending on the topographic position of the piezometer: loops were always anticlockwise (leading to negative values of the HI) for the piezometer located at the top of the hillslope HSS-F5b(Q), mostly anticlockwise for the midslope piezometer HSS-F4(Q) and mostly clockwise (positive values of the HI) for the piezometer in the riparian zone RSS-F1b(Q) (Fig. 5).

In the riparian zone, storage at Q_{mid} was usually lower in the recession period than in the recharge period, especially in dry years, leading to a positive HI. This is due to the fact that the riparian groundwater level increased early at the beginning of the recharge period, before the stream discharge, due to the limited storage capacity of the narrow unsaturated layer in bottomlands, reinforced by groundwater ridging, which is linked to the extent of the capillary fringe. However, the hysteretic loops were narrow, and, for wet years, the storage value during the recession period occasionally exceeded the value in the recession period without modifying the general direction of the hysteresis when looking at the whole pattern (e.g. in 2003–2004, see Fig. 4a). When this occurred at the time of Q_{mid}, it led to a negative HI although absolute values remained small (Fig. 5)

The hillslope groundwater responded later than the stream, due to the deeper groundwater levels and higher unsaturated storage capacity (Rouxel et al., 2011), both introducing a time lag for the recharge and thus for the groundwater response. This led to negative values of the HI as groundwater levels in recession periods were higher than in recharge periods for the same level of discharge (in particular at Q_{mid}). The loops were also wider in the hillslope, leading to high absolute values of the HI (Fig. 5).

Table 3. Prior and posterior distribution of the model parameters.

	C_p [–]	f [–]	k_F [d^{-1}]	k_R [d^{-1}]	k_S [d^{-1}]	L_p [–]	$Q_{L,const}$ [mm d^{-1}]	P_{max} [mm d^{-1}]	$S_{S,p,max}$ [mm]	$S_{Umax,H}$ [mm]	$S_{Umax,R}$ [mm]	β [–]	β_R [–]
Prior distribution	0–1	0.1	0.025–1	0.05–2	0.001–0.05	0–1	0.37	0–4	0–2000	0–1500	0–750	0–100	0–2
Posterior distribution													
M1	0.12/0.63		0.042/0.094		0.031/0.049	0.00/0.07		0.03/0.29		637/1446		10.5/61.5	
M2	0.14/0.55		0.054/0.627		0.041*	0.05/0.34	0.37*	0.27/1.98		722/1461		2.4/36.9	
M3	0.15/0.64	0.1*	0.054/0.619	0.333/1.863	0.041*	0.04/0.27	0.37*	0.34/2.29		686/1442	132/725	13.6/69.7	
M4	0.19/0.64	0.1*	0.054/0.466	0.318/1.857	0.041*	0.04/0.27	0.37*	0.29/2.18		683/1444	120/730	13.0/69.2	0.13/1.86

* Fixed parameter values.

Table 4. Hydrological calibration criteria and evaluation signatures. The performance metrics include the Nash–Sutcliffe efficiency (E_{NS}), the volume error (E_{V}) and the relative error (E_{R}). For all variables and signatures, except for Q, Qlow and GW, the long-term averages were used.

$$E_{\mathrm{NS,X}} = 1 - \frac{\sum_{i=1:n}\sqrt{(X_{\mathrm{obs},i} - X_{\mathrm{sim},i})^2}}{\sum_{i=1:n}\sqrt{\left(X_{\mathrm{obs},i} - \frac{1}{n}\sum_{i=1:n} X_{\mathrm{obs},i}\right)^2}}$$

$$E_{\mathrm{V,X}} = 1 - \frac{\sum_{i=1:n}|X_{\mathrm{obs},i} - X_{\mathrm{sim},i}|}{\sum_{i=1:n} X_{\mathrm{obs},i}}$$

$$E_{\mathrm{R,X}} = \frac{X_{\mathrm{obs}} - X_{\mathrm{sim}}}{X_{\mathrm{obs}}}$$

	Variable/signature	ID	Performance metric	Reference
Calibration	Time series of flow	O1	$E_{\mathrm{NS},Q}$	Nash and Sutcliffe (1970)
		O2	$E_{\mathrm{NS,log}(Q)}$	
		O3	$E_{\mathrm{V},Q}$	Criss and Winston (2008)
	Flow duration curve	O4	$E_{\mathrm{NS,FDC}}$	Jothityangkoon et al. (2001)
			$D_{\mathrm{E,cal}}$	Schoups et al. (2005)
Evaluation	Flow during low-flow period	S1	$E_{\mathrm{NS},Q,\mathrm{low}}$	Freer et al. (2003)
	Groundwater dynamics[a]	S2	$E_{\mathrm{NS,GW}}$	Fenicia et al. (2008a)
	Flow duration curve low-flow period	S3	$E_{\mathrm{NS,FDC,low}}$	Yilmaz et al. (2008)
	Flow duration curve high-flow period	S4	$E_{\mathrm{NS,FDC,high}}$	Yilmaz et al. (2008)
	Groundwater duration curve[a]	S5	$E_{\mathrm{NS,GDC}}$	–
	Peak distribution	S6	$E_{\mathrm{NS,PD}}$	Euser et al. (2013)
	Peak distribution low-flow period	S7	$E_{\mathrm{NS,PD,low}}$	Euser et al. (2013)
	Rising-limb density	S8	$E_{\mathrm{R,RLD}}$	Shamir et al. (2005)
	Declining-limb density	S9	$E_{\mathrm{R,DLD}}$	Sawicz et al. (2011)
	Autocorrelation function of flow[b]	S10	$E_{\mathrm{NS,AC}}$	Montanari and Toth (2007)
	Lag-1 autocorrelation of high-flow period	S11	$E_{\mathrm{R,AC1,Q10}}$	Euser et al. (2013)
	Lag-1 autocorrelation of low-flow period	S12	$E_{\mathrm{R,AC1,low}}$	Euser et al. (2013)
	Runoff coefficient[c]	S13	$E_{\mathrm{R,RC}}$	Yadav et al. (2007)
			D_{E}	Schoups et al. (2005)

[a] Averaged and normalized time series data of the five piezometer were compared to normalized fluctuations in model state variable S_{S} (see Table 1). [b] Describing the spectral properties of a signal and thus the memory of the system, the observed and modelled autocorrelation functions with lags from 1 to 100 d where compared. [c] Note that in catchments without long-term storage changes and intercatchment groundwater flow, long-term average RC equals the long-term average 1-E_{A} (Table 1).

The intermediate behaviour of the midslope piezometer (F4), exhibiting varying patterns throughout the years, reflects the fact that the riparian zone extends spatially towards the hillslope and reaches a larger spatial extension during wet years.

Similar observations have been reported by other authors. For example, anticlockwise hysteresis between groundwater tables and discharge are observed by Gabrielli et al. (2012) in the Maimai catchment, while studies on riparian groundwater or river bank groundwater report clockwise hysteresis on the storm event scale (Frei et al., 2010; Jung et al., 2004). Similar patterns were also observed by Jung et al. (2004), who found that in the inner floodplain and in river bank piezometers, the hysteresis curve between the water table and river stage exhibits a synchronous response, while in the hillslope hysteresis, curves are relatively open as the water table is higher during the recession than during the rising limb.

3.1.2 Observations regarding hillslope: saturated and unsaturated storages vs. flow

Figure 6 shows the 3-dimensional relationship between hillslope saturated storage (HSS), unsaturated storage (HUS) and stream flow (Q) for the year 2010–2011. Four main periods can be identified, similar to what was outlined in recent studies (e.g. Heidbuechel et al., 2012; Hrachowitz et al., 2013a): three characterized the recharge period and the last one the recession period. First, stream flow was close or equal to 0 and was almost exclusively sustained by drainage of the saturated storage, while the unsaturated zone exhibited a significant storage deficit and only minor fluctuations due to transpiration and small summer rain events (dry period). As steadier precipitation patterns set in, here typically around November, the unsaturated-zone storage reached its maximal value relatively quickly, rapidly establishing connectivity with fast-responding flow pathways (wetting period). This led to a relatively rapid increase in stream flow while the saturated storage did not change much until the end of this

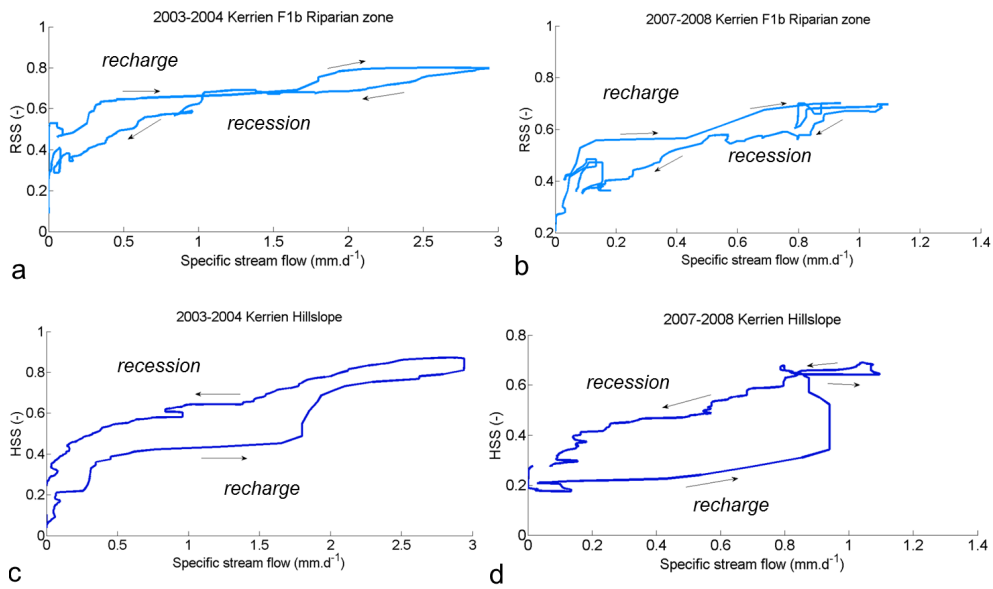


Figure 4. Examples of annual hysteretic loops for saturated-zone storage vs. stream flow which are clockwise in the riparian zone **(a, b)** and anticlockwise in the hillslope **(c, d)** for the wet year 2003–2004 **(a, c)** and the dry year 2007–2008 **(b, d)**.

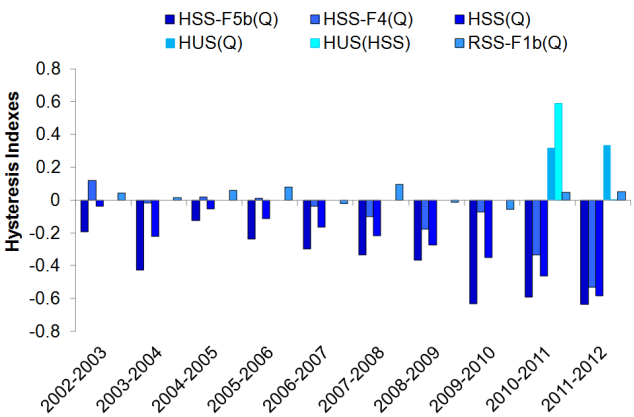


Figure 5. Annual hysteresis indexes (HI) computed for the piezometers in the Kerrien catchment from 2002 to 2012. F5b is located upslope, F4 midslope and F1b downslope in the riparian area. RRS(Q) is the hysteresis between stream flow and riparian saturated-zone storage (measured at F1b). HSS-F5b(Q), HSS-F4(Q) and HSS(Q) are hystereses between stream flow and upslope (at F5b), midslope (at F4) and hillslope (average of F5b and F4) saturated storages respectively. HUS(Q) is the hysteresis between stream flow and hillslope unsaturated storage (HUS) (computed from the average of normalized volumic moisture sensors in profiles sB1 and sB2), and HUS(HSS) is that between the hillslope unsaturated- and saturated-zone storage (average of F5b and F4).

period as incoming precipitation first had to fill the storage deficit in the unsaturated zone before a significant increase in percolation could occur. A further lag was introduced by the time taken for water to percolate and eventually recharge the relatively deep groundwater. As soon as conditions were wet enough to allow for established percolation, the saturated storage eventually also responded, increasing faster than the stream flow (wet period), while unsaturated storage remained full. During the wet period (or high-flow period), no pattern appeared clearly because all storage elements were almost full and the responses of all the compartments were more directly linked to the short-term dynamics of rain events. Finally during the recession period (drying period), unsaturated storage decreased comparatively quickly by drainage and transpiration, while the saturated storage kept increasing for a while by continued percolation from the unsaturated zone before decreasing through groundwater drainage at a relatively slow rate. A similar pattern was also observed for 2011–2012 (not shown).

The unsaturated-zone storage followed a clockwise hysteresis loop with the stream flow and with the saturated-zone storage. The hysteresis indexes (Fig. 5, years 2010–2011 and 2011–2012) reflected these directions and showed that the hysteresis loops were narrower for unsaturated storage than for saturated storage, inducing smaller absolute values of the hysteresis indexes due to the small size of the unsaturated storage compartment compared to the saturated storage compartment.

3.1.3 Interpretation

There are three main hypotheses generally proposed to interpret storage–discharge hystereses in hydrology. The first one is related to the increase in transmissivity with the groundwater level due to the frequently observed exponential decrease in hydraulic conductivity with depth. However, this would lead to systematic clockwise hysteresis loops and can-

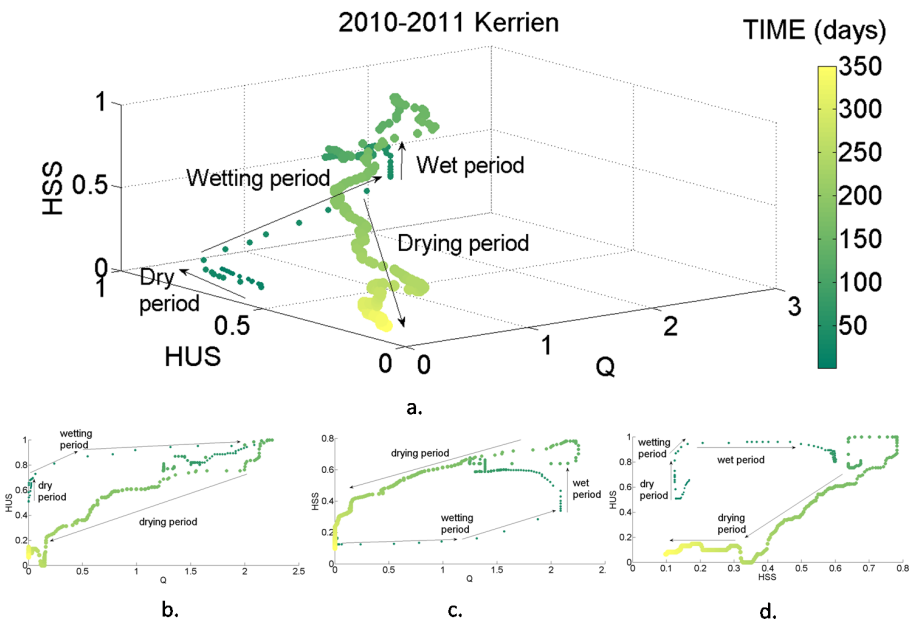


Figure 6. Evolution of stream flow (Q in mm d^{-1}) and normalized hillslope unsaturated storage (HUS) and hillslope saturated storage (HSS) for the water year 2010–2011 (October to September). The size of the dots increases with time. Unsaturated storage (HUS) is computed from the moisture sensors in profiles sB1 and sB2; saturated storage (HSS) is represented using a normalized groundwater table level (computed from two piezometers in the hillslope). **(a)** is the 3-dimensional plot and **(b, c, d)** are the respective 2-dimensional projections of **(a)** on the three plans.

not explain the anticlockwise patterns observed between hillslope saturated storage and stream flow. The second hypothesis proposed by Spence et al., 2010 is that during the recharge period, the groundwater storage not only increases locally (as measured by the piezometric variations), but the spatial extension of connected storage also increases gradually, while during the recession period, the storage decreases homogenously across the entire contribution area. This is likely for riparian groundwater and could explain the clockwise hysteresis observed on this piezometer but cannot explain the anticlockwise hysteresis observed in the hillslope groundwater. The third hypothesis is that dominant hydrological processes are different between recharge and recession periods. For instance, Jung et al. (2004) interpret their clockwise hysteresis in peatlands groundwater as the results of a stepwise filling process during the rising flows (fill and spill mechanism) opposed to a more gradual drainage of the groundwater during the recession combined with the first hypothesis result, similar to what was found by Hrachowitz et al. (2013a). This hypothesis of different hydrological pathways allows an adequate interpretation of the opposite directions of the observed hystereses. The recharge period is characterized by a quick filling of the unsaturated and saturated storages in the riparian zone, which is always close to saturation, while the saturated storage on the hillslope is not yet filling up (wetting period). Thus, the wetting period is characterized by an increase in stream flow, here mainly generated in the riparian zone, and eventual quick flows in the hillslope, while the hillslope unsaturated zone reaches the storage capacity volume. At the beginning of the wet period, hillslope saturated storage fills and starts to contribute to the stream, along with riparian and fast flows. During the recession period (drying period), the hillslope saturated zone is the only compartment which continues to sustain stream flow. If this hypothesis is correct, there are three contributions to stream flow in the wet period, while, during the recession period, hillslope groundwater remains the only contributor to stream flow (cf. Hrachowitz et al., 2013a, see Fig. 7). This can explain the difference between storage values in recharge and recession periods. Finally, the hysteretic hydrological signature is not only related to the amount of stored water in the catchment but rather to where it is stored.

These results are consistent with previous studies: the distinction between riparian groundwater and hillslope groundwater components has also been identified in similar catchments (by Molénat et al. (2008) based on nitrate concentration analysis and by Aubert et al. (2013a) based on a range of solutes) and at other site (by Haught and van Meerveld (2011)) using such Q–S relationships and lag time analysis.

3.1.4 Sensitivity of the HI to initial conditions

Sensitivity to antecedent soil moisture conditions is often cited as an explanation for observed storage–discharge hysteresis and its variability between years. The initial levels of each store will obviously influence the time required to fill

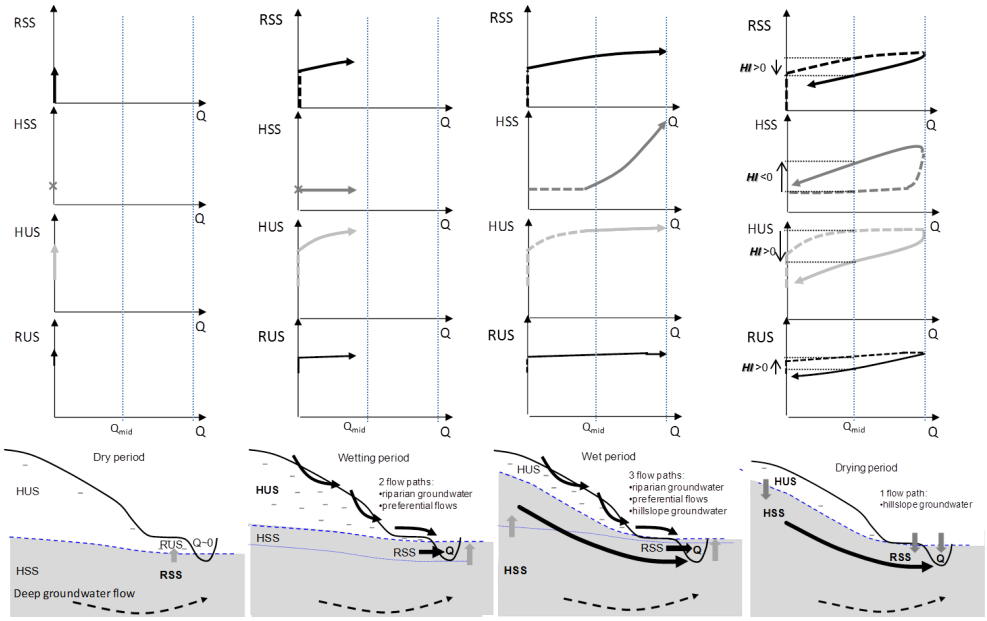


Figure 7. Conceptual scheme of successive mechanisms explaining the annual hysteresis between storages and stream flows. HUS: hillslope unsaturated storage; HSS: hillslope saturated storage; RUS: riparian unsaturated storage; RSS: riparian saturated storage; Q: stream flow. Bold characters indicate compartments with varying storage; grey arrows indicate whether the compartment is filling or emptying; black arrows indicate the water flow paths.

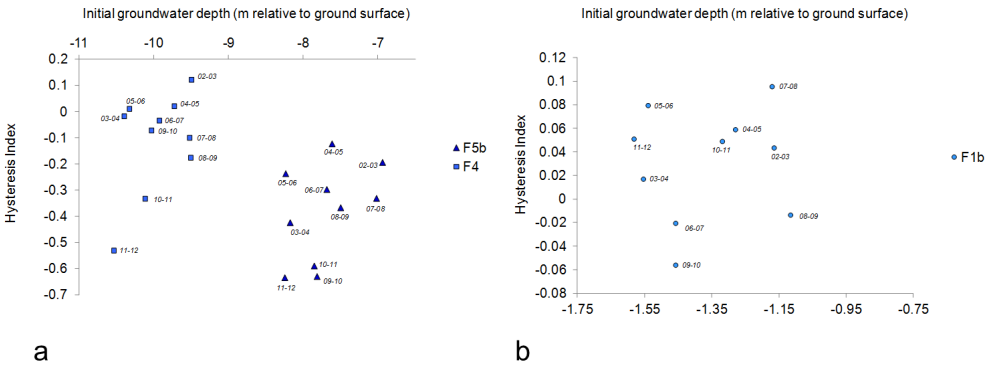


Figure 8. Year-to-year variations, for the 10 monitoring years, of the hysteresis indexes: **(a)** HSS-F5b(Q) and HSS-F4(Q) (HI) versus the initial groundwater table level depth in the corresponding hillslope piezometer (F5b or F4) and **(b)** HSS-F1b(Q) versus the initial groundwater table level depth in the piezometer in the riparian area (F1b).

them and consequently the duration of the successive periods identified in the whole recharge period. As only 2 years of data were available, it was not possible to define a relationship between the initial average soil moisture and the magnitude of the hysteresis indexes. However, the magnitude of the HI was lower for high initial values of average unsaturated-zone storage for both the saturated and unsaturated zones in 2011–2012 (Table 5). The HI for the midslope saturated zone (F4b) seemed to be more sensitive to these initial moisture conditions than the HI for the upslope saturated zone and unsaturated zone. Similarly, the width of the loop (absolute value of the HI) was not very sensitive to initial groundwater levels in the hillslope: although the larger absolute values of the HI were observed for the lower initial water table levels, no clear correlation was observed (Fig. 8).

3.1.5 Sensitivity of the HI to annual rainfall

For the saturated zone, the observed values of the HI were negatively correlated with the total annual rainfall for both the hillslope and the riparian zone, with a more negative slope for the hillslope (Fig. 9). Wet years (i.e. large values of annual rainfall) are generally associated with large values of annual maximal and midpoint stream flows and also with large values of groundwater table level, leading to larger saturated-storage values during the recession period, while

Table 5. Hysteresis indexes (HIs) and initial hillslope unsaturated-storage values (HUS) at the beginning of the water year.

Year	Initial HUS	Hysteresis index (HI)			
		HSS-F5b(Q)	HSS-F4(Q)	HSS(Q)	RSS-F1b(Q)
2010-2011	0.148	−0.591	−0.334	−0.462	0.590
2011–2012	0.026	−0.635	−0.532	−0.583	0.003

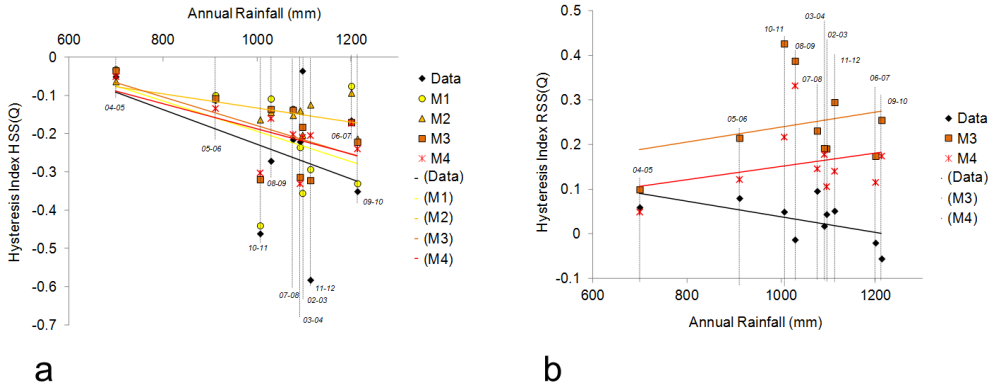


Figure 9. Variations of observed (data) and simulated (M1 to M4) hysteresis index versus annual rainfall for the 10 monitored water years for **(a)** hillslope saturated storage versus discharge HSS(Q) and **(b)** riparian saturated storage vs. discharge RSS(Q). Solid lines indicate the linear regressions.

the storage values during the recharge period do not change much from year to year. Thus, larger storage values at the time of midpoint discharge in the recession period led to smaller values of the HI (i.e larger absolute values for the hillslope, where hystereses are anticlockwise, and smaller absolute value of the HI for the riparian zone, where hystereses are clockwise). In the riparian zone, when rainfall and maximal drainage reached a very high value, it could lead to a saturated-storage value at the time of midpoint discharge in the recession period that was larger than the corresponding value during the recharge period, explaining the inversion of the sign of the HI for RSS(Q) in very wet years.

3.2 Model assessment based on their ability to reproduce the observed hysteresis

3.2.1 Hysteresis simulations

For all years, all models (M1–M4) exhibited a hysteretic relationship between stream flow and storage, as shown in Fig. 10 for the years 2003–2004 and 2007–2008, pertaining to the calibration and validation periods respectively. This means that all tested models introduced a lag time between catchment stores and the stream dynamics. Fig. 11a presents the observed and modelled average and standard deviation of the annual hysteresis indexes, for hillslope saturated storage vs. discharge HSS(Q), hillslope unsaturated storage vs. discharge HUS(Q), hillslope unsaturated storage vs. hillslope saturated storage HUS(HSS) and riparian saturated storage vs. discharge RSS(Q). As riparian saturated storage (RSS) is not modelled in M1 and M2, simulated RSS(Q) was available only for M3 and M4.

For M1, the shape of the simulated hysteresis showed an overestimation of hillslope saturated storage (HSS) and of flow during dry years (e.g. the year 2007–2008 shown in Fig. 10). This was expected as we have seen that the model was unable to reproduce groundwater dynamics and the low signatures during the validation period (Fig. 3 and supplementary material). Simulated HI values were close to the observed ones for HSS(Q) (Fig. 11a). The simulated hysteresis indexes were small and negative for HUS(Q), while the observed values were large and positive. Simulated HI values for HUS(HSS) were also overestimated. These results show that, in model M1, the overestimation of the hillslope saturated storage was partially compensated by the underestimation of the hillslope unsaturated storage. This reveals the poor consistency of the model and explains why it was able to reach good performance in the calibration period but not in the validation period (Fig. 3).

For the model M2, the shape of the hysteresis loops showed a considerable underestimation of HSS and a large underestimation of stream flow in wet years (Fig. 10). Compared to M1, although the introduction of deep losses in M2 led to higher validation performances and better simulation of hydrological signatures (Fig. 3), the simulated HIs (Fig. 11a) worsened, suggesting a poorer model consistency with respect to internal hydrologic processes.

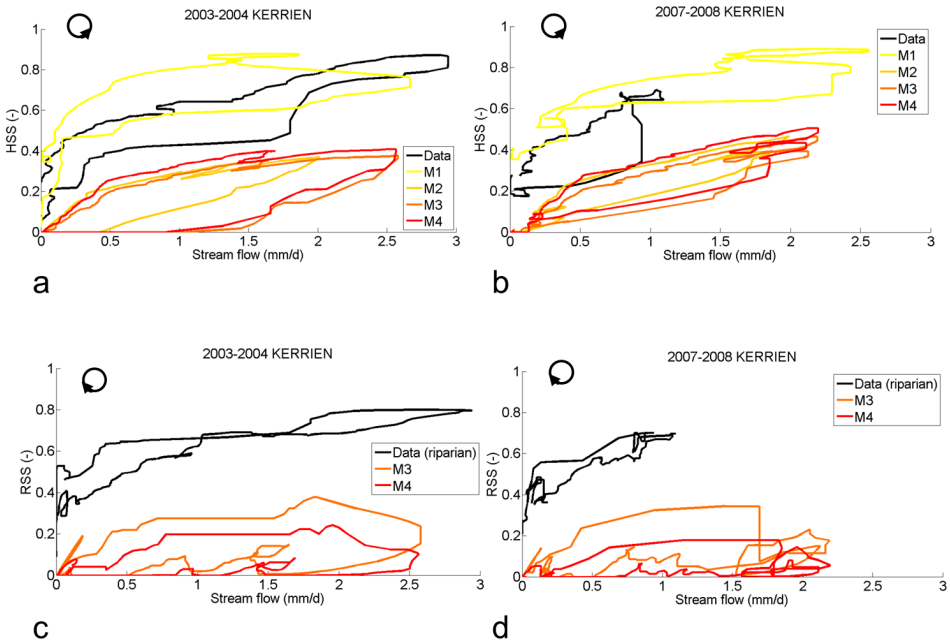


Figure 10. Observed and simulated annual hysteresis between stream flow (Q) and **(a, b)** saturated storage in the hillslope HSS (for observed hysteresis, HSS is the average of F5b and F4) and **(c, d)** saturated storage in the riparian area RSS (for simulated hysteresis, only M3 and M4 represent the riparian area), for the water years **(a, c)** 2003–2004 (wet year, calibration period) and **(b, d)** 2007–2008 (dry year, validation period).

For both models M3 and M4, the introduction of a riparian compartment improved the simulated hysteretic loops, due to a better simulation of stream flow in wet years, but HSS was still largely underestimated (Fig. 10). The mean HI values for HSS(Q) were close to the observed one, but the range of variation was smaller, indicating a reduced sensitivity to climate (Fig. 9). The mean values for HUS(Q) were clearly improved compared to M1 and M2 as the direction of the loop was clockwise as for the observations, although the values were still underestimated. The mean HI values for HUS(HSS) were also greatly improved. The shape of the simulated hysteresis loops between riparian saturated storage (RSS) and stream flow (Q) showed a large underestimation of RSS, especially during the recession period (Fig. 10c, d). This led to simulated HIs for RSS(Q) which are positive, like the observed ones, but also largely overestimated (Fig. 11a). Overall, these results suggest that for models including a riparian component, the underestimation of the hysteresis between HUS and Q was compensated for by an overestimation of the hysteresis between RSS and Q. This highlights that, despite a significant improvement in performances and improved hydrological signature reproduction, these models still involve a certain degree of inconsistency with respect to internal processes. However, M4 provided the most balanced performance considering hysteretic signatures between all storage components and strongly underlines the limitations of overly simplistic model architectures (e.g. M1) and the need for more complete representations of process heterogeneity. The hysteresis index sensitivity to parameter uncertainty increases with the number of parameters from M1 to M2 and then stays in the same range from M2 to M4 (Fig. 11b). This analysis confirms the importance of considering the hysteresis indexes both between saturated and unsaturated storage (HSS and HUS) to avoid accepting a wrong model. For example, considering only the performance regarding the HSS(Q) relationship could lead one to accept model M1, while its performance on HUS is lower and it is not able to reproduce the Riparian compartment hysteresis. For readability purposes, Fig. 11b only illustrates this sensitivity for the different HIs in the year of 2011–2012 but similar behaviour is observed every year. It shows that best behavioural parameter sets (bbp) lead to modelled HI values closer to the observed values than average modelled HI values. Using an additional calibration criterion related to the hysteresis could reduce the sensitivity of the HI to parameter uncertainty and lead to a narrow range of feasible parameter sets.

3.2.2 Sensitivity of modelled hysteresis indexes to annual rainfall

All models were able to represent the decrease in the hysteresis indexes with annual rainfall on the hillslope, the slope of the correlation getting closer to the observed one from M1 to M4 (Fig. 9). The introduction of deep-groundwater losses (M2) led to smaller saturated storage during recharge periods

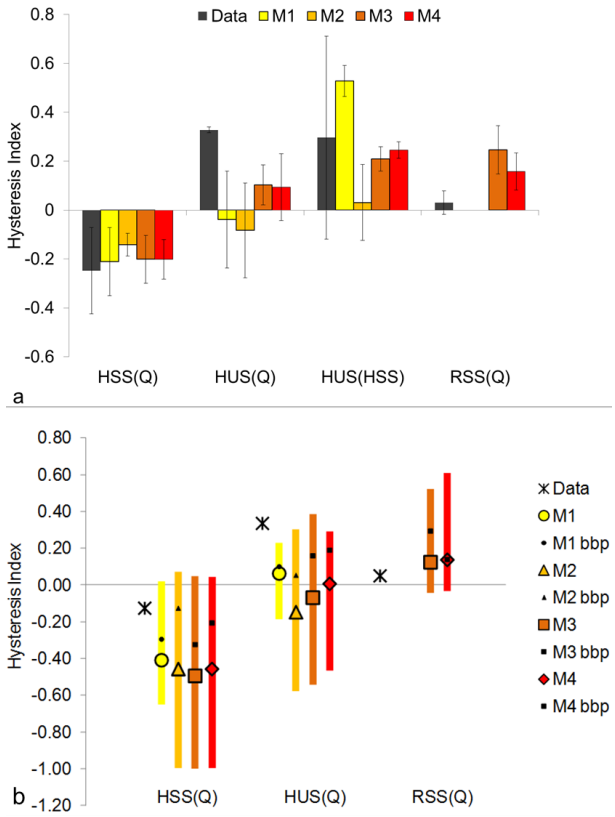


Figure 11. (a) Mean annual hysteresis indexes observed and simulated with the four models M1 to M4 for hillslope saturated storage vs. discharge HSS(Q), hillslope unsaturated storage vs. discharge HUS(Q), hillslope unsaturated storage vs. hillslope saturated storage HUS(HSS) and riparian saturated storage vs. discharge RSS(Q). RSS is only simulated in models M3 and M4. Error bars show the standard deviation for the 10 years for HSS(Q) and RSS(Q) and the values for the 2 available years for HUS(Q) and HUS(HSS). **(b)** Sensitivity of hysteresis index values to parameter uncertainty for the year 2011–2012. Mx bbp indicates the value for best behavioural parameter sets; the circles, triangles, squares and diamonds indicate the mean HI value for the all the behavioural parameter sets, and the corresponding bars indicate its range of variation.

and increased the difference between saturated storage during recharge and recession periods at the time of midpoint discharge. However, as all models tended to overestimate low stream flow values, the slopes of the correlations between annual rainfall and the simulated HI were smaller than for the observed one.

In the riparian zone, the modelled trends were the inverse of the observed one. The modelled recessions were always very sharp (see Hrachowitz et al., 2014), and the simulated riparian storage dried up every year, explaining why saturated storage at the time of midpoint discharge during the recession periods was much greater than during the recharge periods. This led to a general overestimation of HI values, which were even stronger for wet years. This overestimation may be related to an improper conceptualization of the riparian-zone functioning, which is never connected to the hillslope reservoir in the tested models. In reality, during high-flow periods, the observed hydraulic gradient increased along the hillslope, inducing a connection between riparian and hillslope reservoirs which are disconnected during low-flow periods.

3.2.3 Value of such internal signatures for model evaluation

The use of hydrological hysteretic signatures in model assessments led to conclusions that were consistent with the classical hydrological signatures used in Hrachowitz et al. (2014). However, model M2 was less able to reproduce the different hysteretic signatures, whereas it led to a real improvement regarding to the classical signatures in low flows. Considering only the distance between observed and simulated hysteresis indexes on hillslope saturated storage and stream flow would lead one to select model M1. This highlights the fact that using saturated-storage dynamics alone can be deceptive for understanding the system response behaviour and that it is thus crucial to also consider the hysteretic signatures of unsaturated and riparian zones in a combined approach to develop a more robust understanding of the system. Here, hysteretic signatures of the unsaturated and riparian zones provided valuable additional assessment metrics regarding the performance of models M3 and M4 to represent the riparian zone. It was possible to identify when the model failed to represent processes, which processes were mostly compensating for missing ones and therefore why the model may provide some good performance for the wrong reasons. In this regard, the hysteresis index proved to be a useful proxy of hystereses themselves as it exhibited contrasted patterns sensitive to climate and localization within the catchment.

3.3 Perspectives: toward an integrated hydrological-signature-based modelling?

A general issue in model calibration is that, because of the overparameterization of hydrological models and because the objective functions generally only integrate one variable, such as the stream flow, automatic calibration techniques may lead to parameter sets which compensate for internal model errors. These parameter sets are mathematically correct but wrong from a hydrological point of view. The subsequent model should then be considered nonbehavioural (Beven, 2006). For instance, if storage properties are not taken into account well by the model, this is likely to lead to a wrong simulation of storage dynamics in response to precipitation. Thus, the parameterization using traditional objective functions can lead to compensation of these errors in order to simulate a discharge value close to the observed one while the storage is wrong. In such a case, a model able to represent the internal catchment behaviour will generate a

wrong discharge value which is, however, consistent with the storage value and will be rejected in traditional calibration procedures. To handle this issue and in order to select behavioural models, one can use multiple objective functions (Gupta et al., 1998; Seibert and McDonnell, 2002; Freer et al., 2003), including a range of hydrological signatures to be reproduced or additional realism constraints (Kavetski and Fenicia, 2011; Yadav et al., 2007; Yilmaz et al., 2008; Euser et al., 2013; Gharari et al., 2013; Hrachowitz et al., 2014). We argue that, rather than increasing the number of constraints or objective functions which have to be satisfied, an alternative could be to use some objective functions based on a combination of different variables, such as stream flow and the groundwater level, soil moisture or stream concentrations. Among the possible combination of variables, objective functions based on the relative dynamics of storage in different spatial locations, such as riparian versus hillslope, might provide new insights into the catchment-internal processes. We suggest that such combined objective functions would be more constraining for model selection. Therefore, the present study is a first step which aims at highlighting the still underexploited potential of hydrological hysteresis. The next step would be to quantify these relationships through functions or several indexes usable in calibration criteria, such as the hysteresis index proposed in this study. Moreover, such criteria could be used in classification studies. Indeed, some studies in the literature present storage–discharge relationships for different catchments that show patterns that are similar or dissimilar to the ones we observed in the Kerrien catchment (Ali et al., 2011; Gabrielli et al., 2012). This signature may help to classify catchments in terms of dominant processes driving their behaviour.

A remaining difficulty with integrating storage into calibration or evaluation procedure in hydrological modelling is how to measure this storage. McNamara et al. (2011) and Tetzlaff et al. (2011) proposed using all available data from groundwater level monitoring, soil moisture records, water budget, modelling results and so on to estimate the storage in catchments. In this study, we used quite a dense network of piezometers and soil moisture measurements relative to the small size of the catchment. Promising ways to estimate spatial quantification of storage in catchments include remote sensing of soil moisture (Sreelash et al., 2013; Vereecken et al., 2008), gravimetric techniques (Creutzfeldt et al., 2012), geodesy and geophysical methods. The interest in such techniques would be to provide a spatially integrated vision of the catchment water content.

As for the different hydrological variables, the combination of hydrological and chemical variables appears relevant to investigating the hydrochemical behaviour of catchments. Hysteresis patterns between concentration and discharge have been largely documented for storm event characterization (Evans and Davies, 1998; Evans et al., 1999; Taghavi et al., 2011). Some studies also report similar patterns on the annual scale (e.g. Aubert et al., 2013b). Such hysteretic relationships have been observed also between water and chemistry in groundwater (Rouxel et al., 2011; Hrachowitz et al., 2013a), emphasizing a disconnection between water and solute dynamics that simple diffusion or partial mixing processes cannot explain. Stream water chemistry also exhibits particular seasonal cycles with different phasing and with discharge depending on the solutes (Aubert et al., 2013b). This provides extra information on the water pathways within the catchment. These relationships also appear to be powerful in constraining hydrochemical modelling.

4 Conclusions

A method to characterize and partially quantify the relationship between storages in a headwater catchment and stream flow throughout a year has been proposed. It allowed us to then assess the ability of a range of conceptual lumped models to reproduce this catchment-internal signature. Catchment storage has been approximated using a network of piezometric data and several unsaturated-zone moisture profiles to consider the storage in the saturated as well as in the unsaturated zones.

The observations showed that storage–discharge relationships in catchments can be hysteretic, highlighting a successive activation of different hydrological components during the recharge period, while the recession exhibits a fast decrease in unsaturated and riparian storage and a slow decrease in hillslope saturated storage which sustains the stream flow. Four periods have been identified in the hydrological year: (1) first, at the end of the dry period, rainfall starts to refill unsaturated storage; (2) in the wetting period, riparian unsaturated storage is filled and the saturated storage starts to supply the stream while hillslope unsaturated storage is still being replenished; (3) during the wet period, unsaturated storage in the hillslope is also filled and the saturated hillslope storage also feeds the stream. (4) Finally when rainfall declines, flow from the riparian groundwater recedes and, during the recession period, the stream discharge is sustained only by hillslope groundwater. Stream discharge and riparian and hillslope saturated storages exhibited different patterns of hysteresis, with opposite directions of the hysteretic loops.

The tested models were characterized by an increasing degree of complexity and also an increasing consistency, as shown in a previous study using classical hydrologic signatures. In this study, we showed that, if all of the models simulated a hysteretic relationship between storage and discharge, their ability to reproduce the hysteresis index also increased with model complexity. In addition, we suggest that, if classical hydrological signatures help to assess model consistency, the hysteretic signatures also help to identify quickly when and why the models give "right answers for the wrong reasons" and can be used as a descriptor of the internal catchment functioning.

Acknowledgements. The investigations benefited from the support of INRA and CNRS for the Research Observatory ORE AgrHys, and from Allenvi for the SOERE RBV. Data are available at http://geowww.agrocampus-ouest.fr/web/.

Edited by: N. Romano

References

Ali, G. A., L'Heureux, C., Roy, A. G., Turmel, M.-C., and Courchesne, F.: Linking spatial patterns of perched groundwater storage and stormflow generation processes in a headwater forested catchment, Hydrol. Process., 25, 3843–3857, 2011.

Andermann, C., Longuevergne, L., Bonnet, S., Crave, A., Davy, Ph., and Gloaguen, R.: Impact of transient groundwater storage on the discharge of Himalayan rivers, Nat. Geosci., 5, 127–132, 2012.

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., 26, 2206–2210, 2012.

Aubert, A. H., Gascuel-Odoux, C., Gruau, G., Akkal, N., Faucheux, M., Fauvel, Y., Grimaldi, C., Hamon, Y., Jaffrézic, A., Lecoz-Boutnik, M., Molénat, J., Petitjean, P., Ruiz, L., and Merot, P.: Solute transport dynamics in small, shallow groundwater-dominated agricultural catchments: insights from a high-frequency, multisolute 10 yr-long monitoring study, Hydrol. Earth Syst. Sci., 17, 1379–1391, doi:10.5194/hess-17-1379-2013, 2013.

Aubert, A. H., Gascuel-Odoux, C., and Merot, P.: Annual hysteresis of water quality: A method to analyse the effect of intra- and inter-annual climatic conditions, J. Hydrol., 478, 29–39, 2013.

Bergström, S.: The HBV model, in: Computer Models of Watershed Hydrology, edited by: Singh, V. P., Water Resources Publications: Highlands Ranch, CO, 443-476, 1995.

Beven, K.: How far can we go in distributed hydrological modelling?, Hydrol. Earth Syst. Sci., 5, 1–12, doi:10.5194/hess-5-1-2001, 2001.

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320, 18–36, 2006.

Beven, K.: So how much of your error is epistemic? Lessons from Japan and Italy, Hydrol. Process., 27, 1677–1680, 2013.

Butturini, A., Alvarez, M., Bernal, S., Vazquez, E., and Sabater, F.: Diversity and temporal sequences of forms of DOC and NO3-discharge responses in an intermittent stream: Predictable or random succession?, J. Geophys. Res.-Biogeosci., 113, G03016, doi:10.1029/2008JG000721, 2008.

Creutzfeldt, B., Ferré, T., Troch, P., Merz, B., Wziontek, H., and Güntne, A.: Total water storage dynamics in response to climate variability and extremes: Inference from long-term terrestrial gravity measurement, J. Geophys. Res.-Atmos., 117, D08112, doi:10.1029/2011JD016472, 2012.

Criss, R. E. and Winston, W.,E.: Do Nash values have value? Discussion and alternate proposals, Hydrol. Process., 22, 2723–2725, 2008.

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.

Evans, C. and Davies, T. D.: Causes of concentration/discharge hysteresis and its potential as a tool for analysis of episode hydrochemistry, Water Resour. Res., 34, 129–137, 1998.

Evans, C., Davies, T. D., and Murdoch, P. S.: Component flow processes at four streams in the Catskill Mountains, New York, analysed using episodic concentration/discharge relationships, Hydrol. Process., 13, 563–575, 1999.

Everett, D. H. and Whitton, W. I.: A general approach of hysteresis, Trans. Faraday Soc., 48, 749–757, 1952.

Fenicia, F., McDonnell, J. 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, 2008.

Franks, S. W., Gineste, P., Beven, K. J., and Merot, P.: On constraining the predictions of a distributed moder: The incorporation of fuzzy estimates of saturated areas into the calibration process, Water Resour. Res., 34, 787–797, 1998.

Freer, J., Beven, K., and Ambroise, B.: Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, Water Resour. Res., 32, 2161–2173, 1996.

Freer, J., Beven, K., and Peters, N.: Multivariate Seasonal Period Model Rejection Within the Generalised Likelihood Uncertainty Estimation Procedure, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N. and Turcotte, R., American Geophysical Union, Washington, D. C., doi:10.1029/WS006p0069, 69–87, 2003.

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

Frei, S., Lischeid, G., and Fleckenstein, J. H.: Effects of microtopography on surface-subsurface exchange and runoff generation in a virtual riparian wetland - A modeling study, Adv. Water Resour., 33, 1388–1401, 2010.

Gabrielli, C. P., McDonnell, J. J., and Jarvis, W. T.: The role of bedrock groundwater in rainfall–runoff response at hillslope and catchment scales, J. Hydrol., 450–45, 117–133, 2012.

Gascuel-Odoux, C., Weiler, M., and Molenat, J.: Effect of the spatial distribution of physical aquifer properties on modelled water table depth and stream discharge in a headwater catchment, Hydrol. Earth Syst. Sci., 14, 1179–1194, doi:10.5194/hess-14-1179-2010, 2010.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, Hydrol. Earth Syst. Sci., 18, 4839–4859, doi:10.5194/hess-18-4839-2014, 2014.

Giustolisi, O. and Simeone, V.: Optimal design of artificial neural networks by a multiobjective strategy: groundwater level predictions, Hydrol. Sci. J.-J. Sci. Hydrol., 51, 502–523, 2006.

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M.: Towards a comprehensive assessment of model structural adequacy, Water Resour. Res., 48, W08301, doi:10.1029/2011WR011044, 2012.

Gupta, H. V., Wagener, T., and Liu, Y. Q.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol. Process., 22, 3802–3813, 2008.

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, 1998.

Hartmann, A., Wagener, T., Rimmer, A. , Lange, J., Brielmann, H., and Weiler, M.: Testing the realism of model structures to identify karst system processes using water quality and quantity signatures, Water Resour. Res., 49, 3345–3358, doi:10.1002/wrcr.20229, 2013a.

Hartmann, A., Weiler, M., Wagener, T., Lange, J., Kralik, M., Humer, F., Mizyed, N., Rimmer, A., Barberá, J. A., Andreo, B., Butscher, C., and Huggenberger, P.: Process-based karst modelling to relate hydrodynamic and hydrochemical characteristics to system properties, Hydrol. Earth Syst. Sci., 17, 3305–3321, doi:10.5194/hess-17-3305-2013, 2013.

Haught, D. R. W. and van Meerveld, H. J.: Spatial variation in transient water table responses: differences between an upper and lower hillslope zone, Hydrol. Process., 25, 3866–3877, 2011.

Heidbüchel, I., Troch, P. A., Lyon, S. W., and Weiler, M.: The master transit time distribution of variable flow systems, Water Resour. Res., 48, W06520, doi:10.1029/2011WR011293, 2012.

Hopmans, P. and Bren, L. J.: Long-term changes in water quality and solute exports in headwater streams of intensively managed radiata pine and natural eucalypt forest catchments in southeastern Australia, Forest Ecol. Manage., 253, 244–261, 2007.

Hrachowitz, M., Soulsby, C., Tetzlaff, D., and Malcolm, I. A.: Sensitivity of mean transit time estimates to model conditioning and data availabilityle, Hydrol. Process., 25, 980–990, 2011.

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, Hydrol. Sci. J., 58, 1198–1255, 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, 2014.

Jung, M., Burt, T. P., and Bates, P. D.: Toward a conceptual model of floodplain water table response, Water Resour. Res., 40, W12409, doi:10.1029/2003WR002619, 2004.

Kampf, S. K. and Burges, S. J.: Parameter estimation for a physics-based distributed hydrologic model using measured outflow fluxes and internal moisture states, Water Resour. Res., 43, W12414, doi:10.1029/2006WR005605, 2007.

Kavetski, D. and Fenicia, F.: Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, Water Resour. Res., 47, W11511, doi:10.1029/2011WR010748, 2011.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42, W03S04, doi:10.1029/2005WR004362, 2006.

Lamb, R., Beven, K., and Myrabo, S.: Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model, Adv. Water Resour., 22, 305–317, 1998.

Lawler, D. M., Petts, G. E., Foster, I. D. L., and Harper, S.: Turbidity dynamics during spring storm events in an urban headwater river system: The Upper Tame, West Midlands, UK, Sci. Total Environ., 360, 109–126, 2006.

Martin, C., Molénat, J., Gascuel-Odoux, C., Vouillamoz, J.-M., Robain, H., Ruiz, L., Faucheux, M., and Aquilina, L.: Modelling the effect of physical and chemical characteristics of shallow aquifers on water and nitrate transport in small agricultural catchments, J. Hydrol., 326, 25–42, 2006.

McNamara, J. P., Tetzlaff, D., Bishop, K., Soulsby, C., Seyfried, M., Peters, N. E., Aulenbach, B. T., and Hooper, R.: Storage as a Metric of Catchment Comparison, Hydrol. Process., 25, 3364–3371, 2011.

Mohamed, Y. A., Savenije, H. H. G., Bastiaanssen, W. G. M., and van den Hurk, B. J. J. M.: New lessons on the Sudd hydrology learned from remote sensing and climate modeling, Hydrol. Earth Syst. Sci., 10, 507–518, doi:10.5194/hess-10-507-2006, 2006.

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, 2005.

Molénat, J., Gascuel-Odoux, C., Ruiz, L., and Gruau, G.: Role of water table dynamics on stream nitrate export and concentration. in agricultural headwater catchment (France), J. Hydrol., 348, 363–378, 2008.

Molénat, J., Gascuel-Odoux,C., Aquilina, L., and Ruiz, L.: Use of gaseous tracers (CFCs and SF6) and transit-time distribution spectrum to validate a shallow groundwater transport model, J. Hydrol., 480, 1–9, doi:10.1016/j.jhydrol.2012.11.043, 2013.

Moussa, R., Chahinian, N., and Bocquillon, C.: Distributed hydrological modelling of a Mediterranean mountainous catchment - Model construction and multi-site validation, J. Hydrol., 337, 35–51, 2007.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I -A discussion of principles, J. Hydrol., 10, 282–290, 1970.

Nester, T., Kirnbauer, R., Parajka, J., and Bloschl, G.: Evaluating the snow component of a flood forecasting model, Hydrol. Res., 43, 762–779, 2012.

Parajka, J., Naeimi, V., Blöschl, G., Wagner, W., Merz, R., and Scipal, K.: Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale, Hydrol. Earth Syst. Sci., 10, 353–368, doi:10.5194/hess-10-353-2006, 2006.

Rouxel, M., Molénat, J., Ruiz, L., Legout, C., Faucheux, M., and Gascuel-Odoux, C.: Seasonal and spatial variation in groundwater quality along the hillslope of an agricultural research catchment (Western France), Hydrol. Process., 25, 831–841, 2011.

Ruiz, L., Abiven, S., Durand, P., Martin, C., Vertès, F., and Beaujouan, V.: Effect on nitrate concentration in stream water of agricultural practices in small catchments in Brittany: I. Annual nitrogen budgets, Hydrol. Earth Syst. Sci., 6, 497–506, doi:10.5194/hess-6-497-2002, 2002.

Salant, N. L., Hassan, M. A., and Alonso, C. V.: Suspended sediment dynamics at high and low storm flows in two small watersheds, Hydrol. Process., 22, 1573–1587, 2008.

Savenije, H. H. G.: Equifinality, a blessing in disguise?, Hydrol. Process., 15, 2835–2838, 2001.

Schwientek, M., Osenbruck, K., and Fleischer, M.: Investigating hydrological drivers of nitrate export dynamics in two agricultural catchments in Germany using high-frequency data series, Environ. Earth Sci., 69, 381–393, 2013.

Schwientek, M., Osenbruck, K., and Fleischer, M.: Investigating hydrological drivers of nitrate export dynamics in two agricultural catchments in Germany using high-frequency data series, Environ. Earth Sci., 69, 381–393, doi:10.1007/s12665-013-2322-2, 2013.

Seibert, J.: Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, Hydrol. Earth Syst. Sci., 4, 215–224, doi:10.5194/hess-4-215-2000, 2000.

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.

Seibert, J., Rodhe, A., and Bishop, K.: Simulating interactions between saturated and unsaturated storage in a conceptual runoff model, Hydrol. Process., 17, 379–390, 2003.

Sivapalan, M., Bloschl, G., Merz, R., and Gutknecht, D.: Linking flood frequency to long-term water balance: Incorporating effects of seasonality. Water Resour. Res., 41, W06012, doi:10.1029/2004WR003439, 2005.

Spence, C.: A Paradigm Shift in Hydrology: Storage Thresholds Across Scales Influence Catchment Runoff Generation, Geogr. Compass, 4, 819–833, 2010.

Spence, C., Guan, X. J., Phillips, R., Hedstrom, N., Granger, R., and Reid, B.: Storage dynamics and streamflow in a catchment with a variable contributing area, Hydrol. Process., 24, 2209–2221, 2010.

Sreelash, K., Sekhar, M., Ruiz, L., Buis, S., and Bandyopadhyay, S.: Improved Modeling of Groundwater Recharge in Agricultural Watersheds Using a Combination of Crop Model and Remote Sensing, J. Indian Inst. Sci., 93, 189–207, 2013.

Taghavi, L., Merlina, G., and Probst, J. L.: The role of storm flows in concentration of pesticides associated with particulate and dissolved fractions as a threat to aquatic ecosystems Case study: the agricultural watershed of Save river (Southwest of France), Knowledge and Management of Aquatic Ecosystems(400), 2011.

Tetzlaff, D., McNamara, J. P., and Carey, S. K.: Measurements and modelling of storage dynamics across scales Preface, Hydrol. Process., 25, 3831–3835, 2011.

Velleux, M. L., England, J. F., and Julien, P. Y.: TREX: Spatially distributed model to assess watershed contaminant transport and fate, Sci. Total Environ., 404, 113–128, 2008.

Vereecken, H., Huisman, J. A., Bogena, H., Vanderborght, J., Vrugt, J. A., and Hopmans, J. W.: On the value of soil moisture measurements in vadose zone hydrology: A review, Water Resour. Res., 44, W00D06, doi:10.1029/2008WR006829, 2008.

Wagener, T.: Can we model the hydrological impacts of environmental change?, Hydrol. Process., 21, 3233–3236, 2007.

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.

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.

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, 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.