

Water Resources Research[®]

RESEARCH ARTICLE

10.1029/2020WR028835

Key Points:

- Formulation of a routine to address losing stream conditions, important for properly modeling low flows and intermittent streams
- Implementation of the new routine in a bucket-type hydrological model
- For the Panola Mountain Research Watershed the routine led to better simulations of streamflow during drying-down and wetting-up periods

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

M. Staudinger,
maria.staudinger@geo.uzh.ch

Citation:

Staudinger, M., Seibert, J., & van Meerveld, H. J. (2021). Representation of bi-directional fluxes between groundwater and surface water in a bucket-type hydrological model. *Water Resources Research*, 57, e2020WR028835. <https://doi.org/10.1029/2020WR028835>

Received 24 SEP 2020

Accepted 6 JUL 2021

© 2021 The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Representation of Bi-Directional Fluxes Between Groundwater and Surface Water in a Bucket-Type Hydrological Model

M. Staudinger¹ , J. Seibert^{1,2} , and H. J. van Meerveld¹ 

¹Department of Geography, University of Zurich, Zurich, Switzerland, ²Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden

Abstract In most bucket-type hydrological models, water can only flow from the groundwater to the stream and the flux is based on the groundwater storage. However, many catchments have losing stream sections, where streamflow recharges the groundwater. We developed a formulation to represent groundwater recharge by streamwater in a bucket-type model and tested this formulation for the Panola Mountain Research Watershed to demonstrate its function and assess its performance. The upper reach of the Panola catchment is often dry and highly affected by flow from a bedrock outcrop; further downstream the stream is perennial. We simulated streamflow with the fully lumped version of a bucket-type model and compared it to (a) a variant with sub-catchments to more realistically represent the low storage and quick response from the bedrock outcrop and (b) a variant that also includes the bi-directional exchange between the groundwater and the stream. For all three model variants, we compared simulated and observed streamflow and groundwater dynamics. Although the gain in overall model performance by including the bi-directional exchange between the groundwater and the stream was small, the explicit representation of this exchange led to better streamflow simulations during drying-down and wetting-up periods. For Panola the fluxes along the stream appeared less important than subsurface drainage from the upper sub-catchment to the downstream sub-catchment. We recommend considering the bi-directional fluxes between groundwater and the stream in bucket-type hydrological models where these processes are important, and the focus of the simulations is on low flow conditions.

1. Introduction

Hydrological models aim to represent the processes that occur in catchments, but by definition do so in a simplified way. This is especially true for bucket-type models (e.g., the Hydrologiska Byråns Vattenbalansavdelning (HBV) model, see Bergström, 1976 and Lindström et al., 1997), which simulate the transformation of precipitation into streamflow based on a limited number of buckets that represent the storage of water in different parts of the catchment, such as the snow pack, the unsaturated zone, and the saturated zone. In these models, the flux from groundwater storage to the stream is usually computed using a monotonically increasing function and can occur only in one direction, namely from the groundwater toward the stream. For most models, at least theoretically, this implies that the simulated streamflow can never become zero. This is problematic for the simulation of streamflow in intermittent or ephemeral streams, which are very common (Hammond et al., 2021; Messenger et al., 2021; Sauquet et al., 2021; Shanafield et al., 2021; van Meerveld et al., 2020). Equally important, losing stream conditions and decreasing streamflow along the stream network cannot be simulated by these models. In reality, however, there can be a flux from the stream toward the groundwater, so that streamflow decreases downstream (e.g., Huang et al., 2015; McCallum et al., 2014; McMahan & Nathan, 2021; Orłowski et al., 2014; Shanafield & Cook, 2014). Multiple field studies have highlighted the high variability in gaining and losing flow conditions along the stream network and during different flow conditions (Covino & McGlynn, 2007; Doering et al., 2007; Payn et al., 2009; Simpson & Meixner, 2013; Ward et al., 2013; Yu et al., 2013; Zimmer & McGlynn, 2017).

In fully distributed, physically based models, where fluxes between the groundwater and the stream are computed based on head differences, these potentially bi-directional fluxes between the groundwater and the stream are simulated explicitly. For example, Gutiérrez-Jurado et al. (2019) used the HydroGeoSphere model to determine how soil type and rainfall event characteristics affect runoff generation for an intermittent

stream and Querner et al. (2016) used the SIMGRO model to simulate the different hydrological states along a stream network. Ward et al. (2018), similarly, modeled time-variable hydrologic connectivity along headwater mountain streams but used a less complex model that required only information from one stream gauge, estimates of hydrogeologic properties (particularly hydraulic conductivity) and elevation data from a digital terrain model. Stoll and Weiler (2010) used the Hillvi model to explicitly simulate the stream network and showed that this information was useful for model calibration. Others (e.g., De Girolamo et al., 2017; Jaeger et al., 2014) have used semi-distributed models, such as the SWAT model, to simulate streamflow in intermittent streams to predict their response to climate change. However, in most bucket-type hydrological models, simulation of bi-directional fluxes between the groundwater and stream or intermittent stream sections are not possible. A notable exception is the work by Brauer et al. (2014), who simulated the changes in flow directions for lowland catchments. Their WALRUS model includes a flux from stream water to the groundwater and vice-versa. The simulated exchange was based on both the depth of the groundwater table below the soil surface and the surface water level (i.e., the gradient driving the flow) and the average channel depth, which defined the variable contact surface through which the exchange could take place. A constant represented the hydraulic properties and defined the rate of exchange.

Ignoring the possibility of bi-directional fluxes between the groundwater and the stream in bucket-type models is particularly problematic if the models are used to address questions related to low flows or droughts because during these periods even small losses to the groundwater can affect streamflow significantly. Similarly, inter-basin groundwater transfers into the catchment may maintain streamflow during droughts or can cause the stream to dry up earlier (e.g., Welch et al., 2012). Le Moine et al. (2007) modeled inter-basin groundwater flow via subsurface catchment boundaries that differ from the ones derived from surface topography but did not directly simulate groundwater-surface water exchange within a catchment.

To overcome the limitation of a uni-directional flux from the groundwater to the stream in bucket-type hydrological models, we present an exchange routine that can be used to represent the bi-directional fluxes between the groundwater and the stream. We implemented the routine in the HBV model to simulate streamflow in the Panola Mountain Research Watershed (from here on referred to as “Panola”) in Georgia, USA. Numerous studies have focused on runoff generation mechanisms in Panola (Aulenbach & Peters, 2018; Peters & Aulenbach, 2011; Tromp-van Meerveld & Weiler, 2008; Tromp-van Meerveld & McDonnell, 2006b; van Meerveld et al., 2015; Wang, 2011) and tested models for this catchment (Peters, Freer, & Aulenbach, 2003; Peters, Freer, & Beven, 2003). Panola is a suitable test case for this study because the large bedrock outcrop with minimal storage capacity generates streamflow quickly during precipitation events and generates a flood wave that travels along the stream (Peters, Freer, & Aulenbach, 2003). The upper parts of the stream network are dry for extended periods of time, such that streamwater generated at the bedrock outcrop infiltrates into the streambed and recharges the near-stream groundwater during many events.

The aim of this study was to investigate the suitability of the new exchange routine to simulate the bi-directional exchange between the stream and near-stream groundwater in a more realistic way in a bucket-type model. We used Panola as a test case to show how the model works and how adding the exchange routine affects the model simulations (note that the aim was not to create the best possible model for Panola). The HBV model was applied (a) in a fully lumped variant and (b) using sub-catchments that represent different parts of the catchment, and in particular the smaller storage capacity of the bedrock outcrop, and (c) using these sub-catchments and the new exchange routine to additionally simulate the bi-directional fluxes between the groundwater and the stream. The model performance was evaluated for both streamflow and groundwater levels, with a focus on the performance during different seasons. This allowed an assessment of the value of the exchange routine to more realistically represent catchment functioning. The hydrological response of Panola varies strongly seasonally, allowing model evaluation during contrasting conditions (wet, dry, wetting up, and drying up). From previous modeling studies for Panola, it is known that streamflow responses during the transition periods (i.e., wet to dry and dry to wet conditions) are difficult to simulate. Peters, Freer, and Beven (2003), for instance, found a systematic overestimation of streamflow during wet winter periods and underestimation during dry summer periods. Only two sub-catchments (the bedrock outcrop and the upper catchment with the losing stream) were used in the calibration. Flow for the entire catchment was simulated by applying the parameterization of the calibrated sub-catchment and applied to the other sub-catchments for the evaluation period to test if the simulation of losing stream

conditions and inter-basin groundwater flow (i.e., leakage below the gauge) improved the simulation of streamflow for the entire catchment.

We expected that explicit representation of the limited storage capacity for the bedrock outcrop would improve model performance compared to the fully lumped version of the model. We also expected that the new exchange routine would improve the simulations of both streamflow and groundwater levels, particularly during the wetting up and drying down periods. Although we tested the exchange routine only for Panola, a small, well-studied catchment, we assume that the test for this data-rich catchment provides guidance on the value of an explicit representation of the bi-directional exchange between streamflow and groundwater for other, and larger, catchments.

2. Exchange Routine for the HBV Model

2.1. HBV Model Description

The HBV model is a simple bucket-type rainfall runoff model where water storage (e.g., in the snowpack, unsaturated zone, or saturated zone) is represented by buckets that are arranged in a sequence (Bergström, 1976, 1992, 1995; Lindström et al., 1997). Since its development in the 1970s, the HBV model has been used in numerous catchments around the world (Al-Safi & Sarukkalige, 2017; Azmat et al., 2020; Parra et al., 2018; Pool et al., 2019) and different model variants have been developed, which allow, for example, inclusion of lakes or delayed drainage of groundwater storage. The model is usually calibrated based on streamflow data.

Input data for the model are time series of precipitation, temperature, and potential evaporation. The precipitation and air temperature data can be distributed in the model by using elevation zones and lapse rates. Precipitation is divided into rain and snow based on a threshold temperature and snow melt is simulated by a degree-day model. Evaporation and groundwater recharge are simulated as functions of storage in the soil bucket. The standard structure of the HBV model uses nine parameters, plus five parameters if snow processes are included (the list of parameters can be found in Table S1 in the Supporting Information S1). For more details about the model, we refer to Bergström (1976); Bergström (1992, 1995), Lindström et al. (1997) and Seibert and Vis (2012).

2.2. Model Variants

2.2.1. Lumped

The simplest variant of the HBV model is fully lumped, where each storage is represented by a bucket that represents the entire catchment, that is, one soil bucket, one shallow groundwater, and one deeper groundwater bucket represent the unsaturated and saturated zone in the entire catchment. Outflow from the groundwater buckets depends on the storage (i.e., linear reservoir formulation) and is calculated in the response routine. The output from the response routine is distributed over several time steps in a routing routine to obtain the final simulated streamflow.

2.2.2. Distributed-ST

To represent spatially variable storage, the catchment is divided into multiple sub-catchments that are parameterized differently. The routed streamflow of all sub-catchments is summed to obtain the streamflow of the catchment outlet.

2.2.3. Distributed-ER

In the exchange routine (ER) variant of the model, the catchment is divided into at least two sub-catchments that can have a different parameterization. The sub-catchments are linked to each other via the response routine and the exchange routine (described below).

2.3. Exchange Routine

The newly developed groundwater-surface water ER allows for the simulation of bi-directional fluxes between the stream and groundwater. In many catchments around the world, there are some stream reaches

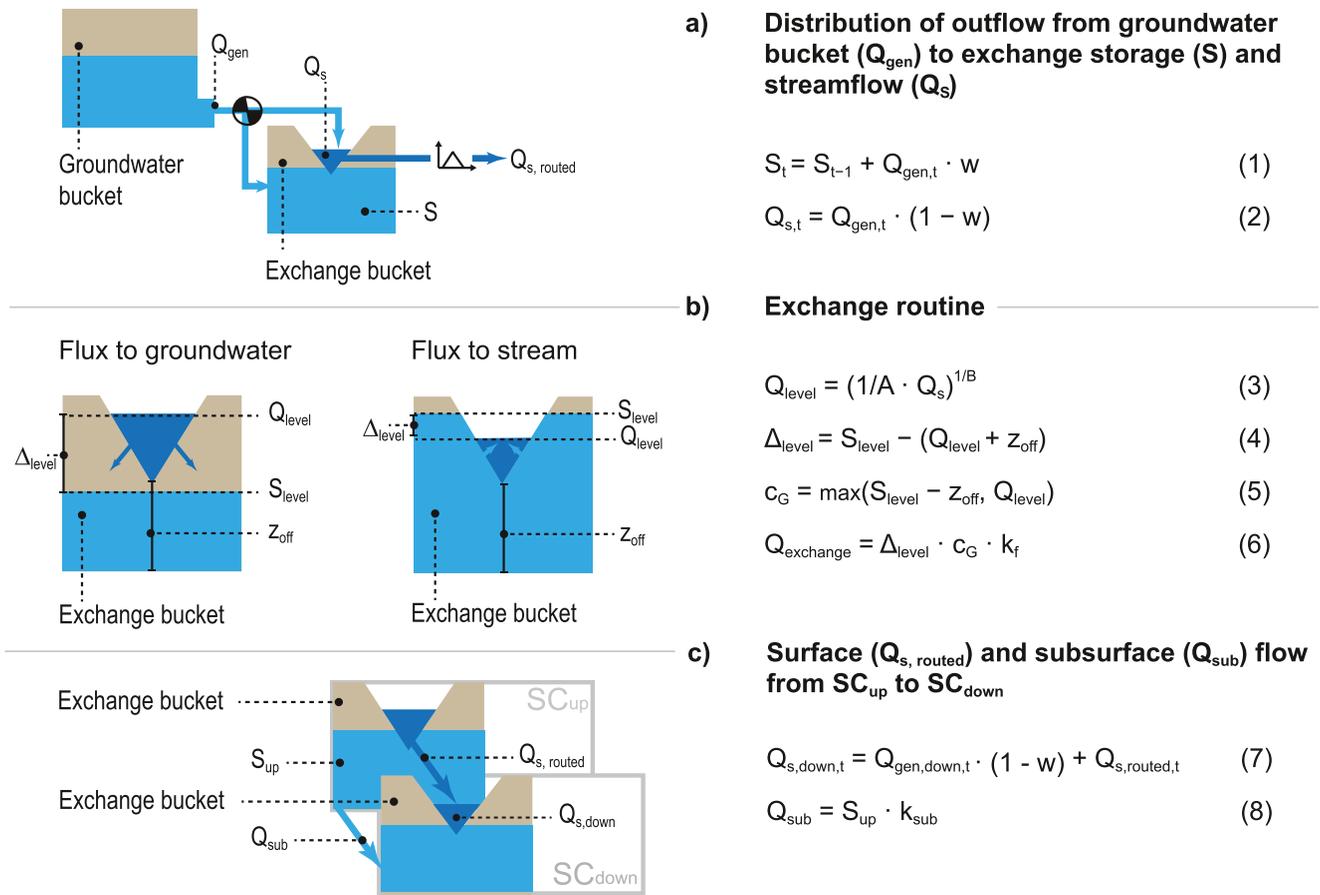


Figure 1. Conceptual diagram of the exchange routine and its parameters. The example shows the minimum setup with a headwater sub-catchment and a downstream sub-catchment. Panel (a) the response routine describes the flux from the groundwater bucket to the stream. The routing routine distributes the generated runoff and the runoff that is, remaining in the stream over the following time steps and is indicated with the triangular weighting function resulting in $Q_{s,routed}$ (mm/time step). The response and routing routine are the same as in the standard model Distributed-ST. w (-) is the fraction of the catchment that participates in the exchange and is indicated by the black-and-white checkered circle. Panel (b) the exchange routine calculates the direction of the flux between the groundwater (storage level S_{level} [mm]) and stream (stream level Q_{level} [mm]) using the inverse rating curve parameters A (1/time step) and B (-) that define the water level in the stream. The combined effect of all resistances and the variability in soil storage and aquifer characteristics that determine the rate of exchange are represented by the parameter k_f (time step/mm). The parameter z_{off} (mm) allows for the simulation of long lasting water deficits. The calculated exchange flux is either added to the stream or the groundwater of the downstream sub-catchment. Panel (c) subsurface drainage between the exchange buckets of the two sub-catchments (SC_{up} and SC_{down}) is calculated using the parameter k_{sub} (1/time step). See Table S1 for the list of parameters and parameter values.

(that can be represented by sub-catchments), where water infiltrates into the streambed and recharges the groundwater during some period of the year. The exchange routine allows for the simulation of these fluxes. The exchange routine also allows for subsurface drainage from an upper sub-catchment to a lower sub-catchment and thus for the simulation of inter-catchment groundwater flow.

The exchange routine (Figure 1b) connects the deeper groundwater storage in an upper sub-catchment to the stream in a lower sub-catchment. The flow generated from the upper sub-catchment partly goes to the exchange bucket in the lower sub-catchment (i.e., recharges the groundwater in the lower sub-catchment) or is added to the streamflow generated in the lower sub-catchment (Figure 1c). It is important to note that the exchange routine does not affect the simulated flow for the upper sub-catchment. The streamflow generated in the upper sub-catchment, however, affects the flow in the downstream sub-catchments because the inflow from the upstream sub-catchment is partly added to the streamflow generated within the downstream sub-catchment and partly added to the groundwater storage in the downstream sub-catchment (Figure 1a).

The exchange routine has three parameters that allow estimation of the gradient between the groundwater and the stream based on an inverse rating curve (rating curve parameters A [-] and B [-]). The combined effect of all resistances and the variability in soil storage and aquifer and streambed characteristics that determine the rate of exchange is represented by parameter k_f . An additional parameter (w) describes the fraction of the sub-catchment that participates in the exchange. Note that this parameter was not calibrated in our study to reduce the number of parameters that need to be calibrated and was instead derived from information about the extent of the riparian zone (see Section 4.2 Model Calibration).

When implemented in the HBV-model, the exchange routine is only effective if there is streamflow from an upstream sub-catchment to a lower sub-catchment. In the upstream sub-catchment, the streamflow comes by definition always from the groundwater, that is, it is determined by the response routine for this sub-catchment. Subsurface drainage from the upstream sub-catchment to the downstream sub-catchment (or an upper reach to a lower reach) is implemented by adding the drainage from the exchange bucket in the upstream sub-catchment to the exchange bucket of the downstream sub-catchment (Figure 1c). This is parameterized in the model by a constant recession coefficient, k_{sub} . This water can leave the catchment below the gauge (i.e., inter-basin groundwater flow), but it can also be forced to return to the river in the same sub-catchment by setting k_{sub} to zero.

There is one more parameter z_{off} , which enables catchment memory and allows simulation of long lasting water deficits (e.g., during multi-year droughts) because it works as a threshold (Figure 1b). Only when the water level in the exchange bucket is exceeded, it is connected to the stream. Hence, a deficit can build up in the exchange bucket that reflects previous dry spells. This parameter was not used in this study (by setting it to zero) but could be useful when long-term data are used and there is information on the position of the riverbed relative to the aquifer.

3. Study Site and Data

3.1. Panola Mountain Research Watershed

The Panola Mountain Research Watershed (Panola) is a 41 ha forested catchment located in Georgia, USA (Figure 2). The total topographic relief in the catchment is 56 m; the average slope is 18%. The climate is humid-continental to subtropical. The annual average precipitation of 1,250 mm/y (water years 1986–2015) is distributed uniformly throughout the year. In winter, synoptic weather systems result in low-intensity and long-duration events, while in summer, most precipitation events are convective. Less than 1% of the precipitation is snow or sleet. The high evapotranspiration demand in summer leads to a highly seasonal streamflow response and an extended period of low baseflow (Figure 3) (Aulenbach & Peters, 2018).

The bedrock in the southwestern part of the catchment (SC1 and SC2 in Figure 2) consists of Panola Granite (Peters, Freer, & Beven, 2003). In other parts of the catchment amphibolite (Clairmont member of the Stonewall) is found as well. Partially vegetated outcrops cover 10% of the area; the largest outcrop (3.6 ha) is located in the southwestern part of the catchment (SC1; Figure 2). The red, clayey soil that developed in the colluvium and residuum is classified as Ultisol. Inceptisols can be found in the colluvium or recent alluvium in highly eroded landscape positions. The saprolite is 0–6 m thick where it overlays granodiorite bedrock and 5–20 m where there are amphibolites (Burns et al., 2001). However, in many locations it is relatively thin. The dynamic storage estimated from a baseflow-storage relationships is about 550 mm (Aulenbach & Peters, 2018).

Hillslopes comprise most of the catchment (>75%) and have relatively shallow soils (generally <2 m). Perched water tables on the hillslopes and significant lateral subsurface stormflow above the soil-bedrock interface occur only during large events (>55 mm) (Tromp-van Meerveld & McDonnell, 2006a, 2006b).

The (<5 m wide) riparian zone in the upper catchment (SC2) is characterized by deep soils (up to 6 m deep) and a permanent water table. Headwater streams in the southeastern parts of the catchment have deep gullies. The riparian zone in the lower catchment area (SC6), where the stream is perennial, is wider and characterized by a more persistent high water table.

Baseflow at the outlet of the catchment is sustained throughout the year, even during longer dry spells. Most of the annual streamflow at the lower gauge is generated from the lower catchment area (roughly SC6 in

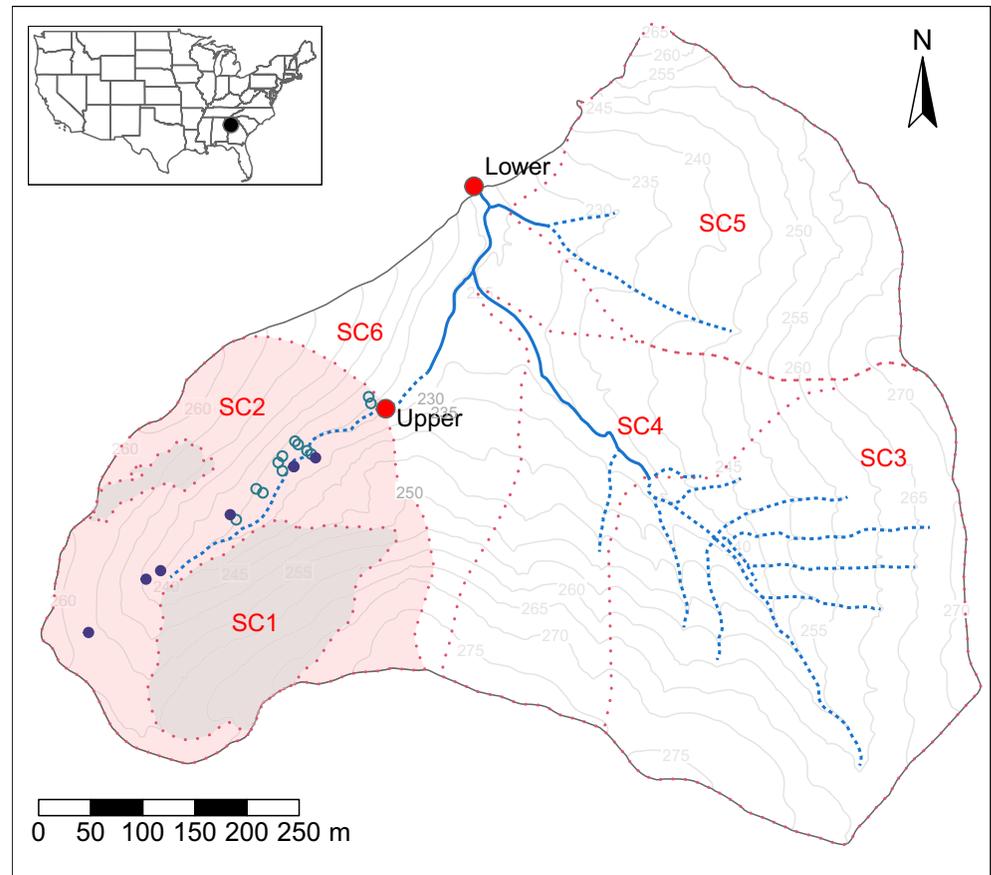


Figure 2. Map of the Panola Mountain Research Watershed, with the location of upper- and lower-gauge (red circles) and the groundwater wells on the hillslope (6 wells, closed violet circles) and in the riparian zone (13 wells, green open circles) used in this study, as well as the location of the sub-catchments for the distributed model variants (dotted red lines). The area of the upper catchment (SC1 plus SC2) is shown in light red, and includes the main bedrock outcrop area (SC1), which is shown in gray. The perennial (solid blue lines) and intermittent reaches (dashed blue lines) of the stream network are shown as well. The bedrock outcrop does not have a stream channel.

Figure 2), where the stream is perennial (Aulenbach et al., 2021). The perennial flow in this area is likely sustained by recharge from the hillslopes through the bedrock and the upstream riparian areas (Aulenbach et al., 2021). The eastern tributary (roughly SC5, in Figure 2) flows seasonally, while the southeastern tributary (SC3 and part of SC4 in Figure 2) above the perennial stream only flows for short periods during and after rainfall events. The stream above the upper gauge (draining both SC1 and SC2) flows in direct response to rainfall events, with flow lasting longer after rainfall events that cause connectivity between the hillslopes and the stream (van Meerveld et al., 2015) and can last up to several weeks during very wet periods (Peters, Freer, & Aulenbach, 2003).

Streamflow responses to rainfall events are flashy (Figure 3). Geochemical analyses showed that streamflow predominantly consists of flow from the bedrock outcrop and the riparian zone (Burns et al., 2001). Peters and Ratcliffe (1998) show that “new” water could contribute 95% of the streamflow at the upper gauge, which comes largely from the bedrock outcrop. Burns et al. (2001) showed that the outcrop contributed 50%–55% of peak streamflow at the upper gauge for one event and 80%–85% for another event. Stormflow from the upper catchment (SC1 and SC2) is about a third larger than stormflow from the other headwater catchments (Aulenbach et al., 2021). For rainfall events larger than about 15 mm, a flood wave propagates quickly downstream from the main bedrock outcrop. Streamflow at the upper gauge (250 m downstream from the outcrop, see Figure 2) peaks 15–20 min after initiation of flow in the channel at the base of the outcrop and about 20 min later at the lower gauge (200 m from the upper gauge). The time to peak flow is longer when the catchment is dry, suggesting more losses to the streambed during dry conditions (Peters,

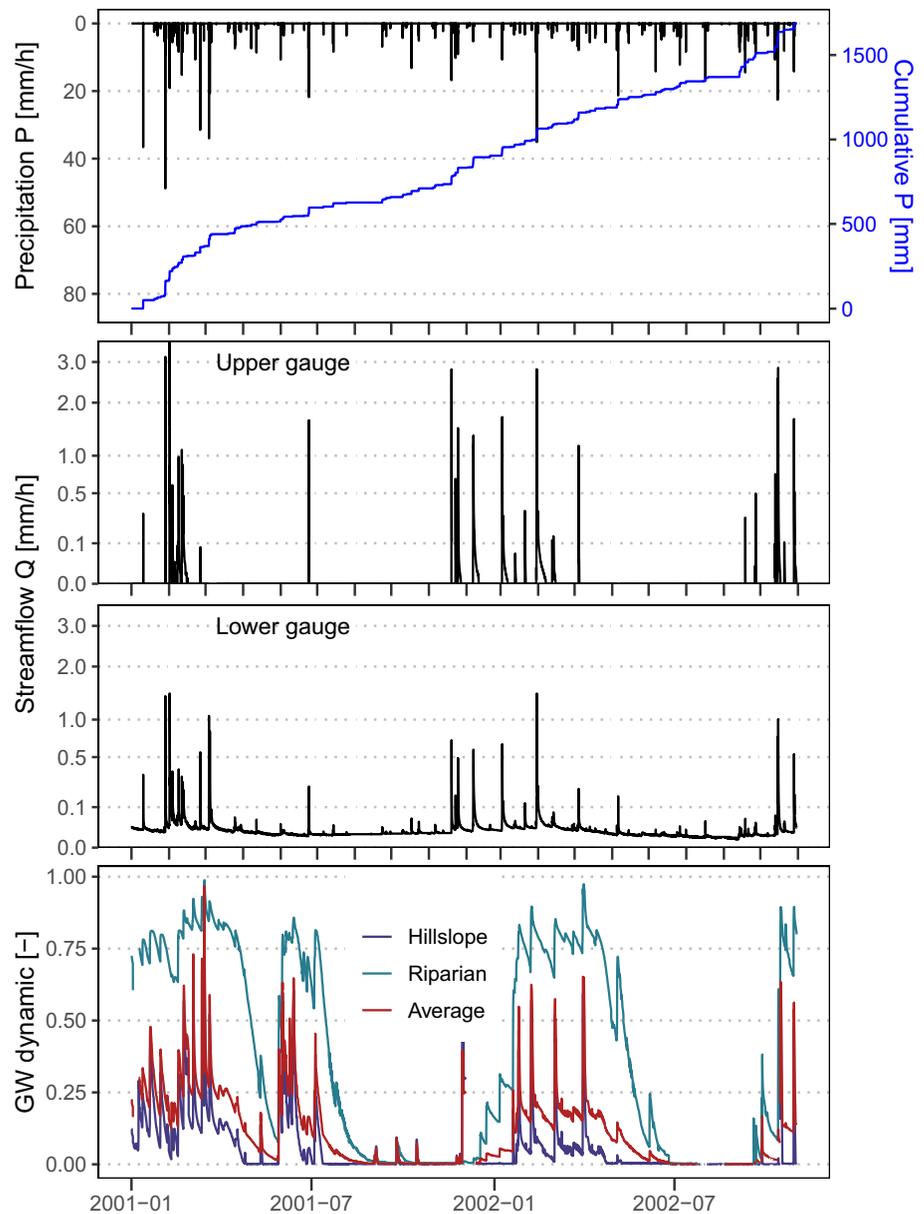


Figure 3. Hourly precipitation (P [mm/h]), cumulative precipitation (P [mm]), and streamflow (Q [mm/h]) at the upper and lower streamflow gauge, as well as the mean scaled groundwater response for riparian and hillslope groundwater wells and the catchment weighted average (0 minimum level, 1 maximum level) for the May 2001–October 2002 study period. Note that streamflow is plotted on a square root scale to better show the variation in low flows, including zero flows.

Freer, & Aulenbach, 2003). The riparian aquifer in the upper catchment receives substantial recharge from the bedrock outcrop. Burns et al. (2001) estimated that for one event recharge of the riparian aquifer by runoff from the outcrop was 7%–20% of rainfall, for another event it was 11%–28% of rainfall.

It is known that groundwater flow occurs below the upper gauge, meaning there are groundwater losses across sub-catchment boundaries, but the amount of groundwater flow below the lower gauge is assumed to be small based on the high and relatively stable solute concentrations in a deep well near the outlet (Tromp-van Meerveld et al., 2007). Groundwater in this deep well is about 26 years old (Burns et al., 2003). The mean transit time of stream water was estimated to be about 4.7 years (Peters et al., 2014). A simple

Table 1

Hydro-Meteorological Conditions for the Calibration and Validation Period and the Long Term Average as Reported in Aulenbach et al. (2021) for the Upper Catchment (SC1 and SC2) and the Entire Catchment

Period	P (mm)	Q_{lower} (mm)	RR_{lower} (-)	Q_{upper} (mm)	RR_{upper} (-)	ZQ (-)
Calibration (May 2001–Oct 2002)	1,674	290	0.17	149	0.09	0.91
Calibration (water year 2002)	907	165	0.18	71	0.08	0.92
Validation (water year 2003)	1,728	552	0.32	278	0.16	0.63
Long term (water years 1986–2015)	1,250	358	0.29	212	0.16	0.79

Note. Precipitation (P), streamflow at the lower gauge (Q_{lower}), runoff ratio at the lower gauge (RR_{lower}), streamflow at the upper gauge (Q_{upper}), runoff ratio at the upper gauge (RR_{upper}), and the fraction of time with zero streamflow at the upper gauge (ZQ).

back of the envelope water balance calculation suggested that flow through the bedrock contributed at least 14%–21% of streamflow (Tromp-van Meerveld et al., 2007).

Panola is a good study location to test the new exchange routine because streamflow data are available for two locations in the catchment: at the outlet of the 10 ha upper catchment, where flow is intermittent and at the outlet of the entire 41 ha catchment, where flow is perennial. In addition, groundwater level data are available for several transects across the riparian zone and up the hillslope (Peters, Freer, & Aulenbach, 2003) that can be used to test the functioning of the different model variants. The large bedrock outcrop and infiltration into the headwater stream are perhaps not so common but infiltration into the stream and losing stream sections occur also in other locations. The application to Panola, therefore, allows us to test the usefulness of the new exchange routine.

3.2. Data

We ran the model for the May 2001 to October 2002 study period at an hourly time step and used a warm-up period of four months (January–April 2001). This period consisted of a late spring transition phase, a summer with low streamflow, a fall and early winter transition phase to high flow conditions and a subsequent drying phase with lower streamflow, followed by a second wetting up period. The period was chosen because of the availability of good quality groundwater level data and field observations during this period. The period was drier than average (Table 1); the water year 2002 was characterized as a drought (D2). However, the period represents the typical streamflow and groundwater dynamics (Freer et al., 1997; Peters, Freer, & Beven, 2003). Water year 2003 was wetter than average and used for the validation.

For this study, we used hourly precipitation and air temperature data measured in the catchment, as well as monthly averages of potential evapotranspiration, calculated based on the hourly air temperature and solar radiation data using the Priestley-Taylor equation with a vapor pressure deficit correction (Aulenbach, 2017; Aulenbach & Peters, 2018). In addition, we used the measured streamflow (averaged to hourly resolution) for the two gauging stations (Staudinger et al., 2021).

We used the hourly groundwater levels measured in wells that are mainly located along transects in the upper catchment (Figure 2). In this study, we only used the groundwater data from the 19 (of the existing 41) wells that passed the initial data quality control, that is, plausible groundwater behavior for the well. The groundwater-level dynamics vary considerably between riparian and hillslope sites (Figure 3). At the hillslope sites, the groundwater responds quickly to rainfall events but the response is short-lived and most wells are dry between events, whereas in the riparian zone the response is dampened and most wells contain water throughout the year (Peters, Freer, & Aulenbach, 2003; van Meerveld et al., 2015). We scaled the groundwater levels based on the minimum and maximum measured level during the study period and calculated the average scaled water level for the riparian zone and the hillslopes to minimize problems with the averaging of censored data (i.e., when the groundwater level drops below the depth of the well) and to avoid an overly strong influence of a particular well on the calculated average groundwater level. The weighted average relative groundwater level was calculated for the upper catchment, where the weights reflect the size of the two landscape units (Equation 1).

$$\overline{G_{catch}} = 0.83 \cdot \overline{G_H} + 0.17 \cdot \overline{G_R} \quad (1)$$

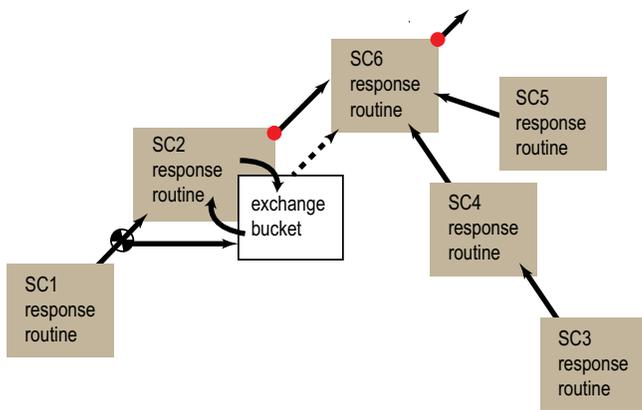


Figure 4. Implementation of the exchange routine in the model for the Panola catchment. The black arrows indicate the streamflow from one sub-catchment to the next, the dashed arrow the subsurface drainage from SC2 to SC6. The exchange routine is indicated with the curved arrows between the response routine and exchange bucket in SC2. The red circles indicate the location of the upper (SC2) and lower (SC6) streamflow gauge. See Figure 2 for the map with the location of the different sub-catchments.

where \overline{G}_{catch} is the weighted mean relative groundwater level used for model calibration, \overline{G}_H is the arithmetic mean of the relative groundwater levels for the hillslope wells and \overline{G}_R is the arithmetic mean of the relative groundwater levels for the riparian wells (Staudinger et al., 2021).

4. Model Application to Panola

4.1. Model Set Up

We separated Panola into six similar sized sub-catchments. The high spatial variability in hydrological responses, and in particular the limited storage and fast response from the bedrock outcrop (Figure 2) was accounted for by treating the outcrop as a sub-catchment with a different model parameterization. Two of these sub-catchments (SC1 and SC2) drain to the upper streamflow gauge. For the distributed-ER model variant, we simulated a flux from SC1 into the exchange bucket of SC2 and subsurface drainage via the exchange bucket from SC2 into SC6, which represents the lower catchment area with the perennial stream that drains to the lower gauge (Figure 1 and 4).

4.2. Model Calibration

We calibrated the three different variants of the HBV model (lumped, distributed-ST, and distributed-ER) separately by matching the observed and simulated streamflow and groundwater levels for sub-catchments SC1 and SC2 only. The model variants were calibrated separately using a genetic algorithm (Seibert, 2000), which is implemented in the HBV-light software (Seibert & Vis, 2012). This means that parameter sets were optimized by an evolution-like procedure starting with a pool of randomly generated parameter sets (see parameter ranges in the Table S1). The calibration procedure was stopped after 5,000 model runs.

For the distributed model variants (distributed-ST and distributed-ER), the parameter FC (maximum soil water storage [mm]) was calibrated separately for the outcrop (SC1) and the remainder of the upper catchment (SC2). We used a much smaller parameter range for FC for the outcrop (0.01–100 mm instead of 200–1,000 mm) to force a fast response. Because snow is rare in the catchment, we did not include the snow routine of the HBV-model. Precipitation and air temperature data were not distributed using elevation zones and lapse rates due to the limited relief at Panola. Thus, in total there were ten parameters that were optimized for the distributed-ST variant, compared to the nine parameters for the lumped variant.

As optimization criteria, we used the Nash-Sutcliffe model efficiency (Nash & Sutcliffe, 1970), NSE_Q , for streamflow and the Pearson correlation coefficient (Pearson, 1896), COR_{GW} , for the groundwater dynamics. The values of NSE_Q range from $-\infty$ to 1, where 1 would be a perfect match between simulation and observation. NSE_Q evaluates both streamflow dynamics and volume, but emphasizes peak flows as it is based on squared errors. COR_{GW} was chosen for the groundwater simulations because the focus was on the dynamics rather than absolute values. COR_{GW} ranges from -1 to 1, with -1 indicating a perfect negative correlation and 1 a perfect positive correlation.

For the calibration of the response routine (all model variants), we used the weighted average relative groundwater level for the catchment (i.e., catchment average groundwater level). We tested if using an arithmetic mean rather than a weighted average based on the size of the landscape unit changed the calibration results but this had little effect and therefore we show only the results based on the weighted average of the scaled groundwater levels.

To test whether the model variants can reproduce the groundwater and streamflow dynamics simultaneously, we performed a series of calibration trials using different weights for the streamflow and groundwater simulations to determine the overall goodness of fit. More specifically, we increased the optimization weight that was given to the streamflow (i.e., NSE_Q) incrementally in steps of 10% from 0% to 100%, and

correspondingly decreased the weight that was given to the groundwater (i.e., COR_{GW}) from 100% to 0%. To consider parameter uncertainty and to reflect the fact that there is not one single optimal parameter set (Beven, 2011), each calibration was repeated 25 times independently, resulting in 25 parameter sets for each model variant and each optimization weight for streamflow.

For the distributed-ER model variant, the exchange routine parameters were also optimized during the calibration (i.e., 13 parameters were optimized). The arithmetic mean of the relative groundwater levels for the riparian zone were used for the calibration of the parameters of the exchange routine. The optimization weight for groundwater (i.e., COR_{GW}) was equally divided between the simulations of the catchment average and riparian groundwater responses. Thus for the calibration with a 10% optimization weight for streamflow, the weight for the riparian groundwater simulation (for the determination of the exchange routine parameters) was 45% and the weight for the catchment average groundwater simulation (for the determination of the response routine parameters) was also 45%.

Parameter w , describing the fraction of the sub-catchment that participates in the exchange, was set to 0.17 to reduce the number of parameters that needed to be calibrated. The fixed value corresponds to the relative area of the riparian zone, assuming that it relates to alluvial storage. The sensitivity of the results to the value for this parameter is shown in Figure S1. Model parameter, z_{off} , was also not used in this study and set to zero because long-term memory effects were not expected for the short period for which the model was calibrated.

4.3. Model Evaluation

The three different model variants were evaluated based on several criteria, for both the entire calibration period and specific seasons (January–March (JFM), April–June (AMJ), July–September (JAS), and October–December (OND)). We also compared the performance of the three model variants to each other. In particular, we compared the performance of the lumped variant to the distributed-ST variant, to assess the effects of spatially variable soil water storage, and compared the distributed-ER variant to the distributed-ST variant to determine the effect of the simulation of groundwater-surface water exchange and subsurface drainage. We, furthermore, assessed the uncertainty of the FC parameter for the different model variants.

Streamflow and groundwater simulation performance: The performance of the streamflow simulation was based on the model efficiency, NSE_Q . The performance of the groundwater simulations was based on the Pearson correlation coefficient COR_{GW} .

For the distributed-ER model variant, the objective function of the groundwater COR_{GW} included both the goodness of fit for the catchment average groundwater (for the calibration of the response routine) and the riparian groundwater (for the calibration of the exchange routine). We, therefore, compared the Pearson correlation coefficients for both the catchment averaged groundwater levels and the averaged riparian groundwater levels.

Even though the focus of the study was to show how the exchange routine affects the model simulations and not on finding the best model for Panola, we also ran the calibrated model variants for the wetter validation period (October 2002–October 2003) and calculated the performance measures for the streamflow and the groundwater for this period (see Figures S3, S4, S5 and S6).

Model flexibility: It is generally assumed that a model with a good performance for both groundwater and streamflow represents reality better, but previous studies have shown that an improvement in the simulation of groundwater levels or storage may lead to a deterioration of the performance of the streamflow simulations (Efstratiadis & Koutsoyiannis, 2010; Seibert, 2000). Therefore, we assessed if there was a trade-off between streamflow and groundwater model performance in scatter plots and checked if there was a pareto-optimum between the two performance measures. These performance measures are calculated from the 10% increments in the optimization weight for streamflow. If there is an increase in performance of one measure while the performance for the other measure also increases, then the model can take advantage of more information.

Streamflow performance for the lower gauge: How well can the model simulate streamflow dynamics at the lower gauge after calibration for the upper gauge? If the processes are similar throughout the catchment and

the model represents these processes well, then the performance should be similar for the upper and lower gauge. If subsurface drainage occurs below the upper gauge and is important for flow at the lower gauge, then the performance of the distributed-ER model variant should be better than for the distributed-ST model for both the upper gauge and the lower gauge because the distributed-ER model variant accounts for this drainage. To test this, the calibrated parameter sets were used to simulate streamflow at the lower gauge. We assumed a similar catchment functioning as SC2 for all the other sub-catchments (SC3-6). Thus, for the distributed-ST and the distributed-ER variants the calibrated parameter values from SC2 were transferred to all other sub-catchments (Figure 2). We admit that catchment functioning for SC3-SC6 is unlikely to be fully similar to that of SC2 (Aulenbach & Peters, 2018; Aulenbach et al., 2021) but we still use this approach because the focus of this analysis is on the effects of the exchange routine on the simulations. Furthermore, all other approaches require the introduction of additional parameters that need to be optimized.

For the distributed-ER model variant, subsurface flow was only allowed to occur via the exchange bucket from SC2 into SC6 (Figure 4). We did not simulate a similar groundwater-surface water exchange for the other sub-catchments because infiltration of streamwater into the riparian aquifer is most important for SC2 because this sub-catchment is affected by the flow from the bedrock outcrop (SC1). Thus, the calibrated parameters for the distributed-ST model variant of SC2 were applied to SC3-6 for both the distributed-ST and the distributed-ER variants.

5. Results

5.1. Model Performance

The difference between the model performance of the lumped model variant and the distributed variants was large (Figure 5a). The lumped model variant resulted in a much poorer performance for the simulation of the streamflow than the spatially distributed model variants (distributed-ST and distributed-ER) that accounted for the difference in the soil storage capacity of the bedrock outcrop and the remainder of the catchment. The distributed-ER variant performed better for streamflow simulations than the distributed-ST variant but the performance for the groundwater simulations was similar (Figure 5a).

The results for the calibrations with the different optimization weights for streamflow indicated a clear trade-off between groundwater and streamflow performance for the lumped model variant and for the distributed-ER variant. The performance for the streamflow simulations for the lumped model variant increased and the performance for the groundwater decreased when the weight for streamflow increased, up to a weight of 50% for streamflow (Figure 5b). When the weight for streamflow in the calibration was more than 50%, the performance for both streamflow and groundwater simulations increased as the weight for streamflow increased. For the distributed-ST variant there was no such systematic increase or decrease visible for the entire period (Figure 5b), but could be seen for the winter period (JFM; Figure 6).

The performance of the streamflow simulations decreased as the catchment dried out (compare JFM [wetting up winter], AMJ [wet to drying spring], JAS [drying up summer]; Figure 6). In summer (JAS), the two distributed model variants performed poorly for the streamflow (Figure 6). Note however, that the performance for streamflow during this period only represents the dynamics of the streamflow in September because there was no streamflow at the upper gauge during July and August. As streamflow ceased in the upper catchment during the dry fall (OND), the model performance for streamflow NSE_Q could not be determined.

We also found large differences when comparing the performances for each single month (see Figure S7) but these results need to be interpreted with care due to the limited explanatory power of a Nash-Sutcliffe efficiency when it is calculated over short periods. The efficiency, furthermore, depends on the duration that the stream was flowing.

The model performance decreased for the validation period (median NSE_Q of 0.21 and 0.35 for the distributed-ST and distributed-ER variants, respectively). The model performance for streamflow was better for the distributed-ER variant than for the distributed-ST variant for all seasons. The differences in model performance between the distributed-ST and distributed-ER model variants were larger for the validation period than for the calibration period (Figure S5).

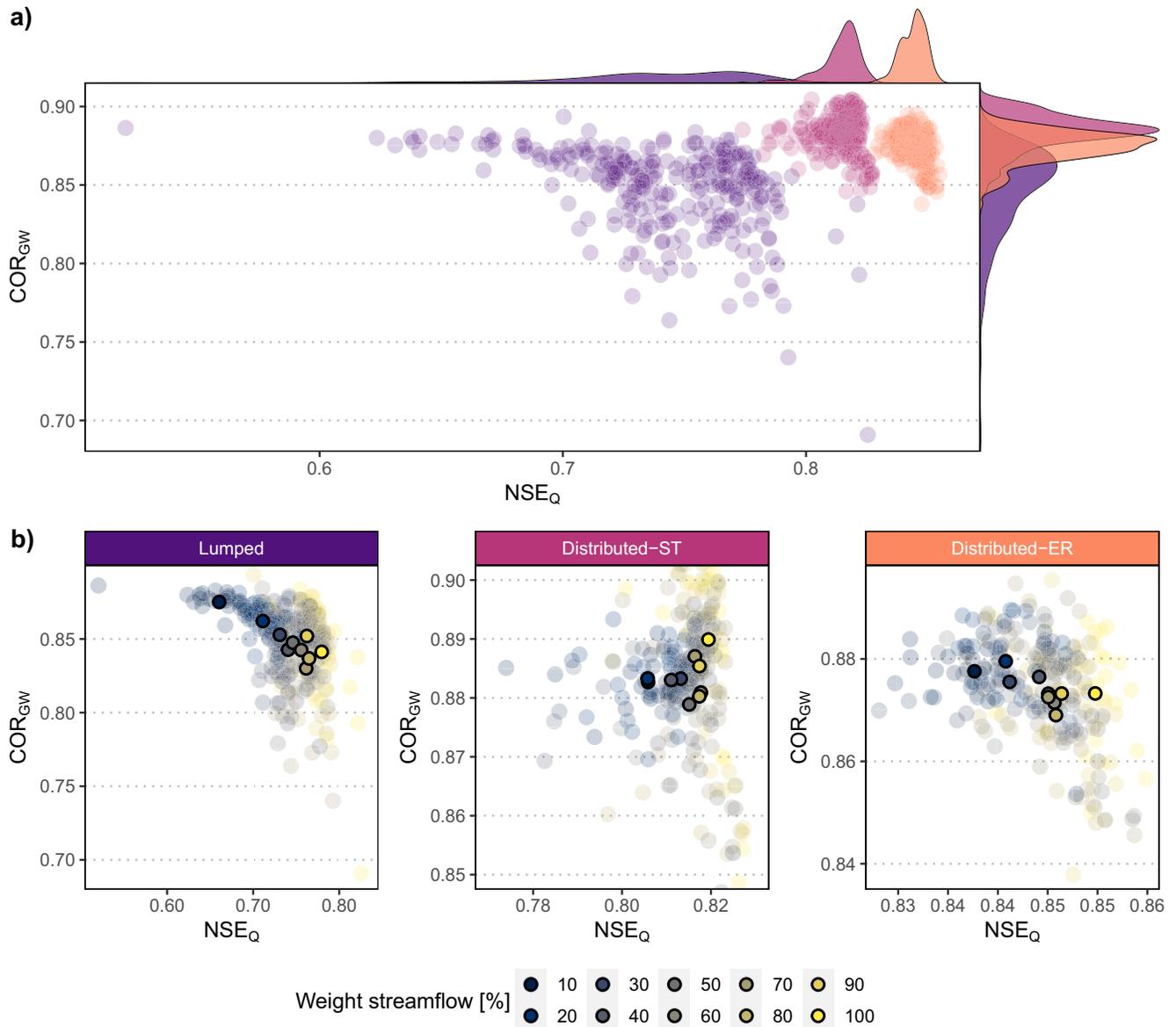


Figure 5. Model performance for streamflow at the upper gauge (NSE_Q) and catchment average groundwater (COR_{GW}) for the three model variants and different weighting schemes. (a) shows the comparison of all model variants with violet dots for the lumped, pink dots for the distributed-ST, and orange dots for the distributed-ER model variant, as well as the marginal distributions of NSE_Q and COR_{GW} for each model variant at the plot margins. (b) zooms into the model performances for each model variant. In these plots the weight of the streamflow objective function in the calibration is indicated by the color gradient: blue to yellow. The circles with the black outline show the mean model performance for the 25 different parameterizations for each weighting scheme.

5.2. Evaluation for the Lower Gauge

As expected, the simulations of the streamflow at the lower gauge were poorer than for the upper gauge when the model parameters derived from the calibration for the streamflow at the upper gauge (100% optimization weight for streamflow and 0% for groundwater) were used to simulate streamflow at the lower gauge without any re-calibration (Figure 7a). This was the case for the full period, as well as for winter, but only partly for spring, and not for summer (JAS; Figure 7b). NSE_Q values for the simulated streamflow at the lower gauge were between 0.4 and 0.6 during summer (JAS) but close to zero for the upper gauge for the distributed-ER variant and even negative for the distributed-ST variant. Note however, that even though the model performance for the streamflow simulation for the lower gauge was higher than for the upper gauge for the summer period, the model performance was poor for both. The slightly better performance for the

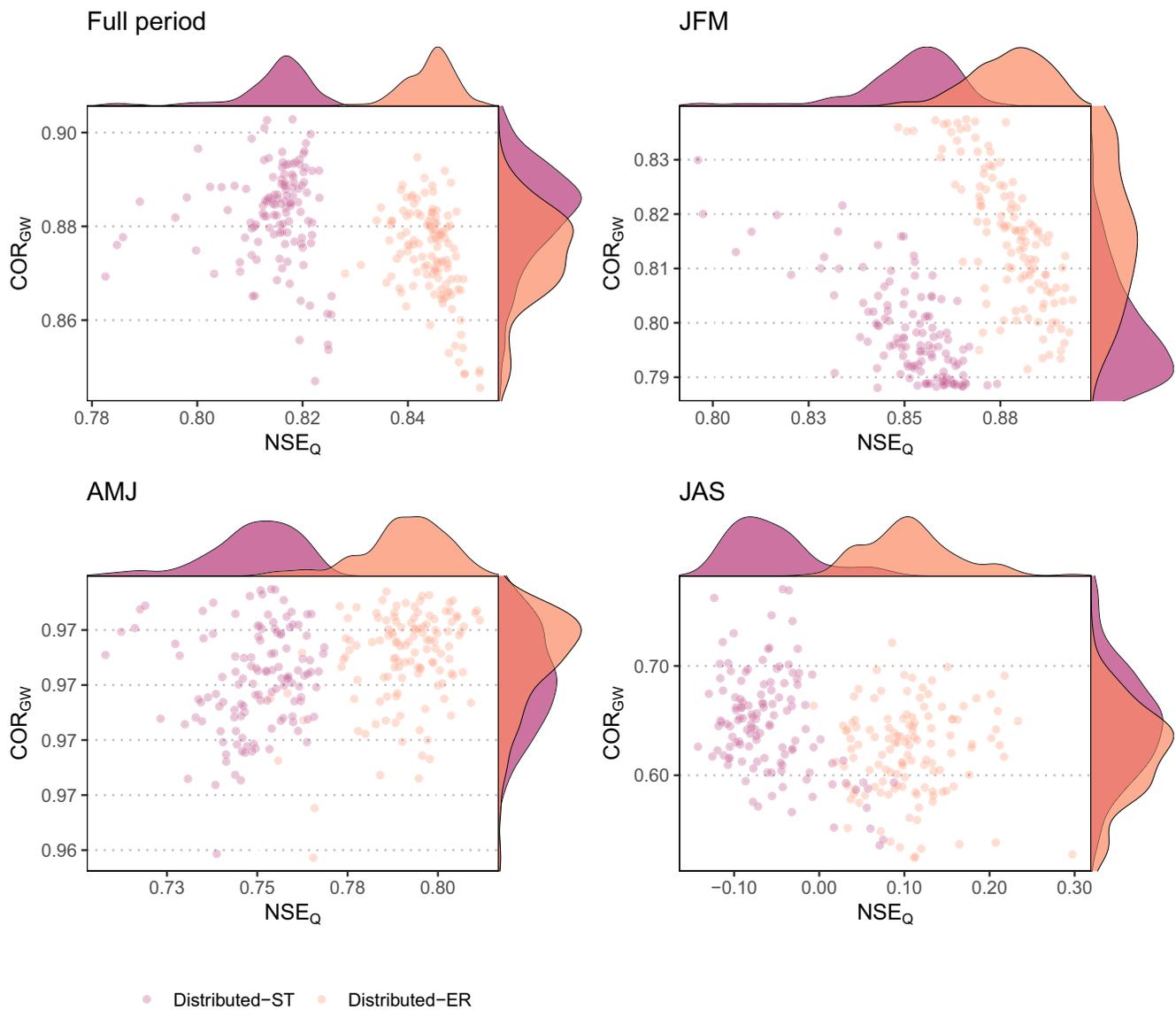


Figure 6. Model performance for streamflow at the upper gauge (NSE_Q) and groundwater dynamics (COR_{GW}) for different seasons for the two distributed model variants for all calibrations with an optimization weight of at least 30% for both streamflow and groundwater. Each circle represents one model simulation. For July–September (JAS), there was only flow during September. Since streamflow ceased in the upper catchment during the dry fall (October–December [OND]), the model performance for streamflow could not be determined for this season. The density plots on the axes highlight the marginal distributions for both NSE_Q and COR_{GW} . There was a similar decrease in model performance as the catchment dried out for the lumped model variant, but since the lumped variant performed poorer (see Figure 5) it is not shown in this comparison. For the results for the individual months, see Figure S7.

lower gauge may be attributed to the months July and August for which there was streamflow for the lower gauge but not at the upper gauge due to the contributions from the other contributing sub-catchments (Figure 7b). The difference in model performance for the two distributed model variants, however, was smaller for the lower gauge than for the upper gauge.

5.3. Example Time Series

The time series of the observed and simulated streamflow in response to a large rainfall event in late March 2002 and the following recession period in April 2002 were selected as an example to illustrate the functioning of the exchange routine for the Panola catchment. During the recession in April, the streamflow simulations (with the 100% optimization weight to streamflow) were better (NSE_Q) for the distributed-ER

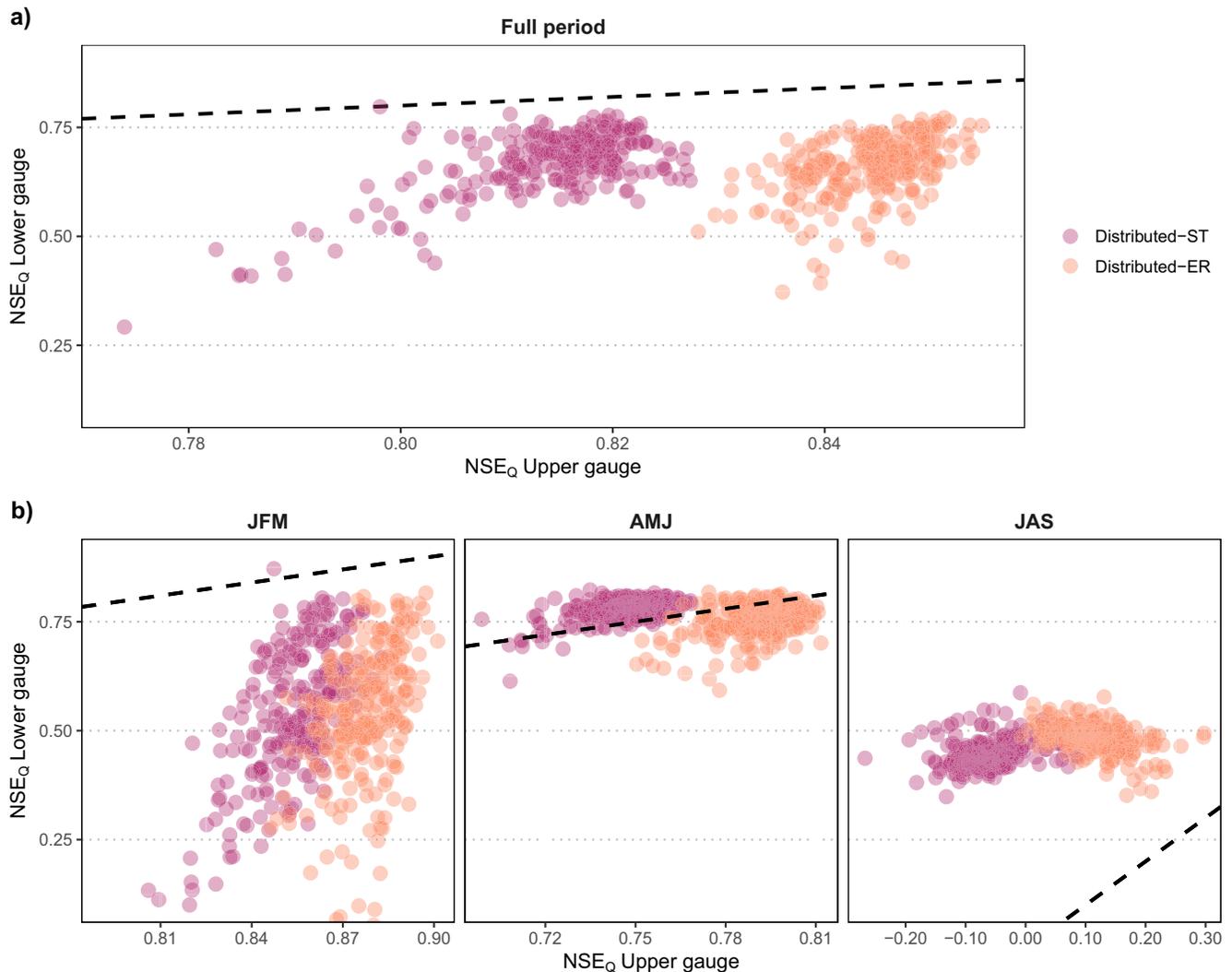


Figure 7. Streamflow performance (NSE_Q) for the upper and lower gauge for all optimization weights for streamflow in the calibration for the full period (a) and seasonally (b). The black dashed line represents the 1:1 line. For the results for the individual months, see Figure S7.

model variant than the distributed-ST variant for both the upper and lower gauge (Figure 8). Even though neither of the model variants simulated the observations particularly well in terms of the size of the peaks, the zero flow occurrence at the upper gauge and its timing were better simulated by the distributed-ER variant than the distributed-ST model variant. The distributed-ST variant simulated a streamflow response also for some rainfall events (e.g., April 25, 2002 in Figure 8), for which such a response was not observed at the upper gauge. This can be explained by water that infiltrated in the streambed, which was not simulated in the distributed-ST variant (Figure 8b).

The baseflow at the lower gauge was simulated better for the distributed-ER variant than the distributed-ST variant, which was a direct consequence of the simulation of subsurface drainage from the upper catchment. The simulated flow for the distributed-ST variant became almost zero, similar to the observed streamflow at the upper gauge, while streamflow at the lower gauge was sustained throughout the study period. During August, October, and November (2001) and during July and August (2002), all the simulated flow from the upper catchment to the lower catchment occurred via subsurface flow (there was no streamflow observed at the upper gauge during this period either). However, these simulated contributions were small in absolute terms (about 1% of the streamflow at the lower gauge). The simulated flux of the water that leaves the upper catchment as subsurface drainage is, however, considerable during recession conditions

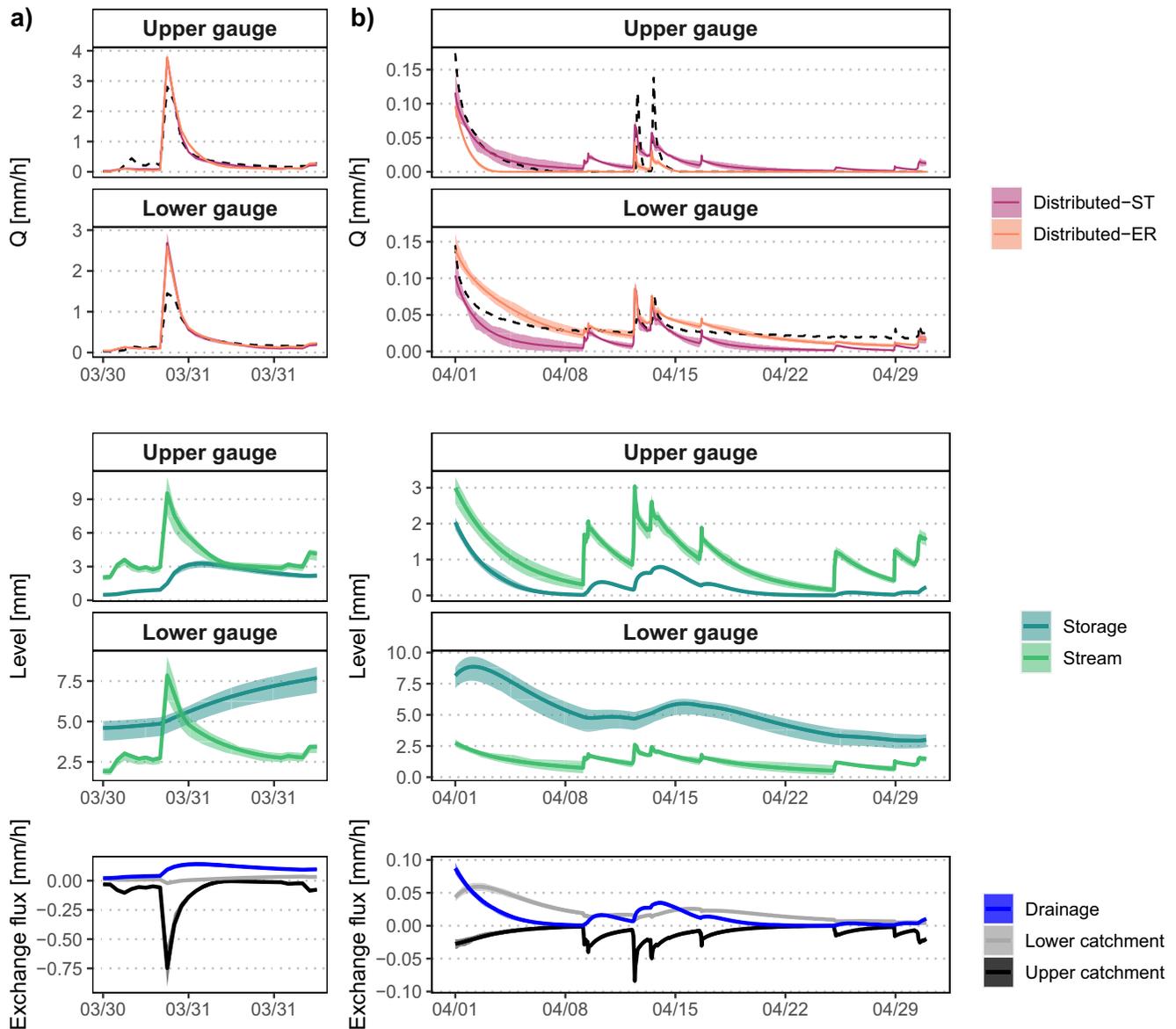


Figure 8. Upper panel: Observed and simulated hydrographs at the upper and lower gauge for the distributed-ST and the distributed-ER model variants for (a) a large event at the end of March 2002 and (b) the recession period in April 2002 for the calibration with a 100% optimization weight for streamflow. The dashed black lines indicate the observed streamflow, the colored lines indicate the average simulated streamflow from all parameter sets, and the shaded areas indicate the range of the simulations for the different parameter sets. Middle panel: Simulated stream stage and storage level in the exchange bucket for the upper (SC2) and the lower (SC6) catchment. Lower panel: The exchange fluxes in the upper and lower catchment (in black and gray lines) together with the subsurface drainage from the upper to the lower sub-catchment (blue line). Negative values indicate a flux from the stream to the groundwater, positive values indicate a flux from the groundwater to the stream. All fluxes are expressed per catchment area (mm/h), but note the different scale for (a) and (b).

(Figure 9), as well as in the wetting up transition period (December 2001 and September 2002) and therefore essential for a good simulation during these periods.

During the large event at the end of March (Figure 8a), the exchange flux from the stream to the groundwater in the upper catchment increased and then decreased to almost zero for a short period of time (for some model parameter sets the flux even reverted, representing a flux from the groundwater to the stream, as can be seen in Figure S2 in the Supporting Information S1). Most of the simulated exchange flux during the study period, however, was from the stream to the groundwater. The simulated subsurface drainage (from SC2 to SC6) caused the riparian groundwater levels in SC2 to be lower than would otherwise be the case, and hence maintained a gradient toward the groundwater. The simulated subsurface drainage to SC6 varied

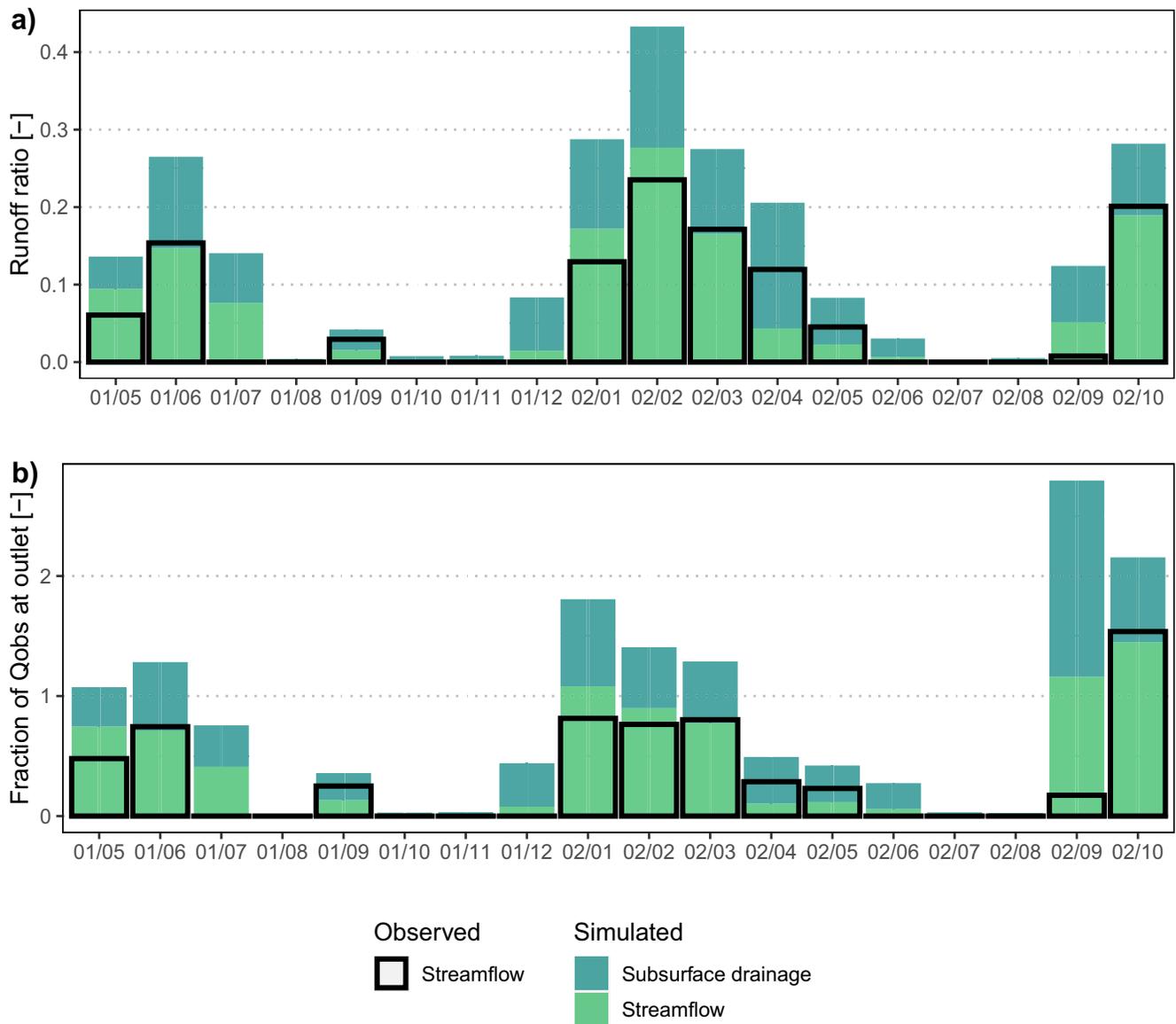


Figure 9. Observed and simulated runoff ratios for the upper catchment (a) and the contribution to the observed and simulated streamflow from the upper gauge to the observed streamflow at the lower gauge (b) for the different study months (yy/mm). The runoff ratio for the simulated subsurface flow that leaves the upper catchment (a) and the contribution of this flow to the lower gauge (b) are shown as well. The simulated results are shown for the calibration period with a 100% optimization weight for streamflow. Comparing simulations to observations can result in estimated contributions from the upper gauge that exceed the observed flow at the lower gauge if the simulated streamflow is overestimated. Note that when streamflow at the upper gauge exceed streamflow at the lower gauge, the fraction of Qobs at the outlet is <1. One would expect streamflow at the upper gauge to be less than the lower gauge (fraction <1) due to contributions from other sub-catchments.

for the 25 parameter sets but was 37%–80% of the sum of the exchange flux and subsurface drainage. If we did not allow for any subsurface drainage from SC2 to SC6 by setting parameter k_{sub} to zero, the simulated flux from the groundwater back into the stream would be larger (see Figure S2).

5.4. Simulation of the Groundwater Dynamics

The distributed-ER model variant was calibrated to the catchment average and riparian average scaled groundwater levels. The simulations represented the average catchment groundwater dynamics, that is, weighted average of hillslope and riparian groundwater levels, better than the riparian groundwater dynamics (Figure 10). The incremental increase in optimization weight for streamflow caused a small decrease in

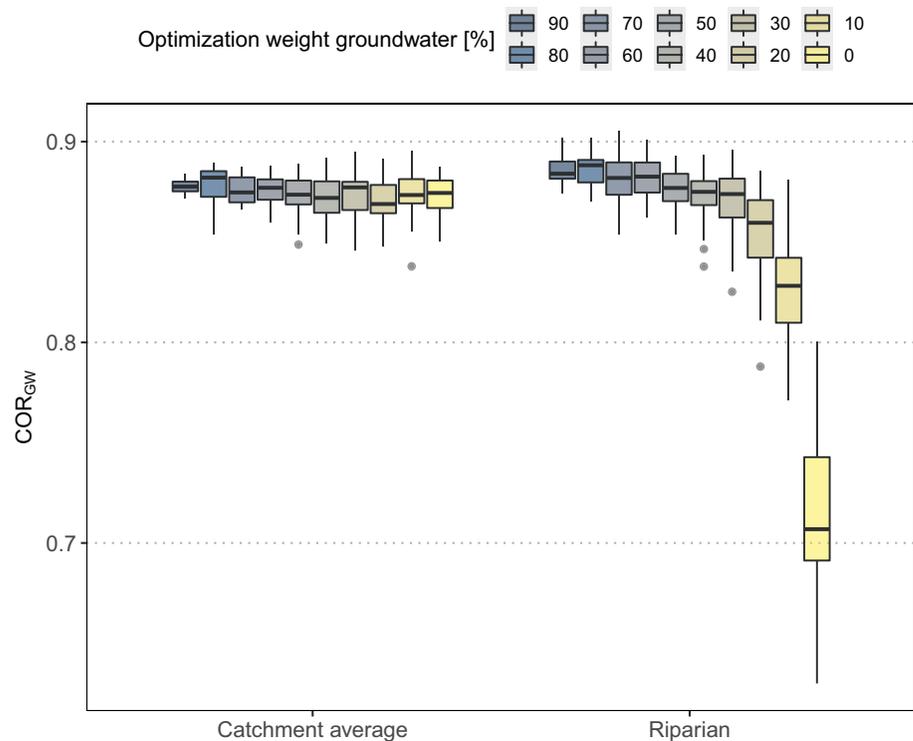


Figure 10. Boxplots of the performance of the simulations of the catchment average groundwater dynamics (used for the calibration of the response routine parameters) and riparian groundwater dynamics (used for the calibration of the exchange routine parameters) for different optimization weights for streamflow (and thus groundwater). Each optimization weight to groundwater (%) corresponds to 100% minus the optimization weight for streamflow.

performance for the catchment average groundwater simulations, but a marked decrease in the performance for the riparian groundwater simulations, particularly when the optimization weight for streamflow was more than 50% and a sharp drop in performance with an optimization weight of more than 80% for streamflow (Figure 10). There were no systematic differences in the performances of the models when a different weighting scheme for the groundwater simulations (i.e., between the area weighted and arithmetic mean) was used in the calibration.

5.5. Uncertainty in the Soil Water Storage Potential

We used the distributed model variants (distributed-ST and distributed-ER) to represent the difference in the soil water storage potential for the bedrock outcrop and the other parts of the catchment. For the outcrop (SC1), the range of FC values was kept very small (0.01–100 mm) to simulate the low storage capacity and the observed fast response to rainfall. From the comparison of the calibrated FC values, we see that the calibration of the distributed-ER variant led to a smaller parameter range for both the outcrop (SC1) and the remainder of the upper catchment (SC2) than the calibration of the distributed-ST and the lumped variants (Figure 11; coefficient of variation lumped: 0.13, distributed-ST: 0.12, distributed-ER 0.06).

6. Discussion

6.1. Advantages and Disadvantages of the Implementation of Bi-Directional Fluxes Between Groundwater and Surface Water in a Bucket Type Hydrological Model

For Panola, the use of the new exchange routine led to only a small improvement in model performance compared to the standard distributed model variant (Figures 5 and 6). Although the overall gain in model performance for the streamflow simulation was small (increase in NSE_Q of about 0.03 for the entire calibration period and up to about 0.10 for the validation period), the streamflow was simulated better using

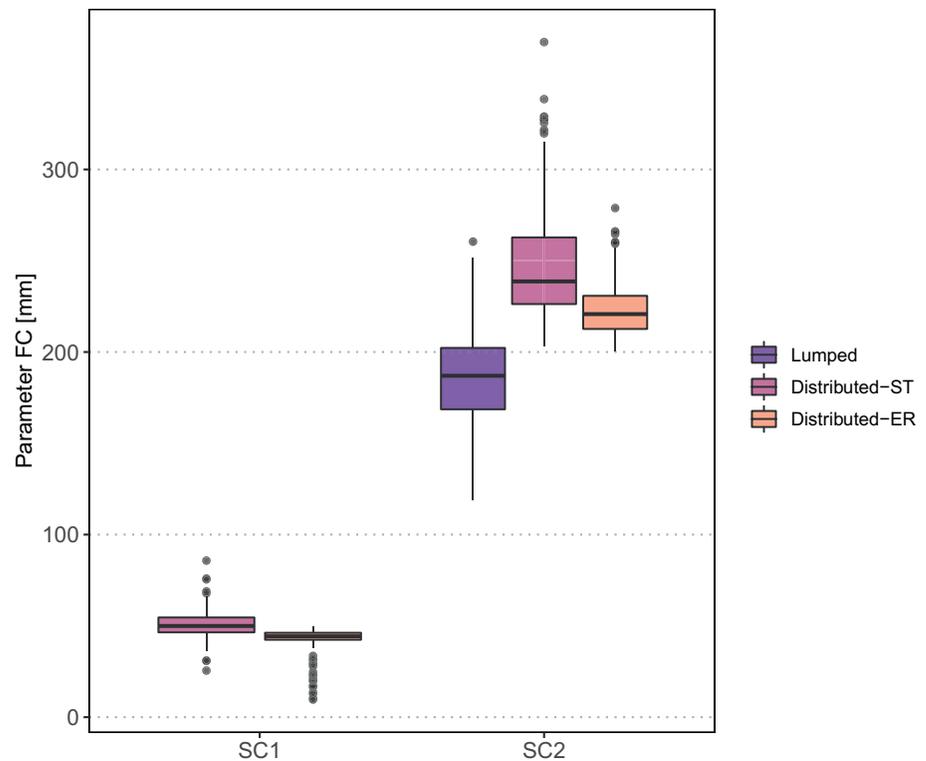


Figure 11. Ranges for parameter FC for the outcrop (SC1) and the remainder of the upper catchment (SC2). The boxplots contain the 25 best parameter values for all calibrations with the different optimization weights for streamflow. For the lumped model variant, the calibration parameter range for FC was restricted to a range between 0.01 to 1,000 mm; for the distributed model variants (distributed-ST and distributed-ER), it was restricted to 0.01–100 mm for SC1 and to 200–1,000 mm for SC2. For boxplots for the other parameters, see Figure S8.

the new model variant for specific periods of time, particularly the drying down and wetting up periods, hydrological situations that previous studies (e.g., Peters, Freer, & Beven, 2003) found difficult to simulate. This highlights the need to carefully analyze model simulations: improvements might average out or are not detectable if we only look at statistical measures calculated over longer time periods.

The benefit of the new exchange routine extends beyond model performance measures. The new routine contributes to making models more realistic, while maintaining the bucket-type modeling approach. This contributes to simulations being right for the right reasons (Kirchner, 2006) and is especially important when simulating extremely wet or dry situations, which might be outside the calibration conditions. Indeed, the validation period used in this study was much wetter than the calibration period (see Table 1). The fact that the performance for the distributed-ER model is better than that of the distributed-ST model during the validation period (Figures S3, S4 and S5) provides confidence that the distributed-ER model variant represents the processes better. Enabling the possibility of bi-directional fluxes between groundwater and streamflow in hydrological models is particularly useful for catchments where there are long stretches with losing flow conditions, or dry streambeds. It is especially useful when the model is used to address questions related to low flows or droughts because less streamflow is generated at locations or times where the stream is losing water to groundwater storage. The subsurface flow to downstream sub-catchments may furthermore be relevant to maintain baseflow in these sub-catchments. This is especially important for simulations during dry conditions (when the relative contribution of this flux is large, even though the actual flux is small) or when the model is applied to address questions related to water quality, temperature or stream habitats. It is also very useful for studies on the hydrology, ecology or bio-geochemistry of intermittent streams.

However, there are also disadvantages of representing the bi-directional groundwater-surface water exchange in a bucket-type model. In order to simulate the bi-directional exchange, we had to add four model

parameters. Even though bucket-type models are meant to be conceptual, the model parameters should still be representative of the processes they intend to represent. Adding model parameters makes the model more flexible but more model parameters also increase the risk of over-parameterization and problematic parameter identifiability (Beven, 1993; Seibert et al., 2019). To reduce the risk of over-parameterization, we can use additional information (as we did in the application to Panola by including groundwater level data in the calibration or pre-defining the value for some parameters (as we did for w by assuming that it is similar to the relative size of the riparian zone). In fact, the parameter FC , representing the soil moisture storage capacity was better constrained by including the exchange routine (Figure 11). We did not observe any compensation effect, that is, the better constraint of one parameter leading to a larger range for another parameter (see Figure S8 in the Supporting Information S1), therefore we assume that this improved parameter identifiability is related to the more suitable model structure for the distributed-ER variant. It is interesting to note that the optimized values for the soil moisture storage capacity (FC) are comparable to the estimated 180 mm of storage for a 1 m deep hillslope soil (Aulenbach & Peters, 2018; Peters & Aulenbach, 2011).

There might be even better constraints possible if we would use the groundwater data in a smarter way. In this study, we simply averaged the scaled groundwater levels measured in the hillslopes and riparian zone. This averages out some of the observed variation between the different groundwater wells. We, furthermore, used a simple correlation coefficient in the calibration, which ended up being very high for both distributed model variants and all seasons. For instance, we could have considered the depth of the wells and hence used a volume-based weighting (rather than a region-based weighting) to represent the average dynamics of the groundwater in the catchment.

Including the exchange routine can provide more options for validation when there are different data available. We could, for instance, compare the exchange fluxes with groundwater level dynamics in different sub-catchments where exchange between the stream and groundwater is expected and needs to be modeled. For the Panola case study, we simulated streamflow at the outlet of the catchment for evaluation (remember data from the lower gauge was not used in the calibration). By including the exchange routine, the simulated streamflow at the outlet was more similar to the observed streamflow than for the standard distributed variant of the model.

6.2. Application to Panola

Depending on the parameterization of the exchange routine, different types of dynamics are possible. We considered the simulation of temporally changing flow directions between the groundwater and the stream an important aspect of the exchange routine but in the application for Panola it turned out that fluxes along the stream, such as subsurface drainage from one sub-catchment to another, were more important. This can be seen from the relatively large simulated subsurface drainage contributions to streamflow at the outlet (Figure 9) and the exchange flux direction (Figure 8). It is beneficial that there are parameterizations of the exchange routine that allow for the simulation of this behavior because leakage below a gauge and inter-basin groundwater flow are common (e.g., Fan, 2019; Käser & Hunkeler, 2016; Le Moine et al., 2007; Welch et al., 2012).

Field observations at Panola have shown that the large bedrock outcrop produces flow during almost all events, while measurements at the trenched hillslope suggest that flow from the hillslopes only occurs during large events (Tromp-van Meerveld & McDonnell, 2006a). We related the simulated exchange fluxes during the large rain event at the end of March to the observations at the trenched hillslope and concluded that the change in exchange flux direction or decrease of the flux occur at very plausible moments. The stream was likely losing during the initial part of the event due to the quick response from the outcrop (i.e., SC1 in our setup). Measurements at the trenched hillslope showed significant flow from the hillslopes later during the event (Tromp-van Meerveld & McDonnell, 2006b). This is the cause for the decrease in the flux toward the groundwater at the end of the large event and for some model parameterizations there was even a switch in the direction of the exchange flux (Figure 8).

Previous studies have also shown that flow from the bedrock outcrop causes significant recharge of the riparian aquifer (Burns et al., 2001) and that the majority of streamflow (on an annual basis) comes from the lower area of the catchment (roughly corresponding to SC6) (Aulenbach et al., 2021). The simulations

for the distributed-ER variant suggest that some of this water leaves the upper catchment via subsurface flow and that this flow contributes to the sustained baseflow in the lower catchment (Figure 9). The simulations also highlight that the stream in the upper catchment is rarely a gaining stream. This explains why mixing-model applications at Panola showed only a very small contribution of hillslope water to streamflow (e.g., Hooper, 2001). This is also in agreement with the measurements at the trenched hillslope that showed that significant subsurface flow occurred only rarely (Tromp-van Meerveld & McDonnell, 2006b) and groundwater measurements at another hillslope that showed that hillslope-riparian zone-stream connectivity was rare (van Meerveld et al., 2015).

The exchange routine further can allow a water deficit to build up if parameter z_{off} is included. This can be beneficial for the simulation of long duration droughts. Because of the short simulation time, we did not do this and set parameter z_{off} to zero. However, it is known that droughts at Panola can affect streamflow (Aulenbach & Peters, 2018) and stream chemistry (Aulenbach, 2020; Peters & Aulenbach, 2009). Therefore, it would be interesting to include this parameter when simulating longer time periods. Perhaps it would have improved the model performance for the distributed-ER model variant for the validation period.

The large bedrock outcrop is a special feature of Panola but there are other headwater catchments with either bedrock outcrops or very shallow soils near the ridges. For other catchments the bi-directional fluxes between the stream and the groundwater may occur in the middle or lower reaches of the catchment. Panola was used in this study as an illustrative example of a catchment where these processes occur. The same processes of streambed infiltration and subsurface drainage can be modeled for middle or lower reaches as well. In other words, while the particular situation of the Panola catchment is perhaps special, it is a suitable application to test the effect of the exchange routine on the model simulations.

6.3. When to Use the Exchange Routine and When Not

The addition of the exchange routine to simulate the bi-directional fluxes between the groundwater and the stream allows us to simulate what happens in intermittent streams, namely that streamflow ceases in certain stretches of the stream and reappears further downstream (Covino & McGlynn, 2007; Doering et al., 2007; Payn et al., 2009; Simpson & Meixner, 2013; Yu et al., 2013; Zimmer & McGlynn, 2017) or groundwater flow from one sub-catchment to another (e.g., Fan, 2019; Le Moine et al., 2007). Because losing stream conditions are common (e.g., Fan, 2019; McMahon & Nathan, 2021) and most streams are intermittent (Datry et al., 2014; Fritz et al., 2013; Hammond et al., 2021; Nadeau & Rains, 2007; Shanafield et al., 2021; van Meerveld et al., 2020), it is useful to explore the benefits and limitations of such an exchange routine. However, as already mentioned, we also introduce more parameters to the model. Due to the problems related to the additional parameters (Jakeman & Hornberger, 1993; Seibert et al., 2019), we will only benefit from the increased model realism if we (a) know that there are considerable losses from the stream to the riparian aquifer, (b) want to study questions related to groundwater-surface water exchange or transmission losses, or want to simulate streamflow during periods when these fluxes are important, and (c) we have suitable data to validate the model simulations and constrain the model parameters (e.g., groundwater levels or flow at different locations in the catchment). If groundwater-surface water exchange is insignificant or if the studied problem does not require this simulation, then this model routine should not be used.

7. Conclusions

In this paper, we propose a new parsimonious model routine to represent bi-directional fluxes between the groundwater and the stream in a bucket type hydrological model. The bi-directional interaction can be used to simulate the fluxes from the stream to riparian groundwater and subsurface drainage from upstream to downstream sub-catchments. Thus the routine allows for the representation of losses from the stream to the riparian groundwater, and as a result the simulation of intermittent streamflow and inter-basin groundwater flow.

The results of the test application for the Panola catchment demonstrate that the routine can make a bucket-type model more realistic. When looking at the overall model performance and comparing a model variant where the exchange routine was included and one where it was not, the improvements were small (increase in NSE_Q of about 0.03 for the calibration period and about 0.10 for the validation period), but for

certain hydrological situations (such as the drying down period) the improvements were notable, mainly because of the flux between sub-catchments via subsurface drainage. The timing of the simulated flux during a large event agreed with field observations. Overall, the results highlight the value of the new routine for catchments with notable fluxes from the stream to the groundwater or inter-basin groundwater flow but the routine should be tested for other catchments as well.

Data Availability Statement

The observational data underlying this work can be obtained from the USGS ScienceBase repository (<https://www.sciencebase.gov/catalog/>).

Acknowledgments

Thanks to Marc Vis (UZH) for discussions and implementation of the exchange routine in the HBV-model and Brent Aulenbach (USGS) for sharing the data (precipitation, streamflow at the upper and lower gauge, groundwater level data, evapotranspiration estimates) used in the simulations. The authors also thank two anonymous reviewers and Claudia Brauer for the thoughtful and detailed comments that helped greatly to improve the manuscript.

References

- Al-Safi, H. I. J., & Sarukkalgale, P. R. (2017). Assessment of future climate change impacts on hydrological behavior of Richmond river catchment. *Water Science and Engineering*, 10(3), 197–208. <https://doi.org/10.1016/j.wse.2017.05.004>
- Aulenbach, B. T. (2017). *Data for estimating monthly water budgets at Panola Mountain Research Watershed, Stockbridge, Ga., water years 1986–2015*. U.S. Geological Survey Data Release. <https://doi.org/10.5066/F7XSS5NV>
- Aulenbach, B. T. (2020). Effects of climate-related variability in storage on streamwater solute concentrations and fluxes in a small forested watershed in the Southeastern United States. *Hydrological Processes*, 34, 189–208. <https://doi.org/10.1002/hyp.13589>
- Aulenbach, B. T., Hooper, R. P., van Meerveld, H. J., Burns, D. A., Freer, J. E., Shanley, J. B., et al. (2021). The evolving perceptual model of streamflow generation at the Panola Mountain Research watershed. *Hydrological Processes*, 35, e14127. <https://doi.org/10.1002/hyp.14127>
- Aulenbach, B. T., & Peters, N. E. (2018). Quantifying climate-related interactions in shallow and deep storage and evapotranspiration in a forested, seasonally water-limited watershed in the Southeastern United States. *Water Resources Research*, 54(4), 3037–3061. <https://doi.org/10.1002/2017WR020964>
- Azmat, M., Wahab, A., Huggel, C., Qamar, M. U., Hussain, E., Ahmad, S., & Waheed, A. (2020). Climatic and hydrological projections to changing climate under CORDEX-South Asia experiments over the Karakoram-Hindukush-Himalayan water towers. *The Science of the Total Environment*, 703, 135010. <https://doi.org/10.1016/j.scitotenv.2019.135010>
- Bergström, S. (1976). *Development and application of a conceptual runoff model for Scandinavian catchments* (Technical Report).
- Bergström, S. (1992). *The HBV model: Its structure and applications*. Swedish Meteorological and Hydrological Institute.
- Bergström, S. (1995). The HBV model. In *Computer models of watershed hydrology* (pp. 443–476).
- Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16(1), 41–51. [https://doi.org/10.1016/0309-1708\(93\)90028-e](https://doi.org/10.1016/0309-1708(93)90028-e)
- Beven, K. (2011). *Rainfall-runoff modelling: The primer* (2nd ed.). New York: John Wiley and Sons.
- Brauer, C. C., Teuling, A. J., Torfs, P. J. J. F., & Uijlenhoet, R. (2014). The Wageningen lowland runoff simulator (WALRUS): A lumped rainfall-runoff model for catchments with shallow groundwater. *Geoscientific Model Development*, 7(5), 2313–2332. <https://doi.org/10.5194/gmd-7-2313-2014>
- Burns, D. A., McDonnell, J. J., Hooper, R. P., Peters, N. E., Freer, J. E., Kendall, C., & Beven, K. (2001). Quantifying contributions to storm runoff through end-member mixing analysis and hydrologic measurements at the Panola Mountain Research Watershed (Georgia, USA). *Hydrological Processes*, 15(10), 1903–1924. <https://doi.org/10.1002/hyp.246>
- Burns, D. A., Plummer, L. N., McDonnell, J. J., Busenberg, E., Casile, G. C., Kendall, C., et al. (2003). The geochemical evolution of riparian ground water in a forested piedmont catchment. *Groundwater*, 41(7), 913–925. <https://doi.org/10.1111/j.1745-6584.2003.tb02434.x>
- Covino, T. P., & McGlynn, B. L. (2007). Stream gains and losses across a mountain-to-valley transition: Impacts on watershed hydrology and stream water chemistry. *Water Resources Research*, 43(10), W10431. <https://doi.org/10.1029/2006WR005544>
- Datry, T., Larned, S. T., & Tockner, K. (2014). Intermittent rivers: A challenge for freshwater ecology. *BioScience*, 64(3), 229–235. <https://doi.org/10.1093/biosci/bit027>
- De Girolamo, A. M., Bouraoui, F., Buffagni, A., Pappagallo, G., & Lo Porto, A. (2017). Hydrology under climate change in a temporary river system: Potential impact on water balance and flow regime. *River Research and Applications*, 33, 1219–1232. <https://doi.org/10.1002/rra.3165>
- Doering, M., Uehlinger, U., Rotach, A., Schlaepfer, D. R., & Tockner, K. (2007). Ecosystem expansion and contraction dynamics along a large Alpine alluvial corridor (Tagliamento River, Northeast Italy). *Earth Surface Processes and Landforms*, 32(11), 1693–1704. <https://doi.org/10.1002/esp.1594>
- Efstratiadis, A., & Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: A review. *Hydrological Sciences Journal*, 55(1), 58–78. <https://doi.org/10.1080/02626660903526292>
- Fan, Y. (2019). Are catchments leaky? *Wiley Interdisciplinary Reviews: Water*, 6(6), e1386. <https://doi.org/10.1002/wat2.1386>
- Freer, J., McDonnell, J., Beven, K., Brammer, D., Burns, D., Hooper, R., & Kendall, C. (1997). Topographic controls on subsurface storm flow at the hillslope scale for two hydrologically distinct small catchments. *Hydrological Processes - Letters*, 11(9), 1347–1352. [https://doi.org/10.1002/\(sici\)1099-1085\(199707\)11:9<1347::aid-hyp592>3.0.co;2-r](https://doi.org/10.1002/(sici)1099-1085(199707)11:9<1347::aid-hyp592>3.0.co;2-r)
- Fritz, K. M., Hagenbuch, E., D'Amico, E., Reif, M., Wigington, P. J., Jr., Leibowitz, S. G., et al. (2013). Comparing the extent and permanence of headwater streams from two field surveys to values from hydrographic databases and maps. *Journal of the American Water Resources Association*, 49, 867–882. <https://doi.org/10.1111/jawr.12040>
- Gutiérrez-Jurado, K. Y., Partington, D., Batelaan, O., Cook, P., & Shanafield, M. (2019). What triggers streamflow for intermittent rivers and ephemeral streams in low-gradient catchments in Mediterranean climates. *Water Resources Research*, 55(11), 9926–9946. <https://doi.org/10.1029/2019WR025041>
- Hammond, J. C., Zimmer, M., Shanafield, M., Kaiser, K., Godsey, S. E., Mims, M. C., et al. (2021). Spatial patterns and drivers of nonperennial flow regimes in the contiguous United States. *Geophysical Research Letters*, 48, e2020GL090794. <https://doi.org/10.1029/2020GL090794>

- Hooper, R. P. (2001). Applying the scientific method to small catchment studies: A review of the panola mountain experience. *Hydrological Processes*, 15(10), 2039–2050. <https://doi.org/10.1002/hyp.255>
- Huang, Y., Chen, X., Chen, X., & Ou, G. (2015). Transmission losses during two flood events in the Platte River, south-central Nebraska. *Journal of Hydrology*, 520, 244–253. <https://doi.org/10.1016/j.jhydrol.2014.11.046>
- Jaeger, K. L., Olden, J. D., & Pelland, N. A. (2014). Climate change poised to threaten hydrologic connectivity and endemic fishes in dryland streams. *Proceedings of the National Academy of Sciences*, 111(38), 13894–13899. <https://doi.org/10.1073/pnas.1320890111>
- Jakeman, A., & Hornberger, G. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, 29(8), 2637–2649. <https://doi.org/10.1029/93wr00877>
- Käser, D., & Hunkeler, D. (2016). Contribution of alluvial groundwater to the outflow of mountainous catchments. *Water Resources Research*, 52(2), 680–697. <https://doi.org/10.1002/2014wr016730>
- Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, 42(3), W03S04. <https://doi.org/10.1029/2005wr004362>
- Le Moine, N., Andréassian, V., Perrin, C., & Michel, C. (2007). How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments. *Water Resources Research*, 43(6), W06428. <https://doi.org/10.1029/2006WR005608>
- Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, 201(1), 272–288. [https://doi.org/10.1016/s0022-1694\(97\)00041-3](https://doi.org/10.1016/s0022-1694(97)00041-3)
- McCallum, A. M., Andersen, M. S., Rau, G. C., Larsen, J. R., & Acworth, R. I. (2014). River-aquifer interactions in a semiarid environment investigated using point and reach measurements. *Water Resources Research*, 50(4), 2815–2829. <https://doi.org/10.1002/2012wr012922>
- McMahon, T. A., & Nathan, R. J. (2021). Baseflow and transmission loss: A review. *WIREs Water*, 8(4), e1527. <https://doi.org/10.1002/wat2.1527>
- Messenger, M., Lehner, B., Cockburn, C., Lamouroux, N., Pella, H., Snelder, T., et al. (2021). Global prevalence of non-perennial rivers and streams. *Nature*, 594, 391–397. <https://doi.org/10.1038/s41586-021-03565-5>
- Nadeau, T.-L., & Rains, M. C. (2007). Hydrological connectivity between headwater streams and downstream waters: How science can inform policy. *Journal of the American Water Resources Association*, 43(1), 118–133. <https://doi.org/10.1111/j.1752-1688.2007.00010.x>
- Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models - Part I - A discussion of principles. *Journal of Hydrology*, 10, 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Orlowski, N., Lauer, F., Kraft, P., Frede, H.-G., & Breuer, L. (2014). Linking spatial patterns of groundwater table dynamics and streamflow generation processes in a small developed catchment. *Water*, 6(10), 3085–3117. <https://doi.org/10.3390/w6103085>
- Parra, V., Fuentes-Aguilera, P., & Muñoz, E. (2018). Identifying advantages and drawbacks of two hydrological models based on a sensitivity analysis: A study in two Chilean watersheds. *Hydrological Sciences Journal*, 63(12), 1831–1843. <https://doi.org/10.1080/02626667.2018.1538593>
- Payn, R. A., Gooseff, M. N., McGlynn, B. L., Bencala, K. E., & Wondzell, S. M. (2009). Channel water balance and exchange with subsurface flow along a mountain headwater stream in Montana, United States. *Water Resources Research*, 45(11), W11427. <https://doi.org/10.1029/2008WR007644>
- Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London - Series A*, 187, 253–318. <https://doi.org/10.1098/rsta.1896.0007>
- Peters, N. E., & Aulenbach, B. T. (2009). Hydrologic pathway contributions to stream fluxes of weathering products at the Panola Mountain Research Watershed, Georgia. In G. D. Carroll (Ed.), *Proceedings of the 2009 Georgia water resources conference. Warnell school of forestry and natural resources*. The University of Georgia Athens.
- Peters, N. E., & Aulenbach, B. T. (2011). Water storage at the Panola Mountain Research Watershed, Georgia, USA. *Hydrological Processes*, 25(25), 3878–3889. <https://doi.org/10.1002/hyp.8334>
- Peters, N. E., Burns, D. A., & Aulenbach, B. T. (2014). Evaluation of high-frequency mean streamwater transit-time estimates using groundwater age and dissolved silica concentrations in a small forested watershed. *Aquatic Geochemistry*, 20(2–3), 183–202. <https://doi.org/10.1007/s10498-013-9207-6>
- Peters, N. E., Freer, J., & Aulenbach, B. T. (2003). Hydrological dynamics of the Panola Mountain Research Watershed, Georgia. *Groundwater*, 41(7), 973–988. <https://doi.org/10.1111/j.1745-6584.2003.tb02439.x>
- Peters, N. E., Freer, J., & Beven, K. (2003). Modelling hydrologic responses in a small forested catchment (Panola Mountain, Georgia, USA): A comparison of the original and a new dynamic TOPMODEL. *Hydrological Processes*, 17(2), 345–362. <https://doi.org/10.1002/hyp.1128>
- Peters, N. E., & Ratcliffe, E. B. (1998). Tracing hydrologic pathways using chloride at the Panola Mountain Research Watershed, Georgia, USA. *Water, Air, and Soil Pollution*, 105(1–2), 263–275. https://doi.org/10.1007/978-94-017-0906-4_25
- Pool, S., Viviroli, D., & Seibert, J. (2019). Value of a limited number of discharge observations for improving regionalization: A large-sample study across the United States. *Water Resources Research*, 55(1), 363–377. <https://doi.org/10.1029/2018WR023855>
- Querner, E., Froebrich, J., Gallart, F., Cazemier, M., & Tzoraki, O. (2016). Simulating streamflow variability and aquatic states in temporary streams using a coupled groundwater-surface water model. *Hydrological Sciences Journal*, 61(1), 146–161. <https://doi.org/10.1080/02626667.2014.983514>
- Sauquet, E., Shanafield, M., Hammond, J. C., Sefton, C., Leigh, C., & Detry, T. (2021). Classification and trends in intermittent river flow regimes in Australia, Northwestern Europe and USA: A global perspective. *Journal of Hydrology*, 597, 126170. <https://doi.org/10.1016/j.jhydrol.2021.126170>
- Seibert, J. (2000). Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Sciences*, 4(2), 215–224. <https://doi.org/10.5194/hess-4-215-2000>
- Seibert, J., Staudinger, M., & van Meerveld, H. I. (2019). Validation and over-parameterization—Experiences from hydrological modeling. In *Computer simulation validation* (pp. 811–834). Springer. https://doi.org/10.1007/978-3-319-70766-2_33
- Seibert, J., & Vis, M. J. P. (2012). Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. *Hydrology and Earth System Sciences*, 16, 3315–3325. <https://doi.org/10.5194/hess-16-3315-2012>
- Shanafield, M., Bourke, S. A., Zimmer, M. A., & Costigan, K. H. (2021). An overview of the hydrology of non-perennial rivers and streams. *WIREs Water*, 8(2), e1504. <https://doi.org/10.1002/wat2.1504>
- Shanafield, M., & Cook, P. G. (2014). Transmission losses, infiltration and groundwater recharge through ephemeral and intermittent streambeds: A review of applied methods. *Journal of Hydrology*, 511, 518–529. <https://doi.org/10.1016/j.jhydrol.2014.01.068>
- Simpson, S. C., & Meixner, T. (2013). The influence of local hydrogeologic forcings on near-stream event water recharge and retention (Upper San Pedro River, Arizona). *Hydrological Processes*, 27(4), 617–627. <https://doi.org/10.1002/hyp.8411>

- Staudinger, M., van Meerveld, H. J., & Aulenbach, B. T. (2021). Precipitation, air temperature, streamflow, and water table levels from selected wells at the Panola Mountain Research Watershed, 2001-2003: U.S. Geological Survey data release [Data set]. <https://doi.org/10.5066/P96549L0>
- Stoll, S., & Weiler, M. (2010). Explicit simulations of stream networks to guide hydrological modelling in ungauged basins. *Hydrology and Earth System Sciences*, *14*, 1435–1448. <https://doi.org/10.5194/hess-14-1435-2010>
- Tromp-van Meerveld, H., & McDonnell, J. (2006a). Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope. *Water Resources Research*, *42*(2), W02410. <https://doi.org/10.1029/2004wr003778>
- Tromp-van Meerveld, H., & McDonnell, J. (2006b). Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research*, *42*(2), W02411. <https://doi.org/10.1029/2004wr003800>
- Tromp-van Meerveld, H., & Weiler, M. (2008). Hillslope dynamics modeled with increasing complexity. *Journal of Hydrology*, *361*(1), 24–40. <https://doi.org/10.1016/j.jhydrol.2008.07.019>
- Tromp-van Meerveld, H. J., Peters, N. E., & McDonnell, J. J. (2007). Effect of bedrock permeability on subsurface stormflow and the water balance of a trenched hillslope at the Panola Mountain Research Watershed, Georgia, USA. *Hydrological Processes*, *21*(6), 750–769. <https://doi.org/10.1002/hyp.6265>
- van Meerveld, H. J., Sauquet, E., Gallart, F., Sefton, C., Seibert, J., & Bishop, K. (2020). Aqua temporaria incognita. *Hydrological Processes*, *34*, 5704–5711. <https://doi.org/10.1002/hyp.13979>
- van Meerveld, H. J., Seibert, J., & Peters, N. E. (2015). Hillslope–riparian-stream connectivity and flow directions at the Panola mountain research watershed. *Hydrological Processes*, *29*(16), 3556–3574. <https://doi.org/10.1002/hyp.10508>
- Wang, D. (2011). On the base flow recession at the Panola mountain research watershed, Georgia, United States. *Water Resources Research*, *47*(3), W03527. <https://doi.org/10.1029/2010WR009910>
- Ward, A. S., Payn, R. A., Gooseff, M. N., McGlynn, B. L., Bencala, K. E., Kelleher, C. A., et al. (2013). Variations in surface water-ground water interactions along a headwater mountain stream: Comparisons between transient storage and water balance analyses. *Water Resources Research*, *49*, 3359–3374. <https://doi.org/10.1002/wrcr.20148>
- Ward, A. S., Schmadel, N. M., & Wondzell, S. M. (2018). Simulation of dynamic expansion, contraction, and connectivity in a mountain stream network. *Advances in Water Resources*, *114*, 64–82. <https://doi.org/10.1016/j.advwatres.2018.01.018>
- Welch, L. A., Allen, D. M., & van Meerveld, H. (2012). Topographic controls on deep groundwater contributions to mountain headwater streams and sensitivity to available recharge. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques*, *37*(4), 349–371. <https://doi.org/10.4296/cwrj2011-907>
- Yu, M. C. L., Cartwright, I., Braden, J. L., & de Bree, S. T. (2013). Examining the spatial and temporal variation of groundwater inflows to a valley-to-floodplain river using ²²²Rn, geochemistry and river discharge: The ovens river, southeast Australia. *Hydrology and Earth System Sciences*, *17*(12), 4907–4924. <https://doi.org/10.5194/hess-17-4907-2013>
- Zimmer, M. A., & McGlynn, B. L. (2017). Bidirectional stream–groundwater flow in response to ephemeral and intermittent streamflow and groundwater seasonality. *Hydrological Processes*, *31*(22), 3871–3880. <https://doi.org/10.1002/hyp.11301>