



## RESEARCH ARTICLE

WILEY

# The relevance of preferential flow in catchment scale simulations: Calibrating a 3D dual-permeability model using DREAM

Luisa Hopp<sup>1</sup> | Barbara Glaser<sup>1,2</sup> | Julian Klaus<sup>2</sup> | Thilo Schramm<sup>1,2</sup>

<sup>1</sup>Department of Hydrology, University of Bayreuth, 95440, Bayreuth, Germany

<sup>2</sup>Catchment and Eco-Hydrology Research Group, Luxembourg Institute of Science and Technology (LIST), 4362, Esch-sur-Alzette, Luxembourg

**Correspondence**

Luisa Hopp, Department of Hydrology, University of Bayreuth, 95440 Bayreuth, Germany.

Email: luisa.hopp@uni-bayreuth.de

**Funding information**

Luxembourg Institute of Science and Technology; Luxembourg National Research Fund, Grant/Award Number: 10189601

**Abstract**

The occurrence of preferential flow in the subsurface has often been shown in field experiments. However, preferential flow is rarely included in models simulating the hydrological response at the catchment scale. If it is considered, preferential flow parameters are typically determined at the plot scale and then transferred to larger-scale simulations. Here, we successfully used the optimization algorithm Differential Evolution Adaptive Metropolis (DREAM) to calibrate a 3D physics-based dual-permeability model directly at the catchment scale. In order to keep computational costs of the optimization routine at a reasonable level, we limited the number of parameters to be calibrated to the ones that had been shown before to be most influential for the simulation of discharge. We also calibrated parameters of the matrix domain and the macropore domain with a fixed parameter ratio between soil layers instead of calibrating every layer separately. These ratios reflected observed depth profiles of soil hydraulic properties at our study site. The dual-permeability parameter sets identified during calibration were able to simulate observed discharge time series satisfactorily but did not outperform a calibrated single-domain reference model scenario. Saturated hydraulic conductivities of the macropore domain were calibrated such that they became very similar to matrix saturated hydraulic conductivities, thereby effectively removing the effect of macropores. This suggests that the incorporation of vertical preferential flow as represented by the dual-permeability approach was not relevant for reproducing the hydrometric response reasonably well in the studied catchment. We also tested the scale-invariance of the calibrated dual-permeability parameter sets by using the parameter sets performing best at catchment scale to simulate plot-scale bromide depth profiles obtained from tracer irrigation experiments. This parameter transfer proved to be not successful, indicating that soil hydraulic parameters are scale-variant, independent of the direction of parameter transfer.

**KEYWORDS**

catchment scale, discharge simulation, dual-permeability model, HydroGeoSphere, model calibration, optimization algorithm, preferential flow

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Hydrological Processes* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Numerous studies have shown that water flow and solute transport (e.g., of nutrients, pesticides and pollutants) in soils can be much faster than would be expected from Darcy–Richards type subsurface flow (Flury, 1996; Gächter, Ngatiah, & Stamm, 1998; Germann, Smith, & Thomas, 1987; Kung et al., 2000; Stamm, Flüher, Gächter, Leuenberger, & Wunderli, 1998). Preferential flow through macropores has been recognized as one of the main reasons for this phenomenon. Soil macropores may originate from desiccation and freezing, growth and decay of roots and mycelia, bioturbation by soil fauna and further sculpting by water may lead to large macropores called ‘soil pipes’ (Bachmair & Weiler, 2011; Beven & Germann, 1982; Beven & Germann, 2013; Coppola, Kutilek, & Frind, 2009; Jones, 2010). Macropores can be vertical or lateral, depending on their origin, and enhance water flow in these directions at different scales (e.g., earthworm burrows at the plot scale in vertical direction; soil pipes in lateral downslope direction at the hillslope scale). Another form of lateral preferential flow that is more relevant beyond the plot scale can occur if the soil profile is characterized by several soil layers with contrasting hydraulic conductivities, where a highly permeable layer underlain by a much less permeable layer experiences transient saturation as response to an event and provides a fast pathway for downslope movement of water (e.g., Klaus & Jackson, 2018; Tromp-van Meerveld & McDonnell, 2006).

In situ detection of macropores by destructive sampling after dye tracing (e.g., Anderson, Weiler, Alila, & Hudson, 2009; Graham, Woods, & McDonnell, 2010; Laine-Kaulio, Backnäs, Koivusalo, & Lauren, 2015; Weiler & Naef, 2003) or non-invasive techniques, such as ground-penetrating radar (Gormally, McIntosh, & Mucciardi, 2011; Gormally, McIntosh, Mucciardi, & McCarty, 2011; Nyquist, Toran, Pitman, Guo, & Lin, 2018), have shown a ubiquitous presence of macropores in different landscapes. Due to this widespread evidence of preferential flow in the subsurface, it seems necessary to represent macropores in hydrological models in order to advance the understanding of subsurface flow behaviour and hydrological threshold processes, catchment runoff generation, leaching of nutrients and contaminants and slope stability mechanisms with related landslide risks (e.g., Klaus & Zehe, 2011; Roulier et al., 2006; Shao, Bogaard, & Bakker, 2014; Shao, Bogaard, Bakker, & Greco, 2015; Weiler & McDonnell, 2007; Zehe, Becker, Bárdossy, & Plate, 2005).

Macropores are characterized by their ability to allow non-equilibrium flow under certain conditions (Beven & Germann, 1982; Jarvis, 2007). Experimental evidence has shown that water flow in macropores often occurs as laminar free-surface film flow that does not require saturation neither of the surrounding matrix nor of the macropore itself (Nimmo, 2012, and references therein). Capillarity-based approaches are not suitable to describe this non-diffusive film flow, and alternative mathematical descriptions have been developed and applied to experimental data with some success (e.g., Di Pietro, Ruy, & Capowiez, 2003; Germann & Di Pietro, 1999; Nimmo, 2010; Peters & Durner, 2008). However, although the concept of diffusive capillarity-driven Darcy–Richards type flow does theoretically not

apply to water flow in macropores in many instances, models based on the Darcy–Richards equation for variably saturated flow still prevail in the hydrological modelling community and have been successfully used to simulate hydrological response (Beven & Germann, 2013). Since a derivation of the geometry and the connectivity of preferential flow pathways and also of the associated hydraulic properties remains challenging, modifications of Darcy-type flow models with approaches such as dual continuum, dual porosity, or dual permeability implement the concept of preferential flow without explicitly describing the geometry of the macropore network (Beven & Germann, 2013; Gerke & van Genuchten, 1993; Köhne, Köhne, & Šimůnek, 2009; Šimůnek, Jarvis, van Genuchten, & Gärdenäs, 2003). All three approaches divide the total pore space into two domains with differing hydraulic properties and an exchange term between the domains. Flow in the macropore domain may be described using also the Darcy–Richards equation or some other equation, such as, for example, a gravity-driven kinematic wave formulation (e.g., the MACRO model, Jarvis & Larsbo, 2012). Such preferential flow models have been implemented successfully to simulate water flow and/or solute transport at the plot scale (Arora, Mohanty, & McGuire, 2012; Köhne et al., 2009; Köhne & Mohanty, 2005; Larsbo, Roulier, Stenemo, Kasteel, & Jarvis, 2005; Roulier et al., 2006) and at the hillslope and catchment scale (Christiansen, Thorsen, Clausen, Hansen, & Refsgaard, 2004; Laine-Kaulio, Backnäs, Karvonen, Koivusalo, & McDonnell, 2014; Laine-Kaulio & Koivusalo, 2018; Roulier et al., 2006; Schaik et al., 2014; Shao et al., 2014; Shao et al., 2015; Yu, Duffy, Baldwin, & Lin, 2014; Zehe et al., 2005). Working with bi- or multimodal soil hydraulic functions in a single-domain setup is another possibility to represent flow in heterogeneous pore systems (e.g., Durner, 1994; Othmer, Diekkrüger, & Kutilek, 1991; Ross & Smettem, 2000).

Hydrological models based on the Darcy–Richards equation require information on hydraulic properties of the pore system as input. There is typically a discrepancy between the scale of the measurements of physical soil characteristics and the scale of hydrological models that study flow and transport at the hillslope or catchment scale. Although there has been progress in using geophysical techniques to derive subsurface hydraulic properties at larger scale (Binley et al., 2015), there remains a large uncertainty in the estimated values. This makes it challenging to identify soil hydraulic parameters for hydrological models in a representative way. These points are even more true for the parameterization of preferential flow in hydrological models (Arora et al., 2012; Arora, Mohanty, & McGuire, 2011; Pechlivanidis, Jackson, McIntyre, & Wheeler, 2011; Šimůnek et al., 2003). There are two options to deal with the limits of available measurement techniques to identify model parameters at hillslope and catchment scales. One is to transfer parameters derived from small-scale field or laboratory observations (measurements or calibrated parameters) to a larger scale (Vereecken, Kasteel, Vanderborght, & Harter, 2007). A direct transfer of parameters implicitly assumes that physical processes and properties are scale-independent (Pechlivanidis et al., 2011). However, studies have shown a dependency of, for example, measured saturated hydraulic conductivity on

the scale of measurement method (Brace, 1980; Brooks, Boll, & McDaniel, 2004; Rovey & Cherkauer, 1995; Schulze-Makuch, Carlson, Cherkauer, & Malik, 1999). Therefore, the transfer of small-scale parameters to a larger scale might be done by scaling the parameters (Cadini, De Sanctis, Bertoli, & Zio, 2013; Nasta, Boaga, Deiana, Cassiani, & Romano, 2019), although the scaling relationship is subject to estimation and uncertainty, and may also depend on properties of the porous media (Schulze-Makuch et al., 1999). The second option is to calibrate the model parameters directly at the catchment scale. Typically, only discharge is available as long time series at catchment scale for calibration and validation. However, studies have shown that discharge data alone is usually not sufficient to constrain the parameter space (Köhne et al., 2009) and should be complemented with distributed observations of internal hydrologic state variables (e.g., Ebel et al., 2007). In addition to the problem of obtaining suitable calibration and validation data, calibration of physically based models at hillslope and catchment scale is computationally expensive (e.g., Ala-aho, Soulsby, Wang, & Tetzlaff, 2017; Wildemeersch, Goderniaux, Orban, Brouyère, & Dassargues, 2014).

The two different options for identifying model parameters at hillslope and catchment scale have been used in different complexity for preferential flow simulations. Often, soil hydraulic parameters in dual-permeability models have been inversely fitted in 1D and subsequently transferred to a larger scale by upscaling approaches (e.g., Cadini et al., 2013; Wang, Bradford, & Šimůnek, 2014). Roulier et al. (2006) calibrated the parameters for preferential flow in a dual-permeability model at plot scale with tracer data and then used these parameter sets for the simulation of pesticide transport in a 1D catchment model. Christiansen et al. (2004) derived preferential flow parameters from small scale measurements and implemented them in a 3D dual-permeability catchment model. Klaus and Zehe (2010, 2011) used some field measurements of the conductance of preferential flow paths and literature data to parameterize a 2D hillslope model with explicit representation of preferential flow structures, but eventually used a Monte Carlo simulation to calibrate these parameters. Laine-Kaulio et al. (2014) defined some parameters based on soil core measurements and literature values and some parameters based on calibration for their 3D physically based dual-permeability model of a forested hillslope. The only study known to us where a physically based 3D preferential flow model was inversely calibrated at catchment scale is from Yu et al. (2014).

We carried out this study in a headwater section (6 ha) of the Weierbach catchment, located in western Luxembourg, where Glaser, Jackisch, Hopp, and Klaus (2019) and Glaser et al. (2016) recently modelled the hydrological response using the 3D physically based model HydroGeoSphere (Therrien, McLaren, Sudicky, & Panday, 2010). The model reproduced the observed hydrograph well, but missed some specific features of the hydrograph, for example, peaks directly after dry conditions and the rise and recession of second, delayed peaks that are typical for the hydrological response at this site during wet conditions (Martínez-Carreras et al., 2016). Glaser et al. (2016) suggested that accounting for preferential flow could improve

the model simulations, which was supported by the findings of Jackisch et al. (2017) who – by the use of plot-scale tracer irrigation experiments – suggested a fine-scale network of inter-aggregate voids that enabled a fast vertical movement of water from surface to subsoil layers. In a subsequent study, Glaser et al. (2019) included preferential flow in the model following the approach to directly transfer parameters derived from small-scale field observations to catchment scale. They calibrated dual-permeability parameters with the tracer experiments by Jackisch et al. (2017) in a plot scale Monte Carlo simulation and subsequently transferred some parameter sets to the 3D representation of the catchment. This incorporation of a dual-permeability approach improved solute transport at plot scale, yet the transferred parameter sets did not improve discharge simulations at catchment scale.

In this study, we built upon the work and setup of Glaser et al. (2019) but calibrated preferential flow parameters against catchment discharge, analogous to Yu et al. (2014). In order to account for vertical preferential flow through the fine-scale inter-aggregate pores, we implemented a dual-permeability approach. In addition, we compared the calibrated model scenarios to distributed data, namely in situ soil moisture and spatial patterns of surface saturation. Subsequently, we tested the scale invariance of hydraulic parameters in reverse direction (i.e., from catchment scale to plot scale) by transferring the calibrated parameter sets to the plot scale model of Glaser et al. (2019), and we evaluated the parameter performance for simulating bromide depth profiles from plot scale tracer experiments (Jackisch et al., 2017). Instead of using a nonlinear optimization algorithm like Yu et al. (2014), we applied the Markov chain Monte Carlo (MCMC) Differential Evolution Adaptive Metropolis (DREAM) algorithm (Guillaume & Andrews, 2012; Vrugt, 2016) to calibrate parameters and estimate their uncertainties. Nonlinear optimization algorithms and MCMC algorithms are both widely used in environmental modelling for the estimation of parametric uncertainties. However, MCMC algorithms tend to provide a more robust assessment of parameter uncertainty through parameter distributions and confidence intervals (Hartig, Calabrese, Reineking, Wiegand, & Huth, 2011; Lu, Ye, & Hill, 2012). They have been frequently and successfully used in hydrological modelling (Arora et al., 2012; Gallagher & Doherty, 2007; Joseph & Guillaume, 2013; Shi, Ye, Finsterle, & Wu, 2012; Vrugt, 2016; Vrugt, ter Braak, Gupta, & Robinson, 2009), yet we are not aware of any study that applied an MCMC algorithm for a 3D physically based catchment model with dual permeability.

The hypothesis that guided our study was: The dual permeability model, calibrated on catchment discharge, will perform better than a single permeability model when simulating the hydrological response of the Weierbach headwater catchment. Our specific research questions were: (a) Is it feasible to use the DREAM algorithm for a calibration of a 3D physics-based hydrological model at catchment scale?, (b) Does the calibrated preferential flow model improve the simulation of the hydrological response, compared to a single permeability setup? and (c) Are the fitted preferential flow parameters suitable to reproduce observed plot-scale tracer profiles?

## 2 | MATERIALS AND METHODS

### 2.1 | Study site description

The Weierbach catchment (Figure 1) is a 42 ha sub-catchment of the Attert catchment located in the northwest of Luxembourg (49,490 N, 5470 E). Its geology is dominated by slate, phyllites, schist and quartzite, and its shallow soils are mainly cambisol, ranker and lithosol (Juillet, Iffy, Pfister, & Hissler, 2011). The climate is temperate and rainfall is mainly driven by eastward moving Atlantic depressions (Pfister, Kwadijk, Musy, Bronstert, & Hoffmann, 2004). Mean annual precipitation is about 900 mm, of which approximately half becomes streamflow (Kavetski, Fenicia, & Clark, 2011; Pfister, McDonnell, Hissler, & Hoffmann, 2010). The catchment is dominated by European beech (*Fagus sylvatica*), sessile oak (*Quercus petraea*) and Norway spruce (*Picea abies*). Model simulations presented in this study were carried out in a 6 ha headwater of the catchment (Figure 1).

Data used in this study include time series of discharge and of soil moisture as well as surface saturation observations. Water stage was measured at a stream gauge (December 2011–April 2014) with a water pressure transducer (ISCO 4120 Flow Logger, 5 min logging interval). Stage was converted to discharge using a power-type rating curve (Glaser et al., 2016). Soil moisture time series (December 2012–April 2013) were provided from five volumetric water content sensors (M1–M5; CS616, Campbell Scientific, 30 min logging intervals; manufacturer's calibration) that were installed horizontally in 10 cm depth each across the riparian zone of the Weierbach headwater (Figure 1). A high-resolution map of surface saturation in the riparian zone in February 2013 obtained with a portable thermal infrared (TIR) camera (Glaser et al., 2016) was used to evaluate the simulated surface saturation. More details about the Weierbach catchment are described elsewhere (e.g., Glaser et al., 2016; Schwab, Klaus, Pfister, & Weiler, 2016; Wrede et al., 2015).

### 2.2 | The catchment model

We used the fully integrated, spatially distributed surface-subsurface model HydroGeoSphere (HGS). HGS simultaneously simulates surface flow and variably saturated subsurface flow. It relies on a modified Richards equation coupled with soil water retention parameters to describe three-dimensional variably saturated flow in porous media

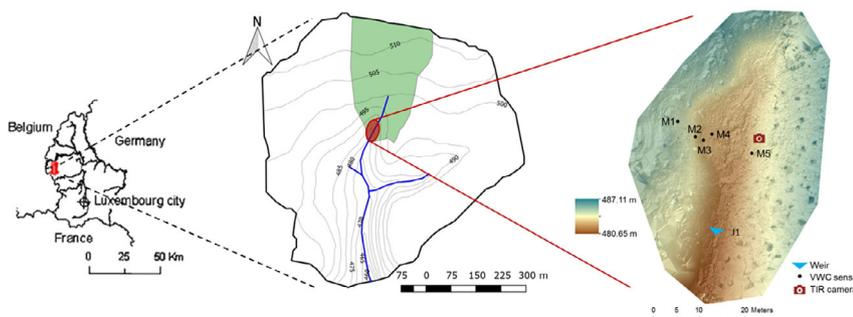
(Therrien et al., 2010), and it uses the two-dimensional diffusive wave approximation of the shallow water equations for surface flow. Surface and subsurface flow can be coupled via a dual-node approach, that is, the interface between the surface and the subsurface is represented as a very thin layer of porous material through which a Darcy flux is driven by hydraulic head differences. Preferential flow can be implemented in HGS via the dual permeability approach by Gerke and van Genuchten (1993). In this approach, the pore system is conceptually separated into two overlaying domains, a macropore domain and a less conductive matrix domain. Water flow in both domains is described by the Richards equation, and the domains are coupled by a first order bidirectional exchange term:

$$C_{dp} \frac{\delta h_{dp}}{\delta t} = \frac{\delta}{\delta z} \left( K_{s_{dp}} \frac{\delta h_{dp}}{\delta z} - K_{s_{dp}} \right) - \frac{\Gamma}{dp} - S_{dp} \quad (1)$$

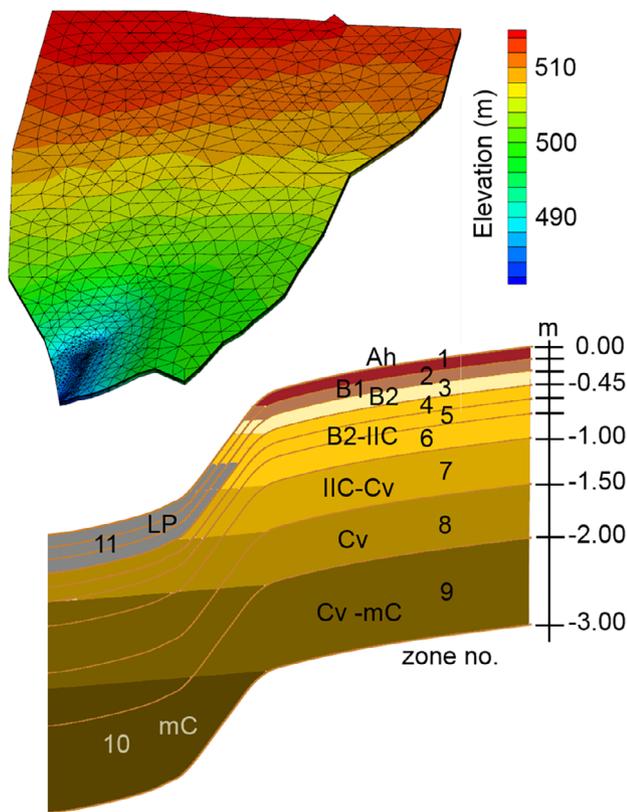
$$C \frac{\delta h}{\delta t} = \frac{\delta}{\delta z} \left( K_s \frac{\delta h}{\delta z} - K_s \right) - \frac{\Gamma}{1-dp} - S \quad (2)$$

Equation (1) describes the flow through macropores (preferential flow), Equation (2) through the soil matrix (matrix flow). Note that, for clarity, the flow equations are shown in their one-dimensional form although HGS simulates three-dimensional water flow in the matrix and in the macropore domain.  $K_s$  and  $K_{s_{dp}}$  are the saturated hydraulic conductivities of the matrix and the macropores, respectively,  $\Gamma$  is the exchange term between matrix and macropores where exchange is driven by pressure head gradients between the two domains,  $dp$  the volumetric fraction of the macropore domain,  $S$  and  $S_{dp}$  are sink terms and  $C$  and  $C_{dp}$  describe the specific soil water capacities in the two domains.

We based the model of the 6 ha headwater of the Weierbach catchment (Figures 1 and 2) on the model setup by Glaser et al. (2016, 2019). Here, we briefly present the most important aspects of the setup. Climate data (precipitation, air temperature at 2 m, wind speed, relative humidity and net radiation) from two meteorological stations approx. 4 km south of the study site were used for input of precipitation and potential evapotranspiration in daily resolution as boundary conditions at the surface domain. Interception and actual evapotranspiration were calculated with HGS via the model of Kristensen and Jensen (1975). LAI values followed a seasonal relationship based on 8 days MODIS MOD15 data for the years 2012–2014 (see Figure 3 top in Glaser et al., 2016). Since the calibration and the validation



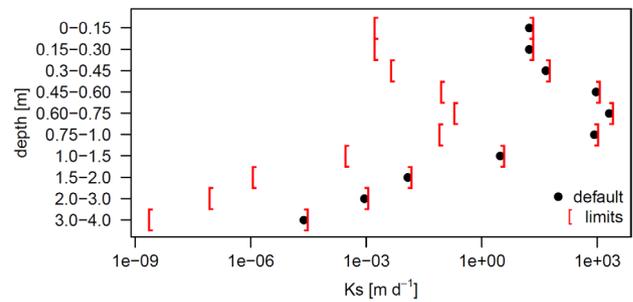
**FIGURE 1** The Weierbach catchment in Luxembourg (left) with the modelled 6 ha headwater region (green) and a close-up of the riparian zone (red/right), showing the positions of the weir, the thermal infrared camera (TIR) and the five volumetric water content (VWC) sensors (M1–M5), installed in 10 cm depth each. Elevations in the riparian zone map are given in metres above sea level



**FIGURE 2** Model domain of the 6 ha headwater catchment with finite element grid (top) and vertical cross-section through the subsurface setup of the hillslope-riparian-stream zone (bottom). Letters and colours indicate the soil horizons in accordance to soil profile and ERT profile information (Ah: topsoil, B1 and B2: subsoil horizons, LP: stagnic soil in the riparian zone, IIC: regolith, Cv: weathered bedrock, mC: solid slate). The numbers indicate the respective soil zone (cf. Table 1). The light brown lines indicate top and bottom sides of the model elements (adapted from Glaser et al., 2016)

period comprised the winter months, LAI values ranged from 0.6 to 1.1 (see Table S1 in the Supporting Information for other transpiration, evaporation and surface parameters). Eleven different soil zones were parameterized in the model, representing topsoil and subsoils at the hillslopes (Ah, B1, B2), a stagnic soil in the riparian zone (LP) and universally underlying transitional layers from subsoil to regolith (IIC), weathered bedrock (Cv) and solid slate (mC) (Figure 2). We implemented the dual-permeability approach with the primary intent to enable vertical preferential flow, in addition to horizontal non-uniform flow pathways as defined by the different soil layers. However, the flow equations are formulated in 3D and also allow other directions of flow, depending on developing gradients.

We defined 'default' soil hydraulic parameters (as starting values for the calibration) for the matrix domain based on field evidence and modelling work described in Glaser et al. (2016) (Table 1 for the hydraulic parameters being subject to calibration; Tables S2 and S3 for the parameters that were not calibrated). Glaser et al. (2016) derived soil hydraulic characteristics for the matrix domain from field



**FIGURE 3** The fixed soil layer ratio of the matrix saturated hydraulic conductivity  $K_s$  with the upper and lower limits for calibration. See Figure S2 in the Supporting Information for soil layer ratios used for the other soil hydraulic parameters in the calibration

and laboratory measurements as well as from electrical resistivity tomography (ERT) data for the different soil layers in the model and refined them in a manual calibration. Hydraulic conductivities were estimated to be highest in middle soil layers (0.45–1.0 m depth) with a minimum at the deepest soil layers (1.75–4 m depth), whereas porosities were highest in topsoil layers, decreasing with depth (Table 1). We chose default parameters for the macropore domain (Table 1; Table S1) from a range of preliminary test simulations with an HGS plot scale model (Glaser et al., 2019) that could reproduce the general preferential flow characteristics of the bromide depth profiles observed in plot scale tracer experiments (Jackisch et al., 2017) (see also section Plot scale simulation below). The stagnic soil in the riparian zone (LP, soil zone 11) was implemented without macropores (Table S3).

As the original dual-permeability model (Glaser et al., 2019) was computationally expensive, for this study we decreased the spatial and temporal resolution of the simulation to adapt it to the calibration approach. By reducing the number of nodes in the grid and modifying the adaptive time-stepping options in HGS, we were able to reduce the runtime by approx. 80%. The adapted nested model grid was composed of nine layers of three-sided prisms with vertical element heights ranging from 0.15 m (top layers) to 1 m (bottom layer) and horizontal element lengths ranging from about 20 m (hillslope) to 1 m and less (riparian zone and stream bed). The resulting coarsened model consisted of 920 nodes for each of the 10 soil layers (compared to 3460 nodes in the original setup) with the highest density of nodes in the riparian zone (Figure 2). By comparing hydrographs simulated with the original grid and with the coarsened grid, we verified that the coarsened resolution of the model was suitable for calibration purposes. We assigned a critical depth boundary to the perimeter of the surface domain, allowing water to exit the system everywhere at the lateral surface edges. All side edges and the bottom of the subsurface domain were no-flow boundaries (cf. Glaser et al., 2016).

### 2.3 | Selection of parameters for calibration

Five parameters were selected for calibration based on a principal component analysis (PCA) that we carried out using differently

Soil zone	Layer (m)	$K_s$ (m d <sup>-1</sup> )	$n$ (m <sup>3</sup> m <sup>-3</sup> )	$K_{s_{dp}}$ (m d <sup>-1</sup> )	$n_{dp}$ (m <sup>3</sup> m <sup>-3</sup> )	$dp$ (-)
1	0–0.15	1.71E+01	0.74	6.795E+01	0.89	0.18
2	0.15–0.3	1.71E+01	0.61	1.044E+02	0.89	0.08
3	0.3–0.45	4.59E+01	0.44	2.800E+02	0.89	0.06
4	0.45–0.6	9.30E+02	0.30	4.115E+03	0.89	0.07
5	0.6–0.75	2.04E+03	0.14	9.905E+03	0.89	0.08
6	0.75–1.0	8.40E+02	0.20	3.575E+03	0.89	0.06
7	1.0–1.5	3.00E+00	0.15	3.160E+01	0.89	0.05
8	1.5–2.0	1.20E–02	0.10	1.620E–01	0.89	0.05
9	2.0–3.0	9.00E–04	0.04	1.415E–02	0.89	0.05
10	3.0–4.0	2.4E–05	0.01	6.220E–04	0.89	0.05

Notes: Soil hydraulic parameters that were not subject to calibration are listed in Table S2 in the Supporting Information. Please refer also to the Supporting Information for a more detailed description on deriving the default parameter values and their depth profiles.

Abbreviations:  $dp$ , percentage macropore domain of total domain;  $K_s$ , matrix saturated hydraulic conductivity;  $K_{s_{dp}}$ , macropore saturated hydraulic conductivity;  $n$ , porosity of matrix domain;  $n_{dp}$ , porosity of macropore domain.

performing macropore parameterizations that had been tested in previous catchment scale simulations (Glaser et al., 2019): saturated hydraulic conductivities of matrix and macropores ( $K_s$  and  $K_{s_{dp}}$ ), porosity of matrix and macropores ( $n$  and  $n_{dp}$ ) and the volumetric proportion of the macropore domain ( $dp$ ). Three different combinations of parameter calibration were compared, selecting for calibration: (a) all five parameters (MacroMat), (b) only macropore parameters (Macro) and (c) only the matrix hydraulic parameters (NoMacro) in a single domain model as the reference scenario (Table 2). The number of soil layers (10; for soil zone 11, i.e., stagnic soil in riparian zone, macropores were not implemented and soil hydraulic parameters were not varied; cf. Table S3) multiplied by the number of parameters for calibration (five, three or two) would lead to vary many degrees of freedom for the DREAM algorithm and possibly to unrealistic depth profiles of the parameters. Therefore, we calibrated parameter values of the different soil layers jointly as one hyperparameter instead of separately by fixing the parameter ratio between the different soil layers for each parameter. This was done to preserve (a) observed depth profiles of the parameters, thus avoiding unrealistic changes between soil layers, and (b) to keep the degrees of freedom for the DREAM algorithm at a feasible number. The fixed parameter ratios were defined by the default parameter set (Table 1) and the upper and lower limits of the parameters for the calibration were chosen to stay in the parameter range of the preliminary test runs being successful at plot scale (Figure 3, Figure S2; see the Supporting Information for a more detailed description on deriving the depth profiles of the default parameter values). Since hydraulic conductivities are known to vary over several orders of magnitude, their calibration was carried out on the common logarithm of  $K_s$ - and  $K_{s_{dp}}$ -hyperparameters. After calibration, the calibrated hyperparameters were multiplied with the parameter values of the default parameter set (cf. Table 1) to obtain the actual parameter values.

The calibration against catchment discharge was performed for the period from December 2013 to April 2014, building up to a total

**TABLE 1** Default parameter values of the five parameters being subject to calibration

**TABLE 2** Number and type of hyperparameters calibrated for the three calibration scenarios and abbreviation for scenarios to be used in the following text

Number of hyperparameters to be fitted	Hyperparameters for calibration	Name of scenario
5	$K_s, n, K_{s_{dp}}, n_{dp}, dp$	MacroMat
3	$K_{s_{dp}}, n_{dp}, dp$	Macro
2 (single domain)	$K_s, n$	NoMacro

of 100 days. The validation period for catchment discharge and soil moisture was from December 2012 to April 2013 (100 days as well). Surface saturation validation was carried out visually against surface saturation derived from a panorama TIR image from Glaser et al. (2016) in February 2013 (see Glaser et al., 2019 for details on the method). As there were no surface saturation data available for the validation period, surface saturation patterns were evaluated for the calibration period only.

In order to define initial conditions for both the calibration and the validation periods, a spin-up period of 1 year was run from December 2012 to November 2013 for the calibration period and from December 2011 to November 2012 for the validation period.

For comparison, we also simulated discharge with the default parameter set in a dual-permeability approach for the calibration as well as validation period.

## 2.4 | DREAM setup

We coupled HGS with the MCMC algorithm DREAM via the R DREAM package (Guillaume & Andrews, 2012), running it on an HPC cluster with 16 CPUs (see Supporting Information). DREAM produces a chain of simulations with parameter draws from a parameter

posterior distribution, where one random draw (model run) depends on the previous draw of the algorithm (Markov property) (Vrugt, 2016). In this study, the performance of the different parameterizations is evaluated with a maximum likelihood function, that is, a logarithmic Gaussian likelihood ( $\log L$ ) calculated on mean daily discharge values. As the parameterizations of subsequent model runs depend on likelihood values of current parameterizations, MCMC algorithms tend to stay in regions of the parameter space with better fit of modelled to measured values (with higher  $\log L$  values). The DREAM-algorithm is a multi-chain method, so a number of Markov chains are simultaneously exploring the parameter space (this study: 15 chains in parallel). This has the advantages that (a) the global optimum is more likely to be reached, (b) the algorithm can be easily parallelized and (c) the distance between the chains in parameter space can be used to autotune jumping parameters, so that convergence can be reached faster (Vrugt, 2016). The algorithm is supplemented with subspace sampling and outlier chain correction to further improve the efficient exploration of the parameter space.

When the parameter distribution does not change over a longer span of the calibration, convergence is reached and the parameter distribution of the Markov chains is assumed to represent the actual parameter distribution. Convergence is commonly evaluated using the Gelman–Rubin convergence criterion  $R$  (Gelman & Rubin, 1992), which compares the variance between the chains to the variance within the chains. A plot of  $R$  against iteration number is called Gelman-plot and allows the evaluation of the convergence (see an example of a Gelman plot in the Supporting Information, Figure S1).  $R$  theoretically converges to 1, and a value much greater than 1 suggests that the simulation should be continued (Brooks & Gelman, 1998). Recommendations on  $R$ -thresholds are 1.1, 1.2 or even higher (Guillaume & Andrews, 2012). Here, we used a threshold of 1.3 for the multivariate  $R$ -value, which can be calculated for a multivariate optimization problem (Brooks & Gelman, 1998). When the multivariate  $R$  stayed above the threshold for a longer period, we alternatively checked  $R$ -values of individual parameters for convergence, since the multivariate  $R$  is considered an approximate maximum of the univariate  $R$  over all variables (Brooks & Gelman, 1998). Prior to estimating the actual parameter distribution, we carried out a burn-in where the first 10% of the chain were discarded and not used for analysis, so that the parameter distribution obtained from the respective MCMC algorithm was independent of the starting point.

## 2.5 | Plot scale simulation

After the MCMC calibration of the catchment model we used *the best performing parameter set (highest  $\log L$ ) from each of the three calibration scenarios* 'MacroMat', 'Macro' and 'NoMacro' to simulate bromide transport at the plot scale, in order to test the reverse transfer of calibrated parameters to plot scale. This test built upon data derived from plot scale irrigation experiments (Jackisch et al., 2017) and previous plot scale simulations with a dual-permeability HGS model (Glaser

et al., 2019). For comparison, we also simulated the irrigation experiments using the default parameter set with dual-permeability (not calibrated). During the tracer (brilliant blue and bromide) irrigation experiment two 1 m<sup>2</sup> plots were irrigated for 1 hr with 50 mm (plots X and XII in Jackisch et al., 2017) and one 1 m<sup>2</sup> plot was irrigated for 1 hr with 30 mm (plot XI in Jackisch et al., 2017). After the end of the irrigation the soil profiles were sampled in 5 cm increments down to a depth of 1 m, and bromide concentrations were determined (for details on the irrigation experiments refer to Jackisch et al., 2017). We used these bromide profiles to validate the concentration profiles simulated with the calibrated parameter sets.

Solute transport in HGS is simulated with the advection–dispersion equation. We simulated bromide transport using a 6 m deep soil column with a horizontal area of 1 m<sup>2</sup>. The grid was defined by 0.25 m<sup>2</sup> quadratic elements with element heights of 1 cm between 0 and 4 m depth and element heights of 5 cm between 4 and 6 m depth. The definition of soil layers (depths and soil hydraulic properties) was identical to the catchment scale simulations described above for the upper 4 m. The lower 2 m of the model domain (4–6 m) served as porous storage ( $K_s = 1 \text{ m d}^{-1}$ ,  $n = 20\%$ ,  $\theta_r = 0.02$ ,  $\alpha = 6 \text{ m}^{-1}$ ,  $\beta = 1.5$ ) in order to prevent ponding of water at the bottom of the upper 4 m. Flow and transport were simulated in the subsurface only (no surface domain), since this can avoid numerical problems, and no surface runoff was observed during the experiments. Bromide transport parameters were set as follows: tortuosity = 0.1, diffusion coefficient =  $1.6 \times 10^{-4} \text{ m}^2 \text{ d}^{-1}$ , longitudinal and transverse dispersivities of  $d_l = 0.05 \text{ m}$  and  $d_t = 0.005 \text{ m}$  for the matrix and  $d_l = 0.1 \text{ m}$  and  $d_t = 0.01 \text{ m}$  for the macropore domain and mass exchange coefficient  $\omega_{ex} = 0.41$  for all soil layers. We partitioned the input fluxes (solute and water) at the upper boundary of the model between the macropore and matrix domain with a ratio of 90:10 (cf. Laine-Kaulio et al., 2014) and assigned no flow boundaries to the sides and the bottom of the model column. Initial saturation was identical for the matrix and macropore domain (corresponding to measurements before the start of the experiment), and the initial bromide concentration was set to zero ( $10^{-15} \text{ kg m}^{-3}$  in order to avoid numerical instabilities) in both domains. Please refer to Glaser et al. (2019) for more details on the HGS plot scale model setup.

## 3 | RESULTS

### 3.1 | Calibration

#### 3.1.1 | Convergence and parameters

The three calibration scenarios differed considerably in maximum  $\log L$  and iterations necessary for convergence (Table 3). The maximum  $\log L$  was highest for the 'NoMacro' scenario. While scenario 'Macro' needed the smallest number of iterations and converged in roughly 2 weeks, scenario 'NoMacro' and 'MacroMat' took 42 and 73 days, respectively, to reach a more relaxed Gelman–Rubin convergence criterion  $R$  threshold.

By using the optimization algorithm DREAM, we were able to explore the whole defined parameter space. Confidence

intervals narrowed down considerably after burn-in for some hyperparameters and stayed broad and non-informative for others (Figure 4).  $K_s$  and  $n$ , that is, the matrix hydraulic hyperparameters, formed narrow confidence intervals in every calibration scenario with saturated hydraulic conductivities tending to high values and porosity values tending to low values within the analysed parameter space. Hyperparameter  $dp$  converged close to the lower limit and did not exceed 0.3 for the two scenarios 'Macro' and 'MacroMat'. This means that as a result of calibration the proportion of the macropore domain to the total domain decreased considerably compared to the initial default parameter set (cf. Table 1). The

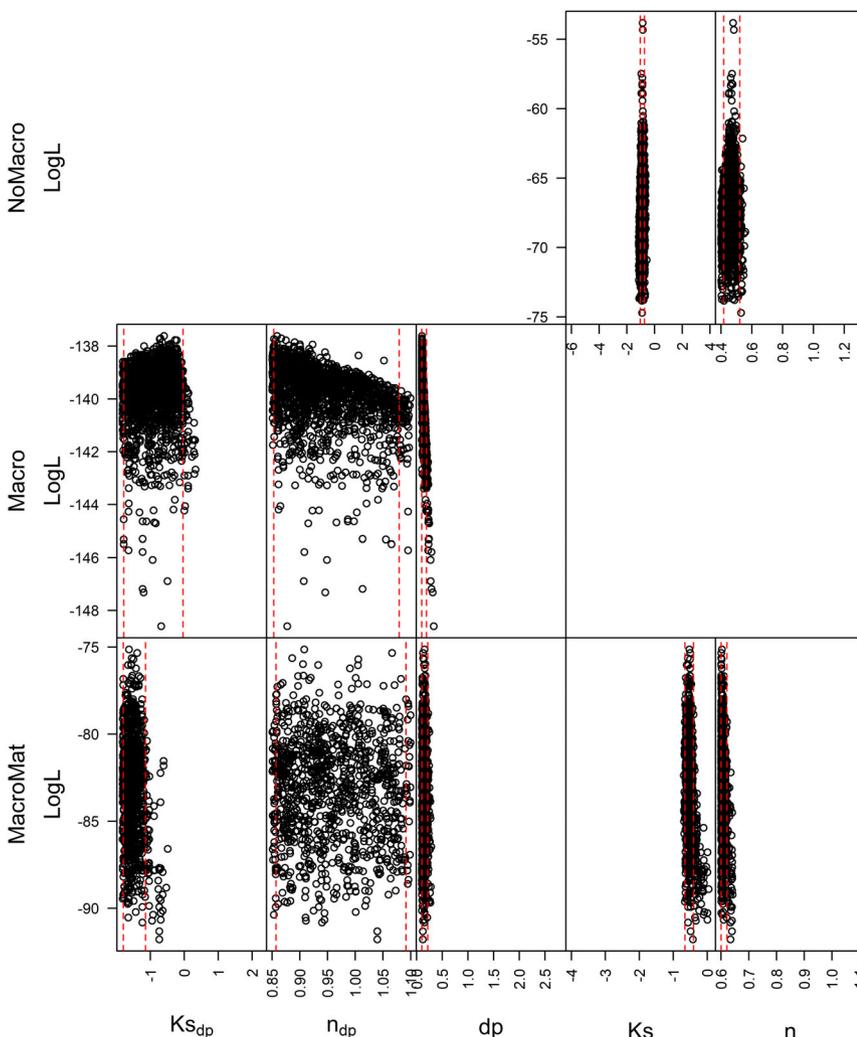
hyperparameter porosity of the macropores  $n_{dp}$  varied independently of  $\log L$  within the whole analysed parameter space for both scenarios 'Macro' and 'MacroMat'. Confidence intervals of hyperparameter  $K_{s_{dp}}$  had an intermediate width, covering the lower half of the analysed parameter space.

After conversion from hyperparameters to parameters, we calculated the effective hydraulic conductivities ( $K_{s_{eff}}$ ), that is, the  $dp$ -weighted average of matrix and macropore conductivities. All three scenarios showed conductivities in the same order of magnitude for the respective soil layers (Figure 5). The 95%-confidence intervals of  $K_{s_{eff}}$  were in general narrow.

Scenario	Max log L	R	Number of iterations	Runtime (days)
MacroMat	-75	1.33	1780	73
Macro	-138	1.07	420	16
NoMacro	-54	1.26	1220	42

Notes: Since 15 chains were run in parallel, iterations have to be multiplied by 15 to get the number of total model evaluations, and runtime has to be multiplied by 16 for summed-up CPU-time for each scenario.

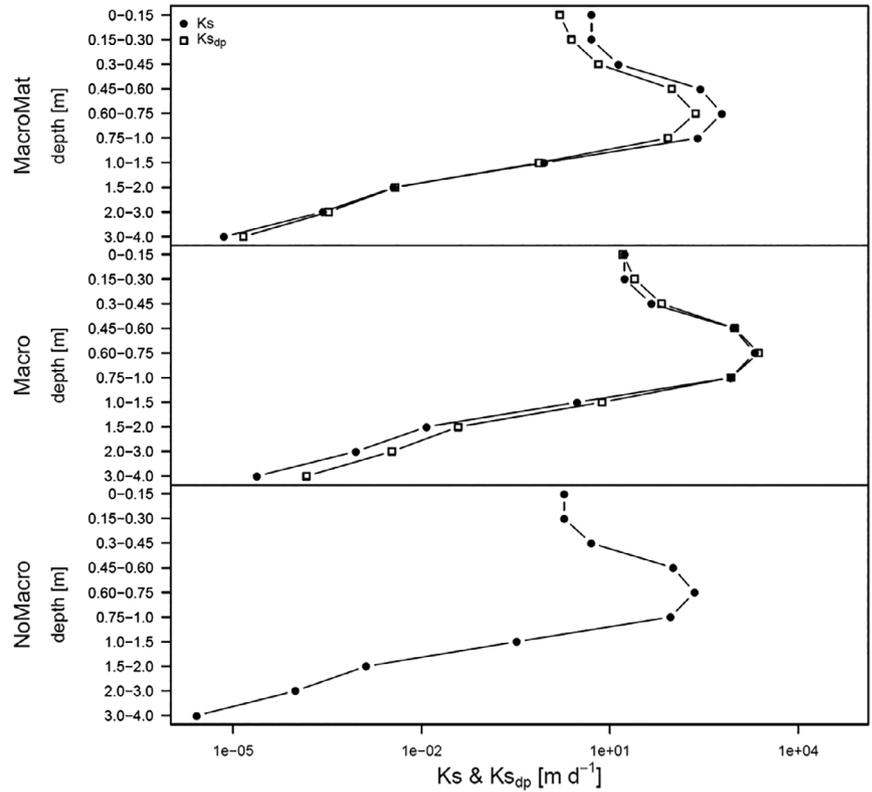
**TABLE 3** Maximum log likelihood  $L$ , multivariate Gelman–Rubin convergence criterion  $R$ , number of DREAM iterations and DREAM runtime for each scenario



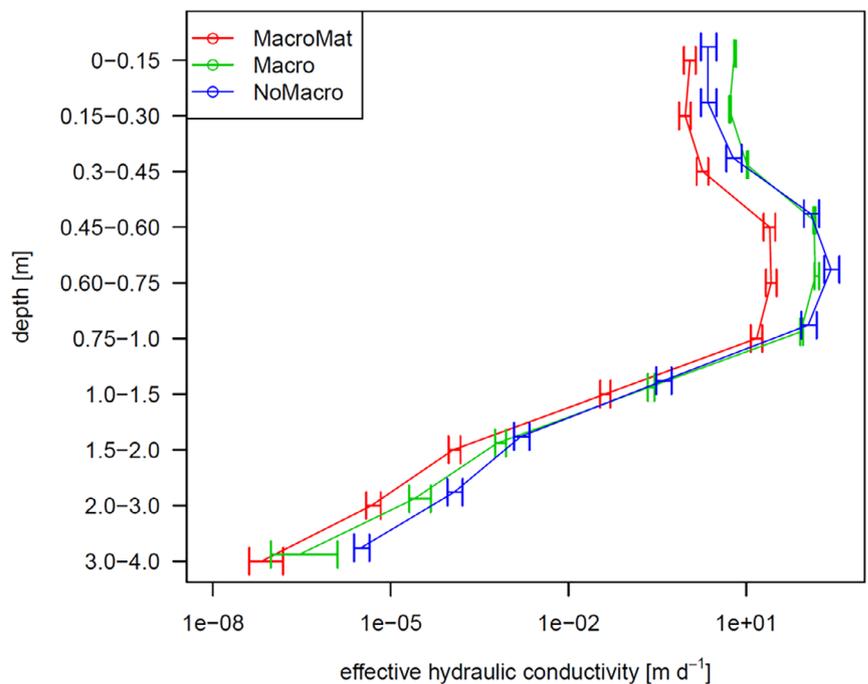
**FIGURE 4** Scatterplots of  $\log L$  versus hyperparameter values (dimensionless) with 95%-confidence intervals (red dashed lines) after burn-in for the respective hyperparameters. One point in the scatterplots represents a single HydroGeoSphere model run. The dimensionless values on the x-axis signify the multiple of the respective default parameter as listed in Table 1 (dimensionless value = 1 means default parameter value). For the parameters  $K_s$  and  $K_{s_{dp}}$  the log of the values was varied (dimensionless value = 0 means default parameter value). The scaling of the x-axis reflects the upper and lower limits of the parameter space allowed for calibration. The range of the parameter space for the 'NoMacro' scenario was set wider to allow for a similar range of effective hydraulic conductivities and porosities as compared to the two other scenarios

Although upper and lower limits for the calibration of  $K_{s,dp}$  were approximately two orders of magnitude higher than the limits for  $K_s$ , both conductivities converged to similar values, generally in the same order of magnitude, in the respective soil layers after calibration (Figure 6). For scenario "MacroMat", matrix conductivities of the best performing model run (highest log  $L$ ) were higher than macropore conductivities in soil layers above 1.5 m

depth and lower than macropore conductivities in soil layers below 1.5 m depth. The scenario "Macro" did not show these reverse conductivities. Instead,  $K_s$  and  $K_{s,dp}$  were very similar in layers above 1 m and diverged more below this depth. The maximum difference between  $K_s$  and  $K_{s,dp}$  occurred in the scenario 'Macro' in the deepest layer and was less than one order of magnitude (Figure 6).



**FIGURE 5** Matrix hydraulic conductivities and macropore hydraulic conductivities for each soil layer for the three different calibration scenarios (best performing model run with highest log  $L$ )



**FIGURE 6** Median effective saturated hydraulic conductivities (i.e., the  $dp$ -weighted average of matrix and macropore conductivities) in the different soil layers for each calibration scenario. Horizontal bars indicate the 95%-confidence intervals

### 3.1.2 | Performance of best parameter sets

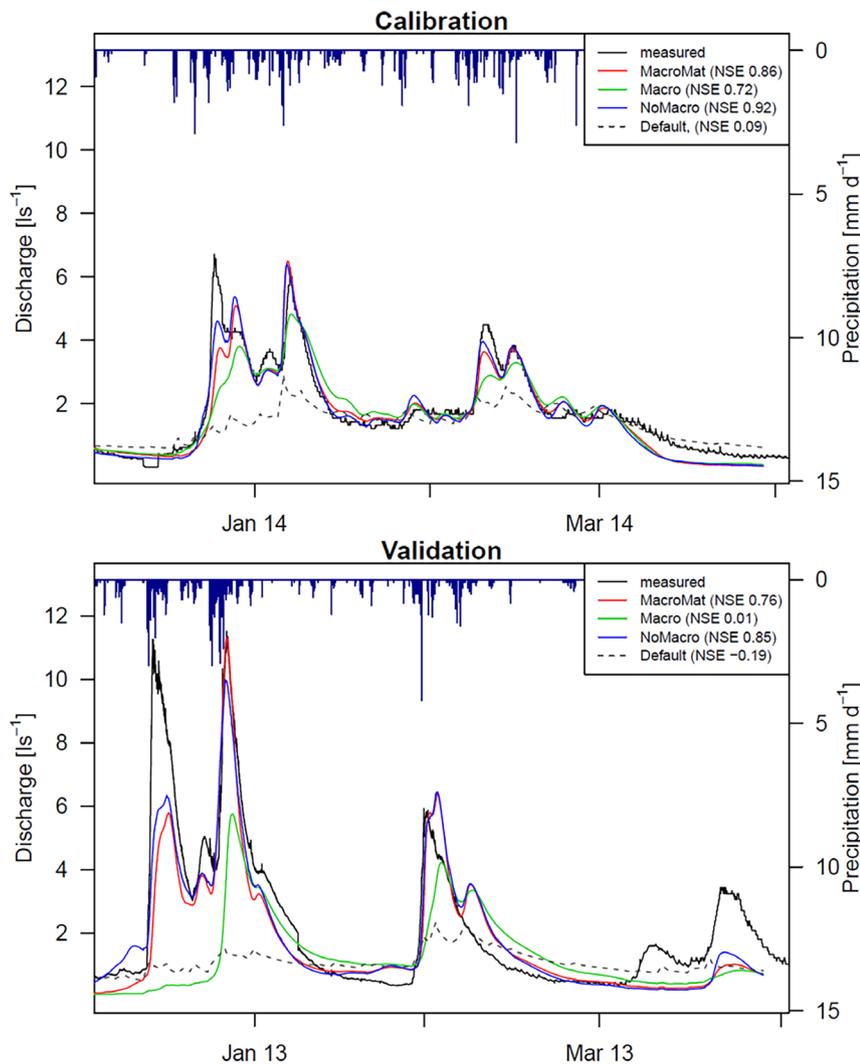
The calibrated parameter values for the model run with maximum log  $L$  for each scenario are listed in Table S4. Nash–Sutcliffe efficiency values (NSE) for the discharge simulation using the parameter set with the maximum log  $L$  of each scenario (Table 3 and Table S4) showed that 'MacroMat' (NSE 0.86) and 'NoMacro' (NSE 0.92) performed quite similar to each other and better than 'Macro' (NSE 0.72). This means that the scenario only calibrated on preferential flow parameters did not describe streamflow as well as the scenarios including matrix hydraulic parameters in the calibration procedure, particularly during peak flow (Figure 7, top). All three calibration scenarios clearly improved the discharge simulation compared to the uncalibrated default parameter set (NSE 0.09).

### 3.2 | Validation

Minor performance differences between the calibration scenarios that occurred during discharge calibration became more pronounced

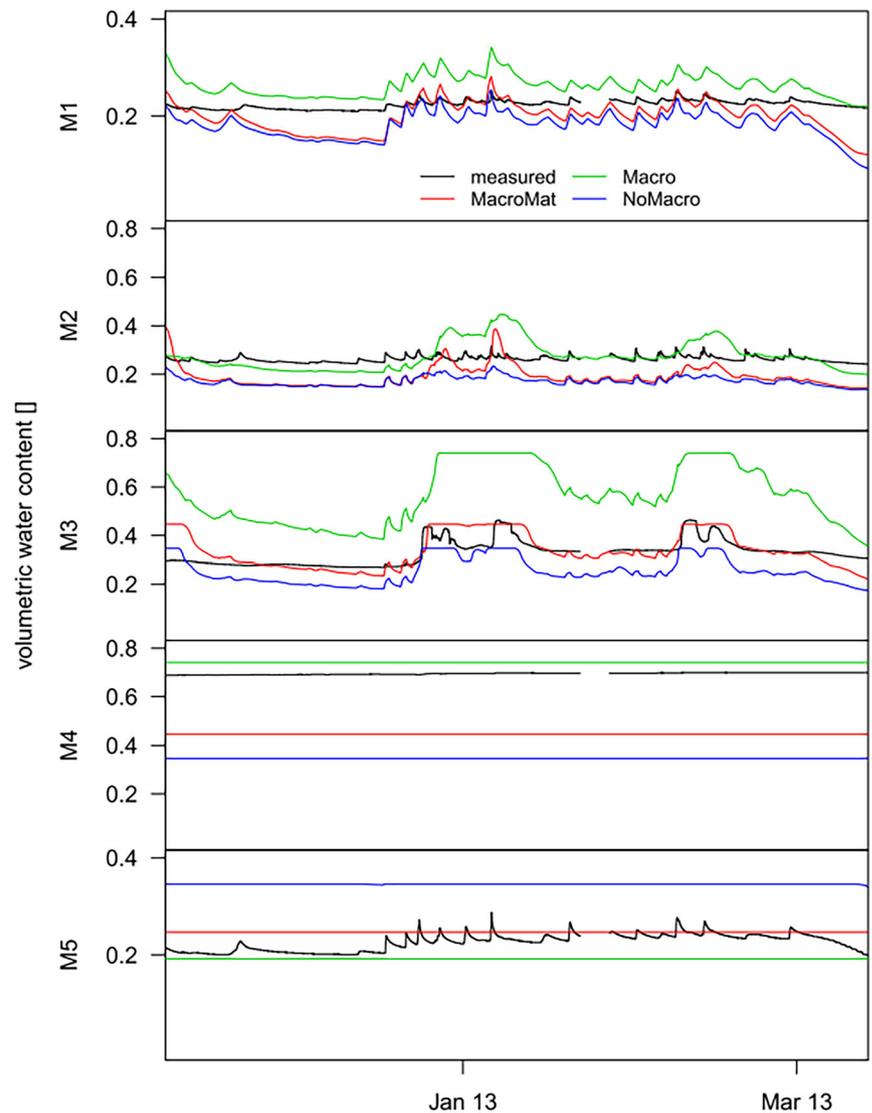
during the validation period (Figure 7, bottom). The scenario 'Macro' was not able to model the discharge during the validation period in a satisfactory way (NSE 0.01) and underestimated discharge considerably, particularly at the beginning of the validation period. This behaviour was most likely related to the high matrix porosity values of the three uppermost soil layers, which were not calibrated in this scenario (Table 1). The other two scenarios again did not differ substantially from each other and simulated discharge well for the validation period (with NSE 0.76 and 0.85, respectively), except for the first discharge peak in December 2012, which was reproduced in terms of dynamics but not in terms of absolute discharge value. This was also the case for the discharge peak in December 2013 during the calibration period. Again, the scenario 'NoMacro' had the highest NSE.

The visual validation comparing observed with simulated soil moisture responses at the five monitoring sites remained inconclusive (Figure 8). Since the soil moisture sensors were not calibrated to the Weierbach soils, the comparison between observed and simulated soil moisture focussed on the dynamics and amplitude, not on absolute values. In some cases, simulated soil moisture showed a higher



**FIGURE 7** Hydrographs of measured (black) and modelled streamflow during the calibration (top) and the validation period (bottom) with corresponding NSE values in the legend. From each scenario the parameter set with the highest log  $L$  was chosen for validation. Additionally, the hydrographs simulated with the default parameter set are shown. Note that snowmelt processes were not included in the model, so the discharge peak in March/April 2013 was not represented by any of the scenarios

**FIGURE 8** Soil moisture time series (10 cm depth) during the validation period for M1 to M5 (see Figure 1 for locations). From each calibration scenario the parameter set with the highest log L was chosen for validation



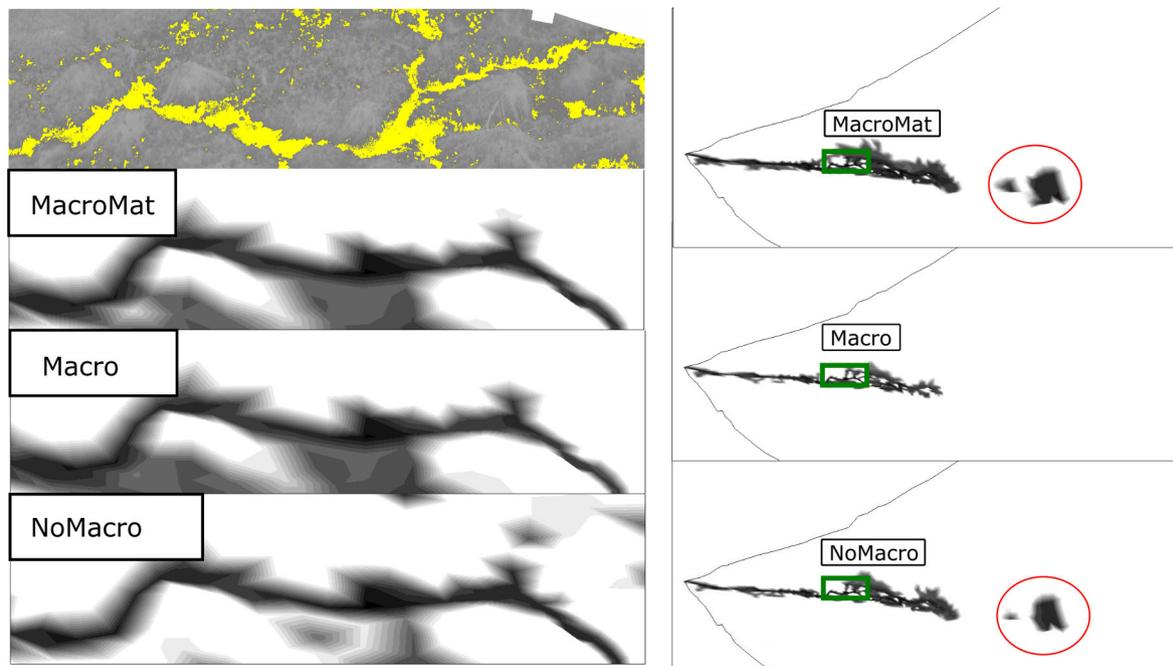
amplitude than observations (M1, M2 and M3), although dynamics were similar between simulations and observations. Soil moisture in the riparian zone, represented by M4, was characterized by an absence of dynamics, indicating saturated conditions. Soil moisture response of M4 was captured well by the simulations with respect to the missing dynamics, remaining at saturation at the respective porosity values (cf. Table S4 in Supporting Information). In contrast, for M5, simulations did not reflect the observed dynamics but showed flat soil moisture time series. M5 in the field was located at the transition between riparian zone and hillslope, whereas in the model M5 was placed still in the riparian zone (soil zone 11, Figure 2), therefore showing the typical absence of dynamics. Overall, the scenario 'NoMacro' agreed best with observations with respect to dynamics and amplitude.

Simulated surface saturation patterns of the three calibration scenarios did not show meaningful differences in performance when compared to a panorama TIR image taken during a wet period in February 2013 in the riparian zone (Figure 9). Evaluated at a larger scale, 'MacroMat' and 'NoMacro' showed a surface saturated area above the stream, which the scenario 'Macro'

did not show. This area is known to be saturated during very wet periods and thus provides additional qualitative validation information.

### 3.3 | Plot scale model

Despite showing differences in simulated catchment discharge, soil moisture and surface saturation, the two dual-permeability scenarios 'Macro' and 'MacroMat' simulated similar bromide depth profiles at plot scale (Figure 10). The fit to observed data was better for the 30 mm irrigation than for the 50 mm irrigation. Although the dual-permeability scenarios simulated bromide depth profiles that showed slightly elevated concentrations at depth, both parameter sets were not able to capture the observed bromide depth profiles with their distinct bromide concentration peaks around a depth of 0.6 m satisfactorily. The single-domain scenario ('NoMacro') simulated a steadily decreasing bromide depth profile for both irrigation rates and thus failed completely to reproduce concentration peaks deeper in the profile. The default parameter set captured the main characteristics of



**FIGURE 9** Surface saturation patterns for the three calibration scenarios (using the parameter set of the simulation with highest log  $L$  per scenario) compared to the surface saturation in the panorama TIR image of February 11, 2013 (left). The observed panorama (top left) is composed of processed, linear grey-scaled (black  $-5^{\circ}\text{C}$ , white  $+20^{\circ}\text{C}$ ) TIR images, and surface saturated pixels estimated based on the water temperature are shown in yellow. In the modelled panoramas, surface saturated pixels are depicted with different shades of grey for different surface water depths (minimum depth  $10^{-7.5}$  m, logarithmic scale). Red circles (right) indicate surface saturation spots, which are known to appear in the field during very wet periods. Green rectangle signifies location of the saturation patterns shown in more detail on the left

the bromide depth profiles, simulating elevated bromide concentrations at a depth of 0.6 m, which fitted well the observations.

## 4 | DISCUSSION

### 4.1 | Applicability and limitations of the calibration approach

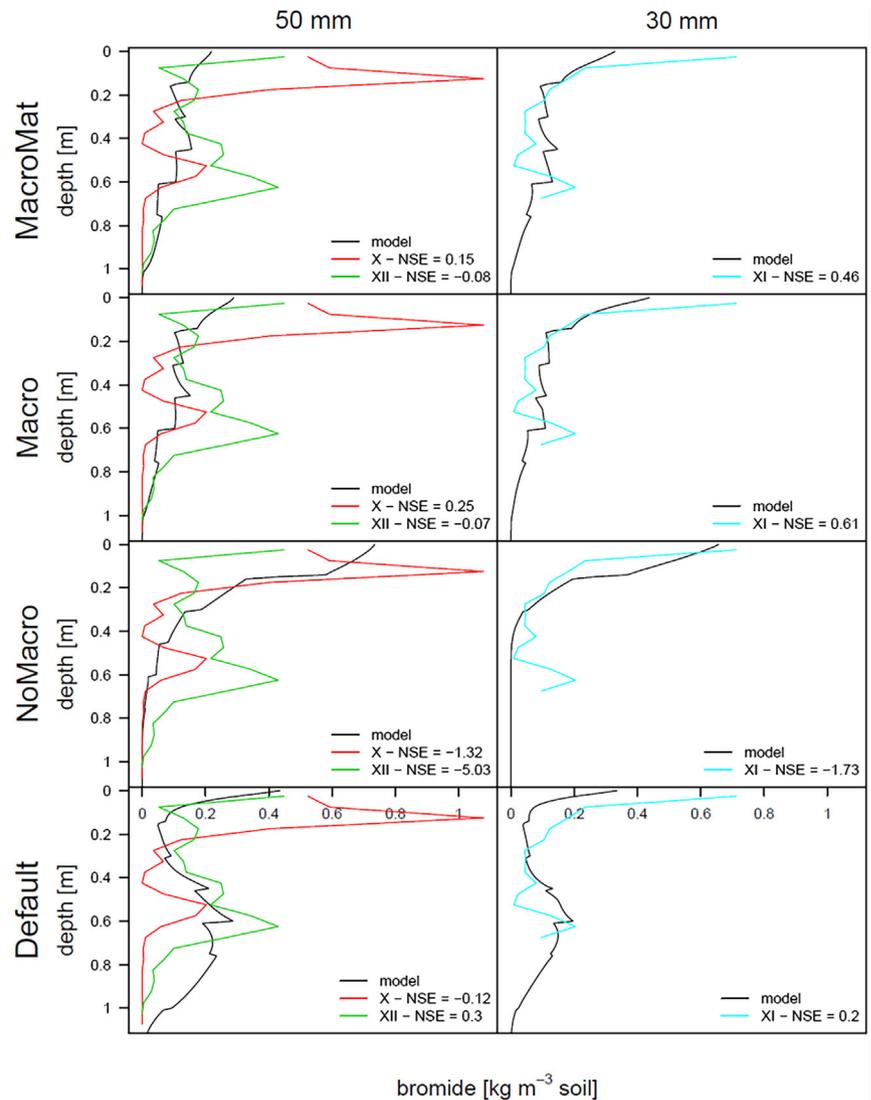
We successfully applied the automatic MCMC calibration algorithm DREAM for calibrating a 3D physically based catchment scale dual-permeability model with 10 soil layers against discharge data. Convergence was reached in all three scenarios, although DREAM runtime and number of iterations necessary for convergence differed substantially between the tested scenarios. The scenario calibrated on all five hyperparameters had the longest runtime with 73 days.

The long runtimes clearly limited the number of parameters to be calibrated and the number of scenarios that could be investigated within a reasonable timeframe. Nonetheless, DREAM proved to be very suitable for our calibration study at the catchment scale. We applied two constraints to balance computational costs and degrees of freedom, and thus, to realize a parameter calibration for a 3D catchment model with multiple soil layers. By doing a PCA on a set of 40 parameter sets that had been assembled during preparatory work for Glaser et al. (2019), we identified the five parameters with the highest influence on discharge simulation and chose those for

calibration in this study. We certainly cannot exclude that the calibration of more or other soil hydraulic parameters, such as the van Genuchten parameters  $\alpha$  and  $\beta$ , could have resulted in a better performing parameter set. However, based on the PCA results, we were confident that we calibrated only those parameters that seemed to be most influential for simulating the hydrometric response at catchment scale.

The second approach that facilitated a successful convergence of the optimization routine and reasonable runtimes was to work with hyperparameters, that is, establish fixed ratios of parameter values between soil layers. However, the primary reason for applying the hyperparameter approach was to preserve observed depth profiles of soil hydraulic parameters and to prevent parameter sets with sharp changes in hydraulic properties between layers, which would be physically unrealistic and not consistent with field evidence. As a secondary effect, this approach also reduced the number of parameters to be fitted. The depth profiles (i.e., ratios) for  $K_s$  and  $n$  of the matrix domain were based on field measurements, and the depth profiles of  $K_{s,dp}$  and  $n_{dp}$  were defined based on previous detailed modelling work (Glaser et al., 2016; 2019) that reproduced in plot scale models the general characteristics of bromide depth profiles obtained in sprinkling experiments (cf. Jackisch et al., 2017). It is possible that discharge simulations could have been improved by not using these fixed ratios. However, to us it was more important to only allow those parameter combinations that reflected realistic depth profiles of hydraulic properties, typical for this site.

**FIGURE 10** Observed (in red, green and blue; Jackisch et al., 2017) and modelled bromide concentrations (using the parameter set of the simulation with highest log L per scenario and the default parameter set, respectively) for the three irrigation plots, that is, X, XI and XII. Plots X and XII received 30 mm of irrigation, plot XI received 50 mm. NSE values show the goodness of fit between measured values and the respective model run



There are not many examples of automated calibration of a coupled surface-subsurface physics-based model in 3D at the catchment scale (e.g., Ala-aho et al., 2017; Hwang et al., 2018; Yu et al., 2014) due to the high computational costs that are particularly associated with combining these models with automated parameter estimation and uncertainty analysis procedures. All studies have limited the number of parameters to be calibrated by performing a priori sensitivity studies and/or simplifying the description of the model domain. The studies mentioned above calibrated 9–12 soil hydraulic parameters simultaneously, and none of them considered multiple soil layers in the subsurface, as we did in this study. Other calibration strategies such as using grid coarsening to identify the hydrological variables most suitable for calibration have also been suggested to make the calibration of physics-based models more efficient (von Gunten et al., 2014; Wildemeersch et al., 2014).

In combination with physically based hydrological models, the PEST suite of algorithms (Doherty, 2009) has often been used for automated parameter estimation and uncertainty analysis (e.g., Ala-aho et al., 2017; Wildemeersch et al., 2014). For instance, the null-

space Monte Carlo (NSMC) analysis within PEST is efficient with respect to its model run requirements and may have provided faster runtimes. However, parameter sets identified with this method may not necessarily be a sample of the posterior probability density function in a formal Bayesian sense (Keating, Doherty, Vrugt, & Kang, 2010). We chose to use DREAM (Vrugt, 2016) as it is theoretically more rigorous and comprises a formal Bayesian methodology, using a formal convergence criterion (Gelman–Rubin criterion  $R$ , cf. section ‘Dream setup’ above). We assume that the gain with respect to faster runtimes by using PEST (or a different optimization algorithm) would not have been sufficient to permit adopting a different calibration strategy with, for example, more parameters to be calibrated.

The goal of our study was not to compare different calibration strategies. As is commonly done in model calibration studies, we decided on one particular model domain setup, chose a particular calibration approach and applied it. Yet, there is a gap in knowledge about adequate calibration strategies for coupled surface-subsurface physics-based models at the catchment scale. Thus, subsequent

studies could implement different calibration strategies for the same problem to evaluate how our a priori decisions influence calibration results and the conclusions we draw from them. Such systematic tests of calibration strategies could, for example, vary the number of parameters to be calibrated, apply fixed or variable parameter ratios between layers, use different macropore representations or vary the number of soil layers to be considered.

## 4.2 | The effect of macropores is calibrated away

Neither of the two dual-permeability model scenarios ('MacroMat', 'Macro') was superior to the performance of the single domain model scenario ('NoMacro'). This is in contrast to the findings of Yu et al. (2014) who could improve the performance of their hydrological model by including a macropore parameterization. During calibration as well as during validation the 'NoMacro' scenario achieved the highest agreement with observed discharge, and evaluating soil moisture time series did not show any meaningful differences between the dual-permeability scenarios and the single permeability setup. All three scenarios simulated similar saturation patterns that agreed reasonably well with TIR observations considering the model grid resolution. However, only the scenarios 'MacroMat' and 'NoMacro' also simulated a saturated area above the stream, which is known to appear during very wet periods. This, together with the fact that the scenario 'Macro' did not perform well during validation, suggested that the calibration of only the macropore parameters, without calibrating also the matrix domain parameters at the same time, was not a successful strategy.

In the best performing calibrated parameter sets, the proportion of the macropore domain to the total domain,  $dp$ , became very small (never above 0.03; cf. Table S3) and  $K_s$  of the macropore domain became very similar to matrix  $K_s$  (Figure 5) so that the effect of a macropore domain with faster flow became insignificant. The fact that the effect of macropores was essentially calibrated out by the optimization algorithm suggested that the implementation of a dual-permeability approach, which we performed in order to represent the fine-scale network of inter-aggregate voids as identified by Jackisch et al. (2017) at the Weierbach site, was not advantageous as compared to a single domain model setup (considering only matrix flow) at the catchment scale. This implies that in this case the  $K_s$  parameter in the single-domain setup could be adjusted to be just as effective as the dual-permeability model to account for the impact of preferential flow in catchment-scale simulations. An effective parameterization of a single-permeability model may not be possible anymore if larger macropores such as slope-parallel soil pipes are to be considered. Also the dual-permeability approach would likely not be appropriate for such cases. Representing the effect of larger macropore features or soil pipes on water flow would call for a different mathematical description altogether, accounting for non-diffusive or even turbulent flow behaviour in a possibly explicit way.

Implementing the hyperparameter approach was part of our calibration strategy. The fixed ratios limited the variability of macropore

characteristics over depth, yet the macropores were connected throughout the profile and the fixed ratios ensured a smooth, physically realistic change of hydraulic properties in general with depth. We do not have an indication or evidence that the fact that the saturated hydraulic conductivities of matrix and macropore domain converged to very similar values was a result of using the hyperparameter approach. The matrix and the macropore parameters were still calibrated independently from each other and could vary within the given limits of several orders of magnitude. Thus, we assume that the calibration of the hyperparameters did not cause the conductivities of the matrix and macropore domain to converge, and it is likely that a similar outcome would be obtained if the macropore and matrix parameters would not only be calibrated independently from each other but also variably over depth. If a different preferential flow approach would have yielded a similar result, remains an open question.

The results of our study corroborate the findings of Glaser et al. (2019). They showed that the transfer of dual-permeability parameters from plot to catchment scale simulations did not improve discharge simulations and reasoned that vertical preferential flow does not seem to be of major relevance for catchment scale runoff generation in the Weierbach catchment. The heterogeneity of soil hydraulic properties and lateral preferential flow implemented already by having soil layers with contrasting saturated hydraulic conductivities seemed to be sufficient to capture the general streamflow generation behaviour and reproduce discharge reasonably well in this catchment. We would expect that this would be the case at all catchments that are characterized by soil layers with strong contrasts in saturated hydraulic conductivities and where shallow subsurface storm flow dominates streamflow generation (e.g., Mirus, 2015). This applies to water flow simulations only and may be different when considering solute transport (cf. Christiansen et al., 2004).

## 4.3 | Suitability of catchment scale parameter sets to describe plot-scale observations

Glaser et al. (2019) demonstrated that dual-permeability parameter sets calibrated at plot scale did not result in a satisfactory simulation of catchment-scale hydrometric response. Here, we applied the reverse procedure and tested if the parameter sets calibrated at catchment scale were also appropriate to simulate the plot scale observations. The starting values of the hydraulic parameters for calibration ('default parameter set') had been assembled based on field evidence and previous modelling work such that a simulation of the plot-scale irrigation experiments reproduced the elevated bromide concentrations at depth (Figure 10). This default parameter set was obviously not able to simulate discharge satisfactorily (Figure 7), and the calibration clearly improved the simulation of the hydrometric response at catchment scale. However, the dual-permeability parameter sets ('Macro' and 'MacroMat' scenarios) performing best at catchment scale performed even worse than the default parameter set when simulating the bromide depth profiles at plot scale. This

suggests that calibrated soil hydraulic parameters are not scale-invariant, independent of the direction of the parameter transfer. Instead they can be seen as effective parameters at the scale of calibration. The scale-variance of soil hydraulic parameters has been shown before for the transfer from the smaller scale to the larger scale (e.g., Brooks et al., 2004; Grayson, Moore, & McMahon, 1992), but to our knowledge not tested in the reverse direction.

## 5 | CONCLUSIONS

We successfully used the optimization algorithm DREAM to calibrate a 3D physics-based dual-permeability model at the catchment scale. In order to keep computational costs of the optimization routine within a reasonable timeframe and, at the same time, to obtain physically realistic depth profiles of soil hydraulic parameters, we limited the number of parameters to be calibrated by doing a PCA prior to the calibration to identify the most influential soil hydraulic parameters and by fixing the parameter ratios between the 10 simulated soil layers, thus calibrating one hyperparameter instead of calibrating parameters for the 10 soil layers independently. The dual-permeability parameter sets identified during calibration were able to simulate observed discharge time series satisfactorily but did not outperform a single-domain reference model scenario. The effect of macropores was calibrated away, which suggests that the incorporation of vertical preferential flow as incorporated by the dual-permeability approach was not relevant for catchment scale runoff generation. The heterogeneity of soil hydraulic properties and lateral preferential flow already implemented by having soil layers with contrasting saturated hydraulic conductivities seemed to be sufficient to reproduce the hydrometric response reasonably well in the studied catchment. We also tested if the dual-permeability parameter sets performing best at catchment scale could be transferred to the plot-scale to describe bromide depth profiles obtained from tracer irrigation experiments. This parameter transfer proved to be not successful, suggesting that soil hydraulic parameters are not scale-independent in both directions. Future work could test the relative importance of vertical preferential flow versus laterally highly conductive subsurface layers also for solute transport.

## ACKNOWLEDGEMENTS

We thank Conrad Jackisch for providing his data of the irrigation experiments for the evaluation of the plot scale simulations. The Luxembourg Institute of Science and Technology (LIST) is gratefully acknowledged for supporting the stay of T.S. at the LIST in 2017. We would like to thank Steve Berg from Aquanty Inc. for his support with HGS. B.G. would like to thank the Luxembourg National Research Fund (FNR) for funding within the framework of the FNR-AFR Pathfinder project (Reference-ID 10189601). Three reviewers are thanked for their constructive comments and suggestions for improving the initial manuscript.

## CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Luisa Hopp  <https://orcid.org/0000-0002-1354-8193>

## REFERENCES

- Ala-aho, P., Soulsby, C., Wang, H., & Tetzlaff, D. (2017). Integrated surface-subsurface model to investigate the role of groundwater in headwater catchment runoff generation: A minimalist approach to parameterization. *Journal of Hydrology*, 547, 664–677. <https://doi.org/10.1016/j.jhydrol.2017.02.023>
- Anderson, A. E., Weiler, M., Alila, Y., & Hudson, R. O. (2009). Dye staining and excavation of a lateral preferential flow network. *Hydrology and Earth System Sciences*, 13, 935–944. <https://doi.org/10.5194/hess-13-935-2009>
- Arora, B., Mohanty, B. P., & McGuire, J. T. (2011). Inverse estimation of parameters for multidomain flow models in soil columns with different macropore densities. *Water Resources Research*, 47, W04512. <https://doi.org/10.1029/2010WR009451>
- Arora, B., Mohanty, B. P., & McGuire, J. T. (2012). Uncertainty in dual permeability model parameters for structured soils. *Water Resources Research*, 48, W01524. <https://doi.org/10.1029/2011WR010500>
- Bachmair, S., & Weiler, M. (2011). New dimensions of hillslope hydrology. In D. Levia, et al. (Eds.), *Forest hydrology and biogeochemistry*. Dordrecht: Springer.
- Beven, K., & Germann, P. (1982). Macropores and water flow in soils. *Water Resources Research*, 18, 1311–1325. <https://doi.org/10.1029/WR018i005p01311>
- Beven, K., & Germann, P. (2013). Macropores and water flow in soils revisited. *Water Resources Research*, 49, 3071–3092. <https://doi.org/10.1002/wrcr.20156>
- Binley, A., Hubbard, S. S., Huisman, J. A., Revil, A., Robinson, D. A., Singha, K., & Slater, L. D. (2015). The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales. *Water Resources Research*, 51, 3837–3866. <https://doi.org/10.1002/2015WR017016>
- Brace, W. F. (1980). Permeability of crystalline and argillaceous rocks. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, 17, 241–251. [https://doi.org/10.1016/0148-9062\(80\)90807-4](https://doi.org/10.1016/0148-9062(80)90807-4)
- Brooks, E. S., Boll, J., & McDaniel, P. A. (2004). A hillslope-scale experiment to measure lateral saturated hydraulic conductivity. *Water Resources Research*, 40, W04208. <https://doi.org/10.1029/2003WR002858>
- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 434–455. <https://doi.org/10.2307/1390675>
- Cadini, F., De Sanctis, J., Bertoli, I., & Zio, E. (2013). Upscaling of a dual-permeability Monte Carlo simulation model for contaminant transport in fractured networks by genetic algorithm parameter identification. *Stochastic Environmental Research and Risk Assessment*, 27, 505–516. <https://doi.org/10.1007/s00477-012-0595-8>
- Christiansen, J. S., Thorsen, M., Clausen, T., Hansen, S., & Refsgaard, J. C. (2004). Modelling of macropore flow and transport processes at catchment scale. *Journal of Hydrology*, 299, 136–158. <https://doi.org/10.1016/j.jhydrol.2004.04.029>
- Coppola, A., Kutilek, M., & Frind, E. O. (2009). Transport in preferential flow domains of the soil porous system: Measurement, interpretation, modelling, and upscaling. *Journal of Contaminant Hydrology*, 104, 1–3. <https://doi.org/10.1016/j.jconhyd.2008.05.011>
- Di Pietro, L., Ruy, S., & Capowiez, Y. (2003). Predicting preferential water flow in soils by traveling-dispersive waves. *Journal of Hydrology*, 278, 64–75. [https://doi.org/10.1016/S0022-1694\(03\)00124-0](https://doi.org/10.1016/S0022-1694(03)00124-0)

- Doherty, J. (2009). *PEST: Model independent parameter estimation, watermark numerical computing*. Corinda, Queensland, Australia. Retrieved from <http://www.pesthomepage.org>
- Durner, W. (1994). Hydraulic conductivity estimation for soils with heterogeneous pore structure. *Water Resources Research*, 30, 211–223. <https://doi.org/10.1029/93WR02676>
- Ebel, B. A., Loague, K., Vanderkwaak, J. E., Dietrich, W. E., Montgomery, D. R., Torres, R., & Anderson, S. P. (2007). Near-surface hydrologic response for a steep, unchanneled catchment near Coos Bay, Oregon: 2. Physics-based simulations. *American Journal of Science*, 307, 709–748. <https://doi.org/10.2475/04.2007.03>
- Flury, M. (1996). Experimental evidence of transport of pesticides through field soils—A review. *Journal of Environmental Quality*, 25, 25–45. <https://doi.org/10.2134/jeq1996.00472425002500010005x>
- Gächter, R., Ngatiah, J. M., & Stamm, C. (1998). Transport of phosphate from soil to surface waters by preferential flow. *Environmental Science & Technology*, 32, 1865–1869. <https://doi.org/10.1021/es9707825>
- Gallagher, M., & Doherty, J. (2007). Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modelling & Software*, 22, 1000–1020. <https://doi.org/10.1016/j.envsoft.2006.06.007>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–472. <https://doi.org/10.1214/ss/1177011136>
- Gerke, H. H., & van Genuchten, M. T. (1993). A dual-porosity model for simulating the preferential movement of water and solutes in structured porous media. *Water Resources Research*, 29, 305–319. <https://doi.org/10.1029/92WR02339>
- Germann, P., & Di Pietro, L. (1999). Scales and dimensions of momentum dissipation during preferential flow in soils. *Water Resources Research*, 35, 1443–1454. <https://doi.org/10.1029/1998WR900112>
- Germann, P. F., Smith, M. S., & Thomas, G. W. (1987). Kinematic wave approximation to the transport of *Escherichia coli* in the vadose zone. *Water Resources Research*, 23, 1281–1287. <https://doi.org/10.1029/WR023i007p01281>
- Glaser, B., Jackisch, C., Hopp, L., & Klaus, J. (2019). How meaningful are plot scale observations and simulations of preferential flow for catchment models? *Vadose Zone Journal*, 18, 180146. <https://doi.org/10.2136/vzj2018.08.0146>
- Glaser, B., Klaus, J., Frei, S., Frenress, J., Pfister, L., & Hopp, L. (2016). On the value of surface saturated area dynamics mapped with thermal infrared imagery for modeling the hillslope-riparian-stream continuum. *Water Resources Research*, 52, 8317–8342. <https://doi.org/10.1002/2015WR018414>
- Gormally, K. H., McIntosh, M. S., & Mucciardi, A. N. (2011). Ground-penetrating radar detection and three-dimensional mapping of lateral macropores: I. calibration. *Soil Science Society of America Journal*, 75, 1226–1235. <https://doi.org/10.2136/sssaj2010.0339>
- Gormally, K. H., McIntosh, M. S., Mucciardi, A. N., & McCarty, G. W. (2011). Ground-penetrating radar detection and three-dimensional mapping of lateral macropores: II. Riparian application. *Soil Science Society of America Journal*, 75, 1236–1243. <https://doi.org/10.2136/sssaj2010.0342>
- Graham, C. B., Woods, R. A., & McDonnell, J. J. (2010). Hillslope threshold response to rainfall: (1) A field based forensic approach. *Journal of Hydrology*, 393, 65–76. <https://doi.org/10.1016/j.jhydrol.2009.12.015>
- Grayson, R. B., Moore, I. D., & McMahon, T. A. (1992). Physically based hydrologic modeling: 1. A terrain-based model for investigative purposes. *Water Resources Research*, 28, 2639–2658. <https://doi.org/10.1029/92WR01258>
- Guillaume, J. H., & Andrews, F. (2012). *Dream: Differential evolution adaptive metropolis*. R Package Version 0.4-2. Retrieved from <http://dream.r-forge.r-project.org/>
- Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T., & Huth, A. (2011). Statistical inference for stochastic simulation models—Theory and application. *Ecology Letters*, 14, 816–827. <https://doi.org/10.1111/j.1461-0248.2011.01640.x>
- Hwang, H.-T., Park, Y.-J., Sudicky, E. A., Berg, S. J., McLaughlin, R., & Jones, J. P. (2018). Understanding the water balance paradox in the Athabasca River basin, Canada. *Hydrological Processes*, 32, 729–746. <https://doi.org/10.1002/hyp.11449>
- Jackisch, C., Angermann, L., Allroggen, N., Sprenger, M., Blume, T., Tronicke, J., & Zehe, E. (2017). Form and function in hillslope hydrology: In situ imaging and characterization of flow-relevant structures. *Hydrology and Earth System Sciences*, 21, 3749–3775. <https://doi.org/10.5194/hess-21-3749-2017>
- Jarvis, N., & Larsbo, M. (2012). MACRO (v5.2): Model use, calibration, and validation. *Transactions of the ASABE*, 55, 1413–1423. <https://doi.org/10.13031/2013.42251>
- Jarvis, N. J. (2007). A review of non-equilibrium water flow and solute transport in soil macropores: Principles, controlling factors and consequences for water quality. *European Journal of Soil Science*, 58, 523–546. <https://doi.org/10.1111/j.1365-2389.2007.00915.x>
- Jones, J. A. A. (2010). Soil piping and catchment response. *Hydrological Processes*, 24, 1548–1566. <https://doi.org/10.1002/hyp.7634>
- Joseph, J. F., & Guillaume, J. H. A. (2013). Using a parallelized MCMC algorithm in R to identify appropriate likelihood functions for SWAT. *Environmental Modelling & Software*, 46, 292–298. <https://doi.org/10.1016/j.envsoft.2013.03.012>
- Juilleret, J., Iffly, J. F., Pfister, L., & Hissler, C. (2011). Remarkable Pleistocene periglacial slope deposits in Luxembourg (Oesling): Pedological implication and geosite potential. *Bulletin de la Société Des Naturalistes Luxembourgeois*, 112, 125–130. Retrieved from [http://www.snl.lu/publications/bulletin/SNL\\_2011\\_112\\_125\\_130.pdf](http://www.snl.lu/publications/bulletin/SNL_2011_112_125_130.pdf)
- Kavetski, D., Fenicia, F., & Clark, M. P. (2011). Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment. *Water Resources Research*, 47, W05501. <https://doi.org/10.1029/2010WR009525>
- Keating, E. H., Doherty, J., Vrugt, J. A., & Kang, Q. (2010). Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality. *Water Resources Research*, 46, W10517. <https://doi.org/10.1029/2009WR008584>
- Klaus, J., & Jackson, C. R. (2018). Interflow is not binary: A continuous shallow perched layer does not imply continuous connectivity. *Water Resources Research*, 54, 5921–5932. <https://doi.org/10.1029/2018WR022920>
- Klaus, J., & Zehe, E. (2010). Modelling rapid flow response of a tile drained field site using a 2D-physically based model: Assessment of “equifinal” model setups. *Hydrological Processes*, 24, 1595–1609. <https://doi.org/10.1002/hyp.7687>
- Klaus, J., & Zehe, E. (2011). A novel explicit approach to model bromide and pesticide transport in connected soil structures. *Hydrology and Earth System Sciences*, 15, 2127–2144. <https://doi.org/10.5194/hess-15-2127-2011>
- Köhne, J. M., Köhne, S., & Šimunek, J. (2009). A review of model applications for structured soils: (a) Water flow and tracer transport. *Journal of Contaminant Hydrology*, 104, 4–35. <https://doi.org/10.1016/j.jconhyd.2008.10.002>
- Köhne, J. M., & Mohanty, B. P. (2005). Water flow processes in a soil column with a cylindrical macropore: Experiment and hierarchical modeling. *Water Resources Research*, 41, W03010. <https://doi.org/10.1029/2004WR003303>
- Kristensen, K. J., & Jensen, S. E. (1975). A model for estimating actual evapotranspiration from potential evapotranspiration. *Hydrology Research*, 6, 170–188. <https://doi.org/10.2166/nh.1975.0012>
- Kung, K.-J. S., Steenhuis, T. S., Klavivko, E. J., Gish, T. J., Bubenzer, G., & Helling, C. S. (2000). Impact of preferential flow on the transport of adsorbing and non-adsorbing tracers. *Soil Science Society of America Journal*, 64, 1290–1296. <https://doi.org/10.2136/sssaj2000.6441290x>

- Laine-Kaulio, H., Backnäs, S., Karvonen, T., Koivusalo, H., & McDonnell, J. J. (2014). Lateral subsurface stormflow and solute transport in a forested hillslope: A combined measurement and modeling approach. *Water Resources Research*, 50, 8159–8178. <https://doi.org/10.1002/2014WR015381>
- Laine-Kaulio, H., Backnäs, S., Koivusalo, H., & Lauren, A. (2015). Dye tracer visualization of flow patterns and pathways in glacial sandy till at a boreal forest hillslope. *Geoderma*, 259–260, 23–34. <https://doi.org/10.1016/j.geoderma.2015.05.004>
- Laine-Kaulio, H., & Koivusalo, H. (2018). Model-based exploration of hydrological connectivity and solute transport in a forested hillslope. *Land Degradation & Development*, 29, 1176–1189. <https://doi.org/10.1002/ldr.2823>
- Larsbo, M., Roulier, S., Stenemo, F., Kasteel, R., & Jarvis, N. (2005). An improved dual-permeability model of water flow and solute transport in the Vadose zone. *Vadose Zone Journal*, 4, 398–406. <https://doi.org/10.2136/vzj2004.0137>
- Lu, D., Ye, M., & Hill, M. C. (2012). Analysis of regression confidence intervals and Bayesian credible intervals for uncertainty quantification. *Water Resources Research*, 48, W09521. <https://doi.org/10.1029/2011WR011289>
- Martínez-Carreras, N., Hissler, C., Gourdol, L., Klaus, J., Juilleret, J., Iffly, J. F., & Pfister, L. (2016). Storage controls on the generation of double peak hydrographs in a forested headwater catchment. *Journal of Hydrology*, 543, 255–269. <https://doi.org/10.1016/j.jhydrol.2016.10.004>
- Mirus, B. B. (2015). Evaluating the importance of characterizing soil structure and horizons in parameterizing a hydrologic process model. *Hydrological Processes*, 29, 4611–4623. <https://doi.org/10.1002/hyp.10592>
- Nasta, P., Boaga, J., Deiana, R., Cassiani, G., & Romano, N. (2019). Comparing ERT- and scaling-based approaches to parameterize soil hydraulic properties for spatially distributed model applications. *Advances in Water Resources*, 126, 155–167. <https://doi.org/10.1016/j.advwatres.2019.02.014>
- Nimmo, J. R. (2010). Theory for source-responsive and free-surface film modeling of unsaturated flow. *Vadose Zone Journal*, 9, 295–306. <https://doi.org/10.2136/vzj2009.0085>
- Nimmo, J. R. (2012). Preferential flow occurs in unsaturated conditions. *Hydrological Processes*, 26, 786–789. <https://doi.org/10.1002/hyp.8380>
- Nyquist, J. E., Toran, L., Pitman, L., Guo, L., & Lin, H. (2018). Testing the fill-and-spill model of subsurface lateral flow using ground-penetrating radar and dye tracing. *Vadose Zone Journal*, 17(17), 170142. <https://doi.org/10.2136/vzj2017.07.0142>
- Othmer, H. B., Diekkrüger, B., & Kutilek, M. (1991). Bimodalporosity and unsaturated hydraulic conductivity. *Soil Science*, 152, 139–150. <https://doi.org/10.1097/00010694-199109000-00001>
- Pechlivanidis, I. G., Jackson, B. M., McIntyre, N. R., & Wheeler, H. S. (2011). Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST Journal*, 13, 193–214. <https://doi.org/10.30955/gnj.000778>
- Peters, A., & Durner, W. (2008). A simple model for describing hydraulic conductivity in unsaturated porous media accounting for film and capillary flow. *Water Resources Research*, 44, W11417. <https://doi.org/10.1029/2008WR007136>
- Pfister, L., Kwadijk, J., Musy, A., Bronstert, A., & Hoffmann, L. (2004). Climate change, land use change and runoff prediction in the Rhine-Meuse basins. *River Research and Applications*, 20, 229–241. <https://doi.org/10.1002/rra.775>
- Pfister, L., McDonnell, J. J., Hissler, C., & Hoffmann, L. (2010). Ground-based thermal imagery as a simple, practical tool for mapping saturated area connectivity and dynamics. *Hydrological Processes*, 24, 3123–3132. <https://doi.org/10.1002/hyp.7840>
- Roulier, S., Baran, N., Mouvet, C., Stenemo, F., Morvan, X., Albrechtsen, H.-J., ... Jarvis, N. (2006). Controls on atrazine leaching through a soil-unsaturated fractured limestone sequence at Brévilles, France. *Journal of Contaminant Hydrology*, 84, 81–105. <https://doi.org/10.1016/j.jconhyd.2005.12.004>
- Ross, P. J., & Smettem, K. R. J. (2000). A Simple Treatment of Physical-Nonequilibrium Water Flow in Soils. *Soil Science Society of America Journal*, 64, 1926–1930. <https://doi.org/10.2136/sssaj2000.6461926x>
- Rovey, C. W., & Cherkauer, D. S. (1995). Scale dependency of hydraulic conductivity measurements. *Groundwater*, 33, 769–780. <https://doi.org/10.1111/j.1745-6584.1995.tb00023.x>
- Schaik, N. L. M. B., Bronstert, A., Jong, S. M., Jetten, V. G., Dam, J. C., Ritsema, C. J., & Schnabel, S. (2014). Process-based modelling of a headwater catchment in a semi-arid area: The influence of macropore flow. *Hydrological Processes*, 28, 5805–5816. <https://doi.org/10.1002/hyp.10086>
- Schulze-Makuch, D., Carlson, D. A., Cherkauer, D. S., & Malik, P. (1999). Scale dependency of hydraulic conductivity in heterogeneous media. *Groundwater*, 37, 904–919. <https://doi.org/10.1111/j.1745-6584.1999.tb01190.x>
- Schwab, M., Klaus, J., Pfister, L., & Weiler, M. (2016). Diel discharge cycles explained through viscosity fluctuations in riparian inflow. *Water Resources Research*, 52, 8744–8755. <https://doi.org/10.1002/2016WR018626>
- Shao, W., Bogaard, T., & Bakker, M. (2014). How to use COMSOL multi-physics for coupled dual-permeability hydrological and slope stability modeling. *Procedia Earth and Planetary Science*, 9, 83–90. <https://doi.org/10.1016/j.proeps.2014.06.018>
- Shao, W., Bogaard, T. A., Bakker, M., & Greco, R. (2015). Quantification of the influence of preferential flow on slope stability using a numerical modelling approach. *Hydrology and Earth System Sciences*, 19, 2197–2212. <https://doi.org/10.5194/hess-19-2197-2015>
- Shi, X., Ye, M., Finsterle, S., & Wu, J. (2012). Comparing nonlinear regression and Markov chain Monte Carlo methods for assessment of prediction uncertainty in vadose zone modeling. *Vadose Zone Journal*, 11, vzj2011.0147. <https://doi.org/10.2136/vzj2011.0147>
- Šimůnek, J., Jarvis, N. J., van Genuchten, M. T., & Gärdenäs, A. (2003). Review and comparison of models for describing non-equilibrium and preferential flow and transport in the vadose zone. *Journal of Hydrology*, 272, 14–35. [https://doi.org/10.1016/S0022-1694\(02\)00252-4](https://doi.org/10.1016/S0022-1694(02)00252-4)
- Stamm, C., Flüher, H., Gächter, R., Leuenberger, J., & Wunderli, H. (1998). Preferential transport of phosphorus in drained grassland soils. *Journal of Environmental Quality*, 27, 515–522. <https://doi.org/10.2134/jeq1998.00472425002700030006x>
- Therrien, R., McLaren, R. G., Sudicky, E. A., & Panday, S. M. (2010). *HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport*. Waterloo, Canada: Groundwater Simulations Group, University of Waterloo. Retrieved from <https://www.ggl.ulaval.ca/fileadmin/ggl/documents/rtherrien/hydrogeosphere.pdf>
- Tromp-van Meerveld, H. J., & McDonnell, J. J. (2006). Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research*, 42, W02411. <https://doi.org/10.1029/2004WR003800>
- Vereecken, H., Kasteel, R., Vanderborght, J., & Harter, T. (2007). Upscaling hydraulic properties and soil water flow processes in heterogeneous soils. *Vadose Zone Journal*, 6, 1–28. <https://doi.org/10.2136/vzj2006.0055>
- von Gunten, D., Wohling, T., Haslauer, C., Merchan, D., Causape, J., & Cirpka, O. A. (2014). Efficient calibration of a distributed pde-based hydrological model using grid coarsening. *Journal of Hydrology*, 519, 3290–3304. <https://doi.org/10.1016/j.jhydrol.2014.10.025>
- Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, 75, 273–316. <https://doi.org/10.1016/j.envsoft.2015.08.013>

- Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V., & Robinson, B. A. (2009). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, 23, 1011–1026. <https://doi.org/10.1007/s00477-008-0274-y>
- Wang, Y., Bradford, S. A., & Šimůnek, J. (2014). Estimation and upscaling of dual-permeability model parameters for the transport of *E. coli* D21g in soils with preferential flow. *Journal of Contaminant Hydrology*, 159, 57–66. <https://doi.org/10.1016/j.jconhyd.2014.01.009>
- Weiler, M., & McDonnell, J. J. (2007). Conceptualizing lateral preferential flow and flow networks and simulating the effects on gauged and ungauged hillslopes. *Water Resources Research*, 43, W03403. <https://doi.org/10.1029/2006WR004867>
- Weiler, M., & Naef, F. (2003). An experimental tracer study of the role of macropores in infiltration in grassland soils. *Hydrological Processes*, 17, 477–493. <https://doi.org/10.1002/hyp.1136>
- Wildemeersch, S., Goderniaux, P., Orban, P., Brouyère, S., & Dassargues, A. (2014). Assessing the effects of spatial discretization on large-scale flow model performance and prediction uncertainty. *Journal of Hydrology*, 51, 10–25. <https://doi.org/10.1016/j.jhydrol.2013.12.020>
- Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., ... Pfister, L. (2015). Towards more systematic perceptual model development: A case study using 3 Luxembourgish catchments. *Hydrological Processes*, 29, 2731–2750. <https://doi.org/10.1002/hyp.10393>
- Yu, X., Duffy, C., Baldwin, D. C., & Lin, H. (2014). The role of macropores and multi-resolution soil survey datasets for distributed surface–subsurface flow modeling. *Journal of Hydrology*, 516, 97–106. <https://doi.org/10.1016/j.jhydrol.2014.02.055>
- Zehe, E., Becker, R., Bárdossy, A., & Plate, E. (2005). Uncertainty of simulated catchment runoff response in the presence of threshold processes: Role of initial soil moisture and precipitation. *Journal of Hydrology*, 315, 183–202. <https://doi.org/10.1016/j.jhydrol.2005.03.038>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Hopp L, Glaser B, Klaus J, Schramm T.

The relevance of preferential flow in catchment scale simulations: Calibrating a 3D dual-permeability model using DREAM. *Hydrological Processes*. 2020;34:1237–1254.

<https://doi.org/10.1002/hyp.13672>