On synergies between spatially-distributed, physically-based simulations and field observations in catchment hydrology

DISSERTATION

zur Erlangung des akademischen Grades einer
Doktorin der Naturwissenschaften
(Dr. rer. nat.)

an der Fakultät für
Biologie, Chemie und Geowissenschaften
der Universität Bayreuth

vorgelegt von

Barbara Glaser
geboren am 7. Juli 1989 in Starnberg

Bayreuth, September 2019

Vollständiger Abdruck der von der Fakultät für Biologie, Chemie und Geowissenschaften der Universität Bayreuth genehmigten Dissertation zur Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften (Dr. rer. nat.).

Dissertation eingereicht am: 18.09.2019

Zulassung durch die Promotionskommission: 30.09.2019

Wissenschaftliches Kolloquium: 30.04.2020

Amtierender Dekan: Prof. Dr. Matthias Breuning

Prüfungsausschuss:

Prof. Dr. Stefan Peiffer (Gutachter)

Dr. Julian Klaus (Gutachter)

Prof. Dr. Jan Fleckenstein (Vorsitz)

Prof. Dr. Oliver Sass

Acknowledgements

Contents

Graphical abstract

Study 1: How meaningful are plot-scale observations and simulations of preferential flow for catchment models?

Preferential flow paths
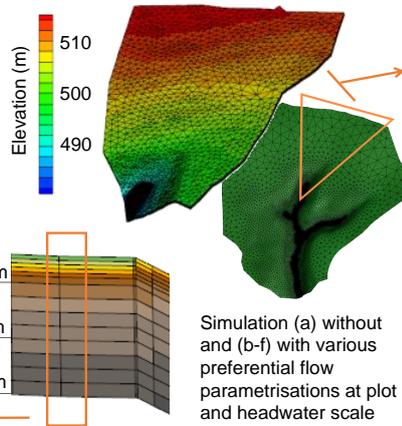
Precipitation (mm/d)
Simulated discharge (l/s)
Measured discharge (l/s)

Simulation (a) without and (b-f) with various preferential flow parametrisations at plot and headwater scale

Field irrigation plot 1-3

Br⁻ (kg/m³)

Generation and occurrence of surface saturation

Stream section of 6 m

TIR panorama

Observation

Normalised frequency of surface saturation
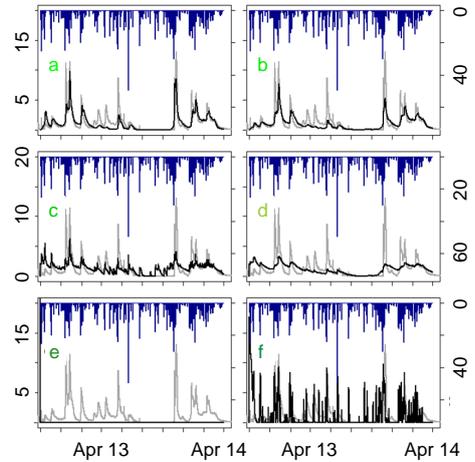
Simulation
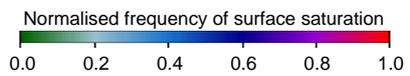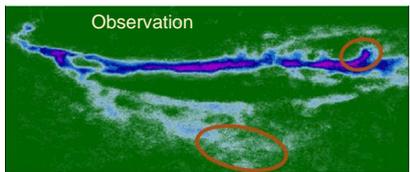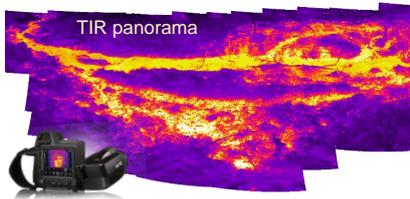
Discrete location of subsurface water exfiltration
★ Streambed
★ Riparian zone
★ Classification unclear

Study 4: Sources of surface water in space and time
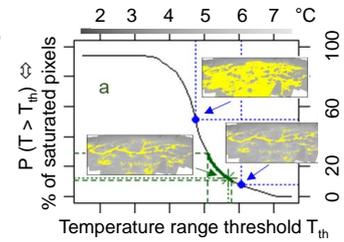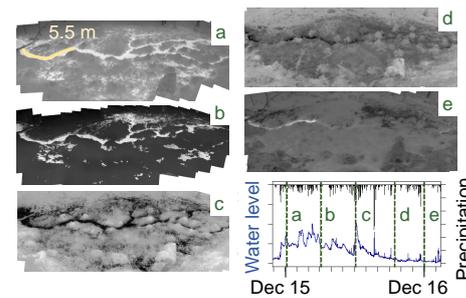
Water level

Absolute saturation range:
Obs: 0.3 - 22.4 %
Sim: 0.0 - 59.7 %

Normalised extent of surface saturation

Study 3: Intra-catchment variability of surface saturation – insights from long-term observations and simulations

Study 2: Mapping surface-saturation dynamics with thermal infrared imagery

Normalised extent of surface saturation

Catchment discharge (l/s)

Immediate delivery paths of surface water

Geographical sources of surface water

Precipitation
Discharge
Volume weighted fractions over 4 days

1st time step
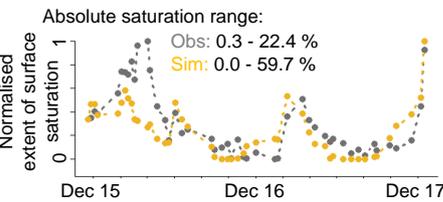
Fractions of initial water sources
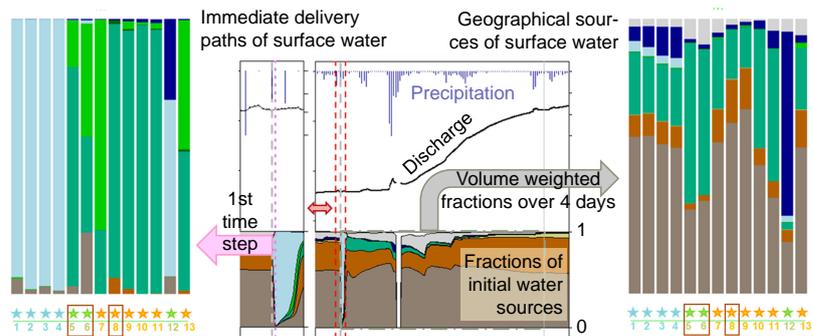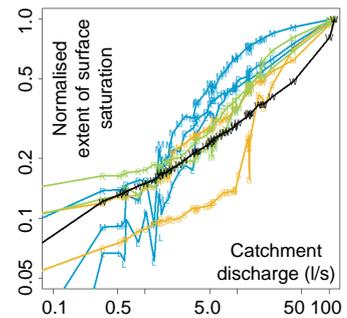
Zusammenfassung

Ein wesentlicher Schwerpunkt im Forschungsbereich der Einzugsgebietshydrologie liegt darauf, die hydromechanischen Prozesse zu untersuchen und zu verstehen, die in natürlichen Wassereinzugsgebieten auftreten. Um Einblicke in die hydrologische Funktionsweise von Hängen und Einzugsgebieten zu erhalten, steht eine Vielzahl verschiedener Observierungs- und Feldmethoden zur Verfügung. Allerdings ist die Anwendbarkeit aller existierenden Messmethoden räumlich oder zeitlich eingeschränkt. Infolgedessen ist auch das Wissen darüber begrenzt, ob und wie sich das Auftreten und die Relevanz verschiedener hydrologischer Prozesse in Zeit und Raum verändern. Eine Möglichkeit, die methodischen Grenzen von Feldbeobachtungen zu überwinden, liegt in der ergänzenden Verwendung hydrologischer Modelle. Jedoch werden hydrologische Modelle nur selten in echter Synergie mit Feldbeobachtungen genutzt und in den existierenden Studien werden zumeist konzeptionelle Speichermodelle verwendet, welche sich nicht besonders gut dafür eignen, die raumzeitliche Variabilität und Relevanz verschiedener Prozesse zu untersuchen.

Die vorliegende Dissertation präsentiert zwei Beispiele für die Anwendung räumlich differenzierter, physikalisch basierter hydrologischer Modelle in Synergie mit verschiedenen herkömmlichen und innovativen Felddaten. Der erste Teil der Arbeit (Studie 1) richtet das Augenmerk auf die Relevanz des häufig in kleinräumigen Feldexperimenten beobachteten präferentiellen Flusses für das Verhalten von Abfluss und Bodenfeuchte auf Einzugsgebietsebene. Die Studie bewertet die Bedeutsamkeit der kleinräumigen Feldbeobachtungen mit Hilfe der Verwendung eines räumlich differenzierten, physikalisch basierten Modells, indem die Ergebnisse von Simulationen mit und ohne Parametrisierungen des präferentiellen Flusses auf kleiner räumlicher Skala und auf Einzugsgebietsebene mit entsprechenden Felddaten verglichen werden. Der zweite Teil der Arbeit (Studie 2 - 4) behandelt die dynamische Entstehung von Oberflächensättigung und deren räumliche Variabilität innerhalb eines Einzugsgebietes. Er besteht aus drei aufeinanderfolgenden Beobachtungs- und Modellierungsstudien, die auf der Idee beruhen, zunächst mit Hilfe eines umfangreichen Datensatzes aus Feldbeobachtungen die internen Strukturen und Prozesse eines Modells auf ihre Widerspruchsfreiheit zur Realität zu überprüfen, bevor das Model dazu verwendet wird, die Interpretation von Felddaten zu unterstützen und deren Limitierungen zu überwinden. Die erste der drei Studien (Studie 2) bewertet die Praxistauglichkeit der Verwendung von Thermalinfrarot-Fotografie während verschiedener Jahreszeiten und hydrologischer Bedingungen sowie an verschiedenen Orten in einem Einzugsgebiet, um auftretende Oberflächensättigung zu kartieren und quantifizieren. Die anschließende Studie (Studie 3) wendet ein räumlich differenziertes, physikalisch basiertes Modell in Kombination mit dem in Studie 2 erhobenen Felddatensatz an. Zum einen bewertet die Studie das Vermögen des Modells, die Variabilität der räumlichen Muster und der zeitlichen Dynamik der Oberflächensättigung innerhalb

eines Einzugsgebietes zu reproduzieren. Zum anderen werden aus den identifizierten Konformitäten und Diskrepanzen zwischen den Beobachtungsdaten und Modellergebnissen Rückschlüsse darauf gezogen, welche Schlüsselfaktoren das raumzeitliche Auftreten der Oberflächensättigung beeinflussen. Die letzte Studie (Studie 4) verwendet das umfassend validierte Modell aus Studie 3, um die Entstehung der Oberflächensättigung in Raum und Zeit zu untersuchen, sowohl im Hinblick auf die unmittelbare Art der Wasserzulieferung als auch im Hinblick auf verschiedene geographische Ursprungsgebiete.

Alle Studien wurden für das 42 Hektar große Einzugsgebiet des Weierbachs durchgeführt, einem fest etabliertem Forschungsgebiet im Westen Luxemburgs mit komplexem hydrologischem Verhalten. Die Simulationen wurden mit dem kombinierten Oberflächen-Untergrund Modell HydroGeoSphere ausgeführt. Studie 1 offenbarte, dass kleinräumige Beobachtungen des präferentiellen Flusses keine aufschlussreichen Informationen über den Einfluss präferentieller Fließwege auf Einzugsgebietsebene liefern. Obwohl vertikale präferentielle Fließwege berücksichtigt werden mussten, um die kleinräumigen Feldbeobachtungen zu simulieren, war es nicht zweckmäßig, die entsprechenden Parametrisierungen auf Einzugsgebietsebene anzuwenden. Stattdessen hat sich gezeigt, dass auf Einzugsgebietsebene der Einfluss schneller lateraler Fließwege gegenüber den vertikalen präferentiellen Fließpfaden überwiegt. Studie 2 zeigte, dass Thermalinfrarot-Fotografie eine verlässliche Methode dafür ist, einen umfangreichen Datensatz über Oberflächensättigungsmuster und -dynamiken zu erheben, wenn einige Vorkehrungen während der Bildaufnahme und der nachfolgenden Bearbeitung berücksichtigt werden. Studie 3 zeigte auf, dass das Modell die beobachteten Muster und Dynamiken der Oberflächensättigung weitestgehend wiedergeben kann, einschließlich lokal variierender Muster im Vorkommen der Oberflächensättigung und verschiedener Verhältnisse zwischen der lokalen Sättigungsausdehnung und dem Abfluss aus dem Einzugsgebiet. Das Modellverhalten legte nahe, dass das Auftreten von Oberflächensättigung hauptsächlich durch Grundwasseraustritt in mikrotopographische Vertiefungen bestimmt wurde, dass es aber zusätzliche Einflussfaktoren und Prozesse geben muss, wie etwa lokale Ungleichheiten im Untergrund, wechselnde Einflüsse von Niederschlag oder Oberstromwasser und hysteretische Entstehungsprozesse. Studie 4 belegte, dass sich die Oberflächensättigung im Einzugsgebiet des Weierbachs überwiegend aus Wasser zusammensetzt, das durch Rückfluss aus dem Untergrund in die Uferzone und das Bachbett geliefert wird. Die raumzeitliche Variabilität der Durchmischung von Wasser aus verschiedenen Untergrundspeichern und des Beitrags von Wasser anderer Herkunft, wie etwa Oberflächenabfluss oder Niederschlag, erwies sich als kleiner als erwartet. Insgesamt zeigten die Ergebnisse der verschiedenen Studien das große Potenzial von Synergien zwischen räumlich differenzierter, physikalisch basierter Modellierung und Feldbeobachtungen, um das momentane Prozessverständnis in der Einzugsgebietshydrologie erheblich voranzubringen, sowohl für spezielle Untersuchungsgebiete als auch für allgemeine Fragestellungen und konzeptionelle Vorstellungen.

Abstract

A fundamental research objective in catchment hydrology is to investigate and understand the hydromechanical processes occurring within natural watersheds. A wide variety of different observation and field methods exists for providing insights into the hydrological functioning of hillslopes and catchments. However, all existing methods are limited in their spatial or temporal application. Consequently, there are limitations in the understanding on if and how the occurrence and relevance of different hydrological processes vary in space and time. A possibility to overcome the methodological limitations of field measurements is the complementary use of hydrological models. Yet hydrological models are only rarely used in real synergy with field observations and the existing studies mostly apply conceptual, bucket-type models, which are not particularly suitable for investigating the spatio-temporal variability and relevance of different processes.

This dissertation presents two examples of applying a spatially-distributed, physically-based hydrological model in synergy with various conventional and innovative field data. The first part of the thesis (Study 1) focuses on the relevance of preferential flow as commonly observed in plot-scale field experiments for explaining the long-term response of catchment discharge and soil moisture. The study assesses the importance of the plot-scale field observations by applying a spatially-distributed, physically-based model and comparing the results of simulations with and without preferential flow parametrisations at plot and catchment scale against appropriate field data. The second part of the thesis (Study 2 - 4) addresses the intra-catchment variability of the dynamic development of surface saturation. It consists of three consecutive observation and simulation studies, following the idea to first employ a comprehensive data set of field observations to verify the consistency of the internal structures and processes of a model with reality before using the model to support the interpretation and overcome the limitations of field investigations. The first of the three studies (Study 2) assesses the practicability of applying thermal infrared imagery during different seasons and hydrological conditions and at various locations across a catchment to map and quantify surface saturation. The following study (Study 3) applies a spatially-distributed, physically-based catchment model in combination with the field data set collected in Study 2. On the one hand, it evaluates the capability of the model to reproduce the intra-catchment variability of the spatial patterns and temporal dynamics of surface saturation. On the other hand, the identified matches and mismatches between observation data and simulation results are used to infer which key factors control the spatio-temporal occurrence of surface saturation. The last study (Study 4) uses the comprehensively validated model of Study 3 to analyse the generation of surface saturation in space and time with regards to the immediate mechanism of water delivery and to different geographical source areas.

All studies were carried out for the 42 ha forested Weierbach catchment, a well-established research catchment in western Luxembourg with complex hydrological behaviour. The simulations were performed with the integrated surface-subsurface model HydroGeoSphere. Study 1 revealed that small-scale observations of preferential flow are not informative to infer the influence of preferential flow paths at catchment scale. Although vertical preferential flow paths needed to be considered to simulate the plot-scale field observations, it was not suitable to apply the appropriate parametrisations at catchment scale. Instead, fast lateral flow paths showed to outweigh the role of vertical preferential flow at the catchment scale. Study 2 showed that thermal infrared imagery is a reliable method for collecting a comprehensive data set of surface saturation patterns and dynamics when considering some precautions for the image acquisition and post-processing. Study 3 demonstrated that the model can mostly reproduce the observed surface saturation patterns and dynamics, including locally varying patterns of surface saturation occurrence and varying relations of the local saturation extent to catchment discharge. The model performance suggested that the occurrence of surface saturation was mainly conditioned by exfiltration of groundwater into micro-topographic depressions, but that there must be additional influencing factors and processes, such as local heterogeneities in the subsurface, varying influences of precipitation or upstream water, and hysteretic formation processes. Study 4 proved that the surface saturation in the Weierbach catchment is largely composed of water that is delivered to the riparian zone and streambed by return flow from the subsurface. The spatio-temporal variability of the mixing of water from different subsurface stores and contributions from other water sources such as overland flow or precipitation showed to be smaller than expected. Overall, the results of the different studies demonstrated the great potential of synergies between spatially-distributed, physically-based modelling and field observations to substantially advance current process understanding in catchment hydrology, both for specific study sites and for general questions and conceptual perceptions.

1 Introduction

1.1 Process research in catchment hydrology

Catchment hydrology developed as a distinct field of research in the middle of the last century. Ever since, one of the fundamental objectives in catchment hydrology is to increase the knowledge about runoff generation processes. Questions commonly asked in this context are (e.g. Hewlett and Hibbert, 1967; McDonnell, 2003): How long does precipitation reside in the catchment before reaching the stream? What are the dominant flow paths of water through the catchment? What geographic areas and water stores connect and contribute to streamflow? Finding answers to these questions is not only motivated by pure scientific interest. It is also highly relevant to resources management (cf. Beven, 2006) for enhancing current management practices (e.g. flood and drought risk assessment, water quality management, erosion controls) and for being able to predict the influence of future climate conditions and land use changes on a water system.

Over the last decades, a vast number of experimental studies has taken efforts to characterize the spatial and temporal water sources, water stores, water flow paths, and transit and residence times in catchments with different climatic, morphologic, and pedo-geologic characteristics. Traditionally, experimental studies in catchment hydrology rely on hydrometric measurements (e.g. water table, soil moisture, discharge) and the use of tracers (e.g. dyes, geochemical elements, isotopes) to investigate runoff generation on individual hillslopes or in headwater catchments. In addition, there has been a continuous development of existing and new measurement techniques over the past decades, enhancing the monitoring capabilities in space and time and enabling deepened insights in the hydrologic functioning of hillslopes and catchments. For example, the application of geophysical methods (cf. Binley et al., 2015), the use of high-frequency in-situ measurement sensors (cf. Rode et al., 2016), and the use of remote sensing data (cf. McCabe et al., 2017) greatly improved the knowledge about potential spatial and temporal variabilities of hydrologic states and fluxes in the subsurface and on the surface. Moreover, it is nowadays more and more common to investigate multiple aspects in parallel, which allows a more holistic view on the processes within a catchment (cf. e.g. Hrachowitz et al., 2013).

In parallel to the continuous development of measurement techniques and applications, the focus of research in catchment hydrology changed over the years as well. Many of the early studies focussed on identifying and understanding different runoff generation mechanisms and the principal spatial and temporal origin of discharge water (e.g. Betson, 1964; Dunne, 1983; Dunne and Black, 1970; Hewlett and Hibbert, 1967; Mosley, 1979; Sklash and Farvolden, 1979). There is still ongoing research on

these aspects today, especially since it remains unclear and under debate which mechanisms can explain the often observed fast activation and release of stored pre-event water during storm events (e.g. Barthold and Woods, 2015; Katsura et al., 2014; Kirchner, 2003). At the same time, however, large emphasis is nowadays put on the fact that catchments are heterogeneous landscapes and thus the water distribution and fluxes are spatially and temporally heterogeneous (cf. Hrachowitz et al., 2013; Troch et al., 2009). This focus on the spatial and temporal variability of processes is reflected in some newer concepts for explaining runoff generation, such as the transmissivity feedback (Bishop, 1991), the fill and spill mechanism (Tromp-van Meerveld and McDonnell, 2006), and the distinction between active and contributing areas (Ambroise, 2004). All three concepts highlight the role of threshold values and connectivity, implying that there are mechanisms that may activate different spatial sources of water under different times and conditions. Despite or maybe also partly thanks to the improved observation methods, the hydrologic community is aware that there are still many difficulties to understand and define which processes occur and prevail during different situations and at different locations (e.g. Beven, 2006; Hrachowitz et al., 2013; Kirchner, 2006; McDonnell et al., 2010). Thus, the spatio-temporal heterogeneity of the relevance of different hydrologic processes remains one of the main current research objectives in catchment hydrology.

1.2 Process-based hydrological modelling

From the moment where experimental studies have started to intensively investigate the processes of runoff generation, there have also been efforts to include the gained process understanding into mathematical models (cf. Clarke, 1973; Freeze and Harlan, 1969). Two main reasons did and still do motivate the development of process-based hydrological models (cf. e.g. Loague, 2010). One reason is the need for predictions of the hydrological response in the future. Classical application examples for this are flood forecasting and the forecast of reactions of a hydrological system to land use change and climate change. The second reason is that it is not possible to investigate any hydrologic process, behaviour, or reaction in its full spatial and temporal extent with field observations. While all measurement techniques are at some point limited, models in principle allow investigations in any desired spatial and temporal detail.

Over the decades, computational power and capabilities constantly improved and the number of modelling studies and models in hydrology increased exceedingly (e.g. Burt and McDonnell, 2015; Loague, 2010). As a consequence, the variety of mathematical models that are nowadays used in hydrology is large, including analytical and numerical models, stochastic and deterministic models, and black-box models as opposite to process-based models. More detailed distinctions between the different model types are often vague and a uniform classification scheme does not exist (e.g. Clarke, 1973; Hrachowitz and Clark, 2017; Kampf and Burges, 2007). Most recently, Hrachowitz and Clark (2017) suggested a classification of process-based models according to i) the degree of complexity or abstraction of process description (physically-based or conceptual), ii) the spatial representation (distributed or lumped), iii) the model architecture (continuum-based or bucket-based), and iv) the

modelling strategy (bottom-up or top-down).

A common debate in the hydrological modelling community relates to the question which strategy of model development is more appropriate for process-based modelling. The top-down strategy is a deductive approach and develops a model by iterative testing and adding of process descriptions for reproducing observation data. This model development is classically done with a lumped, conceptual, bucket-type model. The bottom-up strategy is an inductive approach and does the reverse, i.e. it assumes that different processes observed at one scale can be combined to describe an aggregated behaviour at larger scales. Typical models for this approach are distributed, physically- and continuum-based models. However, Hrachowitz and Clark (2017) elaborated that many models mix the different concepts and that there is a wide transition range between the two extremes of 'competing philosophies' with conceptual, lumped, bucket-type, top-down models on the one side and distributed, physically-based, continuum-based, bottom-up models on the other side. For simplicity, models are nonetheless grouped into these binary opposites throughout the thesis.

The different model types have their own advantages and disadvantages such as complexity, calculation times, parameter identification, or validity of underlying assumptions and the suitability of a model type largely depends on the intended application and an appropriate implementation Fatichi et al. (2016); Hrachowitz and Clark (2017). With regard to the current research interest on the spatio-temporal heterogeneity of activation and relevance of hydrologic processes within a catchment (cf. Section 1.1), spatially-distributed, physically-based models have some clear advantages (cf. Fatichi et al., 2016; Hrachowitz and Clark, 2017; Paniconi and Putti, 2015). First of all and implied in the naming, they allow a spatially distributed representation of hydrologic states and fluxes. Moreover, the parameters of spatially-distributed, continuum-based models are supposed to represent state variables with a physical meaning, which allows the explicit consideration and assessment of the influence of varying structural conditions on the spatio-temporal hydrologic functioning. Finally, spatially-distributed, continuum-based models simulate implicitly the spatial and temporal interplay of different processes. By this, it is pictured automatically how different water sources and runoff generation mechanisms activate and predominate depending on the internal hydrologic state.

1.3 Need for synergy between simulations and observations

The process representation in hydrological models can obviously only be as good as is the current process understanding (e.g. Burt and McDonnell, 2015; Kirchner, 2006). In this context, it has been criticized that the number of experimental studies in comparison to modelling studies is decreasing and that the focus of field observations often shifts from process understanding to the collection of data for model setup and parametrisation (e.g. Beven, 2016; Burt and McDonnell, 2015). While it is good and important to have knowledge from field observations for a proper model setup (e.g. Burt and McDonnell, 2015; Cloke et al., 2003; Hrachowitz et al., 2014; Seibert and McDonnell, 2002), more field investigations need to keep focusing on process understanding (e.g. Barthold and Woods, 2015; Beven, 2016; Burt and McDonnell, 2015). Modelling studies, in turn, should not focus on

optimizing parametrisation and computational aspects, but rather should be used as complement to field studies (see next paragraphs).

A synergy between field observation and simulation studies is considered as one of the best ways to advance the hydrological process understanding and to consequently improve the realism of process-based models (e.g. Dunne, 1983; Kirchner, 2006; Seibert and McDonnell, 2002). A basic idea is to use models as tool for testing different hypotheses and to investigate if or which model structure and processes can explain observations from the real world (cf. Beven, 2012; Clark et al., 2011). In this sense, mismatches between model results and observations are not a failure, but the rejection of a model that gives the wrong answers for the right reasons helps to identify current lacks of process understanding and the need for further field investigations (e.g. Beven, 2007, 2016; Beven and Binley, 2014; Hrachowitz et al., 2013; Loague, 2010). Ideally, the synergy between observations and simulations pursues an iterative loop, where the confrontation of model results with measurements is used to iteratively refine and improve the current process understanding, identify new observation needs, and develop improved model formulations (cf. Clark et al., 2016, 2017).

The approach of using different model setups in conjunction with field observations for testing and adapting different structures and processes follows the deductive top-down philosophy. Consequently, models that are traditionally used for explicit hypothesis testing against field data are conceptual, lumped, bucket-type models (e.g. Fenicia et al., 2014; Hrachowitz et al., 2014; Kavetski et al., 2011; Sivapalan et al., 2003). In principle, it is possible to also use physically-based, spatially-distributed models for such a top-down analysis (cf. Hrachowitz and Clark, 2017), yet this is hardly done. If a physically-based, spatially-distributed model is used for explicit testing of the performance of different model setups, then this is currently mainly done with numerical experiments (e.g. Ameli et al., 2016; Frei et al., 2012; Hopp and McDonnell, 2009; Reaney et al., 2014; Weiler and McDonnell, 2006). In the best case the model used for the numerical experiment is set up based on field experience (cf. definition of virtual experiments in Weiler and McDonnell 2004), but the simulation results obtained with different structures and parameters are not directly evaluated against field observations.

Nonetheless, also physically-based, spatially-distributed models are recognized to have their main value as complements to field observations for supporting field data analysis, identifying poorly understood behaviour, or helping to design future field investigations (e.g. Fatichi et al., 2016; Grayson et al., 1992; Loague, 2010). One of their great potentials is to support the interpretation of field investigations by providing detailed insights into aspects that cannot be observed with the desired spatial and temporal detail in the field (e.g. Camporese et al., 2014; Jeannot et al., 2019; Niedda and Pirastru, 2014; Nippgen et al., 2015; Partington et al., 2013; Weill et al., 2013). In order to obtain reliable insights into the hydrologic functioning, a detailed evaluation of the internal process accuracy and consistency of the model should precede. This aspect has been and still is often trifled with, although awareness clearly increased that mere measurements of integrated catchment discharge are not sufficient for a comprehensive evaluation of model parametrisation (cf. e.g. Hrachowitz et al., 2013; Koch et al., 2015; Schilling et al., 2019). Moreover, it has been

often emphasized within the past years that comparing simulation results with multiple field data, including spatially integrated and distributed data and employing new observation techniques, has great potential to reveal where model representations are adequate and where there are limitations in the current process understanding (e.g. Beven, 2006; Clark et al., 2017; Fatichi et al., 2016; Hrachowitz et al., 2013; Kirchner, 2006; Paniconi and Putti, 2015). Although spatially-distributed, physically-based simulations are very suited for such an analysis, they are currently barely used for a comprehensive evaluation and analysis of model performance with regard to the information entailed in matches and especially also mismatches between simulation results and field observations.

Two hydrological aspects that are often mentioned in the context of the need for more joint observation-simulation studies are preferential flow and surface saturation (e.g. Beven, 2010; Clark et al., 2011; Loague, 2010; McDonnell, 2003). Despite several decades of ongoing research there are still many open questions to answer and the current process understanding might largely profit from a better synergy between observations and simulations in general and by making use of physically-based, spatially-distributed models in particular. The following two sections elaborate on these aspects in further detail.

1.4 Preferential flow paths

Numerous field experiments have shown that water infiltrates and moves through the subsurface in a non-uniform way, following preferential flow paths such as macropores, fractures, or natural soil pipes. While this non-uniformity has been largely neglected for a long time when explaining or simulating runoff generation (cf. Beven and Germann, 2013), it is nowadays recognized as a ubiquitous phenomenon and preferential flow is often presumed to have an important impact on runoff generation (e.g. Beven and Germann, 2013; Weiler, 2017). However, most field observations demonstrated the occurrence of preferential flow with tracer experiments and soil moisture measurements in individual soil pits and only few took the effort to analyse the occurrence of preferential flow across a hillslope or an entire catchment (e.g. Anderson et al., 2009; Liu and Lin, 2015; Wiekenkamp et al., 2016). Moreover, the standard experimental techniques focus on the vertical direction of flow. There are some recent attempts to visualize also lateral flow paths through the soil (e.g. Anderson et al., 2009; Jackisch et al., 2017; Laine-Kaulio et al., 2014; Nyquist et al., 2018), but it remains difficult to assess the connection of these flow paths to the stream – except if they are bared by soil pipe collapses (Wilson et al., 2016). As a consequence, it is an open question what is the actual relevance of the ubiquitously observed (vertical) preferential flow paths at larger scales and how and under which conditions they affect the generation of catchment runoff (cf. Beven, 2018; Beven and Germann, 2013; Weiler, 2017).

These questions are a paramount example for the need to benefit from complementary modelling studies, since the observation techniques are (at least currently) clearly not sufficient to find a comprehensive answer. The questions relate to spatial connection and relevance of small-scale processes at larger scales, following the bottom-up philosophy (cf. Section 1.2). Therefore it might be especially

interesting to rely on spatially-distributed, physically-based models. In addition, these models have the advantage that they implicitly include possible variabilities of the interplay of different processes (cf. Section 1.2). In principle, it is straightforward to apply a spatially-distributed, physically-based model for a hypothesis testing approach (cf. Section 1.3) and to implement a catchment model with and without the consideration of preferential flow paths to compare the simulated discharge of the different model setups with field measurements. Yet so far, only few simulation studies considered a representation of preferential flow paths at catchment scale (e.g. Krzeminska et al., 2013; Kukemilks et al., 2018; Steinbrich et al., 2016; Villamizar and Brown, 2017) and even less explicitly compared the performance of catchment model setups with and without preferential flow paths (Beckers and Alila, 2004; Christiansen et al., 2004; De Schepper et al., 2015; van Schaik et al., 2014; Yu et al., 2014).

One reason for the limited application of models as hypothesis testing tool with regard to preferential flow paths is certainly that there are two difficulties involved in the simulation of preferential flow paths that are a matter of research as well. First, there are ongoing debates on how to adequately represent preferential flow paths in process-based simulations (e.g. Beven and Germann, 2013; Jarvis et al., 2016). Suggested approaches for simulating preferential flow range from accurate descriptions of flow processes within discrete fractures or macropore networks to the abstraction of preferential flow paths as a uniform model domain in coexistence to a soil matrix domain (dual domain approach). Even though there are debates about the most appropriate representation, especially with regard to the physical adequacy of the equations commonly used in the prevailing dual domain approaches, all approaches have been shown to provide useful simulation results (cf. Beven and Germann, 2013; Jarvis et al., 2016). Their adequacy mainly depends on the intended use of the simulation (cf. Section 1.2). For assessing the relevance of locally observed preferential flow on catchment responses, any approach should be adequate.

Second, it is an open question how to identify adequate parameters for any chosen representation of preferential flow (cf. Beven and Germann, 2013). A parametrisation approach that follows the idea that locally observed preferential flow paths are relevant at catchment scale is to identify model parameters from plot-scale observations and to use this information to parametrise the catchment model (cf. Beven and Germann, 2013; Cadini et al., 2013; Wang et al., 2014). Such a parameter transfer from plot to catchment scale is promising for obtaining observation-based parametrisations of a catchment model, but more research is needed to understand how useful the plot-scale observations are for identifying reliable parametrisations and if the approach is sensitive to spatial heterogeneity of preferential flow occurrence. Moreover, further assessing and applying the parameter transfer approach is another way to assess how relevant locally observed preferential flow paths are at larger scale and it will help to understand if including a preferential flow representation in a catchment model is actually needed for runoff simulations.

1.5 Surface saturation

Areas of surface saturation are areas where water is ponding or flowing on the ground surface. Since the development of the partial area and variable source area concept half a century ago (e.g. Betson, 1964; Dunne and Black, 1970; Hewlett and Hibbert, 1967) it is recognized that extent and location of surface saturated areas are limited and depend on surface and subsurface properties, precipitation intensities, and the degree of saturation of the subsurface. In principle, surface saturation develops when the infiltration capacity of a surface and the underlying unsaturated subsurface is exceeded (infiltration excess) or when the underlying subsurface is saturated (saturation excess). In both cases, ongoing water supply can activate overland flow and surface saturated areas can contribute to runoff generation when connected to the stream. Until today, various experiment and modelling studies have focussed on understanding the active contribution of surface saturation to runoff generation (e.g. Ambroise, 2016; Latron and Gallart, 2007; Mengistu and Spence, 2016; Spence and Mengistu, 2019; Weill et al., 2013; Zimmermann et al., 2014). Others highlighted the relevance of surface saturated areas for water quality (e.g. Doppler et al., 2014b; Gburek and Sharpley, 1998; Heathwaite et al., 2005; Megahan and King, 1985), or tested which key controls allow the prediction of the spatial and temporal occurrence of surface saturation (e.g. Ali et al., 2014; Grabs et al., 2009; Güntner et al., 2004). Yet there are still limitations in the current understanding about the spatio-temporal heterogeneous generation of surface saturation and the connection between surface saturated areas and stream runoff.

One of the main problems in advancing the current understanding is that extensive data sets on the spatial and temporal distribution of surface saturation occurrence are missing (cf. Ambroise, 2016; Spence and Mengistu, 2019). Early work by Dunne et al. (1975) described and discussed already numerous approaches for the mapping of surface saturated areas. Besides direct inspection by walking through the area of interest, they suggested topography, soil morphology, vegetation, baseflow, water table elevation, soil moisture, or antecedent moisture conditions as proxies for identifying the spatial or temporal occurrence of surface saturation. Since then, numerous studies evaluated, applied, and advanced the usage of these proxies (e.g. Ali et al., 2014; Doppler et al., 2014a; Grabs et al., 2009; Kulasova et al., 2014). Others mapped surface saturation directly with 'on-off' sensors (e.g. Gburek and Sharpley, 1998; Zimmermann et al., 2014) or manually by walking through the area of interest (e.g. Ali et al., 2014; Latron and Gallart, 2007). However, all these direct or proxy methods are limited in their application in space or time and a standard mapping method is missing. Consequently, obtaining a comprehensive data set on the occurrence of surface saturation during different hydrologic conditions (e.g. runoff events and baseflow conditions) with a high spatial resolution (e.g. distinguishing different locations within a catchment) is difficult.

Simulation studies can partly compensate for the lack of detailed spatial field observations during varying hydrologic conditions, particularly when applying spatially-distributed, physically-based models (cf. Section 1.2). Nonetheless, as stated before (Section 1.3), they should be combined with field observations in order to allow advancing the process understanding in a reliable way. Previous simulation studies that performed an evaluation of model performance in comparison to observed

dynamics or spatial patterns of surface saturation aimed to use the model rather as prediction than as learning tool (e.g. Ali et al., 2014; Grabs et al., 2009; Güntner et al., 2004). Others investigated the spatio-temporal generation of surface saturation with the help of spatially-distributed, physically-based models but spared a detailed validation of the internal model consistency (e.g. Partington et al., 2013; Weill et al., 2013). Studies that analyse the generation and occurrence of surface saturation with an exhaustive synergy between observations and simulations are clearly missing. Ideally, such an approach should include i) an extensive evaluation of the internal model consistency and ii) a detailed analysis of matches and mismatches between observed and simulated patterns and dynamics of surface saturation, before iii) eventually using the model to obtain deeper insights into the processes of surface saturation generation. Two examples that applied such an approach to some extent are the studies by Mengistu and Spence (2016) and Glaser et al. (2016). Both evaluated the simulated surface saturation with observations from near-ground remote sensing data before using the model to learn more about the runoff generation processes.

The usage of remote sensing data is a well-established method for the identification of flooding over large areas and commonly relies on satellite or airborne images with multi-spectral or synthetic aperture radar information (e.g. Chini et al., 2017; de Alwis et al., 2007; Matgen et al., 2006). The usage of photographs for mapping surface saturation on scales of centimetres to few hundreds of metres, however, only came more into focus within the last years. For example, Silasari et al. (2017) and Spence and Mengistu (2016) demonstrated the application of visible light spectrum (VIS) imagery for mapping surface saturation on an agricultural and grassland site by taking VIS images with a camera installed on a weather station mast and a drone, respectively. Glaser et al. (2016) and Pfister et al. (2010) demonstrated the capability of using a handheld thermal infrared (TIR) camera for recurrent mapping of surface saturation within a small riparian area of a forested headwater catchment. The advantage of using photographs compared to other methods is that photographs are in principle non-invasive, spatially and temporally flexible, and a rather direct and intuitive mapping method. This implies high potential of VIS or TIR imagery for improving the understanding of the generation and occurrence of surface saturation during different hydrological conditions and at different locations, both based on field mapping itself and in combination with simulations. Nonetheless, applications of the photography methods in catchment hydrology are still rare and further assessment of the technical possibilities and limitations of the application of VIS and especially TIR imagery are needed before they might establish as standard mapping method.

2 Objectives and structure of the thesis

The overarching goal of my PhD project was to advance the hydrologic process understanding by applying a multi-method approach that combines state of the art and innovative field observations and spatially-distributed, physically-based hydrologic modelling. Specifically, my investigations focussed on the role of preferential flow paths for runoff generation and the generation of surface saturation across space and time in the 42 ha forested Weierbach catchment in Luxembourg. In the sense of a close synergy between observations and simulations, I set up a 3-dimensional, spatially-distributed, physically-based hydrological model of the catchment and confronted it with multiple field observation data to test and evaluate the internal process consistency and to learn more about the current process understanding.

The different studies presented in this thesis were designed to not only provide insights into the functioning of the Weierbach catchment, but to also have a relevance for future investigations beyond the exemplary study site with regard to the applied approaches, developed methods, and gained knowledge and process understanding. All studies followed the overarching goal, yet their research focus and methodology (field observation and / or modelling study) are diverging. Study 1 focussed on understanding the role of preferential flow paths observed at plot scale for simulating long-term catchment responses. Study 2 dealt with mapping of surface saturated areas in space and time. Study 3 explored the spatial and temporal occurrence of surface saturation by analysing matches and mismatches between simulations and observations. Study 4 relied on the simulation of surface saturation to identify how different water sources mix and vary in surface saturated areas in space and time.

Below, I summarize the specific research objectives of the different studies and how the studies connect to each other. In chapter 3, I introduce the study site, the field data, and the model setup. Chapter 4 summarizes the findings and conclusions of the individual studies. Chapter 5 closes with concluding remarks on the asset of synergies between observations and simulations and gives an outlook on possible follow-up studies in the study site. Following this, the manuscripts of the individual studies are presented.

Study 1: How meaningful are plot-scale observations and simulations of preferential flow for catchment models?

Methodology: Combined field observation and modelling study.

Focus: Preferential flow. Process understanding in the Weierbach headwater and beyond.

Study 1 investigated the importance of vertical preferential flow paths observed at plot scale (1 m^2) for explaining and simulating the long-term response of discharge and soil moisture of the Weierbach headwater. The first objective was to assess if it is valuable to inform the parametrisation of preferential flow in a catchment model based on simulations of plot-scale observations of preferential flow. The second objective was to assess how such a parameter transfer from the plot scale to the catchment scale is affected by observations that show spatially heterogeneous occurrence and prevalence of preferential flow. The final objective was to assess the conceptual representation of vertical and lateral preferential flow with a dual-permeability approach for simulating the long-term integrated response of a humid-temperate catchment.

Connection to other studies: The findings of this study influenced the setup of the model used in Study 3 and Study 4.

Study 2: Technical note: Mapping surface-saturation dynamics with thermal infrared imagery

Methodology: Mere field observation study.

Focus: Surface saturation. Method development.

The objective of Study 2 was to assess and demonstrate which conditions, precautions, and image processing methods allow for a successful mapping and quantification of surface saturation with thermal infrared (TIR) imagery over different seasons and hydrological conditions and at various locations across a catchment. The study aimed for a strong methodological focus on the acquisition and processing of TIR images for mapping surface saturation, since experience in using the approach for long-term observations and across an entire catchment was non-existent.

Connection to other studies: The described TIR imagery method was used for providing the surface saturation observation data used in Study 3.

Study 3: Intra-catchment variability of surface saturation – insights from long-term observations and simulations

Methodology: Combined field observation and modelling study.

Focus: Surface saturation. Process understanding in the Weierbach catchment and beyond.

Study 3 investigated the occurrence of surface saturation in the Weierbach catchment with a combined observation and simulation approach. One objective was to identify the spatial and temporal characteristics of surface saturation across a catchment and to explore if a spatially-distributed, physically-based model can satisfactorily reproduce variabilities of surface saturation patterns, dynamics, and frequencies. The second objective was to identify possible explanations and key factors for the intra-catchment variability of the surface saturation characteristics by analysing the matches and mismatches between observations and simulations.

Connection to other studies: The study relied on experience from Study 2 and Study 1 for the observation data and model setup, respectively. The results of this study provided confidence in the internal process accuracy and consistency of the model, which was a necessary prerequisite for the in-depth analyses with the model in Study 4.

Study 4: Sources of surface water in space and time

Methodology: Mere modelling study.

Focus: Surface saturation. Process understanding in the Weierbach catchment and beyond.

The objective of Study 4 was to identify the processes and water sources contributing to the generation of surface saturation in the riparian zone and streambed of the Weierbach catchment. The study relied on spatially-distributed, physically-based simulations and extracted in-depth information on the simulated sources of surface water with a hydraulic mixing cell approach to answer two specific research questions: i) Which immediate flow mechanisms deliver water to the surface? and ii) From which geographical sources is the surface water coming from? Both questions were addressed regarding the relative mixing of different water sources, regarding a possible spatial variability in the riparian areas and along the stream, and regarding a possible temporal variability for different wetness states and phases of wetting or drying.

Connection to other studies: The study complements and assists the interpretation of the observation data of Study 2 and Study 3. The confidence that the applied model could reflect reality originated from the detailed comparison of the simulation results with various data annd spatial observations in Study 3.

3 Materials and methods

3.1 Study site - Weierbach catchment

3.1.1 Physiography

The Weierbach catchment is a 42 ha forested headwater catchment of the Attert basin. It is located at the foothills of the Ardennes Massif in western Luxembourg. Elevation ranges from 458 to 513 m a.s.l. with slopes $< 5°$ at a plateau landscape unit (54 % of the catchment area) and slopes $> 5°$ for the hillslopes (45 % of the catchment area) (cf. Antonelli et al., 2019a; Martínez-Carreras et al., 2016). The hillslopes form a central v-shaped stream valley, a v-shaped tributary stream valley in the east, and a short stream branch in the west of the central stream valley. A narrow, flat, and well-defined riparian zone borders the streambed and widens at some particular locations, especially in the source areas of the three stream branches. Dominating tree species are European beech (Fagus sylvatica) and Sessile oak (Quercus petraea) for large parts of the catchment. Smaller parts, especially the south-east of the catchment, are dominated by Norway spruce (Picea abies) and Douglas spruce (Pseudotsuga menziessii). The riparian zone is covered by ferns, mosses, and herbaceous plants (e.g. Dryopteris carthusiana, Chrysosplenium oppositifolium).

The geology of the catchment is characterized by Devonian slate with some inclusions of phyllites and quartzites (Gourdol et al., 2018; Juilleret et al., 2011; Moragues-Quiroga et al., 2017). Highly fractured bedrock starts at 1.40 m depth in average and fractures mainly close down to a depth of 5 m (Gourdol et al., 2018). The subsolum above the fractured bedrock is a regolithic saprock(cf. Juilleret et al., 2016) and is characterized by i) a paralithic layer that mainly consists of rock fragments from decomposed bedrock (in average from 0.9 to 1.4 m below the surface) and ii) a regolithic layer (in average from 0.5 to 0.9 m below the surface) consisting of rock debris mixed with a sandy-loamy soil matrix that stems from periglacial loess deposits (cf. Gourdol et al., 2018). The regolithic layer developed as basal layer from Pleistocene Periglacial Slope Deposits (cf. Moragues-Quiroga et al., 2017). Topsoil (0 to 0.05 m) and subsoil (0.05 m to 0.5 m) developed as loamy, stony Cambisol (classified as leptic, ruptic, humic, dystric, endoskeletic, siltic) in the upper layer of the Pleistocene Periglacial cover bed (Juilleret et al., 2011, 2016; Moragues-Quiroga et al., 2017). In the riparian zone, a clay-loam Leptosol (organic, stagnic) replaces the Cambisol and directly overlies the paralithic layer and bedrock (Glaser et al., 2016).

The climate is oceanic-continental and dominated by atmospheric circulations and temperate air masses from the Atlantic (Carrer et al., 2019; Pfister et al., 2017). Mean annual precipitation is

around 950 mm without apparent seasonality and without substantial amounts of snow (Carrer et al., 2019; Pfister et al., 2017). The average yearly runoff coefficient is 0.5 (Carrer et al., 2019), but the runoff coefficient varies pronouncedly over the seasons. In winter months, when potential evapotranspiration (PET) is low, runoff coefficients are highest (cf. Pfister et al., 2017). In summer months, when PET is high, the runoff coefficient is commonly low (cf. Pfister et al., 2017). In some years, the stream intermittently dries out completely. In other years, some streamflow persists at the outlet of the catchment, but the upstream sections are intermittently dry. During the dry months, runoff is generated as immediate response to precipitation, showing a sharp, short-lasting hydrograph peak. During wet conditions, the hydrograph shows an additional, broad second peak that starts to appear few hours after the first peak, reaches its maximum 20h to 40h after the first peak and produces considerably more runoff than the first peak (e.g. Angermann et al., 2017; Glaser et al., 2016; Martínez-Carreras et al., 2016; Wrede et al., 2015).

3.1.2 Hydrological process understanding

The Weierbach catchment has been established as experimental research catchment for more than 10 years. Several studies investigated the water sources based on hydrograph separations with natural isotopic, chemical and biological tracers (Krein et al., 2006; Martínez-Carreras et al., 2015; Schwab et al., 2017; Wrede et al., 2015). Others applied artificial tracers and irrigation experiments to visualize and analyse the water flow paths through the subsurface of the catchment (Angermann et al., 2017; Jackisch et al., 2017; Scaini et al., 2017, 2018). Recent projects also addressed the water storage and travel time characteristics by applying several statistical and modelling approaches for interpreting field observations (Carrer et al., 2019; Martínez-Carreras et al., 2016; Pfister et al., 2017; Rodriguez and Klaus, 2019). Besides, the catchment has often been used as testing ground for new concepts (e.g. Carrer et al., 2019; Schwab et al., 2016) and new measurement techniques for tracing water and sediment sources, water flow paths and hydrological connectivity(e.g. Jackisch et al., 2017; Klaus et al., 2015; Martínez-Carreras et al., 2015, 2010; Pfister et al., 2009, 2010; Scaini et al., 2017). Conceptual, bucket-type models were used as hypothesis testing tool for identifying the dominant runoff generation processes (Fenicia et al., 2014; Kavetski et al., 2011; Wrede et al., 2015) and in combination with measurements of stable water isotopes for identifying the storage age and travel time distribution (Rodriguez and Klaus, 2019) of the catchment. During my master thesis, I implemented a physically-based, spatially-distributed model of the catchment in strong consultation with field investigations to learn more about the runoff generation processes (Glaser et al., 2016).

Despite this large amount of studies and experimental data in the catchment, the runoff generation processes and dominating water sources, storages, and flow paths in the Weierbach catchment remain under debate. There is a general consent that the broad discharge peaks during wet conditions – and thus most discharge volume – are largely generated by lateral subsurface flow of pre-event water and that this lateral subsurface flow path activates once a certain subsurface storage threshold is exceeded (e.g. Glaser et al., 2016; Martínez-Carreras et al., 2016; Rodriguez and Klaus, 2019; Scaini

et al., 2018; Schwab et al., 2018; Wrede et al., 2015). Yet the spatial origin and extent of this subsurface flow path is still not clarified. Current suggestions (e.g. Glaser et al., 2016; Martínez-Carreras et al., 2016; Rodriguez and Klaus, 2019; Schwab et al., 2017; Wrede et al., 2015) comprise groundwater flow in fractured bedrock, (perched) groundwater flow in the regolithic saprock, or preferential flow paths through the subsurface and the relevance of the different landscape units (i.e. plateau vs. hillslopes vs. riparian zone) is debated, including controversy about the existence of two separate groundwater systems. The first, sharp, short-lasting discharge peaks are known to consist of a mixture of event and pre-event water with high amounts of young water (e.g. Martínez-Carreras et al., 2015; Rodriguez and Klaus, 2019; Wrede et al., 2015). The peak generation is commonly related to a direct input of precipitation and throughfall to the stream as well as to a rainfall-driven activation of flow paths that quickly transfer water from the near-stream (riparian) zone to the stream (e.g. Fenicia et al., 2014; Glaser et al., 2016; Klaus et al., 2015; Martínez-Carreras et al., 2015, 2016; Rodriguez and Klaus, 2019; Scaini et al., 2018; Schwab et al., 2018; Wrede et al., 2015). In addition to saturation excess overland flow from the riparian zone (e.g. Klaus et al., 2015) and shallow subsurface flow through the riparian soil (e.g. Schwab et al., 2017), preferential subsurface flow from the hillslopes to the stream (e.g. Angermann et al., 2017; Glaser et al., 2016; Martínez-Carreras et al., 2016) has been suggested as possible activated flow contributing to the generation of the sharp, short-lasting discharge peaks. However, all of these suggested flow paths have so far barely been investigated in a comprehensive way regarding their effective contribution to runoff generation.

The circumstance that the hydrologic functioning of the Weierbach catchment is still under debate makes the catchment an ideal candidate for applying a combined simulation and observation approach. There is already a large experimental knowledge and collection of field data that can be used for the model setup and evaluation and at the same time the simulation can help to investigate the aspects that are still not fully understood. For example, there have been several field experiments that showed the occurrence of preferential flow in the soils and periglacial deposits of the Weierbach catchment (Angermann et al., 2017; Jackisch et al., 2017; Scaini et al., 2017, 2018). Consequently, preferential flow paths have been hypothesized to contribute to runoff generation in the Weierbach catchment (see above). However, the actual connection and contribution of preferential flow paths to streamflow is difficult to assess with field experiments and complementing the observations with simulations can help to clarify the effective role of preferential flow paths on runoff generation (cf. Study 1). Similarly, saturation excess overland flow has been suggested to contribute to discharge generation based on the observation that riparian zones in the Weierbach catchment often show high extents of surface saturation. Yet it has never been systematically investigated how the surface saturation develops in space and time. In order to clarify the effective contribution of the surface saturated areas to runoff generation, it is first and foremost necessary to understand the behaviour and generation of the surface saturation itself. This can be best achieved with a systematic mapping of the surface saturated areas across space and time (cf. Study 2) in combination with spatially-distributed modelling that allows for a profound analysis of the generation of surface saturation (cf. Study 3 and Study 4).

3.2 Field observations

3.2.1 Mapping of surface saturation

The main experimental focus of my PhD project was to map surface saturation patterns in the Weierbach catchment with the novel thermal infrared (TIR) imagery technique. On one side, I aimed to further test and advance the technique itself (Study 2). On the other side, I aimed to collect a comprehensive data set of surface saturation patterns and their dynamics for investigating the nature of surface saturation generation and its intra-catchment variability in a joint observation - simulation approach (Study 3) and for evaluating the internal process consistency of the model (Study 3) before using it for complementary simulation investigations (Study 4). I shared the mapping work with another PhD student, who focussed on the analysis of observed saturation dynamics in relation to the establishment of connectivity between hillslopes and streamflow for different stream sections (see co-authored manuscripts (Antonelli et al., 2019a,b). We mapped the surface saturation in several selected riparian areas of the Weierbach catchment with a handheld TIR camera with weekly to biweekly recurrence frequency from November 2015 to January 2018. Each time, we took overlapping TIR images or a TIR video of the selected riparian areas. During the growing season, the field work occasionally involved cutting the herbaceous vegetation and ferns in the riparian areas in order to maintain an uncovered view to the ground. Back in the office, we processed the the raw TIR images and videos to panorama images of surface saturation as described in Study 2. Eventually, we obtained a data set that showed the evolution of surface saturation patterns at various locations across the catchment and for various hydrological conditions (Study 2, Study 3, Antonelli et al. 2019a).

Compared to previous work by Pfister et al. (2010), Frentress (2015), and the mapping work I did for my master thesis (Glaser et al., 2016), the long-term mapping campaign largely broadened the experience in applying TIR imagery for mapping surface saturation during diverse meteorological and hydrological conditions and at locations with diverse characteristics (e.g. vegetated with trees, dry areas, smooth topography). We also tested the usage of TIR imagery for high-frequency mapping with cameras that were permanently installed in the stream source areas of the Weierbach catchment. These TIR cameras were programmed to automatically take a picture every 15 to 240 minutes between February 2017 and December 2017. Although the cameras did not run continuously due to some technical difficulties in the setup and with the power supply, the high-frequency mapping resulted in a huge data set of TIR videos showing the high-frequency evolution of surface saturation. Since the high-frequency variation of surface saturation dynamics was far less pronounced than expected and a quantification of the information of the videos was far from being straightforward, the images eventually did not find their way into a manuscript. However, the work with the permanently installed cameras also widened the experience and knowledge about TIR mapping, which is indirectly included in Study 2.

3.2.2 Other data used for the model setup and evaluation

Thanks to the long history of research in the Weierbach catchment, I could draw on a huge existing collection of field data for the model setup and evaluation. Studies in the Weierbach commonly rely on a digital elevation model (DEM) of the catchment that has been extracted from 10 m contour lines of a topographic map. I additionally relied on a high-resolution DEM of the streambed and adjunct riparian zones that has been specifically acquired for my work with ground-based LiDAR (light detection and ranging) measurements. For the setup of the model mesh (Study 1, Study 3, and Study 4), I used an adapted DEM where the height information of the high-resolution LiDAR DEM and the coarse topographic map DEM was merged and interpolated into a 0.1 m raster (resolution of interpolated heights: 0.1 mm).

Soil profiles, core drillings and electrical resistivity tomography profiles have been collected over the past years to describe the pedo-lithological structure across the catchment. I used this information in previous work (Glaser et al., 2016) to define the subsurface structure and parametrisation of the 6 ha headwater model used in Study 1. Analogously, I adapted the subsurface structure and parametrisation of the 42 ha catchment model used in Study 3 and Study 4 based on the existing field data and the knowledge and experience of my colleagues. In Study 1, I additionally used data from irrigation experiments that were performed on a hillslope south of the catchment by Jackisch et al. (2017) for parametrising and testing the preferential flow simulations. In particular, I relied on Br^- concentrations and Brilliant Blue stains sampled in three different irrigation plots.

Meteorological data were available from measurement stations in and nearby the catchment, maintained by technicians from LIST and the agricultural administration of Luxembourg (ASTA). I used the meteorological data for compiling the forcing inputs precipitation and potential evapotranspiration for the simulations (Study 1, Study 3, and Study 4). In case of data gaps, I estimated relationships between data of neighbouring stations. Moreover, I extracted the dynamic leaf area index (LAI) for the catchment from MODIS satellite data. Due to high computational demands of simulations with a realistic LAI during the growing season, I evaluated the effect of a realistic LAI on discharge and saturation simulations in comparison to simulations with a LAI of 0 and eventually used the LAI information in Study 1 only.

Continuous monitoring of groundwater level, volumetric water content and water stage has been initiated and maintained since many years at various locations across the catchment. I solely did some additional manual discharge measurements in turns with another PhD student in order to establish rating curves for converting the continuously measured water stage into discharge values. I used the hydrometric data from the existing monitoring network in addition to the mapped surface saturation to confront and compare the simulation results with multiple data types from multiple measurement locations (previous work in Glaser et al. (2016), Study 1, Study 3). This multi-data evaluation of the simulations was an essential prerequisite and component of my PhD work, since it highlights the strengths and weaknesses of a model in a differentiated and distributed way. Only if the internal model consistency is approved to a certain degree, the model can be used as learning tool as done in Study 1, Study 3 and Study 4.

3.3 HydroGeoSphere Model

3.3.1 3-dimensional integrated hydrologic surface subsurface model

I used for all simulations the model code HydroGeoSphere (Aquanty Inc.). HydroGeoSphere (HGS) is a spatially-distributed, physically-based, continuum-based model that solves the 3-dimensional Richards equation for variably saturated, transient subsurface flow

$$\frac{\partial S}{\partial t} = \nabla \cdot (K(S)\nabla h) \pm Q$$

where $S = \frac{\theta}{n}$ is the saturation of the subsurface medium, defined by the actual volumetric water content θ in relation to porosity n, $K(S) = k_r(S)K_S$ is the actual hydraulic conductivity, expressed as a saturation-dependent relative proportion $k_r(S)$ of the saturated hydraulic conductivity tensor K_S, h is the hydraulic head composed of the pressure head ψ and the elevation head z, and Q represents any exchange flux from sources or to sinks outside of the subsurface medium. For defining a relationship between the hydraulic head h and saturation S and $k_r(S)$, I relied on the Mualem - van Genuchten functions as provided in HGS

$$S = S_r + \frac{(1 - S_r)}{\left[1 + |\alpha\psi|^{\beta}\right]^{\nu}} \text{ for } \psi < 0; \ \ S = 1 \text{ for } \psi \geq 0$$

$$k_r(S) = \left(\frac{S - S_r}{1 - S_r}\right)^{0.5} \left[1 - \left(1 - \left(\frac{S - S_r}{1 - S_r}\right)^{\frac{1}{\nu}}\right)^{\nu}\right]^2$$

with $S_r = \frac{\theta_r}{n}$ being the residual saturation, α and β being empirical parameters that are normally estimated with water retention experiments, and $\nu = 1 - \frac{1}{\beta}$.

Overland flow can be simulated with the diffusion-wave approximation of the 2-dimensional Saint-Venant equation

$$\frac{\partial h_o}{\partial t} = \nabla \cdot (d_o K_o \nabla h_o) \pm Q_o$$

where h_o is the water surface elevation, d_o is the water depth, K_o is the tensor of surface conductance, and Q_o represents the exchange of surface water with any external source or sink. The surface conductance tensor K_o can be expressed by different equations, I relied on the formulation derived from the Manning equation

$$K_o = d_o^{\frac{2}{3}} n_o^{-1} \frac{\partial h_o}{\partial s}^{-0.5}$$

with n_o being the tensor of empirical Manning roughness coefficients and s describing the coordinate

along the direction of the maximum local slope.

Surface and subsurface flow are simulated simultaneously and coupled with a common node or a dual node approach. I applied the dual node approach, which simulates exchange between the surface and subsurface domain of the model via Darcy flow through a possibly thin coupling layer. The simultaneous simulation of surface and subsurface flow is the key characteristic for so-called integrated surface subsurface hydrologic models (ISSHMs), a specific type of process-based models that has been developed and increasingly assessed and applied over the last 20 years (cf. Ebel et al., 2009; Furman, 2008; Kollet et al., 2017; Maxwell et al., 2014; Paniconi and Putti, 2015; Sebben et al., 2013).

The big asset of ISSHMs is their integrative consideration of different processes and components of the water cycle, including interactions and exchanges with and across the land surface. This implies for example that the stream network does not need to be defined as a fixed boundary in the model, but that it can evolve naturally during the simulation. Also, it is not necessary to explicitly specify different surface-subsurface exchange processes such as infiltration excess or saturation excess as distinct mechanisms (cf. Paniconi and Putti, 2015). Instead, these processes automatically develop in the simulation depending on the current conditions. Both aspects were essential for the comprehensive analysis of surface saturation generation in Study 3 and Study 4.

A further aspect that is commonly considered in ISSHMs is evapotranspiration. The common trend is to couple ISSHMs with land surface and atmospheric models (e.g. Noah-MP, CLM, WRF) that simulate physical plant and atmosphere processes according to latest state-of-the-art formulations. Development into that direction is also undertaken for HGS (Davison et al., 2015, 2018), yet the standard approach in HGS is to simulate dynamic actual evapotranspiration internally, following the process-based, conceptual formulation of Kristensen and Jensen (1975). In this approach, the simulation of actual evapotranspiration comprises canopy evaporation and interception, plant transpiration, and evaporation from soil or water surfaces. The amount and partitioning of the components is calculated based on the potential evapotranspiration (input variable), water availability (precipitation input and simulated soil moisture distribution), and several vegetation parameters (e.g. rooting depth, leaf area index). Since evapotranspiration and the partitioning of its different components was not a main focus of my work, I assumed this approach to be adequate.

In addition to surface flow, subsurface flow in a porous medium, and evapotranspiration, I implemented solute transport and preferential flow in the simulations of Study 1. Solute transport is simulated in HGS with the advection-dispersion equation. For the consideration of vertical preferential flow, I relied on the dual-permeability formulation, which divides the subsurface in two interacting domains with differing hydraulic properties (especially differing hydraulic conductivities) that are both described with the Richards equation. The necessary spatial and temporal discretization for the numeric solution of the different equations is done in HGS with an unstructured finite element grid and with time steps that can be automatically adapted to the arising gradients. The discretised set of non-linear equations is implicitly linearised with the Newton-Raphson technique and the equations are solved concurrently in an iterative scheme. A special feature in HGS is that the coefficient

matrix of the control volume finite element discretization can be manipulated to mimic a finite difference discretization. I applied this approach for all my simulations, since it has been proven to deliver similar results with faster simulation time.

3.3.2 Model application to the Weierbach catchment

The models used in Study 1, Study 3 and Study 4 were developed from a HGS model of a 6 ha sub-area of the Weierbach catchment (from now on referred to as headwater) that I implemented for my master thesis (Glaser et al., 2016). This base setup of the Weierbach headwater consisted of several subsurface layers with differing porosities and hydraulic conductivities, representing the different soil, subsolum and bedrock characteristics of the catchment. In the riparian zone, a different subsurface setup accounted for the observed shallow, organic, stagnic soil that directly overlies the fractured bedrock. A nested triangular model mesh with a spatial resolution ranging from > 10 m at the hillslopes to < 0.25 m in the riparian zone ensured to account for micro-topography in the riparian zone while maintaining a reasonable computational demand. I assigned the parameters for the subsurface, surface, and evapotranspiration based on field measurements and experience, literature values, and some manual trial and error calibration. The validation of the base headwater model against discharge, soil moisture and surface saturation measurements provided confidence in the model structure and parametrisation and demonstrated the performance of the base model setup in simulating distributed and internal consistent hydrologic behaviour of the Weierbach headwater (cf. Glaser et al., 2016).

In Study 1, I incorporated the dual-permeability approach into the setup of the 6 ha headwater model in order to analyse the impact of vertical preferential flow on runoff generation and on the internal process consistency of the model. The idea was to base the dual-permeability parametrisation on Br⁻ depth profiles that had been extracted from three different plot-scale (1 m^2) irrigation experiments in the vicinity of the catchment and that showed clear evidence for vertical preferential flow (cf. Jackisch et al., 2017). In order to search for parameter sets that could reproduce the observed Br⁻ profiles, I applied a brute-force Monte Carlo (MC) algorithm to a plot-scale model that I parametrised identical to the hillslopes of the headwater model, extended by dual-permeability and transport simulation. I tested with the MC simulations different approaches for translating field observations of preferential flow (e.g. Brilliant Blue patterns, type of macropores, fast hydraulic conductivities) into realistic parameters for preferential flow simulations and searched for well-performing parameter sets for the three different Br⁻ plot profiles individually. Subsequently, I transferred several well-performing and some low-performing parameter sets of the three different simulated plot profiles to the headwater model and tested the effect of the different parametrisations on the long-term simulation of soil moisture and discharge over 18 months. The performance of the base headwater model (cf. Glaser et al., 2016) served as reference for simulations without vertical preferential flow.

In Study 3, I extended the setup of the 6 ha headwater model to the entire Weierbach catchment (42 ha). Based on the knowledge gained from Study 1 and a follow-up study (Hopp et al., 2019) of a

master student, I decided that a dual-permeability representation of preferential flow was dispensable for the catchment model. The main change in the subsurface structure and parametrisation of the catchment model in comparison to the headwater model of Glaser et al. (2016) was an adapted model depth and the addition of distinct vegetation parameters for the coniferous forest area of the catchment. Except of that, I set up the model in the same way as the headwater model, including a nested mesh with high resolution of the micro-topography in the riparian areas, including the same different subsurface layers as in the headwater model, and including as only spatial heterogeneity a differing subsurface structure for the riparian zone and streambed than for the hillslopes. Particularly, I did not perform any additional parameter calibration for the catchment model. I only checked for the capability of the catchment model to simulate discharge, soil moisture, and groundwater table as observed from October 2015 to January 2018 at various locations distributed across the catchment before confronting the model with the distributed observations of surface saturation dynamics in the riparian zones (cf. Study 2, Antonelli et al. 2019a) to investigate the nature of and controls on the intra-catchment variability of surface saturation.

The model of Study 3 proved to be internally consistent and reliable in simulating the temporal and spatial occurrence of surface saturation from October 2015 to January 2018. Thus, I relied in Study 4 on the identical model setup as used in Study 3. I only extended the simulation with the hydraulic mixing cell approach developed by Partington et al. (2011; 2013) to track and identify the origin of the simulated surface water. The application of the hydraulic mixing cell approach requires the definition of initial source areas, which I assigned according to the zonation of the different surface and subsurface parameters. Water that is initially stored in the model domain is then labelled according to the respective source areas, water that newly enters the model domain throughout the simulation is labelled as precipitation. The label of the water does not change when passing through a model cell of another source area, but when water from different source areas enters the same cell, a mixing ratio of the different sources is calculated and further tracked and adapted throughout the simulation. In order to prevent a possible complete replacement of initially stored water by precipitation, I split the simulation period from October 2015 to January 2018 into 64 subsequent simulation periods and initialised and ran the sub-periods individually. Finally, I selected several locations within the streambed and riparian zone of an upstream, midstream and downstream section of the stream and several time periods reflecting different initial wetness states and phases of drying or wetting to analyse the spatial and temporal variability of the mixing of different geographical sources and mechanisms of immediate delivery of surface water.

4 Key findings and conclusions

4.1 Preferential flow paths – How meaningful are plot-scale observations and simulations of preferential flow for catchment models? (Study 1)

The simulations presented in Study 1 revealed that the role of vertical preferential flow differs depending on the scale of interest. The Br^- depth profiles that were sampled from three irrigation experiments at plot scale (1 m^2 plots) could only be simulated when vertical preferential flow was considered in the simulations. By contrast, the influence of vertical preferential flow on long-term simulations of discharge and soil moisture response at headwater scale (6 ha) showed to be minor. None of the preferential flow parametrisations that proved to be adequate at plot scale could improve the discharge and soil moisture simulation at headwater scale compared with the reference simulation without vertical preferential flow. Instead, several of the parametrisations being adequate at plot scale decreased the performance of discharge and soil moisture simulation at headwater scale and model parametrisations that failed at plot scale did not perform systematically worse at headwater scale than successful plot-scale parametrisations.

Concerning runoff generation in the Weierbach catchment, Study 1 suggested that fast lateral flow paths largely outweighed the relative importance of vertical preferential flow observed at plot scale. This conclusion was deduced from the subsurface structure and model performance of the reference model without vertical preferential flow in comparison with the headwater simulations that included preferential flow. The subsurface structure of the reference model allowed spatially uniform fast infiltration and fast lateral flow at the interface of different subsurface layers with contrasting hydraulic conductivities. While this structure seemed to be important for the discharge simulation, the additional inclusion of vertical preferential flow with a dual-permeability approach did not further improve the simulation of discharge and soil moisture. A follow-up study (Hopp et al., 2019) supported this interpretation by showing that a direct calibration of dual-permeability parameters at the headwater scale basically removed the effect of vertical preferential flow for the Weierbach catchment.

Beyond the site-specific conclusions, the findings of Study 1 also have implications for studies outside of the Weierbach catchment. A minor role of vertical preferential flow on runoff generation than suggested by plot-scale observations is likely to be found in other catchments and landscapes as well, especially for physiographic settings with similar subsurface characteristics as for example

landscapes characterized by glacial till. This means that it should be decided case by case if an improved internal model consistency by including vertical preferential flow is worth concomitant parametrisation efforts and uncertainties and additional computational demands. Moreover, the findings of Study 1 highlighted that it is problematic to identify preferential flow parameters at plot scale for using them in catchment models. First, it was not possible to identify unique parameters for the plot-scale simulations and thus it turned out to not be reliable to transfer one optimized plot-scale parameter set to the catchment scale. Second, a direct parameter transfer is conceptually not valid in a case like the presented, where different structures and processes are relevant at different scales. Only the spatial heterogeneity between different irrigation plots showed not to be a major problem for a parameter transfer from plot to catchment scale. In conclusion, Study 1 suggested that a preferential flow parametrisation for a catchment model needs to be evaluated at catchment scale directly. The follow-up study by Hopp et al. (2019) showed the same for the reverse direction, i.e. parameters calibrated at catchment scale could not be applied successfully for plot-scale simulations.

4.2 Dynamic generation and spatial occurrence of surface saturation

4.2.1 Technical note: Mapping surface-saturation dynamics with thermal infrared imagery (Study 2)

The first part of Study 2 was based on a literature review and the experience gained from the field campaign in the Weierbach catchment and described and assessed which conditions and procedures allow for a successful recurrent mapping of surface saturation. Main identified interferences for the usability of TIR images were i) an insufficient temperature contrast between surface water and surroundings, ii) temporary view obstructions, such as vegetation or snow covering the ground surface, iii) partial exposure to sunlight, and iv) fog or raindrops. Consequently, a weather-dependent and site-specific planning of image acquisition was suggested to minimise the discard of acquired images. Furthermore, Study 2 emphasized the importance of consistent field of views for the comparison of images from different acquisition times and proposed a workflow how to ensure this for panoramas taken with a handheld TIR camera. Finally, several example images demonstrated that TIR imagery can map the temporal evolution of surface saturation over different seasons and during different hydrological conditions and that the TIR images can identify differences of spatial patterns and dynamics of surface saturation between and within various locations across a catchment.

In the second part, Study 2 described and assessed different image post-processing methods for transforming TIR images into binary saturation maps, i.e. for converting the temperature information of a pixel into the binary information 'saturated' or 'unsaturated'. The comparison of several image classification techniques showed that a manual image classification is most reliable for creating binary saturation maps, but that some other, more objective and automatable approaches are valuable options as well. Especially one approach resulted in saturation maps that matched well

the manual classification and visual assessment of the TIR images. This approach combined an automatic decomposition of the TIR image pixels into two pixel class distributions (i.e. saturated vs unsaturated pixels) with a region-growing algorithm. The main difficulty for all tested automatic image processing methods was that the characteristics of the TIR images contrasted for different acquisition times between very wet and very dry conditions and between surface water being the warmest or coldest material in the mapped area. Therefore, subsequent studies that employed the TIR images collected in the Weierbach catchment (i.e. Study 3, Study 4, Antonelli et al. 2019a, Antonelli et al. 2019b) relied on manual image classification. Nonetheless, the presented automatic image classification approaches are considered advantageous for data sets of images that have a fixed vantage point and that were collected during time spans with similar conditions.

4.2.2 Intra-catchment variability of surface saturation – insights from long-term observations and simulations (Study 3)

The catchment model that I set up in Study 3 satisfactorily simulated the discharge, soil moisture and groundwater dynamics observed at several locations distributed across the catchment, especially in the riparian zone and vicinity. Thus, the internal consistency of the model was declared reliable for a detailed analysis of the spatial and temporal occurrence of surface saturation across the riparian zones of the Weierbach catchment in synergy with the mapped patterns, dynamics and frequencies of surface saturation (cf. Study 2, Antonelli et al. 2019a). The spatially distributed extent of surface saturation varied between and within the investigated areas and the simulation reproduced the observed patterns in great detail. Since the spatial setup of the model was rather homogeneous and the main variability was micro-topography, the good model performance suggested micro-topography to be a key control for the spatial occurrence of surface saturation. The temporal increase and decrease of surface saturation extent was observed and simulated to be synchronous across the different investigated riparian areas. A statistical analysis of the observed saturation dynamics by Antonelli et al. (2019a) suggested that the temporal synchronicity reflected a catchment-wide influence of groundwater dynamics and the model substantiated this assumption by revealing that the simulated patterns and dynamics of surface saturation evolved in line with the simulated patterns and dynamics of groundwater reaching the surface.

At the same time, Study 3 suggested that surface saturation in the Weierbach catchment is not solely generated by groundwater exfiltration into local topographic depressions. The evaluation of the simulated surface saturation patterns and dynamics highlighted some shortcomings of the model, despite the generally good agreement between observations and simulations. For example, the simulated surface saturation was less persistent than observed and the simulated spatial occurrence of surface saturation matched the observations better in some of the investigated areas than in others. Moreover, the observed power law relationships between catchment discharge and extent of surface saturation showed some variability between the different areas that was not captured to the same extent with the model. These identified mismatches between observations and simulations highlighted that only additional controlling factors that were not considered in the model

setup, such as spatial heterogeneity of the subsurface structure or hysteretic processes, can explain the full variability of the observed surface saturation characteristics in the Weierbach catchment. Furthermore, the observed and appropriately simulated spatial variability of saturation frequencies was discussed to represent varying mixtures of different water sources, including for example contributions from direct precipitation and upstream areas in addition to water exfiltrating from the subsurface. Study 4 picked up on this aspect and investigated in detail the water sources of the simulated surface saturation.

Despite the particular focus on the Weierbach catchment, several results of Study 3 are also of interest for studies outside of the Weierbach catchment. First, the confrontation of the simulation with the comprehensive data set of surface saturation patterns and dynamics (cf. Study 2) demonstrated that ISSHMs such as HGS are useful and reliable tools to simulate and analyse the spatially distributed and temporally variable generation of surface saturation. Moreover, the simulation results highlighted that micro-topography is an important aspect for surface saturation generation and that it should be in general considered when predicting areas that are prone to flooding. Finally, the study demonstrated that the relationship between surface saturation and catchment discharge can vary for distinct areas within a catchment to a similar extent as observed between catchments with different morphologic and topographic features.

4.2.3 Sources of surface water in space and time (Study 4)

The hydraulic mixing cell analysis of Study 4 revealed that surface saturation in the Weierbach catchment was in large parts composed of a mixture of water originating from the fractured bedrock, riparian soil, and subsolum of the hillslopes. The mixture of geographical sources of surface water was quite homogeneous within and between the three investigated stream sections and adjacent riparian areas. Moreover, water from the subsurface stores and thus pre-event water remained the predominant component of surface water for all analysed wetness states and phases of wetting and drying, although the mixing ratio of water from different subsurface stores shifted with increasing wetness towards a higher contribution of water from upstream and uphill source areas. A clear spatial difference of water sources existed between the locations in the streambed and riparian zone regarding the immediate delivery of the surface water. In the riparian zone, the surface water was mainly delivered by direct exfiltration from the underlying subsurface. A predominance of overland flow from neighbouring surface saturated areas was limited to some distinct locations or very wet conditions. In contrast, the surface water in the streambed was largely delivered by streamflow that was generated upstream. Some relevant contribution of direct exfiltration of water from the underlying fractured bedrock into the streambed was observed during dry conditions, but as soon as streamflow was considerably activated, the relative contribution of exfiltration of subsurface water was negligible.

The identified geographical sources and immediate delivery paths of surface water indicated that surface saturation in the riparian zone of the Weierbach catchment was homogeneously generated and maintained by return flow. Contrary to what was hypothesized in Study 3, the model did not

simulate discrete locations of subsurface water exfiltration, temporary high contributions of direct precipitation, or local influences of streamflow extending into the riparian zone. Some heterogeneity occurred within the riparian zone regarding the relevance of overland flow and the mixing ratio of water originating from the riparian soil versus other subsurface stores. However, this spatial variability was small and it did not relate to the surface saturation frequency or other visually apparent features. The generation of surface saturation in the streambed and thus streamflow was related to return flow as well. This return flow was suggested to occur all along the stream and to increase its extent towards uphill source areas and upper soil layers with increasing wetness. The contribution of riparian surface water to streamflow was small relative to upstream water contributions. Yet it can be assumed that overland flow from the riparian zone to the stream occurred, eventually being nothing else than intermittent streamflow outside of the streambed.

The small simulated contribution of precipitation and thus event water to streamflow generation in the Weierbach catchment is in accordance to the commonly observed dominance of pre-event water in event runoff. The fact that the model could reproduce this behaviour suggested a pressure wave that induces an increase and extent of return flow as the mechanism underlying a fast activation and release of stored pre-event water with the onset of precipitation. This is probably the finding of Study 4 with the largest relevance for studies outside the Weierbach catchment, yet all other identified processes and water sources may be relevant at other field sites as well, especially in humid temperate forests. Moreover, although the spatio-temporal variability of sources of surface water was small in the riparian zone of the Weierbach catchment, the identified spatial variations might influence microhabitats, the biogeochemical activity, or water sampling campaigns and it is likely that the delivery and mixing of sources of surface water is more heterogeneous in riparian zones or floodplains with different sizes, landscape characteristics or climatic conditions. Finally, the study demonstrated the asset of using a hydraulic mixing cell approach to specify and quantify the sources of surface water and encourages more studies to apply this or similar approaches to complement and interpret field observations with the help of in-depth analyses of the simulated processes.

5 Synthesis and outlook

5.1 Asset of synergy of spatially-distributed, physically-based simulations and observations

The objective of my thesis was to advance the understanding about the relevance of preferential flow paths on runoff generation and the spatial and temporal generation of surface saturation within a catchment by combining diverse field observations and spatially-distributed, physically-based modelling in a multi-method approach.

Study 1 tested the information content of plot-scale observations of vertical preferential flow for a realistic model parametrisation at the catchment scale. By evaluating various parametrisations at plot and catchment scale against field observations, the continuum-based HGS model was basically used as hypothesis testing tool, an approach that is typically only followed with conceptual top-down models or in virtual experiments that lack a direct comparison with field observations. The findings of Study 1 demonstrated that there is a clear value of also applying continuum-based models for hypothesis testing, especially since they implicitly allow for changes in the relevance of different processes under varying conditions and because it is possible to test the effect of identical model structures and process parametrisations at different scales. Accordingly, the simulation tests of Study 1 did not only help to improve the confidence in the model and the process understanding for the specifically studied Weierbach catchment. They also revealed that observations of vertical preferential flow at plot scale do not automatically imply that vertical preferential flow has an important control on the long-term runoff generation. Instead, other structures and flow paths may outweigh the role of vertical preferential flow at the catchment scale. This conclusion has relevance way beyond the Weierbach catchment. It demonstrates that the typical investigations of preferential flow paths at plot scale are not sufficient to assess and understand the impact of preferential flow paths on runoff generation and more research efforts need to focus at the larger catchment scale.

In Study 2 to 4, I investigated the spatio-temporal variability of the occurrence and generation of surface saturation within a catchment with a framework of consecutive observation and simulation studies. The studies followed the idea of using comprehensive and new observation data for testing the internal consistency of a model in order to i) improve the current process understanding by identifying matches and mismatches between observation and simulation, and ii) subsequently use the evaluated model for the investigation of aspects that cannot be analysed in such a detail in the field.

Study 2 focussed on the collection of observation data but the field campaign for mapping surface saturation was not only a prerequisite for providing a comprehensive data set for evaluating the internal model consistency. The study also highlighted the great potential of TIR imagery as a novel and flexible approach for mapping surface saturation in general. I am convinced that the presented technical and methodological aspects will help to further advance and promote TIR imagery to be applied in different areas and at different spatial and temporal scales. This might eventually help to overcome measurement limitations and to obtain comprehensive data sets on the spatial and temporal distribution of surface saturation across various catchments and landscapes, information that is currently largely missing and clearly limits the knowledge about the spatio-temporal generation of surface saturation and its relation to runoff generation.

The analysis of matches and mismatches between observed and simulated surface saturation characteristics presented in Study 3 was certainly of primary interest for understanding the hydrologic processes in the specifically studied Weierbach catchment. Yet it also highlighted some aspects that are of more general interest, such as the sound usability of ISSHMs for simulating surface saturation, the importance of micro-topography for the spatial occurrence of surface saturation, and the finding that the intra-catchment variability of the relationship between surface saturation extent and discharge can be similar to the variability observed between catchments with different topographies and morphologies. This clearly demonstrated the value of a detailed analysis of matches and mismatches between spatially-distributed, physically-based simulations and observations to test and challenge current process understanding.

Study 4 showed that the application of a simulation approach like the hydraulic mixing cell simulations can be very useful for in-depth analyses of hydrologic processes and aspects that are difficult to observe in the field. In principle, Study 4 could have been performed without the preceding work and findings of Study 2 and Study 3, but the detailed assessment of the model performance and its internal process consistency against the comprehensive field information on surface saturation was necessary to ensure that results and conclusions of Study 4 were actually meaningful. As for the findings of Study 3, the results of Study 4 are above all relevant for the process understanding in the Weierbach catchment, but the study nonetheless revealed some interesting aspects for hydrological research beyond the specific study site. For example, the simulations suggested return flow into the streambed and precipitation-induced pressure waves as main streamflow generation processes and it is likely that these processes also explain streamflow generation in other catchments where a predominance of pre-event water from different subsurface stores was observed in event runoff. Moreover, Study 4 was a first endeavour to asses the spatio-temporal variability of the generation of surface saturation in a riparian zone and while the identified processes and sources are likely to be similar in other humid temperate forests, more studies need to follow to investigate the variability of sources of surface saturation in riparian zones and floodplains with larger size, other landscape characteristics, or different climatic conditions.

In sum, the findings of the different studies of my thesis demonstrated that the different applied approaches for using a spatially-distributed, physically-based model as complement to field observa-

tions entail potential for substantial contribution to scientific progress. In order to exhaust the full potential of physically-based, spatially-distributed models in synergy with observations to advance hydrologic process understanding, I think far more studies should set up and confront their models with multiple observation data in the sense of a real learning tool than currently done. Certainly, it is worth considering more often to apply also physically-based, spatially-distributed models in synergy with field observations in the sense of hypothesis testing. Moreover, studies should especially follow more often the sequence of first using multiple and distributed data for testing the internal model consistency and for identifying current limitations of the process understanding, before subsequently – if the evaluation was satisfactory – using the model for detailed analyses that go beyond the possibilities of the field observations.

Finally, based on my experience from the studies of my thesis, I would suggest that more often one and the same person should work as experimentalist and modeller for a real synergy between simulations and observations. A model setup always requires some experimental knowledge and at the same time the work with a model always implicitly includes its usage as a learning tool. The gained knowledge from the field work and 'virtual experience' creates a holistic understanding of the hydrologic functioning that is very difficult to obtain with the same detail from a simple 'dialogue' between experimentalists and modellers. This is certainly particularly useful for the investigation of a specific study site, but - as the findings of my thesis showed - the eventually gained knowledge can also be of interest for other study sites and landscapes. In the final section, I give further examples for studies that could and should rely on a synergy between observations and simulations. The examples focus on process understanding in the Weierbach catchment, but all suggested investigations could also be applied in the same or a similar way for other study sites and they have the potential to advance the hydrologic process understanding beyond a specific study site.

5.2 Next steps in the Weierbach catchment

Possible follow-up studies in the Weierbach catchment may use the already intensively tested model in combination with additional field data to further improve the process understanding for that specific catchment. Although my work advanced the process understanding with regard to the role of preferential flow paths at different scales and the spatio-temporal generation of surface saturation, there are several aspects that I did not address in my thesis. For example, Study 1 showed that the non-uniformity of water flow through the subsurface has less impact on the volumetric discharge generation in the Weierbach catchment than often assumed, but it remained unclear if and how preferential flow paths affect the transport of solutes to the catchment outlet. This question might be investigated by including solute transport or particle tracking in the catchment model and comparing results of simulations with and without preferential flow to water chemistry and isotope samples. Moreover, I did not specifically analyse the impact of preferential flow paths on the generation of the characteristic sharp, short-lasting peaks of the Weierbach as immediate runoff response to precipitation. Simulations analogous to Study 1 but for single events and with

high temporal resolution of model input and validation data might help to eventually clarify the often suggested contribution of preferential subsurface flow to these characteristic 'first' discharge peaks.

At the same time, event-based simulations could also investigate the evolution of surface saturation during single events and the contribution of surface saturation excess overland flow to 'first' and 'second' discharge peaks. Study 2 to Study 4 largely improved the understanding of the spatio-temporal generation of surface saturation in the riparian zone of the Weierbach catchment and especially revealed that surface saturation is predominantly generated by return flow of water that was previously stored in the subsurface, while a contribution of direct precipitation to the generation of surface saturation was generally small. Also, the simulation results of Study 4 suggested that contributions of overland flow from the riparian zone to streamflow generation were small. However, the focus of the studies of my thesis was on the long-term generation and occurrence of surface saturation across the catchment and I did not explicitly analyse the evolution of surface saturation during single events and the contribution of overland flow to the particular 'first' and 'second' discharge peaks. Therefore, a future study may further analyse these aspects with similar approaches as followed in Study 3 and Study 4. In particular, it might be interesting to compare the simulated occurrence of surface saturation during single events to the collected high-resolution (15 min) TIR images, where preliminary analyses showed that surface saturation is not very reactive within single events, thus suggesting in line with the findings from Study 4 that the contribution of surface saturation to runoff generation is smaller than often suggested. Moreover, event runoff sources – including saturation excess overland flow – could be easily identified and quantified with the hydraulic mixing cell simulations (cf. Study 4) and be compared to inferences drawn from existing geochemistry and isotope samples about runoff sources. Finally, a study similar to Study 1 might assess how the inclusion or omission of vertical preferential flow paths influences the generation of surface saturation within single events and for the longer term.

In order to enter an ultimate loop of alternating observations and simulations for testing and improving current process understanding, a next step could be to design a sampling campaign for analysing different water sources of surface saturation based on chemical and isotopic water samples. The simulation results of Study 4 could be used to identify sampling locations and sampling times with particular expected similarities or differences of water sources. Once sampled and analysed, the field data could then again be used analogous to Study 3 to further test the internal consistency of the model and to identify which processes or structural features might be neglected in the model. Furthermore, future studies could apply the HGS model of the Weierbach catchment to address completely different aspects than preferential flow and surface saturation. For example, forward particle tracking could be used in order to estimate transient travel time and residence time distributions and the results could be confronted with existing isotopic data and estimations based on storage selection functions. Virtual experiments could investigate how runoff generation evolves under changing climate or land use or when accounting for spatial heterogeneities of the subsurface characteristics or for the often discussed existence of two different groundwater systems.

References

Ali, G., Birkel, C., Tetzlaff, D., Soulsby, C., McDonnell, J.J., Tarolli, P., 2014. A comparison of wetness indices for the prediction of observed connected saturated areas under contrasting conditions. Earth Surf. Process. Landforms 39, 399–413. doi:`10.1002/esp.3506`.

de Alwis, D.a., Easton, Z.M., Dahlke, H.E., Philpot, W.D., Steenhuis, T.S., 2007. Unsupervised classification of saturated areas using a time series of remotely sensed images. Hydrol. Earth Syst. Sci. 11, 1609–1620. doi:`10.5194/hess-11-1609-2007`.

Ambroise, B., 2004. Variable'active' versus'contributing' areas or periods: a necessary distinction. Hydrol. Process. 18, 1149–1155. doi:`10.1002/hyp.5536`.

Ambroise, B., 2016. Variable water-saturated areas and streamflow generation in the small Ringelbach catchment (Vosges Mountains, France): the master recession curve as an equilibrium curve for interactions between atmosphere, surface and ground waters. Hydrol. Process. 30, 3560–3577. doi:`10.1002/hyp.10947`.

Ameli, A., Amvrosiadi, N., Grabs, T., Laudon, H., Creed, I., McDonnell, J., Bishop, K., 2016. Hillslope permeability architecture controls on subsurface transit time distribution and flow paths. J. Hydrol. 543, 17–30. doi:`10.1016/j.jhydrol.2016.04.071`.

Anderson, A.E., Weiler, M., Alila, Y., Hudson, R.O., 2009. Dye staining and excavation of a lateral preferential flow network. Hydrol. Earth Syst. Sci. 13, 935–944. doi:`10.5194/hess-13-935-2009`.

Angermann, L., Jackisch, C., Allroggen, N., Sprenger, M., Zehe, E., Tronicke, J., Weiler, M., Blume, T., 2017. Form and function in hillslope hydrology : characterization of subsurface flow based on response observations. Hydrol. Earth Syst. Sci. 21, 3727–3748. doi:`10.5194/hess-21-3727-2017`.

Antonelli, M., Glaser, B., Teuling, A.J., Klaus, J., Pfister, L., 2019a. Saturated areas through the lens: 1. Spatio-temporal variability of surface saturation documented through Thermal Infrared imagery. Hydrol. Process. Under review, HYP–19–0453.

Antonelli, M., Glaser, B., Teuling, A.J., Klaus, J., Pfister, L., 2019b. Saturated areas through the lens: 2. Spatio-temporal variability of streamflow generation and its relationship with surface saturation. Hydrol. Process. Under review, HYP–19–0455.

Barthold, F.K., Woods, R.A., 2015. Stormflow generation: A meta-analysis of field evidence from small, forested catchments. Water Resour. Res. 51, 3730–3753. doi:`10.1002/2014WR016221`.

Beckers, J., Alila, Y., 2004. A model of rapid preferential hillslope runoff contributions to peak flow generation in a temperate rain forest watershed. Water Resour. Res. 40, 1–19. doi:`10.1029/2003WR002582`.

Betson, R.P., 1964. What is watershed runoff? J. Geophys. Res. 69, 1541–1552. doi:`10.1029/JZ069i008p01541`.

Beven, K., 2007. Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process. Hydrol. Earth Syst. Sci. 11, 460–467. doi:`10.5194/hess-11-460-2007`.

Beven, K., 2012. Causal models as multiple working hypotheses about environmental processes. Comptes Rendus Geosci. 344, 77–88. doi:`10.1016/j.crte.2012.01.005`.

Beven, K., 2016. Advice to a young hydrologist. Hydrol. Process. 30, 3578–3582. doi:`10.1002/hyp.10879`.

References

Beven, K., 2018. A Century of Denial: Preferential and Nonequilibrium Water Flow in Soils, 1864-1984. Vadose Zo. J. 17, 180153. doi:`10.2136/vzj2018.08.0153`.

Beven, K., Binley, A., 2014. GLUE: 20 years on. Hydrol. Process. 28, 5897–5918. doi:`10.1002/hyp.10082`.

Beven, K., Germann, P., 2013. Macropores and water flow in soils revisited. Water Resour. Res. 49, 3071–3092. doi:`10.1002/wrcr.20156`.

Beven, K.J., 2006. Streamflow Generation Processes, in: McDonnell, J.J. (Ed.), IAHS Benchmark Papers in Hydrology Series No. 1. IAHS Press, Wallingford, 432 pages.

Beven, K.J., 2010. Preferential flows and travel time distributions: defining adequate hypothesis tests for hydrological process models. Hydrol. Process. 24, 1537–1547. doi:`10.1002/hyp.7718`.

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 Resour. Res. 51, 3837–3866. doi:`10.1002/2015WR017016`.

Bishop, K.H., 1991. Episodic increase in stream acidity, catchment flow pathways and hydrograph separation. Phd thesis. University of Cambridge.

Burt, T.P., McDonnell, J.J., 2015. Whither field hydrology? The need for discovery science and outrageous hydrological hypotheses. Water Resour. Res. 51, 5919–5928. doi:`10.1002/2014WR016839`.

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. Stoch. Environ. Res. Risk Assess. 27, 505–516. doi:`10.1007/s00477-012-0595-8`.

Camporese, M., Penna, D., Borga, M., Paniconi, C., 2014. A field and modeling study of nonlinear storage-discharge dynamics for an Alpine headwater catchment. Water Resour. Res. 50, 806–822. doi:`10.1002/2013WR013604`.

Carrer, G.E., Klaus, J., Pfister, L., 2019. Assessing the Catchment Storage Function Through a Dual-Storage Concept. Water Resour. Res. 55, 476–494. doi:`10.1029/2018WR022856`.

Chini, M., Hostache, R., Giustarini, L., Matgen, P., 2017. A Hierarchical Split-Based Approach for Parametric Thresholding of SAR Images: Flood Inundation as a Test Case. IEEE Trans. Geosci. Remote Sens. 55, 6975–6988. doi:`10.1109/TGRS.2017.2737664`.

Christiansen, J.S., Thorsen, M., Clausen, T., Hansen, S., Refsgaard, J.C., 2004. Modelling of macropore flow and transport processes at catchment scale. J. Hydrol. 299, 136–158. doi:`10.1016/j.jhydrol.2004.04.029`.

Clark, M.P., Bierkens, M.F.P., Samaniego, L., Woods, R.a., Uijlenhoet, R., Bennett, K.E., Pauwels, V.R.N., Cai, X., Wood, A.W., Peters-Lidard, C.D., 2017. The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism. Hydrol. Earth Syst. Sci. 21, 3427–3440. doi:`10.5194/hess-21-3427-2017`.

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

Clark, M.P., Schaefli, B., Schymanski, S.J., Samaniego, L., Luce, C.H., Jackson, B.M., Freer, J.E., Arnold, J.R., Moore, R.D., Istanbulluoglu, E., Ceola, S., 2016. Improving the theoretical underpinnings of process-based hydrologic models. Water Resour. Res. 52, 2350–2365. doi:`10.1002/2015WR017910`.

Clarke, R., 1973. A review of some mathematical models used in hydrology, with observations on their calibration and use. J. Hydrol. 19, 1–20. doi:`10.1016/0022-1694(73)90089-9`.

Cloke, H., Renaud, J.P., Claxton, A., McDonnell, J., Anderson, M., Blake, J., Bates, P., 2003. The effect of model configuration on modelled hillslope–riparian interactions. J. Hydrol. 279, 167–181. doi:`10.1016/S0022-1694(03)00177-X`.

Davison, J.H., Hwang, H.T., Sudicky, E.a., Lin, J.C., 2015. Coupled atmospheric, land surface, and subsurface modeling: Exploring water and energy feedbacks in three-dimensions. Adv. Water Resour. 86, 73–85. doi:`10.1016/j.advwatres.2015.09.002`.

Davison, J.H., Hwang, H.T., Sudicky, E.A., Mallia, D.V., Lin, J.C., 2018. Full Coupling Between the Atmosphere, Surface, and Subsurface for Integrated Hydrologic Simulation. J. Adv. Model. Earth Syst. 10, 43–53. doi:`10.1002/2017MS001052`.

De Schepper, G., Therrien, R., Refsgaard, J.C., Hansen, A.L., 2015. Simulating coupled surface and subsurface water flow in a tile-drained agricultural catchment. J. Hydrol. 521, 374–388. doi:`10.1016/j.jhydrol.2014.12.035`.

Doppler, T., Honti, M., Zihlmann, U., Weisskopf, P., Stamm, C., 2014a. Validating a spatially distributed hydrological model with soil morphology data. Hydrol. Earth Syst. Sci. 18, 3481–3498. doi:`10.5194/hess-18-3481-2014`.

Doppler, T., Lück, A., Camenzuli, L., Krauss, M., Stamm, C., 2014b. Critical source areas for herbicides can change location depending on rain events. Agric. Ecosyst. Environ. 192, 85–94. doi:`10.1016/j.agee.2014.04.003`.

Dunne, T., 1983. Relation of field studies and modeling in the prediction of storm runoff. J. Hydrol. 65, 25–48. doi:`10.1016/0022-1694(83)90209-3`.

Dunne, T., Black, R.D., 1970. Partial area contributions to storm runoff in a small New England watershed. Water Resour. Res. 6, 1296–1311. URL: `http://soilandwater.bee.cornell.edu/Research/VSA/papers/DunneWRR70.pdf`.

Dunne, T., Moore, T.R., Taylor, C.H., 1975. Recognition and prediction of runoff-producing zones in humid regions. Hydrol. Sci. Bull. 20, 305–327.

Ebel, B.A., Mirus, B.B., Heppner, C.S., VanderKwaak, J.E., Loague, K., 2009. First-order exchange coefficient coupling for simulating surface water-groundwater interactions: parameter sensitivity and consistency with a physics-based approach. Hydrol. Process. 23, 1949–1959. doi:`10.1002/hyp.7279`.

Fatichi, S., Vivoni, E.R., Ogden, F.L., Ivanov, V.Y., Mirus, B., Gochis, D., Downer, C.W., Camporese, M., Davison, J.H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., Tarboton, D., 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. J. Hydrol. 537, 45–60. doi:`10.1016/j.jhydrol.2016.03.026`.

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

Freeze, R., Harlan, R., 1969. Blueprint for a physically-based, digitally-simulated hydrologic response model. J. Hydrol. 9, 237–258. doi:`10.1016/0022-1694(69)90020-1`.

Frei, S., Knorr, K.H., Peiffer, S., Fleckenstein, J.H., 2012. Surface micro-topography causes hot spots of biogeochemical activity in wetland systems: A virtual modeling experiment. J. Geophys. Res. Biogeosciences 117, G00N12. doi:`10.1029/2012JG002012`.

Frentress, J., 2015. The role of near-stream zones on flow, chemistry and isotopic composition at the headwater scale. Phd thesis. Oregon State University.

Furman, A., 2008. Modeling Coupled Surface–Subsurface Flow Processes: A Review. Vadose Zo. J. 7, 741–756. doi:`10.2136/vzj2007.0065`.

Gburek, W.J., Sharpley, A.N., 1998. Hydrologic Controls on Phosphorus Loss from Upland Agricultural Watersheds. J. Environ. Qual. 27, 267–277. doi:`10.2134/jeq1998.00472425002700020005x`.

Glaser, B., Klaus, J., Frei, S., Frentress, 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 Resour. Res. 52, 8317–8342. doi:`10.1002/2015WR018414`.

References

Gourdol, L., Clément, R., Juilleret, J., Pfister, L., Hissler, C., 2018. Large-scale ERT surveys for investigating shallow regolith properties and architecture. Hydrol. Earth Syst. Sci. Discuss. doi:`10.5194/hess-2018-519`.

Grabs, T., Seibert, J., Bishop, K., Laudon, H., 2009. Modeling spatial patterns of saturated areas: A comparison of the topographic wetness index and a dynamic distributed model. J. Hydrol. 373, 15–23. doi:`10.1016/j.jhydrol.2009.03.031`.

Grayson, R.B., Moore, I.D., McMahon, T.a., 1992. Physically based hydrologic modeling: 2. Is the concept realistic? Water Resour. Res. 26, 2659–2666. doi:`10.1029/92WR01259`.

Güntner, A., Seibert, J., Uhlenbrook, S., 2004. Modeling spatial patterns of saturated areas: An evaluation of different terrain indices. Water Resour. Res. 40, W05114. doi:`10.1029/2003WR002864`.

Heathwaite, A., Quinn, P., Hewett, C., 2005. Modelling and managing critical source areas of diffuse pollution from agricultural land using flow connectivity simulation. J. Hydrol. 304, 446–461. doi:`10.1016/j.jhydrol.2004.07.043`.

Hewlett, J.D., Hibbert, A.R., 1967. Factors affecting the response of small watersheds to precipitation in humid areas, in: Sopper, W.E., Lull, H. (Eds.), Int. Symp. For. Hydrol., Pergamon Press, Oxford. pp. 275–290. URL: `http://coweeta.ecology.uga.edu/publications/851.pdf`.

Hopp, L., Glaser, B., Klaus, J., Schramm, T., 2019. Calibration of a three dimensional dual-permeability model at the catchment scale using DREAM. Hydrol. Process. Under review, HYP–19–0345.

Hopp, L., McDonnell, J., 2009. Connectivity at the hillslope scale: Identifying interactions between storm size, bedrock permeability, slope angle and soil depth. J. Hydrol. 376, 378–391. doi:`10.1016/j.jhydrol.2009.07.047`.

Hrachowitz, M., Clark, M.P., 2017. HESS Opinions: The complementary merits of competing modelling philosophies in hydrology. Hydrol. Earth Syst. Sci. 21, 3953–3973. doi:`10.5194/hess-21-3953-2017`.

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

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., Fenicia, F., Freer, J., Gelfan, A., Gupta, H., Hughes, D., Hut, R., Montanari, A., Pande, S., Tetzlaff, D., Troch, P., Uhlenbrook, S., Wagener, T., Winsemius, H., Woods, R., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrol. Sci. J. 58, 1198–1255. doi:`10.1080/02626667.2013.803183`.

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. Hydrol. Earth Syst. Sci. 21, 3749–3775. doi:`10.5194/hess-21-3749-2017`.

Jarvis, N., Koestel, J., Larsbo, M., 2016. Understanding Preferential Flow in the Vadose Zone: Recent Advances and Future Prospects. Vadose Zo. J. 15, 160075. doi:`10.2136/vzj2016.09.0075`.

Jeannot, B., Weill, S., Eschbach, D., Schmitt, L., Delay, F., 2019. Assessing the effect of flood restoration on surface–subsurface interactions in Rohrschollen Island (Upper Rhine river – France) using integrated hydrological modeling and thermal infrared imaging. Hydrol. Earth Syst. Sci. 23, 239–254. doi:`10.5194/hess-23-239-2019`.

Juilleret, J., Dondeyne, S., Vancampenhout, K., Deckers, J., Hissler, C., 2016. Mind the gap: A classification system for integrating the subsolum into soil surveys. Geoderma 264, 332–339. doi:`10.1016/j.geoderma.2015.08.031`.

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. URL: `http://www.snl.lu/publications/bulletin/SNL_2011_112_125_130.pdf`.

Kampf, S.K., Burges, S.J., 2007. A framework for classifying and comparing distributed hillslope and catchment hydrologic models. Water Resour. Res. 43, W05423. doi:10.1029/2006WR005370.

Katsura, S., Kosugi, K., Yamakawa, Y., Mizuyama, T., 2014. Field evidence of groundwater ridging in a slope of a granite watershed without the capillary fringe effect. J. Hydrol. 511, 703–718. doi:10.1016/j.jhydrol.2014.02.021.

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 Resour. Res. 47, W05501. doi:10.1029/2010wr009525.

Kirchner, J.W., 2003. A double paradox in catchment hydrology and geochemistry. Hydrol. Process. 17, 871–874. doi:10.1002/hyp.5108.

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

Klaus, J., Wetzel, C.E., Martínez-Carreras, N., Ector, L., Pfister, L., 2015. A tracer to bridge the scales: On the value of diatoms for tracing fast flow path connectivity from headwaters to meso-scale catchments. Hydrol. Process. 29, 5275–5289. doi:10.1002/hyp.10628.

Koch, J., Jensen, K.H., Stisen, S., 2015. Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, Fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study. Water Resour. Res. 51, 1225–1246. doi:10.1002/2014WR016607.

Kollet, S., Sulis, M., Maxwell, R.M., Paniconi, C., Putti, M., Bertoldi, G., Coon, E.T., Cordano, E., Endrizzi, S., Kikinzon, E., Mouche, E., Mügler, C., Park, Y.J., Refsgaard, J.C., Stisen, S., Sudicky, E., 2017. The integrated hydrologic model intercomparison project, IH-MIP2: A second set of benchmark results to diagnose integrated hydrology and feedbacks. Water Resour. Res. 53, 867–890. doi:10.1002/2016WR019191.

Krein, A., Salvia-Castellvi, M., Iffly, J.F., Barnich, F., Matgen, P., Bos, R.V., Hoffmann, L., Kies, A., Pfister, L., 2006. Uncertainty in chemical hydrograph separation, in: Uncertainties in the 'monitoring-conceptualisation-modelling' sequence of catchment research, pp. 21–26. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.138.7065&rep=rep1&type=pdf#page=32.

Kristensen, K.J., Jensen, S.E., 1975. A model for estimating actual evapotranspiration from potential evapotranspiration. Nord. Hydrol. 6, 170–188. doi:10.2166/nh.1975.012.

Krzeminska, D.M., Bogaard, T.A., Malet, J.P., van Beek, L.P.H., 2013. A model of hydrological and mechanical feedbacks of preferential fissure flow in a slow-moving landslide. Hydrol. Earth Syst. Sci. 17, 947–959. doi:10.5194/hess-17-947-2013.

Kukemilks, K., Wagner, J.F., Saks, T., Brunner, P., 2018. Physically based hydrogeological and slope stability modeling of the Turaida castle mound. Landslides 15, 2267–2278. doi:10.1007/s10346-018-1038-5.

Kulasova, A., Beven, K.J., Blazkova, S.D., Rezacova, D., Cajthaml, J., 2014. Comparison of saturated areas mapping methods in the Jizera Mountains, Czech Republic. J. Hydrol. Hydromechanics 62, 160–168. doi:10.2478/johh-2014-0002.

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 Resour. Res. 50, 8159–8178. doi:10.1002/2014WR015381.

Latron, J., Gallart, F., 2007. Seasonal dynamics of runoff-contributing areas in a small mediterranean research catchment (Vallcebre, Eastern Pyrenees). J. Hydrol. 335, 194–206. doi:10.1016/j.jhydrol.2006.11.012.

Liu, H., Lin, H., 2015. Frequency and Control of Subsurface Preferential Flow: From Pedon to Catchment Scales. Soil Sci. Soc. Am. J. 79, 362–377. doi:10.2136/sssaj2014.08.0330.

Loague, K., 2010. Rainfall-Runoff Modelling, in: McDonnell, J.J. (Ed.), IAHS Benchmark Papers in Hydrology Series No. 4. IAHS Press, Wallingford, 506 pages.

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. J. Hydrol. 543, 255–269. doi:10.1016/j.jhydrol.2016.10.004.

Martínez-Carreras, N., Udelhoven, T., Krein, A., Gallart, F., Iffly, J.F., Ziebel, J., Hoffmann, L., Pfister, L., Walling, D.E., 2010. The use of sediment colour measured by diffuse reflectance spectrometry to determine sediment sources: Application to the Attert River catchment (Luxembourg). J. Hydrol. 382, 49–63. doi:10.1016/j.jhydrol.2009.12.017.

Martínez-Carreras, N., Wetzel, C.E., Frentress, J., Ector, L., McDonnell, J.J., Hoffmann, L., Pfister, L., 2015. Hydrological connectivity inferred from diatom transport through the riparian-stream system. Hydrol. Earth Syst. Sci. 19, 3133–3151. doi:10.5194/hess-19-3133-2015.

Matgen, P., El Idrissi, A., Henry, J.B., Tholey, N., Hoffmann, L., de Fraipont, P., Pfister, L., 2006. Patterns of remotely sensed floodplain saturation and its use in runoff predictions. Hydrol. Process. 20, 1805–1825. doi:10.1002/hyp.5963.

Maxwell, R.M., Putti, M., Meyerhoff, S., Delfs, J.O., Ferguson, I.M., Ivanov, V., Kim, J., Kolditz, O., Kollet, S.J., Kumar, M., Lopez, S., Niu, J., Paniconi, C., Park, Y.J., Phanikumar, M.S., Shen, C., Sudicky, E.A., Sulis, M., 2014. Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks. Water Resour. Res. 50, 1531–1549. doi:10.1002/2013WR013725.

McCabe, M.F., Rodell, M., Alsdorf, D.E., Miralles, D.G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N.E.C., Franz, T.E., Shi, J., Gao, H., Wood, E.F., 2017. The future of Earth observation in hydrology. Hydrol. Earth Syst. Sci. 21, 3879–3914. doi:10.5194/hess-21-3879-2017.

McDonnell, J.J., 2003. Where does water go when it rains? Moving beyond the variable source area concept of rainfall-runoff response. Hydrol. Process. 17, 1869–1875. doi:10.1002/hyp.5132.

McDonnell, J.J., McGuire, K., Aggarwal, P., Beven, K.J., Biondi, D., Destouni, G., Dunn, S., James, A., Kirchner, J., Kraft, P., Lyon, S., Maloszewski, P., Newman, B., Pfister, L., Rinaldo, A., Rodhe, A., Sayama, T., Seibert, J., Solomon, K., Soulsby, C., Stewart, M., Tetzlaff, D., Tobin, C., Troch, P., Weiler, M., Western, A., Wörman, A., Wrede, S., 2010. How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis. Hydrol. Process. 24, 1745–1754. doi:10.1002/hyp.7796.

Megahan, W.F., King, P.N., 1985. Identification of critical areas on forest lands for control of nonpoint sources of pollution. Environ. Manage. 9, 7–17. doi:10.1007/BF01871440.

Mengistu, S.G., Spence, C., 2016. Testing the ability of a semidistributed hydrological model to simulate contributing area. Water Resour. Res. 52, 4399–4415. doi:10.1002/2016WR018760.

Moragues-Quiroga, C., Juilleret, J., Gourdol, L., Pelt, E., Perrone, T., Aubert, A., Morvan, G., Chabaux, F., Legout, A., Stille, P., Hissler, C., 2017. Genesis and evolution of regoliths: Evidence from trace and major elements and Sr-Nd-Pb-U isotopes. Catena 149, 185–198. doi:10.1016/j.catena.2016.09.015.

Mosley, M.P., 1979. Streamflow generation in a forested watershed, New Zealand. Water Resour. Res. 15, 795–806. doi:10.1029/WR015i004p00795.

Niedda, M., Pirastru, M., 2014. Field investigation and modelling of coupled stream discharge and shallow water-table dynamics in a small Mediterranean catchment (Sardinia). Hydrol. Process. 28, 5423–5435. doi:10.1002/hyp.10016.

Nippgen, F., McGlynn, B.L., Emanuel, R.E., 2015. The spatial and temporal evolution of contributing areas. Water Resour. Res. 51, 4550–4573. doi:10.1002/2014WR016719.

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 Zo. J. 17, 170142. doi:10.2136/vzj2017.07.0142.

Paniconi, C., Putti, M., 2015. Physically based modeling in catchment hydrology at 50: Survey and outlook. Water Resour. Res. 51, 7090–7129. doi:10.1002/2015WR017780.

Partington, D., Brunner, P., Frei, S., Simmons, C.T., Werner, A.D., Therrien, R., Maier, H.R., Dandy, G.C., Fleckenstein, J.H., 2013. Interpreting streamflow generation mechanisms from integrated surface-subsurface flow models of a riparian wetland and catchment. Water Resour. Res. 49, 5501–5519. doi:10.1002/wrcr.20405.

Partington, D., Brunner, P., Simmons, C.T., Therrien, R., Werner, a.D., Dandy, G.C., Maier, H.R., 2011. A hydraulic mixing-cell method to quantify the groundwater component of streamflow within spatially distributed fully integrated surface water-groundwater flow models. Environ. Model. Softw. 26, 886–898. doi:10.1016/j.envsoft.2011.02.007.

Pfister, L., Martínez-Carreras, N., Hissler, C., Klaus, J., Carrer, G.E., Stewart, M.K., McDonnell, J.J., 2017. Bedrock geology controls on catchment storage, mixing, and release: A comparative analysis of 16 nested catchments. Hydrol. Process. 31, 1828–1845. doi:10.1002/hyp.11134.

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. Hydrol. Process. 24, 3123–3132. doi:10.1002/hyp.7840.

Pfister, L., McDonnell, J.J., Wrede, S., Hlúbiková, D., Matgen, P., Fenicia, F., Ector, L., Hoffmann, L., 2009. The rivers are alive: on the potential for diatoms as a tracer of water source and hydrological connectivity. Hydrol. Process. 23, 2841–2845. doi:10.1002/hyp.7426.

Reaney, S.M., Bracken, L.J., Kirkby, M.J., 2014. The importance of surface controls on overland flow connectivity in semi-arid environments: results from a numerical experimental approach. Hydrol. Process. 28, 2116–2128. doi:10.1002/hyp.9769.

Rode, M., Wade, A.J., Cohen, M.J., Hensley, R.T., Bowes, M.J., Kirchner, J.W., Arhonditsis, G.B., Jordan, P., Kronvang, B., Halliday, S.J., Skeffington, R.a., Rozemeijer, J.C., Aubert, A.H., Rinke, K., Jomaa, S., 2016. Sensors in the Stream: The High-Frequency Wave of the Present. Environ. Sci. Technol. 50, 10297–10307. doi:10.1021/acs.est.6b02155.

Rodriguez, N.B., Klaus, J., 2019. Catchment travel times from composite StorAge Selection functions representing the superposition of streamflow generation processes. Water Resour. Res. Under review, 2019WR024973R.

Scaini, A., Audebert, M., Hissler, C., Fenicia, F., Gourdol, L., Pfister, L., Beven, K.J., 2017. Velocity and celerity dynamics at plot scale inferred from artificial tracing experiments and time-lapse ERT. J. Hydrol. 546, 28–43. doi:10.1016/j.jhydrol.2016.12.035.

Scaini, A., Hissler, C., Fenicia, F., Juilleret, J., Iffly, J.F., Pfister, L., Beven, K., 2018. Hillslope response to sprinkling and natural rainfall using velocity and celerity estimates in a slate-bedrock catchment. J. Hydrol. 558, 366–379. doi:10.1016/j.jhydrol.2017.12.011.

van Schaik, N.L.M.B., Bronstert, A., de Jong, S.M., Jetten, V.G., van 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. Hydrol. Process. 28, 5805–5816. doi:10.1002/hyp.10086.

Schilling, O.S., Cook, P.G., Brunner, P., 2019. Beyond Classical Observations in Hydrogeology: The Advantages of Including Exchange Flux, Temperature, Tracer Concentration, Residence Time, and Soil Moisture Observations in Groundwater Model Calibration. Rev. Geophys. 57, 146–182. doi:10.1029/2018RG000619.

Schwab, M., Klaus, J., Pfister, L., Weiler, M., 2016. Diel discharge cycles explained through viscosity fluctuations in riparian inflow. Water Resour. Res. 52, 8744–8755. doi:10.1002/2016WR018626.

Schwab, M.P., Klaus, J., Pfister, L., Weiler, M., 2017. How runoff components affect the export of DOC and nitrate: a long-term and high-frequency analysis. Hydrol. Earth Syst. Sci. Discuss. doi:10.5194/hess-2017-416.

Schwab, M.P., Klaus, J., Pfister, L., Weiler, M., 2018. Diel fluctuations of viscosity-driven riparian inflow affect streamflow DOC concentration. Biogeosciences 15, 2177–2188. doi:10.5194/bg-15-2177-2018.

Sebben, M.L., Werner, A.D., Liggett, J.E., Partington, D., Simmons, C.T., 2013. On the testing of fully integrated surface – subsurface hydrological models. Hydrol. Process. 27, 1276–1285. doi:10.1002/hyp.9630.

Seibert, J., McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. Water Resour. Res. 38, 23–1–23–14. doi:10.1029/2001WR000978.

Silasari, R., Parajka, J., Ressl, C., Strauss, P., Blöschl, G., 2017. Potential of time-lapse photography for identifying saturation area dynamics on agricultural hillslopes. Hydrol. Process. 31, 3610–3627. doi:10.1002/hyp.11272.

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

Sklash, M.G., Farvolden, R.N., 1979. The role of groundwater in storm runoff. J. Hydrol. 43, 45–65. doi:10.1016/0022-1694(79)90164-1.

Spence, C., Mengistu, S.G., 2016. Deployment of an unmanned aerial system to assist in mapping an intermittent stream. Hydrol. Process. 30, 493–500. doi:10.1002/hyp.10597.

Spence, C., Mengistu, S.G., 2019. On the relationship between flood and contributing area. Hydrol. Process. 33, 1980–1992. doi:10.1002/hyp.13467.

Steinbrich, A., Leistert, H., Weiler, M., 2016. Model-based quantification of runoff generation processes at high spatial and temporal resolution. Environ. Earth Sci. 75, 1423. doi:10.1007/s12665-016-6234-9.

Troch, P.a., Carrillo, G.a., Heidbüchel, I., Rajagopal, S., Switanek, M., Volkmann, T.H.M., Yaeger, M., 2009. Dealing with Landscape Heterogeneity in Watershed Hydrology: A Review of Recent Progress toward New Hydrological Theory. Geogr. Compass 3, 375–392. doi:10.1111/j.1749-8198.2008.00186.x.

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

Villamizar, M., Brown, C., 2017. A modelling framework to simulate river flow and pesticide loss via preferential flow at the catchment scale. CATENA 149, 120–130. doi:10.1016/j.catena.2016.09.009.

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. J. Contam. Hydrol. 159, 57–66. doi:10.1016/j.jconhyd.2014.01.009.

Weiler, M., 2017. Macropores and preferential flow — a love-hate relationship. Hydrol. Process. 31, 15–19. doi:10.1002/hyp.11074.

Weiler, M., McDonnell, J., 2004. Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology. J. Hydrol. 285, 3–18. doi:10.1016/s0022-1694(03)00271-3.

Weiler, M., McDonnell, J.J., 2006. Testing nutrient flushing hypotheses at the hillslope scale: A virtual experiment approach. J. Hydrol. 319, 339–356. doi:10.1016/j.jhydrol.2005.06.040.

Weill, S., Altissimo, M., Cassiani, G., Deiana, R., Marani, M., Putti, M., 2013. Saturated area dynamics and streamflow generation from coupled surface-subsurface simulations and field observations. Adv. Water Resour. 59, 196–208. doi:10.1016/j.advwatres.2013.06.007.

Wiekenkamp, I., Huisman, J., Bogena, H., Lin, H., Vereecken, H., 2016. Spatial and temporal occurrence of preferential flow in a forested headwater catchment. J. Hydrol. 534, 139–149. doi:10.1016/j.jhydrol.2015.12.050.

Wilson, G.V., Rigby, J.R., Ursic, M., Dabney, S.M., 2016. Soil pipe flow tracer experiments: 1. Connectivity and transport characteristics. Hydrol. Process. 30, 1265–1279. doi:10.1002/hyp.10713.

Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., Savenije, H.H.G., Uhlenbrook, S., Kavetski, D., Pfister, L., 2015. Towards more systematic perceptual model development: a case study using 3 Luxembourgish catchments. Hydrol. Process. 29, 2731–2750. doi: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. J. Hydrol. 516, 97–106. doi:`10.1016/j.jhydrol.2014.02.055`.

Zimmermann, B., Zimmermann, A., Turner, B.L., Francke, T., Elsenbeer, H., 2014. Connectivity of overland flow by drainage network expansion in a rain forest catchment. Water Resour. Res. 50, 1457–1473. doi:`10.1002/2012WR012660`.

Study 1: How meaningful are plot-scale observations and simulations of preferential flow for catchment models?

Authors: Barbara Glaser, Conrad Jackisch, Luisa Hopp, Julian Klaus

BG, LH and JK designed and directed the study. CJ and field engineers from LIST provided the field data. BG performed the simulation and the data analysis. BG, CJ, LH and JK discussed and interpreted the results. BG prepared the figures and tables. BG prepared the manuscript with input from all co-authors. BG is the corresponding author.

Own contribution in %:

- Study concept and design:	65
- Field data acquisition:	0
- Simulation:	100
- Data analysis, figures, and tables:	100
- Interpretation of the results:	70
- Preparation of the manuscript:	80

Core Ideas

- **We tested a range of dual-permeability parameterizations at plot and catchment scale.**
- **Well-performing parameters at plot scale did not clearly improve catchment simulation.**
- **Vertical preferential flow was important for simulating plot-scale observations.**
- **At catchment scale, it appeared more important to consider fast lateral subsurface flow.**
- **This showed that different nonuniform flow processes are critical at different scales.**

B. Glaser and J. Klaus, Catchment and Eco-Hydrology Research Group, Luxembourg Institute of Science and Technology, L-4362 Esch/Alzette, Luxembourg; B. Glaser and L. Hopp, Dep. of Hydrology, Univ. of Bayreuth, 95447 Bayreuth, Germany; C. Jackisch, Chair of Hydrology, Institute of Water Resources and River Basin Management, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany. C. Jackisch currently at: Dep. of Landscape Ecology and Environmental Systems Analysis, Institute of Geoecology, Technische Universität Braunschweig, Langer Kamp 19c, 38106 Braunschweig, Germany. *Corresponding author (barbara.glaser@list.lu).

How Meaningful are Plot-Scale Observations and Simulations of Preferential Flow for Catchment Models?

Barbara Glaser,* Conrad Jackisch, Luisa Hopp, and Julian Klaus

Despite ubiquitous field observations of nonuniform flow processes, preferential flow paths are rarely considered in hydrological models, especially at catchment scale. In this study, we investigated the extent to which plot-scale observations of preferential flow paths are informative for rainfall–runoff simulations at larger scales. We used data from three plot-scale irrigation experiments in the Weierbach catchment (Luxembourg) to identify preferential flow parameterizations via a Monte Carlo simulation with HydroGeoSphere. Subsequently, we tested whether these parameter sets could be used directly to simulate the hydrological response of the Weierbach headwater with a HydroGeoSphere catchment model. The Monte Carlo simulations showed that the different depth profiles of Br⁻ tracer observed in irrigation experiments could be reproduced when vertical preferential flow was simulated with a dual-permeability approach. However, it was not possible to identify unique parameter values for preferential flow. The direct transfer of a range of different dual-permeability parameter sets to the catchment model revealed that the variability of simulated hydrometric catchment responses (discharge and soil moisture over 18 mo) was independent of the variability among the three irrigation experiments. More importantly, the dual-permeability approach did not improve the match between simulated and observed discharge and soil moisture responses compared with the single-domain reference model, where multiple soil layers with differing hydraulic conductivities had already been implemented. This suggests that including structures that allow nonuniform lateral flow was more important for reproducing the hydrological response in the Weierbach catchment than the vertical preferential flow observed at plot scale.

Abbreviations: 3D, three-dimensional; MC, Monte Carlo; NSE, Nash–Sutcliffe efficiency.

It is often criticized that the majority of hydrological models neglect or poorly represent preferential flow processes and therefore miss an important feature of subsurface processes (e.g., Beven and Germann, 2013; Weiler, 2017). At the same time, the integrated effect of preferential flow paths on catchment response remains under discussion (Beven and Germann, 2013; Weiler, 2017). To date, many experimental studies at plot (<5 m^2) and hillslope (5 m^2–1 ha) scales have shown that vertical and lateral preferential flow have a crucial impact on the timing and quantity of water flow and solute transport (e.g., Vogel et al., 2006; Rosenbom et al., 2009; Anderson et al., 2009; van Schaik et al., 2010; Klaus et al., 2013, 2014; Laine-Kaulio et al., 2014; Jackisch et al., 2017; Scaini et al., 2017). Yet, direct observations of preferential flow pathways at catchment scale (>1 ha) are rather scarce. Recently, Wilson et al. (2016) observed that networks of large soil pipes can effectively connect the hillslope areas with the catchment outlet via lateral preferential flow. Other work at catchment scale has relied on soil moisture sensor networks (Liu and Lin, 2015; Wiekenkamp et al., 2016) to analyze the spatiotemporal occurrence of preferential flow across two forested catchments (7.9 ha and 38.5 ha, respectively). These two studies demonstrated that the occurrence of vertical preferential flow was highly variable in space across the two catchments investigated. Based on their field results, Liu and Lin (2015) identified a "hidden" preferential flow network in the subsurface. However, such

an experimental setup does not allow the relevance of small-scale variabilities in preferential flow patterns and in local subsurface flow networks to be analyzed for the rainfall–runoff response at the catchment outlet.

Modeling approaches can help to bridge the gap between observations of preferential flow at plot and hillslope scale and the understanding of the effect of preferential flow on integrated catchment responses. However, as is the case for the experimental investigations, there is a discrepancy between the number of simulations that include preferential flow at plot and hillslope scale (e.g., Weiler and McDonnell, 2007; Klaus and Zehe, 2010, 2011; Laine-Kaulio et al., 2014; Frey et al., 2016; Kukemilks et al., 2018a; Reck et al., 2018) and the number of simulations that include preferential flow at catchment scale (e.g., Krzeminska et al., 2013; Steinbrich et al., 2016; Villamizar and Brown, 2017; Kukemilks et al., 2018b). Consequently, the number of modeling studies that have explicitly analyzed the effect of preferential flow on catchment response by comparing simulations with and without preferential flow is limited (Beckers and Alila, 2004; Christiansen et al., 2004; Zhang et al., 2006; van Schaik et al., 2014; Yu et al., 2014; De Schepper et al., 2015). An increased number of such studies could be a valuable asset for the currently limited understanding of the circumstances, processes, and degree to which preferential flow has a relevant effect on integrated catchment response and thus requires explicit parameterization and simulation.

In practical terms, two main challenges hinder preferential flow concepts from being included more often in distributed, physically based catchment-scale simulations: (i) finding a proper mathematical process description that is adequate for the scale, and (ii) finding adequate model parameters for the process description (cf. Beven and Germann, 2013; Jarvis et al., 2016). While numerous plot-scale studies have focused on advancing accurate descriptions of flow and exchange processes within explicitly implemented macropores and fractures (e.g., Vogel et al., 2006; Scheibe et al., 2015; Jackisch and Zehe, 2018), explicit implementations of discrete macropore and fracture geometries for an entire catchment are (currently) not feasible from a computational and parameterization point of view (Jarvis et al., 2016). Instead, the most common representations of preferential flow processes in existing catchment simulations are dual-domain approaches, which separate the subsurface into two interacting matrix and preferential flow domains with differing hydraulic properties (e.g., Kordilla et al., 2012; van Schaik et al., 2014; Wang et al., 2014; Yu et al., 2014; De Schepper et al., 2015; Jarvis et al., 2016; Steinbrich et al., 2016; Villamizar and Brown, 2017). The physical adequacy of the most commonly used equations within the dual-domain approach (e.g., Darcy equation, Richards equation, Green-Ampt infiltration) for catchment scale has been debated (cf. Beven and Germann, 2013; Jarvis et al., 2016). Aside from this fact, the implicit representation of the subsurface structure with two coexisting bulk domains is a strong simplification of the real preferential flow network. Yet, many of the studies that included preferential flow processes for simulating catchment responses have simplified the spatial representation of

the catchment even more, e.g., by proportionally combining the outputs of several representative one-dimensional dual-permeability simulations to a catchment response (Wang et al., 2014; Villamizar and Brown, 2017), performing a two-dimensional dual-permeability simulation for a representative cross-section through the catchment (e.g., Kordilla et al., 2012), or using spatially distributed three-dimensional (3D) dual-permeability simulations with lumped formulations for lateral (Krzeminska et al., 2013) or groundwater flow (van Schaik et al., 2014). The question of which spatial simplifications of preferential flow are necessary and suitable for appropriate catchment-scale simulations with acceptable computational costs remains open.

Regardless of the representation chosen for preferential flow, the identification of an adequate parameterization to simulate preferential flow is challenging (Beven and Germann, 2013). Most existing catchment modeling approaches rely on parameter calibration or a parameterization based on literature values. Only a few studies have parameterized preferential flow (at least partly) based on measurements of study site characteristics such as saturated hydraulic conductivity, water retention curve, crack density, and fracture aperture (e.g., Kordilla et al., 2012; Steinbrich et al., 2016; Loritz et al., 2017; Kukemilks et al., 2018b). The problem of parameter calibration in physically based models is that (i) the optimized parameters may not capture the real physics of the system, (ii) parameters may not uniquely converge due to equifinality, and (iii) calibration may require a large number of model runs, with the latter being mainly restrictive for physically based, distributed, 3D catchment simulations with long computational times. Therefore, the calibration of physically based 3D catchment models including preferential flow barely exists today (exception, Yu et al., 2014). A direct parameterization based on field observations is challenging because measurements can only capture local characteristics of a catchment and it is hardly feasible to perform sufficient measurements to fully assess the heterogeneity within a catchment.

A promising approach for identifying an observation-based parameterization of preferential flow for catchment simulations is to derive parameters in simulations of detailed plot-scale observations and to then use these parameters for informing catchment-scale simulations (e.g., Vogel and Roth, 2003; Beven and Germann, 2013; Cadini et al., 2013). Previous studies have realized such an upscaling of preferential flow simulations from plot to hillslope or catchment scale but have lacked validation of the plot- and catchment-scale simulations against field data (Cadini et al., 2013; Wang et al., 2014). Van Schaik et al. (2010) used data from plot-scale irrigation experiments to parameterize three soil profile models. They used these parameterized models to simulate plot-scale water balance and eventually compared this to the water balance observed at catchment scale. However, later work by van Schaik et al. (2014) relied on a different model requiring different parameters for 3D simulations of their catchment. For all these studies, it is difficult to fully assess the value of detailed simulations at plot scale for parameterizing catchment simulations. Moreover, the question of how spatial heterogeneity within

a catchment affects the simulation results at catchment scale in such an approach remains open: Does it matter how representative the used plot-scale observation of preferential flow is for the entire catchment? Or does the effect of locally different characteristics "smooth out" with scale, as observed, for example, for initial saturation (Zehe and Blöschl, 2004) and saturated hydraulic conductivities (Meyerhoff and Maxwell, 2011)?

We conducted a study on the transferability of plot-scale-derived preferential flow parameterization to catchment rainfall–runoff simulations for the Weierbach headwater (Luxembourg). We parameterized the preferential flow based on plot-scale observations during irrigation experiments and subsequently transferred the parameters to catchment-scale simulations. We used the integrated hydrological surface–subsurface model HydroGeoSphere (Therrien et al., 2010), which allowed us to use the same preferential flow representation (dual-permeability approach) at plot and catchment scales. By doing this, we hypothesized that model parameters and processes are scale invariant and that a direct extension of the model from plot to catchment scale is possible. We constrained the plot-scale parameters with observations from three irrigation experiments (dye tracer patterns, Br$^-$ concentration profiles). We then tested the transfer of several preferential flow parameter sets from plot scale to the catchment-scale model and validated the catchment simulation results with discharge and soil moisture observations. The aim of this approach was to assess (i) the value of plot-scale irrigation experiments for identifying parameter sets for a realistic dual-permeability simulation at catchment scale, (ii) the effect of the spatial heterogeneity of the occurrence and prominence of vertical preferential flow (and thus the importance of the representativeness of the used plot-scale observation) on the parameter transfer, and (iii) the spatial (vertical and lateral) and conceptual representation of preferential flow with a dual-permeability approach for capturing the integrated signals of a humid-temperate catchment in long-term simulations.

Please note that, when referring to our simulations, we use the term *preferential flow* for describing the parameterization of nonuniform flow with a dual-permeability approach, focusing on the vertical preferential flow component that is introduced with this approach. In addition, the model setup consists of multiple soil layers with contrasting hydraulic conductivities, which enable a development of nonuniform lateral subsurface flow. This lateral flow can also be interpreted as preferential flow but is referred to here as fast or nonuniform lateral (subsurface) flow.

Study Site and Previous Work

The simulation of preferential flow is based on field investigations that were conducted in and near the Weierbach catchment, a forested 42-ha experimental catchment in the foothills of the Ardennes massif in the west of Luxembourg (Fig. 1). The area is characterized by shallow soils overlying periglacial deposit layers and Devonian slate bedrock (Juilleret et al., 2011; Moragues-Quiroga et al., 2017). Catchment runoff shows

distinct differences between dry and wet catchment states, with single-peak hydrographs during dry and double-peak hydrographs (Martínez-Carreras et al., 2016) under wet conditions. The processes contributing to the distinct streamflow behavior have been under investigation for many years (Fenicia et al., 2014; Wrede et al., 2015; Klaus et al., 2015; Martínez-Carreras et al., 2016; Schwab et al., 2017). The common perception of these studies mainly relates the sharp, short-lasting (single-peak) hydrographs to flow paths on the surface and in the shallow (top)soil and the second, delayed peaks to the exceedance of storage thresholds, connecting deeper (ground)water.

Several studies around the Weierbach have investigated preferential flow. Irrigation experiments by Jackisch et al. (2017) on three 1-m^2 plots in the direct vicinity of the catchment (same vegetation and pedolithological structure, Fig. 1) showed fast vertical infiltration and preferential flow through a network of interaggregate pores. Scaini et al. (2017, 2018) observed that vertical subsurface flow and vertical preferential pathways dominated fluxes in the top 2 to 3 m during two irrigation experiments on a 64-m^2 hillslope section in the Weierbach catchment. In addition, Scaini et al. (2017, 2018) found indications of a fast connection between hillslope and stream, which they related to a fast lateral flow in the fractured bedrock 2 to 3 m below the surface. In a nearby catchment with the same pedolithological structure, Angermann et al. (2017) demonstrated for a hillslope that preferential flow paths developed in unsaturated soils shortly after the onset of intense irrigation or precipitation. Based on their investigations, they inferred fast lateral preferential flow through distinct paths at the interface of the periglacial deposit layer as one of the main processes in the hillslope, potentially contributing to the sharp, short-lasting (single) hydrograph peaks. In line with the field investigations, Glaser et al. (2016) suggested that incorporating preferential flow formulations could help to improve the performance of their physically based 3D single-domain HydroGeoSphere model of the 6-ha headwater of the Weierbach (Fig. 1). As highlighted in Fig. 1, their simulations notably missed some specific hydrograph responses of the broad, delayed peaks.

In this study, we relied on the 6-ha headwater HydroGeoSphere model of Glaser et al. (2016) and the information derived from the three 1-m^2 plot irrigation experiments of Jackisch et al. (2017) as the basis for the parameterization of preferential flow simulations at plot and catchment scale (see below). In addition, soil moisture was monitored from October 2012 to April 2014 on a hillslope in the catchment (Fig. 1) with time-domain reflectometers (Campbell CS650) installed horizontally at the 10-, 20-, 40-, and 60-cm depths. The stream level was measured at the outlet of the catchment (ISCO 4120 Flow Logger) and transformed into discharge via a rating curve. Additional discharge measurements (ISCO 4120 Flow Logger, transformation via a rating curve) were performed at the outlet of the upper 6-ha headwater region of the catchment (Fig. 1) starting in spring 2013.

The irrigation experiments were described in detail by Jackisch et al. (2017). Here, we briefly summarize the information relevant for
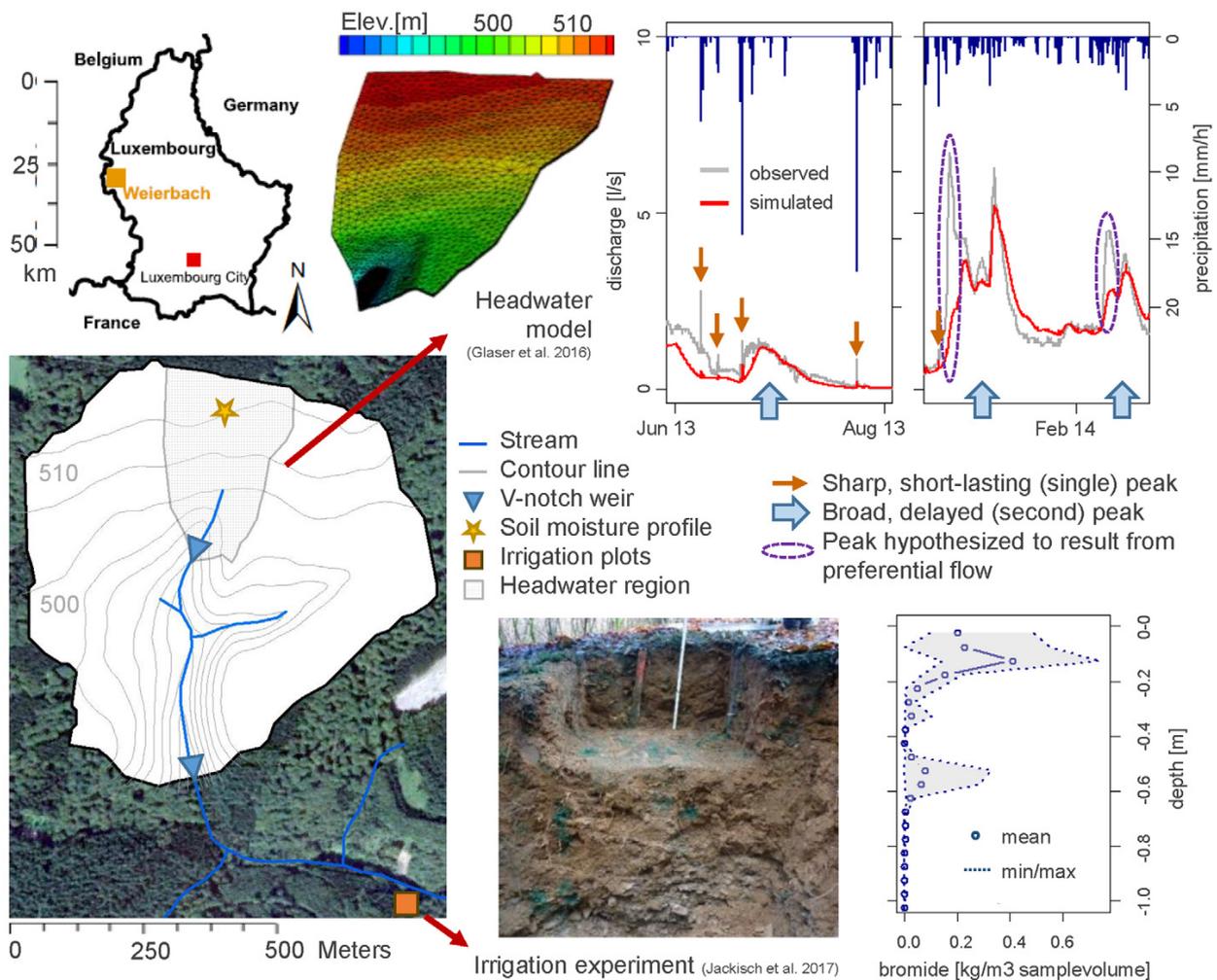
Fig. 1. Weierbach catchment and surroundings (left), model grid and hydrographs of the 6-ha headwater model from Glaser et al. (2016) (top right), and example excavation plot and Br⁻ depth profile of the irrigation experiments from Jackisch et al. (2017) (bottom right).

this study. Three 1-m² plots were irrigated for 1 h with 30 mm (Plot 2) or 50 mm (Plots 1 and 3) of water mixed with Br^- ($5\,g\,L^{-1}$ KBr) and the dye tracer Brilliant Blue ($4\,g\,L^{-1}$). At each plot, soil moisture was continuously measured with a time-domain reflectometry tube probe (IMKO IPH/T3) at 10-cm depth increments down to a depth of 1.2 m. Each plot was excavated 24 h after irrigation and analyzed for Brilliant Blue patterns (three vertical faces, five to seven horizontal faces) and vertical profiles of Br^- recovery (five vertical profiles with horizontal and vertical sample spacing of 5 cm down to a maximum depth of 1 m). The observed soil moisture responses, Brilliant Blue patterns, and Br^- concentrations consistently indicated a fast and pronounced nonuniform vertical transport down to the upper boundary of the periglacial deposit layer (starting at a depth of approximately 0.6 m, Fig. 1). Apart from that, the monitored soil moisture depth profiles and vertical patterns of dye and Br^- showed a high spatial heterogeneity within and between the different irrigated plots (Jackisch et al., 2017).

Modeling Approach
HydroGeoSphere model

HydroGeoSphere (Aquanty, 2015) is a 3D integrated hydrological surface—subsurface model code that can simultaneously solve a modified form of the Richards equation for transient flow in the subsurface domain and the diffusion wave approximation of the two-dimensional Saint-Venant equation for flow in the surface domain. Solute transport can be implemented with an advection–dispersion equation. Preferential flow can be considered via flow in discrete fractures or via a dual-permeability approach (Therrien et al., 2010; Aquanty, 2015).

In this study, we incorporated preferential flow with the dual-permeability approach. In this approach, two subsurface continua coexist. One continuum represents water flow and solute transport in the soil matrix (hereafter called the matrix domain); the other continuum represents flow and solute transport in the macropores (hereafter called the macropore domain). The two domains are

volumetrically partitioned on the total subsurface volume and parameterized independently with different soil hydraulic parameters (used here: saturated hydraulic conductivity K_s, porosity n, residual saturation θ_r, van Genuchten α, and van Genuchten β). Exchange between the matrix and macropore domains follows a dual-node approach, i.e., the interface between the two domains is represented as a very thin layer of porous material through which a Darcy flux is driven by respective hydraulic head differences between the two domains. Consequently, using the dual-permeability approach requires the definition of soil hydraulic parameters for the matrix and macropore domains as well as for the exchange interface. Additionally, fluid and mass exchange parameters for the exchange interface (fex, mex) and a parameter defining the macropore domain percentage on the total subsurface volume (pct, summing up to 100% with the matrix domain percentage) are needed.

The plot-scale and catchment-scale simulations of this study build on a modified HydroGeoSphere model of the 6-ha headwater region of the Weierbach as described by Glaser et al. (2016). They used the model to simulate coupled surface and subsurface flow from October 2010 to August 2014 with hourly precipitation and potential evapotranspiration forcing. A nested model grid (area of 6 ha, depth of 3 m) 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 10 m (hillslope) to 0.25 m and less (riparian zone and stream bed) (Fig. 2, left). Eleven matrix domain zones were parameterized in the grid with differing hydraulic characteristics, representing a humic, dystric, skeletic, and siltic Cambisol at the hillslopes (Ah, B1, B2), a stagnic soil in the riparian zone (LP), and universally underlying layers of transition from subsoil to regolith (IIC), weathered bedrock (Cv), and solid slate (mC) (Fig. 2, middle). The matrix domain zones were implemented as laterally homogeneous layers all over the catchment with the exception of the outcropping of the soil layers and the overlying stagnic soil in the riparian zone.

Model parameterization relied on field observations, including electrical resistivity tomography and measurements of soil hydraulic parameters from soil samples, literature values, and trial and error calibration of porosities, hydraulic conductivities, and evapotranspiration parameters. The measurements used for parameterization did not explicitly exclude macroporous structures. The highly saturated hydraulic conductivities and the porosities of the soil and regolith layers (Table 1, Soil Zones 1–6) suggest that a macropore influence was already implicitly included in the matrix domain parameterization. This reflection was considered in the following dual-permeability parameterization, where we explicitly implemented vertical preferential flow by distinguishing between soil matrix and macropores. To incorporate the dual-permeability approach, we adapted the previous model as described below and simulated the plot-scale irrigation experiments and the hydrological catchment response for a period of 18 mo (October 2012–April 2014). For additional information on the basic model setup (e.g., evapotranspiration parameters, numerical controls), see Glaser et al. (2016).

Plot-Scale Model Setup

We simulated the three plot-scale irrigation experiments of Jackisch et al. (2017) in a horizontal 1-m² soil column of 6-m depth implemented in HydroGeoSphere. The grid was defined by 0.25-m² quadratic elements with element heights of 1 cm between the 0- and 4-m depths and element heights of 5 cm between the 4- and 6-m depths. Flow and transport were simulated in the subsurface only (no surface domain) because this can avoid numerical problems, and no surface runoff was observed during the experiments. We set up the matrix domain of the model column identically to the hillslope structure of Glaser et al. (2016), i.e., the soil type depth profile and parameterization of the hillslopes (Soil Zones 1–10; cf. Fig. 2, Table 1) was applied to the upper 4 m of the model column (Fig. 2). The lower 2 m of the model column (4–6 m) served as porous storage ($K_s = 1$ m d^{-1}, $n = 20\%$, $\theta_r = 0.02$,
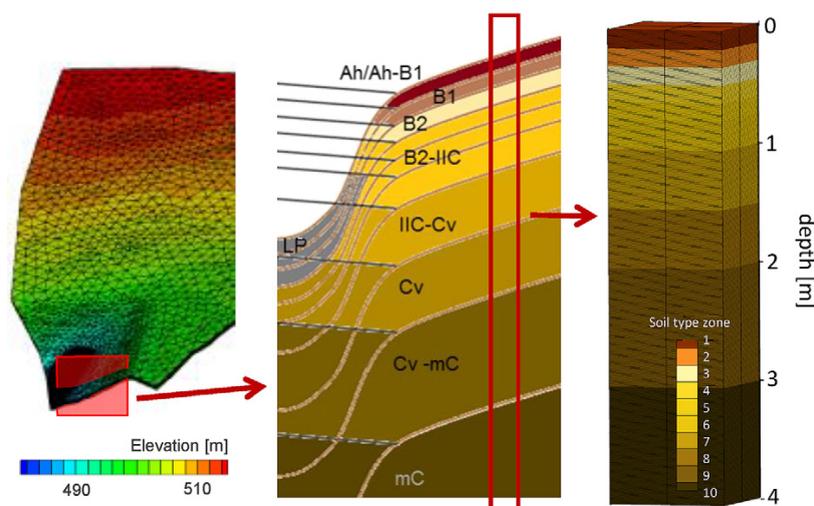

Fig. 2. Schematic cross-section through the subsurface setup showing the soil zones (middle) as defined in the headwater model grid of Glaser et al. (2016) (left) and in the plot-scale model grid (right) used for simulating the irrigation experiments. The cross-section has been modified from Glaser et al. (2016).

Table 1. Soil hydraulic parameters from the single-domain model of Glaser et al. (2016) including residual saturation (θ_r), van Genuchten parameters α and β, porosity (n), and saturated hydraulic conductivity (K_s). In this study, these values were used for the parameters of the matrix domain or as effective model parameters (for K_s and n only), respectively. (Table modified from Glaser et al., 2016).

Soil zone	θ_r	α	β	n	K_s
		m^{-1}			m d^{-1}
1 (Ah/Ah-B1)	0.12	6.6	1.46	0.74	1.71×10^1
2 (B1)	0.10	22.1	1.42	0.61	1.71×10^1
3 (B2)	0.10	22.1	1.42	0.45	4.59×10^1
4 (B2-IIC)	0.10	22.1	1.42	0.30	9.30×10^2
5 (B2-IIC)	0.10	22.1	1.42	0.15	2.04×10^3
6 (B2-IIC)	0.02	6.0	1.50	0.20	8.40×10^2
7 (IIC-Cv)	0.02	6.0	1.50	0.15	3.00×10^0
8 (Cv)	0.02	6.0	1.50	0.10	1.20×10^{-2}
9 (Cv-mC)	0.02	6.0	1.50	0.05	9.00×10^{-4}
10 (mC)	0.02	6.0	1.50	0.01	2.40×10^{-5}
11(LP)	0.10	22.1	1.42	0.61	7.80×10^0

α = 6 m^{-1}, β = 1.5) to prevent water from ponding at the bottom of the upper 4 m.

Preferential flow was incorporated in the upper 4 m by defining a macropore domain with the same soil zone layering as for the matrix domain. This required preferential flow parameters (K_s, n, θ_r, α, β, fex, mex, and pct) for 10 different soil type zones (cf. soil type zones of the matrix domain at the hillslopes, Fig. 2). These parameters were explored within a Monte Carlo approach (see below). Bromide transport was simulated with a tortuosity of 0.1, a diffusion coefficient of 1.6×10^{-4} m^2 d^{-1}, and longitudinal and transverse dispersivities of d_l = 0.05 m and d_t = 0.005 m for the matrix and d_l = 0.1m and d_t = 0.01 m for the macropore domain (cf. Gelhar et al., 1992; Rosenbom et al., 2009; Leistra and Boesten, 2010; Laine-Kaulio et al., 2014).

Following Laine-Kaulio et al. (2014), 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. We assigned no-flow boundaries to the sides and the bottom of the model column. Initial saturation was identical for the matrix and macropore domains (see Supplemental Material S1 for values); the initial Br$^-$ concentration was set to zero (10^{-15} kg m^{-3} to avoid numerical instabilities) in both domains.

Plot-Scale Simulation of Preferential Flow with a Monte Carlo Approach

We performed 20,000 brute-force Monte Carlo (MC) simulations for identifying the best-performing parameter sets for the three irrigation experiments. A total of 12,000 MC runs were simulated with an irrigation rate of 50 mm h^{-1} (representing Plots 1 and 3) and 8000 simulations were performed with an irrigation

rate of 30 mm h^{-1} (representing Plot 2). The MC runs differed in the parameterization of preferential flow (parameters of the macropore domain and parameters for defining the partitioning and exchange between the matrix and macropore domain, cf. Table 2). Transport and matrix parameter values (Table 1) were kept constant with the exception of saturated hydraulic conductivity and porosity of the matrix domain. These two parameters were adapted in half of the MC simulations to account for the assumption that values used in the single-domain model implicitly included the effect of preferential flow (cf. high values in Table 1).

In total, 12 preferential flow parameters were needed for each of the 10 different soil type zones. Consequently, independent value variations for each parameter and soil layer would have resulted in 120 values to be modified per model run. To keep a reasonable number of parameter variations, we varied only 8 of the 12 preferential flow parameters. Precisely, the values of the hydraulic parameters θ_r, α, and β of the interface between the macropore and matrix domain and the residual saturation θ_r of the macropore domain were not varied in the MC runs (Table 2, no value variation). In addition, we reduced the number of necessary parameter variations by assigning a global value for all 10 soil layers for some parameters (fex, mex, and n) (Table 2, global values in depth profile).

The parameters α, β, K_s, and pct were subject to separate value assignment for different soil layers (Table 2, layer-specific values in depth profile). To avoid unrealistic hydraulic and capillary jumps and barriers, the value assignment was constrained to the form (value ratio) of several predefined nonuniform depth profiles. To define realistic depth profiles of parameters [hereafter labeled dp$_{xxx}$(z)], we applied different approaches (hereafter labeled DP$_{xxx}$) as multiple working hypotheses (Clark et al., 2011) that rely on field observations and their different interpretations (Table 2, Supplemental Table S1). Details on the different approaches DP$_{xxx}$ and the pre-defined depth profiles dp$_{xxx}$(z) obtained are described in Supplemental Material S1.

Briefly, different van Genuchten α and β values were assigned to the different soil layers based on two different macropore categories: biopores (b) and fractures (f). The macropore categories b and f were assigned to the different soil layers based on the characteristics observed in the excavated irrigation plots (DP$_{cat}$, dp$_{cat}$; Table 2, Supplemental Table S1). The depth profile dp$_{pct}$ for the macropore percentage (pct) was parameterized with four different approaches DP$_{pct}$ A to D (Table 2). The approaches DP$_{pct}$ A and DP$_{pct}$ B represented two different ways of interpreting the Brilliant Blue patterns observed at the irrigation plots, approach DP$_{pct}$ C assigned a constant percentage with depth, and approach DP$_{pct}$ D relied on results from preliminary test simulations (Supplemental Material S1, Supplemental Table S1). The saturated hydraulic conductivity of the macropores (K_s) was assigned separately to the different soil layers following two different approaches (DP$_{K_s}$ A and B). The K_s value was calculated either based on macropore apertures (apert, which required the definition of macropore aperture depth profiles dp$_{apert}$, DP$_{K_s}$ A, Table 2, Supplemental Table

Table 2. Definition and constraints of macropore and interface parameters in the MC simulations. Parameters were assigned globally or separately to different soil layers z. Variations of separately assigned values were dependent on several fixed depth profiles that were pre-defined with various approaches DP$_{xxx}$ A to D, partly requiring some additional adjunct parameters. The resulting predefined depth profiles dp$_{xxx}$ are shown in Supplemental Table S1. Here, information is given on how the depth profiles and adjunct parameters were used for the definition of parameter values. Parameter values were modified either by randomly sampling them from a uniform distribution U(a,b) or based on a randomly sampled modification factor F$_{xxx}$.

Parameter†	Value variation	Global or layer-specific values in soil layers z‡	Depth profile predefinition method§	Adjunct parameters	Parameter calculation¶	Value constraint¶	Rationale for constraint range
Macropore domain							
n	yes	g				$n \sim U(0.75,0.99)$	ensure $n > n_{\text{original}}$
θ_r	no	g				0.01	
$\alpha(z)$ (m⁻¹)	yes	l	DP$_{\text{cat}}$	$\text{dp}_{\text{cat}}(z) \rightarrow \{b,f\}$		$\alpha_b \sim U(2.5,30.0)$; $\alpha_f \sim U(1.0,10.0)$	literature (Köhne et al., 2002; Rosenbom et al., 2009; Kordilla et al., 2012; Laine-Kaulio et al., 2014)
$\beta(z)$	yes	l	DP$_{\text{cat}}$	$\text{dp}_{\text{cat}}(z) \rightarrow \{b,f\}$		$\beta_b \sim U(1.1,5.0)$; $\beta_f \sim U(1.1,5.0)$	
pct(z)	yes	l	DP$_{\text{pct}}$ A–D		$F_{\text{pct}} \times \text{dp}_{\text{pct}}(z)$, $F_{\text{pct}} = 1/X$ or $F_{\text{pct}} = X$	$X \sim U(1.0,2.5)$	ensure $0 < \text{pct} < 0.95$
$K_s(z)$ (m d⁻¹)	yes	l	DP$_{Ks}$ A–B, DP$_{Ks}$ A	apert(z) (m), DP$_{\text{apert}}$ A–D	$\text{apert}(z)^2 \times 9.81/12 \times 86{,}400 \times 10^6$; $\text{apert}(z) = 10^{F_{\text{apert}}} \times \text{dp}_{\text{apert}}(z)$	$F_{\text{apert}} \sim U(-2,1)$	arbitrary, ensure $10^{-7} < \text{apert} < 0.05$
			DP$_{Ks}$ B	$K_{s\text{-matrix}}(z)$ (m d⁻¹)	$\{K_{s\text{-original}}(z) - [1 - \text{pct}(z)] \times K_{s\text{-matrix}}(z)\}/\text{pct}(z)$; $K_{s\text{-matrix}}(z) = 10^{F_{\text{matKs}}} \times K_{s\text{-original}}(z)$	$F_{\text{matKs}} \sim U(-4,0)$	$K_{s\text{-matrix}} < K_{s\text{-original}}$, lower boundary arbitrary
Interface							
fex (m⁻²)	yes	g			$10^{F_{\text{fex}}}$	$F_{\text{fex}} \sim U(-5,2)$	arbitrary
mex (d⁻¹)	yes	g			$10^{F_{\text{mex}}}$	$F_{\text{mex}} \sim U(-3,4)$	arbitrary
$K_s(z)$ (m d⁻¹)	yes	l			$10^{F_{\text{intKs}}} \times K_{s\text{-original}}(z)$	$F_{\text{intKs}} \sim U(-3,3)$	arbitrary
$\alpha(z)$ (m⁻¹)	no	l			$\alpha_{\text{original}}(z)$		
$\beta(z)$	no	l			$\beta_{\text{original}}(z)$		
$\theta_r(z)$	no	l			$\theta_{r\,\text{original}}(z)$		

† n, porosity; θ_r, residual saturation; α, β, van Genuchten parameters; K_s, saturated hydraulic conductivity; pct, percentage of total subsurface domain comprised of macropores; fex, flow exchange coefficient; mex, mass exchange coefficient

‡ g, global; l, layer specific.

§ DP$_{\text{cat}}$, method for defining a depth profile dp$_{\text{cat}}(z)$ of macropore types (biopores b or fractures f) of macropore depth profile dp$_{\text{cat}}(z)$, which were subject to separate value assignments for α and β; DP$_{\text{pct}}$, four different methods A–D for predefining different depth profiles dp$_{\text{pct}}(z)$ for pct; DP$_{Ks}$, two methods A and B for defining $K_s(z)$; A relies on predefined depth profiles dp$_{\text{apert}}(z)$ of macropore apertures apert(z) (following four predefinition methods DP$_{\text{apert}}$ A–D), B relies on maintaining effective hydraulic conductivities as $K_{s\text{-original}}(z)$, thus requiring adapted matrix conductivities $K_{s\text{-matrix}}(z)$.

¶ X_{original}, matrix parameter in the headwater model of Glaser et al. (2016) (cf. Table 1).

S1) or as a function of varied matrix conductivities $K_{s\text{-matrix}}$, maintaining the same effective hydraulic conductivities as originally assigned to the matrix domain (cf. Table 1), but via a different contribution of the matrix and macropores to the effective hydraulic conductivity. In the latter case (DP_{K_s} B, Table 2, Supplemental Material S1), matrix porosities were also adapted in such a way that effective porosities equaled the porosity values as originally assigned to the matrix domain (Table 1).

Ultimately, we varied eight macropore parameters in the MC simulation (Table 2, value variation). In half of the MC runs, we additionally varied matrix conductivities and porosities (see above and Supplemental Material S1). Some of the value variations were conditioned by variations of adjunct parameters (macropore category b or f, macropore aperture apert, matrix hydraulic conductivity $K_{s\text{-matrix}}$) and different methods (DP_{cat}, DP_{pct} A–D, DP_{K_s} A–B, DP_{apert} A–D) for defining parameter depth profiles $dp_{xxx}(z)$ (see Table 2, Supplemental Material S1). The values for the varied parameters n, α, and β were randomly altered based on a uniform distribution U(a,b) across a specific constraint range (a,b) as defined in Table 2. The remaining varied parameters K_s (of the macropore domain, matrix domain, and interface), pct, fex, and mex were calculated as functions of so-called parameter modification factors F_{xxx}. These modification factors F_{xxx} were introduced to allow random sampling for the parameter values across several orders of magnitude. The modification factors F_{xxx} were randomly chosen from a uniform distribution U(a,b) across a constraint range (a,b), which we individually defined for each parameter modification factor by following the rationale indicated in Table 2.

Furthermore, we used two different depth distributions of initial saturation in the MC simulations. Both depth distributions reflected the conditions at the beginning of the irrigation experiment. One profile reflected the soil moisture observed at several locations close to the irrigation plots (DP_{sat} A, Supplemental Table S1). The second profile was derived from simulated soil moisture from Glaser et al. (2016) (DP_{sat} B, Supplemental Table S1). The two different initial saturation depth profiles were divided evenly among the different MC runs (Supplemental Table S2).

Evaluation of Plot-Scale Simulations

The plot-scale simulations were evaluated based on the Br$^-$ concentrations. We evaluated the simulated depth profiles of Br$^-$ with the average observed Br$^-$ profiles (i.e., the average profiles of the five individual vertical profiles for each irrigation plot, corrected with the corresponding recovery rates, assuming a uniform error with depth) of the different irrigation plots (i.e., comparing the 8000 simulations with 30 mm h^{-1} irrigation intensity with the profile of Plot 2 and the 12,000 simulations with 50 mm h^{-1} irrigation intensity with the profiles of Plots 1 and 3). We used a combination of two Nash–Sutcliff efficiencies (NSEs) as the model quality criterion. We calculated one NSE for the concentration profile across the full extent of observed depths (NSE$_{total}$). We calculated the second NSE (NSE$_{Br\text{-peak}}$) for a specific depth section between

0.32 and 0.82 m where Br$^-$ showed increased concentration for all irrigation plots. This second NSE$_{Br\text{-peak}}$ was specifically used for assessing the model performance in terms of preferential transport down to the periglacial deposit layer (cf. Fig. 1) without biasing the assessment with the influence of high Br$^-$ concentrations originating from a uniform transport front in the soil matrix in the upper 0.3 m. The MC simulations were ranked according to each of the two NSEs, and the resulting two rank numbers for each model run were summed to a final performance rank number.

Transfer from Plot- to Catchment-Scale Simulations

We selected the 10 best performing parameter sets, the parameter set at the first quartile, the median, and the third quartile of performance for each of the three irrigation plots to be used for catchment-scale modeling. The resulting subset of 39 parameter sets was intended to reflect the diversity of plot-scale preferential flow parameter sets. Based on this, we tested the effect of variable preferential flow parameterizations on long-term catchment-scale simulations.

Each of the 39 selected parameter sets was used for a simulation of the hydrometric catchment response (i.e., soil moisture and discharge, no solute transport) for the 6-ha Weierbach headwater from October 2012 to April 2014. We set up a single-domain reference model based on the headwater model of Glaser et al. (2016). Compared with their model, we reduced the temporal resolution (daily instead of hourly meteorological input data, see below and Supplemental Material S2) and the spatial resolution (coarsened horizontal grid in the riparian zone and vicinity) to save computational costs once the dual-permeability approach was added to the reference model. We added the dual-permeability approach by defining a macropore domain for all soil type zones except the stagnic soil layer of the riparian zone (because its soil structure did not indicate preferential flow in the field).

To check whether a daily rather than an hourly resolution of meteorological input data would affect the study conclusions, we compared catchment simulations with hourly input and daily input data for six different parameter sets. To do this, we selected the parameter set for the single-domain reference model and five dual-permeability parameter sets that produced different discharge responses at catchment scale. Details on the comparison are given in Supplemental Material S2. In summary, visual comparison and model efficiency (NSE) showed that the reduced input time step had no relevant effect.

The coarsening of the spatial resolution affected only the horizontal grid spacing in the riparian zone and its vicinity and still ensured a nested grid, with finer grid cells in the riparian zone and stream bed compared with the hillslopes. A comparison between the original headwater model of Glaser et al. (2016) and the reference model showed that this grid coarsening did not visibly affect the discharge simulation (data not shown). However, the adaptation of the grid required a new spin-up for defining initial conditions, which we performed during the period from October 2010 to October 2012 (loop of three repetitions).

The performance of the catchment-scale simulations was evaluated with discharge and soil moisture measurements (Fig. 1) from October 2012 to April 2014. The soil moisture observed generally responded to incoming precipitation with fast and strong soil moisture increases in the topsoil and an increasingly damped response with depth. This behavior was consistent with other measurement profiles in the Weierbach catchment (data not shown), yet some differences exist between absolute moisture values. Hence, we based the catchment model evaluation for soil moisture on the temporal soil moisture dynamic at different depths by calculating the Spearman correlation between moisture observations and moisture simulations. Discharge performance was evaluated by calculating the overall NSE.

⬥ Results

Plot-Scale Simulations

Monte Carlo Simulation of Irrigation Experiments

Simulating the three plot-scale irrigation experiments without dual permeability reproduced the transport front down to the 0.3-m depth, but there was no Br$^-$ transport into deeper soil layers (Fig. 3a). The 20,000 Monte Carlo simulations with varying preferential flow parameters resulted in diverse depth transport of Br$^-$. Approximately 5% of the simulations (8.5% for Plot 1, 4.2% for Plot 2, 2.1% for Plot 3) modeled the Br$^-$ depth profiles observed with NSE$_{total}$ > 0 and NSE$_{depth}$ > 0 (NSE for characteristic Br$^-$ peaks between 0.32 and 0.82 m). Approximately 40% of the simulations (28% of Plot 1, 64% of Plot 2, 26% of Plot 3) did not reproduce any Br$^-$ peaks in deeper soil layers (no local maxima with concentrations $\geq$0.05 kg m^{-3} in the 0.32–0.82-m depth) and in that regard did not perform better than a single-domain model.

The model performance of the 20,000 parameter sets was barely sensitive to the values of single parameters and the related modification factors F_{xxx} or depth profile predefinition methods DP$_{xxx}$, (cf. Table 2), respectively. Mass exchange coefficient mex (F_{mex}), matrix conductivity $K_{s\text{-matrix}}$ (F_{matKs}, DP$_{Ks}$), macropore percentage pct (DP$_{pct}$, F_{pct}), and initial saturation sat (DP$_{sat}$) showed some relation to model performance, but this did not allow the identification of unique, well-performing parameter values (Supplemental Fig. S2).

Characteristics of the Parameter Sets Transferred to Catchment Scale

The 10 best-performing preferential flow parameter sets of each irrigation plot (later used for the catchment-scale simulations) showed some variability within the simulation results and
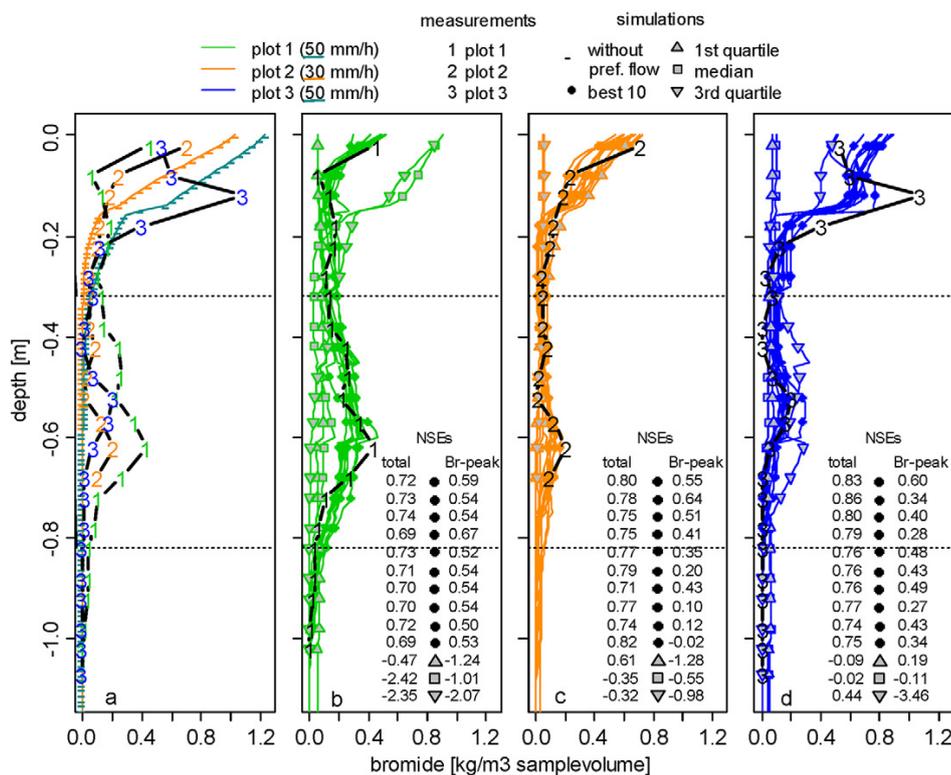
Fig. 3. Observed Br$^-$ depth profiles of the three irrigation plots 1 to 3 (average profiles of five sampled vertical profiles, cf. Jackisch et al., 2017) in comparison to (a) simulations of the irrigation experiments without preferential flow and (b–d) simulations with a dual-permeability approach, showing the results of the 10 best, first quartile, median, and third quartile performing parameter sets of the Monte Carlo simulations for (b) Plot 1 (50 mm h^{-1} irrigation rate), (c) Plot 2 (30 mm h^{-1} irrigation rate) and (d) Plot 3 (50 mm h^{-1} irrigation rate).

the observations were not fully captured (Fig. 3b–3d). Nonetheless, the intraplot variability of the responses was low compared with the interplot variability, which was captured very well. The parameter sets with the first quartile, median, and third quartile of performance for each irrigation plot represented various nonfitting results, with simulations of more or less uniform Br⁻ depth profiles (e.g., Plot 2, third quartile, Fig. 3c), simulations with a transport front down to the 0.3-m depth (e.g., first quartile of Plot 2, Fig. 3c), or simulations with Br⁻ peaks in mismatching depths (e.g., Plot 3, third quartile, Fig. 3d).

The parameter values of the 39 parameter sets that were eventually transferred to catchment scale reflected the low parameter identifiability of the 20,000 parameter sets. The parameters that were not identifiable (n, α, β, apert, $K_{s\text{-int}}$, fex, cf. Supplemental Fig. S2) spread across most of the sampled parameter space for the 39 transferred parameter sets (Fig. 4). The resulting water retention curves also covered a large range of the tested shapes of retention curves (Supplemental Fig. S3). The parameter values and water

retention curves of the first, median, and third quartile model runs of each irrigation plot were not clearly separated from the values for the 10 best model runs (Fig. 4, Supplemental Fig. S3).

The values of the parameters that showed a relation to model performance (mex, $K_{s\text{-matrix}}$, pct, initial saturation sat; cf. Supplemental Fig. S2) extended only across a constrained part of the parameter space for the 10 best model runs of each irrigation plot, and the parameter values of the first, median, and third quartile model runs were more clearly separated from the best performing parameter sets. The modification factor for the mass exchange coefficient (F_{mex}) was spread across <50% of the sampled parameter space for the 10 best runs of all three irrigation plots (Fig. 4). Matrix conductivity was reduced for only two out of the 10 best runs of all three irrigation plots (F_{matKs}, Fig. 4, DP_{Ks} A [no variation of $K_{s\text{-matrix}}$], Supplemental Table S3). The definition of initial saturation was based on soil moisture observations for the vast majority of the 10 best runs of all three irrigation plots (DP_{sat} A, Supplemental Table S3). For pct, the restriction to a part of the parameter space was less distinctive. The depth profile predefinition method DP_{pct} A (Brilliant Blue stains correspond to macropores) showed a tendency to be more common in the first quartile, median, and third quartile model runs than in the 10 best runs (Supplemental Table S3). In addition, the modification factor F_{pct} was limited to half of the sampled parameter space for the 10 best simulations of Plot 3 (Fig. 4).

Catchment-Scale Simulations
Discharge and Soil Moisture Responses

The reference catchment model adequately matched the observed discharge from October 2012 to April 2014 (NSE of 0.63). Nevertheless, the model showed clear limitations in reproducing some specific hydrograph responses (Fig. 5a, cf. also previous modeling by Glaser et al. [2016], Fig. 1). Several of the dual-permeability, catchment-scale simulations matched the observed hydrograph similarly to the reference simulation, with no improvement in the representation of the missing hydrograph features (Fig. 5b and 6, NSE > 0.5). All other dual-permeability simulations at catchment scale resulted in notably damped (Fig. 5c and 5d), flashy (Fig. 5c and 5f), or underestimated (Fig. 5e and 5f) discharge behavior and clearly could not reproduce the observed catchment discharge (Fig. 6, NSE < 0.5).

The differing performance of the simulated hydrographs was not related to the model performance of the parameter sets at plot scale. As an example, the top 10 performing parameter sets of Plot 1 showed a tendency to strongly underestimate catchment discharge (NSEs < −0.5, Fig.
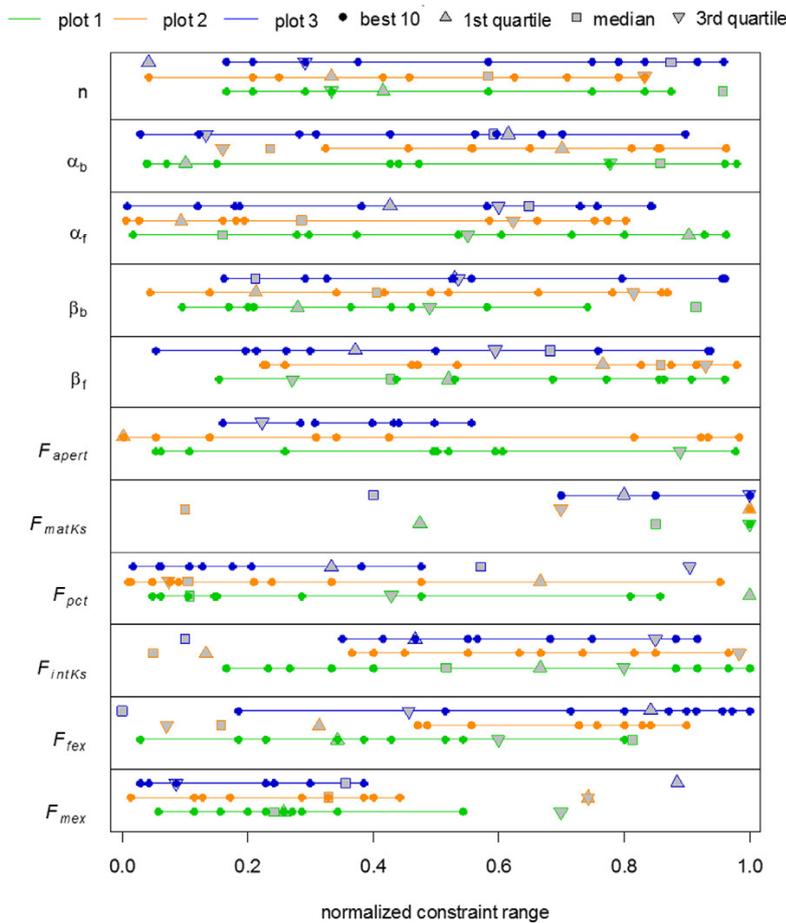

Fig. 4. Distribution of the parameter values and parameter modification factors of the 10 best, first quartile, median, and third quartile performing parameter sets of the MC simulations of the three irrigation plots. The values were normalized to the respective pre-constraint value ranges (Table 2), with 0 corresponding to the lowest and 1 to the highest constraint value.
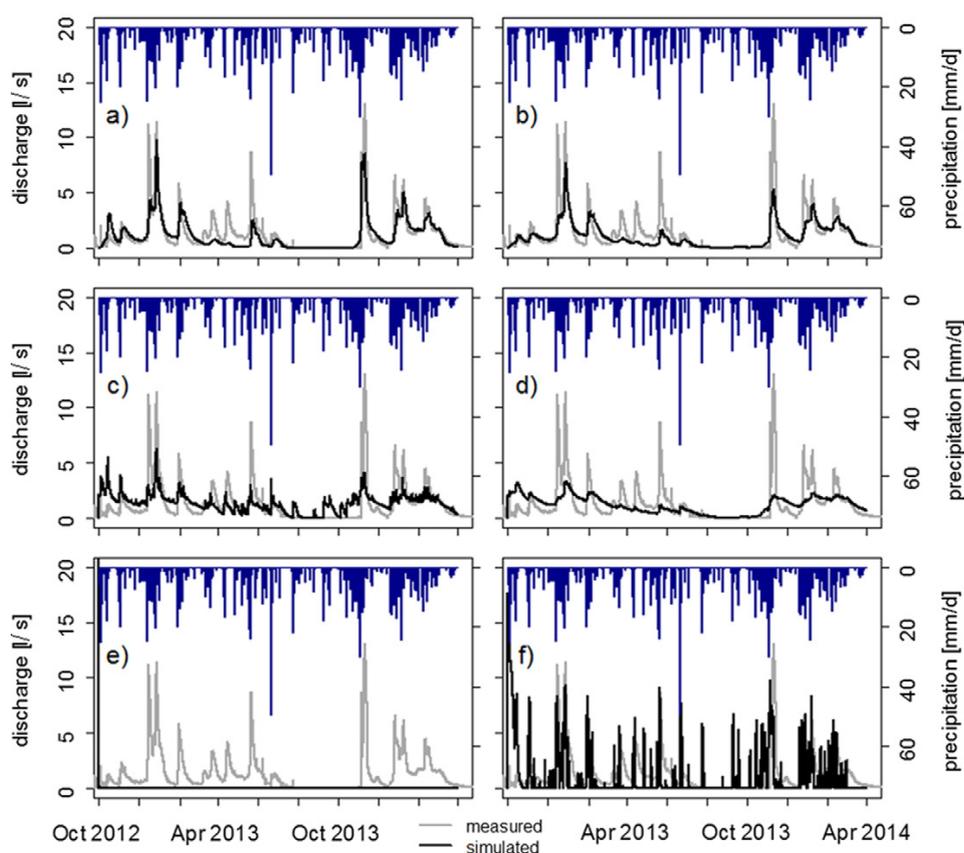
Fig. 5. Observed and simulated discharge for six simulations at catchment scale: (a) the reference simulation, (b–f) simulations with dual-permeability parameter sets from (b,d–f) the 10 best performing and (c) the first quartile performing parameter sets for Irrigation Plot 1.

6). However, the different hydrographs shown in Fig. 5b, 5d, 5e, and 5f all resulted from a parameter set belonging to the top 10 performing parameter sets of Plot 1. Also, the parameter set with the first quartile performance for Plot 1 resulted in a simulated catchment hydrograph with higher model efficiency than many simulations with parameter sets that belonged to the 10 best performing parameter sets at plot scale (Fig. 5c and 6). Thus, the performance of a distinct parameter set at plot scale was no predictor for its performance at catchment scale.

Soil moisture simulated with the reference model was similar to soil moisture observed in respect of the characteristic fast responses and a reduced responsiveness with increasing depth (Fig. 7a). The coefficients of correlation between simulated and observed soil moisture were >0.6 in all depths (Supplemental Fig. S4). Several of the dual-permeability simulations matched the soil moisture depth profiles observed in a similar way to the reference simulation with correlation coefficients > 0.6 for most depths (Fig. 7b and 7d, Supplemental Fig. S4). The parameter sets that resulted in such a similar soil moisture behavior were—for the most part—the same parameter sets that showed well-simulated hydrographs (NSE > 0.5, cf. Fig. 6, Supplemental Fig. S4). The remaining parameter sets resulted in soil moisture dynamics with

a poorer match of observed soil moisture (Fig. 7c, 7e and 7f, correlation coefficients <0.6 in most depths, Supplemental Fig. S4).

Parameter Sensitivity

The analysis of the parameter variations revealed a clear relation between model performance (hydrograph NSE) and the modification factor of the macropore aperture depth profiles F_{apert}. This factor modified the predefined macropore apertures that were used for the conductivity determination method DP_{Ks} A (cf. Table 2). The positive correlation identified between F_{apert} and hydrograph NSE (Fig. 8a) indicates that the model performed better with smaller macropore apertures and thus lower macropore conductivities at catchment scale, while such an effect was not observed at plot scale (cf. Fig. 4, Supplemental Fig. S2). The macropore conductivities that were determined based on method DP_{Ks} B, and thus relied on reduced matrix conductivities ($F_{matKs} < 0$), yielded mid-level performance for the catchment simulations (Fig. 8b). This was consistent with plot-scale behavior (cf. Supplemental Fig. S2). The effective hydraulic conductivities were very similar to the hydraulic conductivities of the reference model in the upper seven soil zones for all model runs resulting in hydrograph NSEs > 0 (Fig. 8c). In the lowest three soil zones, the effective conductivities were more variable (Fig.
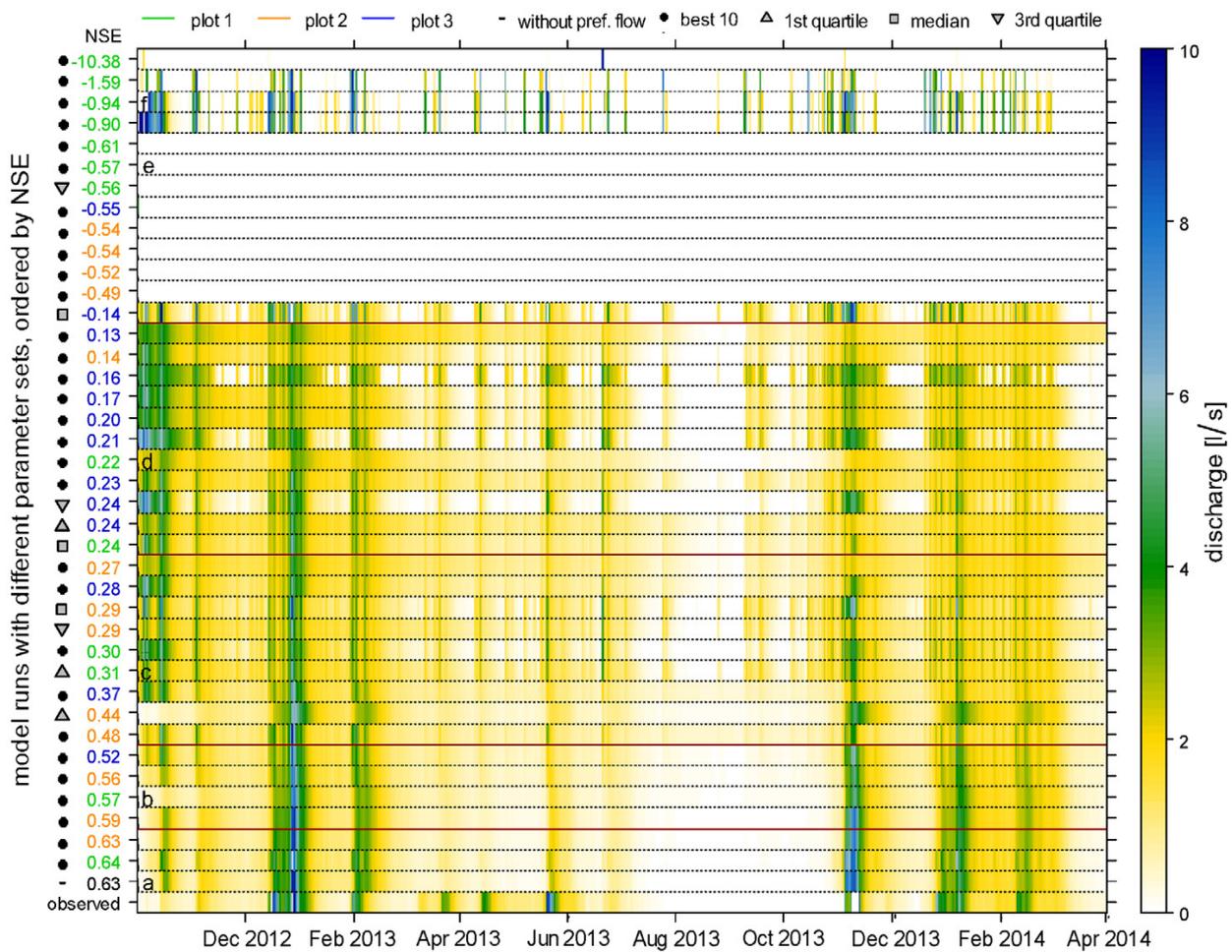
Fig. 6. Heat map showing observed and simulated discharge at catchment scale. Each row corresponds to one hydrograph; the magnitude of discharge is indicated with a color scale. The bottom row is the observed discharge, rows a through f represent the hydrographs of Fig. 5. The hydrographs are sorted according to their Nash–Sutcliffe efficiency (NSE). The colors of the NSE values indicate which irrigation plot was simulated with the parameter set at plot scale. The symbols correspond to the performance of the parameter sets at plot scale.

8d), but in all soil zones (1–10) the effective conductivities were distinctly higher when the model run resulted in a hydrograph NSE < 0. None of the other varied parameters showed any relation between their value and the model performance at catchment scale (Supplemental Fig. S5). This includes the parameters that showed some effect on plot-scale performance (i.e., macropore percentage and initial saturation, cf. Supplemental Fig. S2; note that the mass exchange coefficient was irrelevant for the catchment-scale model because only hydrometric responses were simulated).

◆ Discussion and Conclusions
Value of Plot-Scale Observations for Parameterizing Dual-Permeability Catchment Simulations
Non-uniqueness of Plot-Scale Parameter Values

One aim of this study was to test whether the plot-scale irrigation experiments contained valuable information for

parameterizing a dual-permeability catchment model. The MC simulation allowed us to find various preferential flow parameterizations that matched the plot-scale observations well. However, different preferential flow parameter sets resulted in equally good simulations, and we could not identify unique well-performing parameter values. Moreover, none of the different, partly contrasting approaches for predefining parameter depth profiles (DP_{pct}, DP_{Ks}, DP_{apert}) resulted in better simulation results than other approaches. This shows that it was not meaningful to identify the shape of the parameter depth profiles from our field observations.

The pronounced parameter non-uniqueness is consistent with other studies modeling preferential flow (e.g., Klaus and Zehe, 2010; Arora et al., 2012). Search algorithms for inverse parameter estimation that are more sophisticated than the applied brute-force Monte Carlo could allow a more efficient identification of optimal plot-scale parameterizations. However, this cannot solve the problem of parameter non-uniqueness and insensitivity (cf.
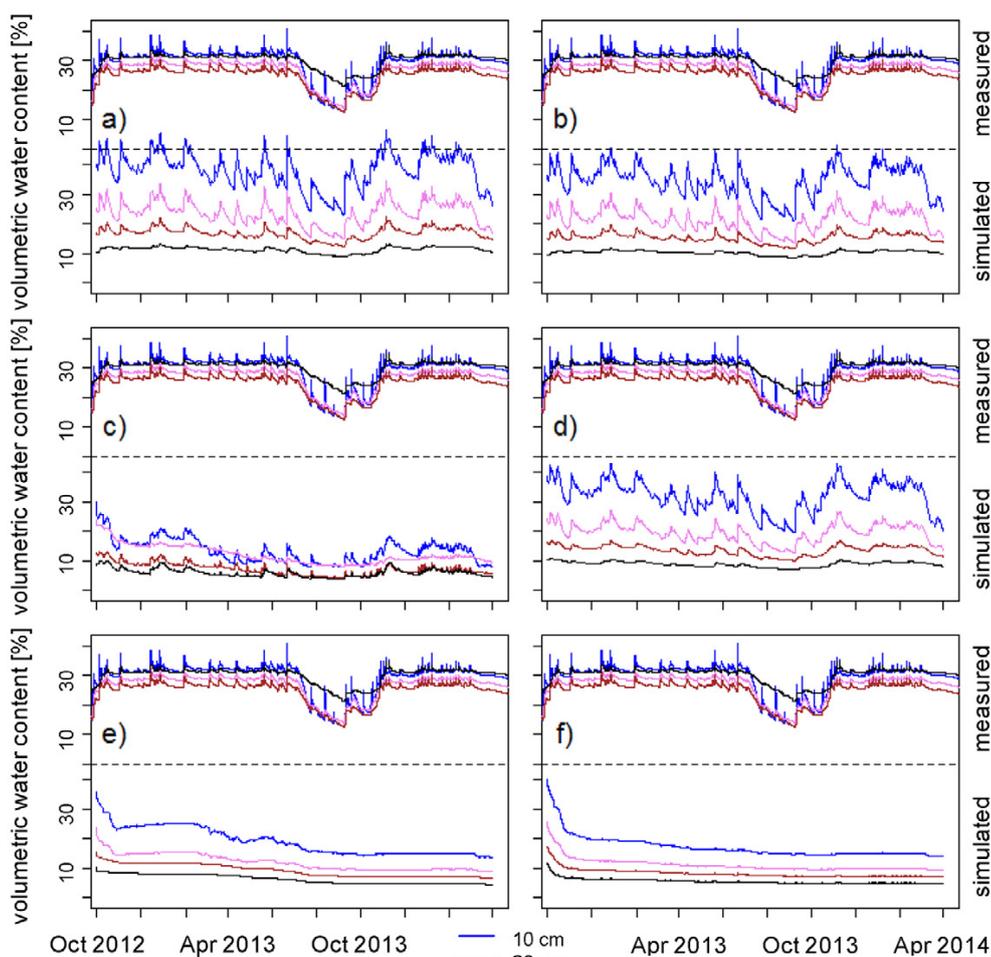
Fig. 7. Observed (top) and simulated (bottom) soil moisture at different depths for six simulations at catchment scale: (a) the reference simulation, (b–f) simulations with dual-permeability parameter sets from (b,d–f) the 10 best performing and (c) the first quartile performing parameter sets for Irrigation Plot 1 (cf. Fig. 5, Supplemental Fig. S4).

Jarvis et al., 2007; Arora et al., 2011, 2012). Additional observation data could help to better constrain the parameter values. For example, Larsbo and Jarvis (2005, 2006) analyzed the information content of solute concentrations in effluent fluxes, resident solute concentrations in the soil, drain flow, and soil water content to identify several parameters of a dual-permeability model. Their results showed that a combination of multiple observation data and measurements with high frequencies at the beginning of irrigation experiments were best for conditioning the parameterization. Furthermore, applying not only conservative tracers (as, e.g., Br⁻) but also several reactive solutes with differing transport characteristics, such as weak or strong vs. intermediate sorptivity, could give information about the partitioning of fluxes between matrix and macropore flow (McGrath et al., 2009, 2010) and thus could help to better constrain the model parameters.

Nevertheless, the aim of this study was not to perform a detailed parameter identification and sensitivity analysis for dual-permeability simulations of the plot-scale irrigation experiments

but rather to explore model performance at catchment scale for several parameter sets that were performing well at plot scale. In this context, we observed that different parameter sets with very similar, good fits at plot scale (i.e., the 10 best parameter sets) resulted in clearly different simulations of catchment discharge and soil moisture (cf. Fig. 6, Supplemental Fig. S4). This shows that a parameter set identified as optimal at plot scale does not necessarily perform well at catchment scale and that transferring one optimal parameter set from plot to catchment scale is problematic. Consequently, the robustness of the transferred parameter set needs to be validated against catchment-scale data or a direct parameter calibration at catchment scale is needed.

Effect of Spatial Heterogeneity of Vertical Preferential Flow on Catchment-Scale Simulations

Simulated catchment discharge and soil moisture showed similar variability among simulations independently of the irrigation experiment that was used for identifying dual-permeability
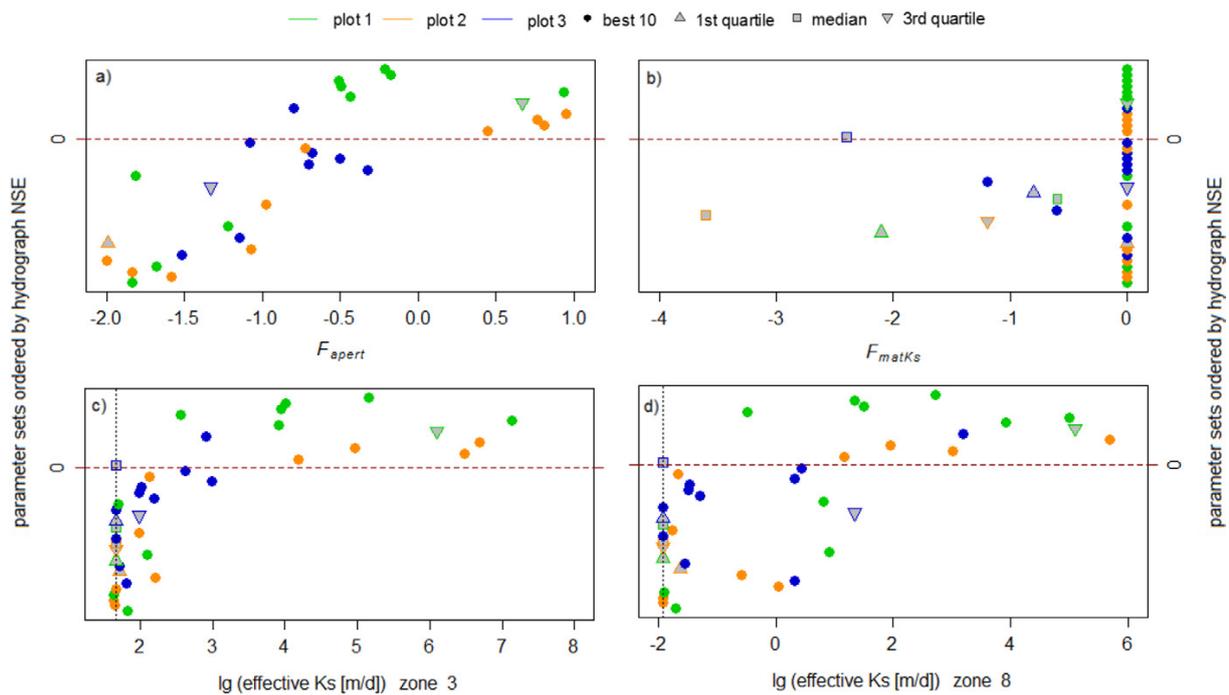
Fig. 8. Distribution of the parameter modification factors for (a) macropore aperture (F_{apert}) and (b) matrix hydraulic conductivity (F_{matKs}) as well as of the effective hydraulic conductivity in (c) soil type Zone 3 and (d) soil type Zone 8 (cf. Fig. 1) within the catchment-scale simulations. Data points are sorted along the y axis according to the hydrograph Nash–Sutcliffe efficiency (NSE) of the parameter set (cf. Fig. 6). Data points above the red line correspond to parameter sets that resulted in a hydrograph NSE < 0, data points below the red line correspond to parameter sets that resulted in a hydrograph NSE > 0. The colors of the data points indicate which irrigation plot was simulated with the parameter set at plot scale. The symbols correspond to the performance of the parameter sets at plot scale.

parameter sets (cf. Fig. 6, Supplemental Fig. S4). Because the interplot variability among the three irrigation plots originated from a different degree of vertical preferential flow, it seems that the effect of spatial heterogeneity of vertical preferential flow between the plot-scale experiments averaged out at catchment scale. Thus, it was possible to condense spatial heterogeneity into average, effective values for the catchment scale. On the one hand, this is consistent with earlier studies that concluded that the effect of the spatial heterogeneity of initial soil moisture (Zehe and Blöschl, 2004), of hydraulic conductivity fields (Meyerhoff and Maxwell, 2011), or of pipe flow networks (Weiler and McDonnell, 2007) averaged out for the integrated hydrological catchment response. On the other hand, this is different from what was expected in studies discussing the role of the small-scale heterogeneity of preferential flow within a catchment (e.g., Christiansen et al., 2004; Graham and Lin, 2011; Liu and Lin, 2015; Wiekenkamp et al., 2016).

One might argue that a spatially homogeneous model setup works well only for simulating discharge. Certainly, it is not possible to capture the observed interplot variability among the three irrigation plots with average, effective values that are assigned homogeneously for all the catchment. However, the quality of the soil moisture evaluated as an internal response of the catchment model was also independent of the irrigation plot that was used to determine the model parameters. Potentially, the soil moisture simulation

may behave differently at various locations, which could lead to an improved realism at locations other than the soil moisture locations evaluated. Yu et al. (2014) demonstrated that simulated groundwater tables matched field observations well only when preferential flow and spatial subsurface heterogeneity were considered. Accordingly, it could be evaluated if and how the simulation of soil moisture would improve if a spatially more heterogeneous distribution of soil types in the model would be accounted for. Such a setup could potentially also result in a better prediction of the hydrograph response. However, including more spatial heterogeneity also increases the number of parameters and thus the degrees of freedom, which in turn increases the risk of overfitting the model. This is especially a problem if spatially heterogeneous parameters cannot be related to obvious differences observed in the landscape, such as different soil types. In the Weierbach catchment, such spatial differences in soil types are, with the exception of the riparian zone, not observable.

Representation of Preferential Flow at Plot and Catchment Scales

Adequacy of the Dual-Permeability Approach

Our study is one among several that has successfully applied the dual-permeability approach for simulating plot-scale observations of solute transport (e.g., Roulier et al., 2006; Arora et al., 2011; Cadini et al., 2013; Wang et al., 2014). The dual-permeability

approach allowed us to simulate the Br⁻ depth profiles for the three different irrigation experiments, whereas a single-domain approach did not reproduce the Br⁻ peak observed in deeper soil layers. In contrast to the plot scale, the best performing simulations with a dual-permeability approach at catchment scale, when transferring plot-scale parameters to the catchment-scale model, were very similar to the results of the reference simulation with a single permeability domain. This suggests that the incorporation of the dual-permeability approach did not improve the representation of the processes that are relevant for simulating the observed soil moisture and discharge response of the Weierbach catchment. Even more, the model performance for soil moisture and discharge decreased for many of our tested dual-permeability parameter sets.

Such a performance decrease when applying a dual-permeability model has not to our knowledge been reported in the literature so far, as it does not appear in cases where the parameters are calibrated at catchment scale. In line with our findings, Christiansen et al. (2004) and De Schepper et al. (2015) showed that the performance of their discharge and groundwater head simulations did not notably improve when including preferential flow in their models. Other studies could—at least partly—improve the simulation of internal, distributed (water tables and soil moisture) and integrated (runoff) responses when preferential flow was incorporated (dual permeability and other approaches) in their catchment models (Beckers and Alila, 2004; Zhang et al., 2006; van Schaik et al., 2014; Yu et al., 2014).

Obviously, the choice of a modeling approach for simulating preferential flow processes depends on the underlying question. If one is interested in the spatial patterns of preferential flow or in detailed analyses of exchange processes between macropores and soil matrix, an explicit implementation of macropore or fracture geometries and distributions (e.g., Vogel et al., 2006; Rosenbom et al., 2009; Klaus and Zehe, 2011; Jackisch and Zehe, 2018) and a conceptually different description of the pore-scale processes (e.g., Beven and Germann, 2013; Scheibe et al., 2015; Jackisch and Zehe, 2018) is necessary. In our case, we aimed to test a direct parameter transfer from plot to catchment scale. This requires a model approach that treats the processes at plot and catchment scales in the same way. The dual-permeability approach as implemented in HydroGeoSphere allows this. Because the dual-permeability approach was adequate for reproducing the three plot-scale irrigation experiments, we assume that it was also an adequate approach for the catchment scale. However, we did not test whether a different approach for simulating vertical preferential flow processes would have improved our simulation results at catchment scale.

Vertical vs. Lateral Flow in the Weierbach and Beyond

The lack of improved model efficiency in this study when incorporating preferential flow with a dual-permeability approach may be explained with a different relative importance of vertical and lateral flow at the different scales. At plot scale, vertical preferential flow (incorporated with dual permeability) was necessary to simulate the Br⁻ depth profiles observed. At catchment scale, model performance was highest for parameter sets where

the influence of vertical preferential flow (i.e., macropore conductivities) was low and homogeneous fast vertical infiltration was ensured with effective hydraulic conductivities, with values being similar to the conductivities of the single-domain reference model (cf. Fig. 8). Moreover, the model performed better when the matrix conductivity was not too low. This is manifested by the fact that the parameter sets with retained effective hydraulic conductivities but reduced matrix conductivities (cf. DP_{K_s} B) resulted in catchment simulations with mid-level performance. This relation between model performance and parameter values indicates that the conceptual setup of the reference model was not improved by the incorporation of the dual-permeability approach. The reference model was composed of multiple soil layers with contrasting hydraulic conductivities (highest conductivities in Soil Zones 4–6, the depth where Br⁻ peaks were observed), allowing fast lateral subsurface flow along the interfaces between specific soil horizons. Thus, the modeling results indicate that uniform fast vertical flow in the unsaturated zone combined with connected fast lateral subsurface flow are the flow processes that mainly control the hydrometric response in the Weierbach catchment.

The role of fast lateral (preferential) flow on runoff generation has been widely observed (e.g., Weiler and McDonnell, 2007; Anderson et al., 2009; Yu et al., 2014; Laine-Kaulio et al., 2014; Wilson et al., 2016). Our results suggest that the relative importance of fast lateral flow on runoff generation largely outweighs the relative importance of vertical preferential flow. The likely reason for this not having been previously reported is that the few studies that explicitly compared catchment runoff simulations with and without preferential flow did not introduce vertical preferential flow and fast lateral flow through certain soil structures separately (i.e., layers, fractures, and macropores) (Beckers and Alila, 2004; Zhang et al., 2006; Yu et al., 2014) or were performed in a different climate (van Schaik et al., 2014). Only Christiansen et al. (2004) and De Schepper et al. (2015) compared a single-permeability domain model that included multiple soil layers with differing hydraulic conductivities (allowing for nonuniform lateral subsurface flow) with an equivalent dual-permeability model (maintaining similar effective hydraulic conductivities but allowing additional vertical preferential flow) for a climate similar to our study site. Consistent with our results, they also could not show a clear improvement of the overall runoff simulation.

The approach to derive catchment model parameters from detailed plot-scale simulations implies that the properties and processes that are relevant at small (plot) scale are also critical at catchment scale. According to our interpretation of our results, this seems not to be the case, since the results suggest that vertical and lateral flow play a different role at plot and catchment scales. Thus, a transfer of parameters for vertical preferential flow is not only problematic due to parameter non-uniqueness (cf. above) but also due to the fact that catchment response may be controlled by different process combinations than the plot-scale response. Certainly, our interpretation relies only on results derived with one specific modeling approach for one specific catchment. Yet, the

interpretation is consistent with the conceptual idea that processes and structures can have a different role at different scales (cf. Vogel and Roth, 2003), which is also supported by direct process observations (e.g., Jackisch et al., 2017). This concept suggests that details of structures that are important for processes at small scales can be integrated as averaged, effective descriptions in the representation of a structure that is relevant at larger scale. With respect to our results, this would mean that small-scale vertical preferential flow features can be integrated as averaged, effective descriptions for the modeling of the hydrometric response at larger scales. In other words, it would mean that the often discussed integrated effect of preferential flow on runoff and soil moisture at catchment scale (Beven and Germann, 2013; Weiler, 2017) could, in our case, be reflected in a combination of fast vertical infiltration and fast lateral subsurface flow in certain soil layers under certain conditions.

Limitations and Needs of Further Research

Limitations of the Modeling Approach

Our findings and their interpretation result from one specific modeling approach, i.e., a dual-permeability approach that was applied to a spatially homogeneous model setup. This means that a generalization of the results is limited. We especially do not intend to state that vertical preferential flow does not play any role in the distribution of soil moisture and on runoff generation. It is possible that a different model setup or conceptualization (e.g., governing equations of preferential flow, spatial variability) could have improved the hydrometric catchment response. Moreover, the dual-permeability approach improved the internal realism of the catchment simulations because it was able to reproduce the plot-scale observations. Yet, there are indications that the relative importance of vertical preferential flow for simulating hydrometric responses at catchment scale was less pronounced than suggested by the plot-scale observations. Instead, it was more important to properly account for a general fast infiltration in combination with nonuniform lateral flow. Therefore, depending on the model application, it is important to decide whether it is necessary to incorporate vertical preferential flow, as it comes with an additional parameterization effort and additional computational costs.

One limitation of our approach is the number of tested parameter sets at catchment scale. The total number of 39 simulations with different parameter sets for preferential flow is high compared with other modeling studies (e.g., Beckers and Alila, 2004; Weiler and McDonnell, 2007; van Schaik et al., 2014; Yu et al., 2014) and allowed us to analyze the effect of the parameter values spreading across the whole parameter space on catchment response. However, the applied approach, where several plot-scale parameter sets were transferred directly to catchment scale, does not fully analyze the parameter space at catchment scale. One may argue that other parameter sets may improve the integrated hydrometric response at catchment scale while also performing well at plot scale. Thus, further research on an inverse calibration of a physically based 3D dual-permeability catchment model with a subsequent validation of the identified parameters at plot scale is needed.

Another limitation is that we simulated solute transport only at plot scale, although solute transport is often one of the main reasons for incorporating preferential flow. However, we think that Br^- transport and water flow are similar enough (cf. Zehe and Blöschl, 2004) to investigate the transferability of parameter sets from plot scale to catchment models. Certainly, the impact of the dual-permeability approach on solute transport toward the catchment outlet and on catchment travel times remains unclear. Using a numerical experiment, Christiansen et al. (2004) found that the incorporation of preferential flow paths had a significant effect on the transport of reactive solutes, while the effect on the transport of conservative solutes (such as Br^-) was small. Relating this to the present study, it might be that the simulation of solute concentrations at the catchment outlet is impacted by a dual-permeability approach (positively or negatively). An analysis of this requires appropriate field data to validate the correctness of simulated solute transport at catchment scale, which was available neither in this study nor in the study of Christiansen et al. (2004). Hence, future work should evaluate the effect of the proposed approach on solute transport at catchment scale, as our results are restricted to the hydrometric response.

Generalizing our Results to Other Landscapes

We performed our study for one particular catchment. The subsurface structure of this catchment is characterized by shallow soils, highly permeable periglacial layers, and fractured slate. Preferential flow probably occurs in a particular network of interaggregate pores (Jackisch et al., 2017) and along imbricated clasts and fractures (Scaini et al., 2017). Plot and hillslope field studies further suggested that there is a substantial vertical preferential flow component (Jackisch et al., 2017; Scaini et al., 2017). Nonetheless, for the particular pedolithological structure of the Weierbach catchment, the performance of the simulations of catchment discharge and soil moisture did not improve when we explicitly accounted for vertical preferential flow. In line with this, for a catchment with a similar pedolithological structure, Loritz et al. (2017) showed that it was sufficient to include fast infiltration and connected lateral subsurface flow paths in a representative hillslope model for simulating the rainfall–runoff behavior.

In catchments with different physiographic settings, a more explicit representation of vertical preferential flow may have a stronger influence on hydrometric catchment responses. In agricultural soils with fine matrix textures and high amounts of biopores, such as earthworm burrows (e.g., Klaus et al., 2013), the influence of vertical preferential flow, as opposed to nonuniform lateral flow, on runoff generation may be much higher. Glacial till soils may be more similar to the structure of the Weierbach catchment, and Jansson et al. (2005) made conclusions in line with our study when comparing simulations of soil moisture in a glacial till soil with a one-dimensional single-domain and a two-domain model. Catchments with a climate different to Luxembourg can be more prone to a high importance of vertical preferential flow on runoff generation. This was shown by van Schaik et al. (2014), who improved hydrometric responses for simulations with preferential flow under semiarid conditions. Runoff

generation in catchments with slow velocities and short travel distances of lateral subsurface flow (cf. Klaus and Jackson, 2018) may also be less dominated by lateral subsurface flow, and the role of vertical preferential flow may become more important. Nonetheless, our findings highlight that the observation of vertical preferential flow in plot or hillslope experiments is not necessarily critical for improving catchment simulations of the hydrological response. Instead, some features that are easier to determine (e.g., multiple soil layers with contrasting hydraulic properties and effective conductivities) may be more important for understanding the structures and processes that are critical for water flow at catchment scale.

Supplemental Material

The supplemental material comprises a detailed description of the parameter depth profiles used in the MC simulations (Supplemental Material S1), the comparison of discharge simulations with hourly and daily meteorological input data (Supplemental Material S2) and supplemental result data (Supplemental Material S3) showing the distribution of all parameters compared to the performance of all plot-scale and catchment-scale simulations, the used water retention curves, and the correlations between simulated and observed soil moisture.

References

Angermann, L., C. Jackisch, N. Allroggen, M. Sprenger, E. Zehe, J. Tronicke, et al. 2017. Form and function in hillslope hydrology: Characterization of subsurface flow based on response observations. Hydrol. Earth Syst. Sci. 21:3727–3748. doi:10.5194/hess-21-3727-2017

Anderson, A.E., M. Weiler, Y. Alila, and R.O. Hudson. 2009. Dye staining and excavation of a lateral preferential flow network. Hydrol. Earth Syst. Sci. 13:935–944. doi:10.5194/hess-13-935-2009

Aquanty. 2015. HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. User guide. Aquanty Inc., Waterloo, ON, Canada.

Arora, B., B.P. Mohanty, and J.T. McGuire. 2011. Inverse estimation of parameters for multidomain flow models in soil columns with different macropore densities. Water Resour. Res. 47:W04512. doi:10.1029/2010WR009451

Arora, B., B.P. Mohanty, and J.T. McGuire. 2012. Uncertainty in dual permeability model parameters for structured soils. Water Resour. Res. 48:W01524. doi:10.1029/2011WR010500

Beckers, J., and Y. Alila. 2004. A model of rapid preferential hillslope runoff contributions to peak flow generation in a temperate rain forest watershed. Water Resour. Res. 40:W03501. doi:10.1029/2003WR002582

Beven, K., and P. Germann. 2013. Macropores and water flow in soils revisited. Water Resour. Res. 49:3071–3092. doi:10.1002/wrcr.20156

Cadini, F., J. De Sanctis, I. Bertoli, and E. Zio. 2013. Upscaling of a dual-permeability Monte Carlo simulation model for contaminant transport in fractured networks by genetic algorithm parameter identification. Stochastic Environ. Res. Risk Assess. 27:505–516. doi:10.1007/s00477-012-0595-8

Christiansen, J.S., M. Thorsen, T. Clausen, S. Hansen, and J.C. Refsgaard. 2004. Modelling of macropore flow and transport processes at catchment scale. J. Hydrol. 299:136–158. doi:10.1016/j.jhydrol.2004.04.029

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

De Schepper, G., R. Therrien, J.C. Refsgaard, and A.L. Hansen. 2015. Simulating coupled surface and subsurface water flow in a tile-drained agricultural catchment. J. Hydrol. 521:374–388. doi:10.1016/j.jhydrol.2014.12.035

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

Frey, S.K., H.T. Hwang, Y.J. Park, S.I. Hussain, N. Gottschall, M. Edwards, and D.R. Lapen. 2016. Dual permeability modeling of tile drain management influences on hydrologic and nutrient transport characteristics in macroporous soil. J. Hydrol. 535:392–406. doi:10.1016/j.jhydrol.2016.01.073

Gelhar, L.W., C. Welty, and K.R. Rehfeldt. 1992. A critical review of data on field-scale dispersion in aquifers. Water Resour. Res. 28:1955–1974. doi:10.1029/92WR00607

Glaser, B., J. Klaus, S. Frei, J. Frentress, L. Pfister, and L. Hopp. 2016. On the value of surface saturated area dynamics mapped with thermal infrared imagery for modeling the hillslope–riparian–stream continuum. Water Resour. Res. 52:8317–8342. doi:10.1002/2015WR018414

Graham, C.B., and H.S. Lin. 2011. Controls and frequency of preferential flow occurrence: A 175-event analysis. Vadose Zone J. 10:816–831. doi:10.2136/vzj2010.0119

Jackisch, C., L. Angermann, N. Allroggen, M. Sprenger, T. Blume, J. Tronicke, and E. Zehe. 2017. Form and function in hillslope hydrology: In situ imaging and characterization of flow-relevant structures. Hydrol. Earth Syst. Sci. 21:3749–3775. doi:10.5194/hess-21-3749-2017

Jackisch, C., and E. Zehe. 2018. Ecohydrological particle model based on representative domains. Hydrol. Earth Syst. Sci. 22:3639–3662. doi:10.5194/hess-22-3639-2018

Jansson, C., B. Espeby, and P-E. Jansson. 2005. Preferential water flow in a glacial till soil. Nord. Hydrol. 36:1–11. doi:10.2166/nh.2005.0001

Jarvis, N., J. Koestel, and M. Larsbo. 2016. Understanding preferential flow in the vadose zone: Recent advances and future prospects. Vadose Zone J. 15(12). doi:10.2136/vzj2016.09.0075

Jarvis, N., M. Larsbo, S. Roulier, A. Lindahl, and L. Persson. 2007. The role of soil properties in regulating non-equilibrium macropore flow and solute transport in agricultural topsoils. Eur. J. Soil Sci. 58:282–292. doi:10.1111/j.1365-2389.2006.00837.x

Juilleret, J., J.F. Iffly, L. Pfister, and C. Hissler. 2011. Remarkable Pleistocene periglacial slope deposits in Luxembourg (Oesling): Pedological implication and geosite potential. Bull. Soc. Nat. Luxemb. 112(1):125–130. http://www.snl.lu/publications/bulletin/SNL_2011_112_125_130.pdf.

Klaus, J., and C.R. Jackson. 2018. Interflow is not binary: A continuous shallow perched layer does not imply continuous connectivity. Water Resour. Res. 54:5921–5932. doi:10.1029/2018WR022920

Klaus, J., C.E. Wetzel, N. Martínez-Carreras, L. Ector, and L. Pfister. 2015. A tracer to bridge the scales: On the value of diatoms for tracing fast flow path connectivity from headwaters to meso-scale catchments. Hydrol. Processes 29:5275–5289. doi:10.1002/hyp.10628

Klaus, J., and E. Zehe. 2010. Modelling rapid flow response of a tile-drained field site using a 2D physically based model: Assessment of "equifinal" model setups. Hydrol. Processes 24:1595–1609. doi:10.1002/hyp.7687

Klaus, J., and E. Zehe. 2011. A novel explicit approach to model bromide and pesticide transport in connected soil structures. Hydrol. Earth Syst. Sci. 15:2127–2144. doi:10.5194/hess-15-2127-2011

Klaus, J., E. Zehe, M. Elsner, C. Külls, and J.J. McDonnell. 2013. Macropore flow of old water revisited: Experimental insights from a tile-drained hillslope. Hydrol. Earth Syst. Sci. 17:103–118. doi:10.5194/hess-17-103-2013

Klaus, J., E. Zehe, M. Elsner, J. Palm, D. Schneider, B. Schröder, et al. 2014. Controls of event-based pesticide leaching in natural soils: A systematic study based on replicated field scale irrigation experiments. J. Hydrol. 512:528–539. doi:10.1016/j.jhydrol.2014.03.020

Köhne, J.M., S. Köhne, and H.H. Gerke. 2002. Estimating the hydraulic functions of dual-permeability models from bulk soil data. Water Resour. Res. 38(7). doi:10.1029/2001WR000492

62

Kordilla, J., M. Sauter, T. Reimann, and T. Geyer. 2012. Simulation of saturated and unsaturated flow in karst systems at catchment scale using a double continuum approach. Hydrol. Earth Syst. Sci. 16:3909–3923. doi:10.5194/hess-16-3909-2012

Krzeminska, D.M., T.A. Bogaard, J.-P. Malet, and L.P.H. van Beek. 2013. A model of hydrological and mechanical feedbacks of preferential fissure flow in a slow-moving landslide. Hydrol. Earth Syst. Sci. 17:947–959. doi:10.5194/hess-17-947-2013

Kukemilks, K., J.-F. Wagner, T. Saks, and P. Brunner. 2018a. Conceptualization of preferential flow for hillslope stability assessment. Hydrogeol. J. 26:439–450. doi:10.1007/s10040-017-1667-0

Kukemilks, K., J.-F. Wagner, T. Saks, and P. Brunner. 2018b. Physically based hydrogeological and slope stability modeling of the Turaida castle mound. Landslides 15:2267–2278. doi:10.1007/s10346-018-1038-5

Laine-Kaulio, H., S. Backnäs, T. Karvonen, H. Koivusalo, and J.J. McDonnell. 2014. Lateral subsurface stormflow and solute transport in a forested hillslope: A combined measurement and modeling approach. Water Resour. Res. 50:8159–8178. doi:10.1002/2014WR015381

Larsbo, M., and N. Jarvis. 2005. Simulating solute transport in a structured field soil: Uncertainty in parameter identification and predictions. J. Environ. Qual. 34:621–634. doi:10.2134/jeq2005.0621

Larsbo, M., and N. Jarvis. 2006. Information content of measurements from tracer microlysimeter experiments designed for parameter identification in dual-permeability models. J. Hydrol. 325:273–287. doi:10.1016/j.jhydrol.2005.10.020

Leistra, M., and J.J.T.I. Boesten. 2010. Measurement and computation of movement of bromide ions and carbofuran in ridged humic-sandy soil. Arch. Environ. Contam. Toxicol. 59:39–48. doi:10.1007/s00244-009-9442-4

Liu, H., and H. Lin. 2015. Frequency and control of subsurface preferential flow: From pedon to catchment scales. Soil Sci. Soc. Am. J. 79:362–377. doi:10.2136/sssaj2014.08.0330

Loritz, R., S.K. Hassler, C. Jackisch, N. Allroggen, L. van Schaik, J. Wienhöfer, and E. Zehe. 2017. Picturing and modeling catchments by representative hillslopes. Hydrol. Earth Syst. Sci. 21:1225–1249. doi:10.5194/hess-21-1225-2017

Martínez-Carreras, N., C. Hissler, L. Gourdol, J. Klaus, J. Juilleret, J.F. Iffly, and L. Pfister. 2016. Storage controls on the generation of double peak hydrographs in a forested headwater catchment. J. Hydrol. 543:255–269. doi:10.1016/j.jhydrol.2016.10.004

McGrath, G.S., C. Hinz, and M. Sivapalan. 2009. A preferential flow leaching index. Water Resour. Res. 45:W11405. doi:10.1029/2008WR007265

McGrath, G., C. Hinz, and M. Sivapalan. 2010. Assessing the impact of regional rainfall variability on rapid pesticide leaching potential. J. Contam. Hydrol. 113:56–65. doi:10.1016/j.jconhyd.2009.12.007

Meyerhoff, S.B., and R.M. Maxwell. 2011. Quantifying the effects of subsurface heterogeneity on hillslope runoff using a stochastic approach. Hydrogeol. J. 19:1515–1530. doi:10.1007/s10040-011-0753-y

Moragues-Quiroga, C., J. Juilleret, L. Gourdol, E. Pelt, T. Perrone, A. Aubert, et al. 2017. Genesis and evolution of regoliths: Evidence from trace and major elements and Sr–Nd–Pb–U isotopes. Catena 149:185–198. doi:10.1016/j.catena.2016.09.015

Reck, A., C. Jackisch, T.L. Hohenbrink, A. Zangerlé, and L. van Schaik. 2018. Impact of temporal macropore dynamics on infiltration: Field experiments and model simulations. Vadose Zone J. 17:170147. doi:10.2136/vzj2017.08.0147

Rosenbom, A.E., R. Therrien, J.C. Refsgaard, K.H. Jensen, V. Ernstsen, and K.E.S. Klint. 2009. Numerical analysis of water and solute transport in variably-saturated fractured clayey till. J. Contam. Hydrol. 104:137–152. doi:10.1016/j.jconhyd.2008.09.001

Roulier, S., N. Baran, C. Mouvet, F. Stenemo, X. Morvan, H.J. Albrechtsen, et al. 2006. Controls on atrazine leaching through a soil–unsaturated fractured limestone sequence at Brévilles, France. J. Contam. Hydrol. 84:81–105. doi:10.1016/j.jconhyd.2005.12.004

Scaini, A., M. Audebert, C. Hissler, F. Fenicia, L. Gourdol, L. Pfister, and K.J. Beven. 2017. Velocity and celerity dynamics at plot scale inferred from artificial tracing experiments and time-lapse ERT. J. Hydrol. 546:28–43. doi:10.1016/j.jhydrol.2016.12.035

Scaini, A., C. Hissler, F. Fenicia, J. Juilleret, J.-F. Iffly, L. Pfister, and K.J. Beven. 2018. Hillslope response to sprinkling and natural rainfall using velocity and celerity estimates in a slate-bedrock catchment. J. Hydrol. 558:366–379. doi:10.1016/j.jhydrol.2017.12.011

Scheibe, T.D., W.A. Perkins, M.C. Richmond, M.I. McKinley, P.D.J. Romero-Gomez, M. Oostrom, et al. 2015. Pore-scale and multiscale numerical simulation of flow and transport in a laboratory-scale column. Water Resour. Res. 51:1023–1035. doi:10.1002/2014WR015959

Schwab, M.P., J. Klaus, L. Pfister, and M. Weiler. 2017. How runoff components affect the export of DOC and nitrate: A long-term and high-frequency analysis. Hydrol. Earth Syst. Sci. Discuss. doi:10.5194/hess-2017-416

Steinbrich, A., H. Leistert, and M. Weiler. 2016. Model-based quantification of runoff generation processes at high spatial and temporal resolution. Environ. Earth Sci. 75:1423. doi:10.1007/s12665-016-6234-9

Therrien, R., R.G. McLaren, E.A. Sudicky, and S.M. Panday. 2010. HydroGeo-Sphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. Groundwater Simulations Group, Univ. of Waterloo, Waterloo, ON, Canada.

van Schaik, N.L.M.B., A. Bronstert, S.M. De Jong, V.G. Jetten, J.C. Van Dam, C.J. Ritsema, and S. Schnabel. 2014. Process-based modelling of a headwater catchment in a semi-arid area : The influence of macropore flow. Hydrol. Processes 28:5805–5816. doi:10.1002/hyp.10086

van Schaik, N.L.M.B., R.F.A. Hendriks, and J.C. van Dam. 2010. Parameterization of macropore flow using dye-tracer infiltration patterns in the SWAP model. Vadose Zone J. 9:95–106. doi:10.2136/vzj2009.0031

Villamizar, M.L., and C.D. Brown. 2017. A modelling framework to simulate river flow and pesticide loss via preferential flow at the catchment scale. Catena 149:120–130. doi:10.1016/j.catena.2016.09.009

Vogel, H.-J., I. Cousin, O. Ippisch, and P. Bastian. 2006. The dominant role of structure for solute transport in soil: Experimental evidence and modelling of structure and transport in a field experiment. Hydrol. Earth Syst. Sci. 10:495–506. doi:10.5194/hess-10-495-2006

Vogel, H., and K. Roth. 2003. Moving through scales of flow and transport in soil. J. Hydrol. 272:95–106. doi:10.1016/S0022-1694(02)00257-3

Wang, Y., S.A. Bradford, and J. Šimůnek. 2014. Estimation and upscaling of dual-permeability model parameters for the transport of *E. coli* D21g in soils with preferential flow. J. Contam. Hydrol. 159:57–66. doi:10.1016/j.jconhyd.2014.01.009

Weiler, M. 2017. Macropores and preferential flow: A love–hate relationship. Hydrol. Processes 31:15–19. doi:10.1002/hyp.11074

Weiler, M., and J.J. McDonnell. 2007. Conceptualizing lateral preferential flow and flow networks and simulating the effects on gauged and ungauged hillslopes. Water Resour. Res. 43:W03403. doi:10.1029/2006WR004867

Wiekenkamp, I., J.A. Huisman, H.R. Bogena, H.S. Lin, and H. Vereecken. 2016. Spatial and temporal occurrence of preferential flow in a forested headwater catchment. J. Hydrol. 534:139–149. doi:10.1016/j.jhydrol.2015.12.050

Wilson, G.V., J.R. Rigby, M. Ursic, and S.M. Dabney. 2016. Soil pipe flow tracer experiments: 1. Connectivity and transport characteristics. Hydrol. Processes 30:1265–1279. doi:10.1002/hyp.10713

Wrede, S., F. Fenicia, N. Martínez-Carreras, J. Juilleret, C. Hissler, A. Krein, et al. 2015. Towards more systematic perceptual model development: A case study using 3 Luxembourgish catchments. Hydrol. Processes 29:2731–2750. doi:10.1002/hyp.10393

Yu, X., C. Duffy, D.C. Baldwin, and H. Lin. 2014. The role of macropores and multi-resolution soil survey datasets for distributed surface–subsurface flow modeling. J. Hydrol. 516:97–106. doi:10.1016/j.jhydrol.2014.02.055

Zehe, E., and G. Blöschl. 2004. Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions. Water Resour. Res. 40:W10202. doi:10.1029/2003WR002869

Zhang, G.P., H.H.G. Savenije, F. Fenicia, and L. Pfister. 2006. Modelling subsurface storm flow with the representative elementary watershed (REW) approach: Application to the Alzette River basin. Hydrol. Earth Syst. Sci. 10:937–955. doi:10.5194/hess-10-937-2006

Supplemental material

S1 Definition of nonuniform depth profiles

We constrained the value variations for the parameters with soil layer specific values (α, β, K_s, and pct) to a number of pre-defined depth profiles. For each of the parameters we defined several different depth profiles $dp_{xxx}(z)$ (Table S1) following different approaches (DP_{xxx} A-D) that rely on field observations and their different interpretation as described in the following paragraphs. The soil layer specific parameter variations in the MC simulations were then constrained in such a way that the characteristic shapes (value ratios) of the pre-defined depth profiles were maintained, evenly distributing the different pre-defined depth profiles of one parameter to the total of MC runs (Table S2).

$\underline{DP_{cat} \text{ for } \alpha \text{ and } \beta:}$

For the definition of nonuniform profiles of van Genuchten α and β we followed the approach of Rosenbom et al. (2009) and differentiated two categories of macropores: biopores (*b*) and fractures (*f*). We defined one depth profile dp_{cat} for the macropore category for each irrigation plot (DP_{cat}, Table S1 and Table S2). Based on the characteristics observed within the excavated irrigation plots, biopores *b* were assigned to the upper three to five soil layers (characterized by roots in the excavated profiles) and fractures *f* (characterized by stones and periglacial deposits in the excavated profiles) to the soil zones below (Table S1). For each MC run, one value for α and one for β was randomly assigned to biopores (α_b, β_b), and one value each was randomly chosen for fractures (α_f, β_f) (Table 2).

$\underline{DP_{K_s} \text{ for } K_s:}$

Nonuniform depth profiles of saturated hydraulic conductivities K_s of the macropores were assigned following two different approaches (DP_{K_s} A and B), both requiring additional adjunct parameters (Table 2, Table S2).

The first approach (DP_{K_s} A, Table 2) was to calculate K_s from macropore apertures (apert) using the cubic law for fluid flow in fractures (e.g. Witherspoon et al., 1980; Wang et al., 2015). This required the definition of macropore aperture depth profiles (dp_{apert}). We predefined four different aperture depth profiles (DP_{apert} A-D, Table S1) and randomly chose one for each MC run (Table S2) for varying the aperture values along the fixed depth profile ratios (F_{apert}, Table 2). Approach DP_{apert} A was to define the aperture depth profile in such a way that the shape of the resulting conductivity depth profile (calculated with the

1

cubic law) followed the shape of the matrix conductivity profile (with a conductivity peak in the periglacial deposit layers, Soil Zone 4-6, Table S1). In approach DP_{apert} B, we assigned two different apertures (Table S1) for the two different macropore categories (cf. dp_{cat}, Rosenbom et al. (2009)). For the third aperture depth profile, we assigned constant macropore apertures (and thus conductivities) over depth (DP_{apert} C, Table S1). The fourth profile (DP_{apert} D, Table S1) was defined according to profiles that had resulted in good model performance in preliminary test runs (data not shown).

The second approach for defining non-uniform depth profiles of K_s (DP_{Ks} B, Table 2) was to assume that the conductivity and porosity depth profiles used in the model of Glaser et al. (2016) (Table 1) were representative for the total soil, i.e. that their values (based on lab and field measurements) already combined the effect of soil matrix and macropore conductivities and porosities. In order to implement this approach, it was necessary to allow to vary the matrix conductivities and matrix porosities for the respective MC runs (note that all other matrix parameters remained unchanged). We reduced the matrix conductivities of Glaser et al. (2016) (Table1) by a factor between 10^{-4} and 10^{0} (Table 2). Macropore conductivities were then calculated from the reduced matrix conductivities and the chosen macropore percentages pct (see below) in that way that the obtained effective conductivities equaled the conductivities from Glaser et al. (2016) (Table 1). In the same manner, we adapted the matrix porosities (Table 1) by calculating them from the chosen macropore porosities and macropore percentages pct (see below) in such a way that the obtained effective porosities equaled the porosities from Glaser et al. (2016). In case the computations resulted in matrix porosities < 0, we set them to 0.

DP_{pct} for pct:

We used four predefined depth profiles dp_{pct} of macropore percentages pct for each irrigation plot (DP_{pct} A-D, Table S1, Table S2). The definition approaches DP_{pct} A and DP_{pct} B were plot specific (Table S1), based on the Brilliant Blue stains observed in the excavated soil profiles (cf. Figure 1). In approach DP_{pct} A, we assumed that the amount of Brilliant Blue stains corresponds to the amount of macropores. In approach DP_{pct} B, we assumed that more Brilliant Blue adsorbed to the soil when water flow slowed down due to a decrease of macropores below. Approach DP_{pct} C was to assign constant percentages over depth (Table S1). In approach DP_{pct} D (Table S1), we defined values according to profiles that had shown good simulation results in preliminary test runs (data not shown).

2

Table S1. Values for the different soil zones (cf. Figure 2, Table 1) for the predefined non-constant depth profiles $dp_{xxx}(z)$ that were defined following different approaches DP_{xxx} A-D and were used for MC simulation parameter constraints (cf. Table 2).

Soil zone	†DP_{cat} Plot1	†DP_{cat} Plot2	†DP_{cat} Plot3	DP_{apert} (values in m) A	DP_{apert} B	DP_{apert} C	DP_{apert} D	DP_{pct} Plot1 A	Plot1 B	Plot2 A	Plot2 B	Plot3 A	Plot3 B	C	D	DP_{sat} A	B
1	b	b	b	0.0005	0.003	0.0025	0.005	0.150	0.100	0.100	0.050	0.250	0.200	0.100	0.375	0.354	0.484
2	b	b	b	0.0005	0.003	0.0025	0.005	0.050	0.150	0.050	0.100	0.025	0.100	0.100	0.100	0.440	0.406
3	b	b	b	0.0008	0.003	0.0025	0.005	0.100	0.100	0.025	0.100	0.025	0.100	0.100	0.050	0.556	0.390
4	b	b	f	0.003	‡0.003	0.0025	0.003	0.050	0.050	0.100	0.050	0.010	0.050	0.100	0.100	0.772	0.406
5	b	f	f	0.005	‡‡0.0001	0.0025	0.003	0.025	0.025	0.200	0.100	0.050	0.010	0.100	0.100	0.994	0.526
6	f	f	f	0.004	0.0001	0.0025	0.001	0.010	0.010	0.050	0.050	0.025	0.010	0.100	0.100	1.000	0.730
7	f	f	f	0.00025	0.0001	0.0025	0.0005	0.010	0.010	0.010	0.010	0.010	0.010	0.100	0.100	1.000	0.920
8	f	f	f	0.000025	0.0001	0.0025	0.0003	0.010	0.010	0.010	0.010	0.010	0.010	0.100	0.100	1.000	1.000
9	f	f	f	0.00001	0.0001	0.0025	0.0001	0.010	0.010	0.010	0.010	0.010	0.010	0.100	0.100	1.000	1.000
10	f	f	f	0.00001	0.0001	0.0025	0.00005	0.010	0.010	0.010	0.010	0.010	0.010	0.100	0.100	1.000	1.000

† Macropore category b=biopores, f=fractures, ‡ 0.0001 for Plot 3, ‡‡ 0.003 for Plot 1

Table S2. Combination of the different depth profile predefinition approaches (DP_{xxx} A-D, cf. Table S1 and Table 2) within the MC runs. One small box equals 500 model runs, DP_{pct} CD indicates that 250 model runs made use of DP_{pct} C, 250 model runs made use of DP_{pct} D.

| | Plot 1 (irrigation rate 50 mm h⁻¹) | | | | | | Plot 2 (irrigation rate 30 mm h⁻¹) | | | | | | Plot 3 (irrigation rate 50 mm h⁻¹) | | | | | |
	Plot1 A	Plot1 B	Plot2 A	Plot2 B	Plot3 A	Plot3 B	Plot1 A	Plot1 B	Plot2 A	Plot2 B	Plot3 A	Plot3 B	Plot1 A	Plot1 B	Plot2 A	Plot2 B	Plot3 A	Plot3 B
DP_{cat}	A		B		A		A		B		A		A		B		B	
DP_{Ks}	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
DP_{apert}	U(A,B,C,D)						U(A,B,C,D)						U(A,B,C,D)					
DP_{pct}	A B	CD	A B	CD	A B	CD	A B	C D	A B	C D	A B	C D	A B	CD	A B	CD	A B	CD
DP_{sat}	A		B		A		B		A		B		B		A		B	

S2 Effect of daily vs hourly meteorological input data on discharge simulation

In order to ensure that our findings are not a result of a mismatch between the temporal resolution of input data for plot and catchment scale simulations (i.e. event scale simulations for 24 h with hourly input data vs. response simulations over 18 mo with daily input data), we ran six of the catchment scale simulations twice: once with daily meteorological input data and once with hourly input data. The six parameter sets were chosen in such a way that they covered the range of different hydrographs simulated with the daily input data (cf. Fig. 5). More specifically, we re-ran with hourly input data the single domain parameterization and the 5 dual permeability parameterizations that produced the hydrographs presented in Fig. 5.

These six additional simulations clearly showed that the input data resolution had only a minor impact on the behavior of catchment runoff (Fig. S1). The difference between the discharge modelled with daily and hourly input data was smaller than the difference between the observed and modelled runoff (hourly and daily input). Furthermore, the difference between the discharge modelled with daily and hourly input data was also smaller than the differences induced by the different parameter sets for the six selected runs. Thus, the choice of a daily input data resolution for optimizing the calculation times did not impact the results presented in this study.

4
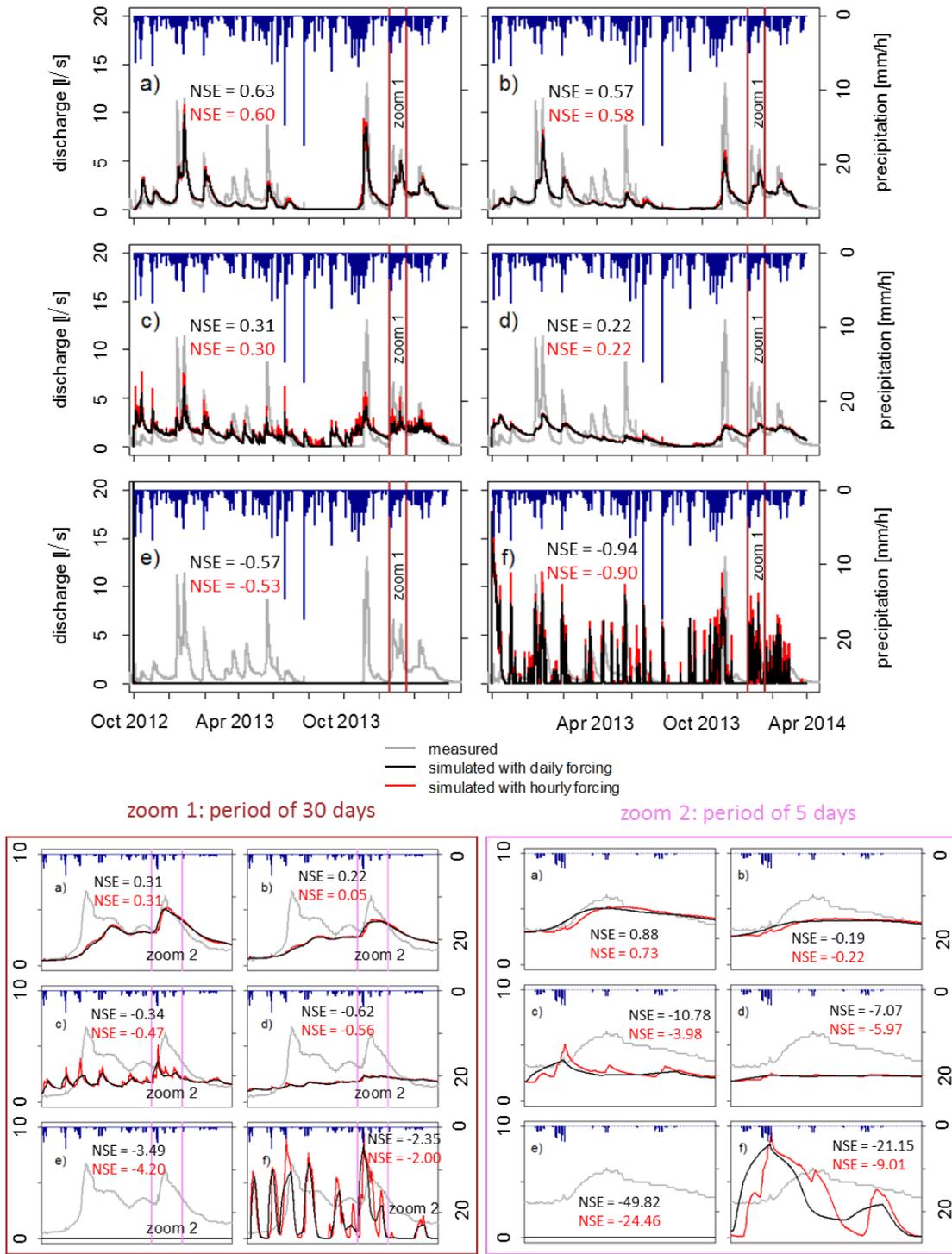
Fig. S1. Comparison of the effect of daily vs hourly meteorological input data on the simulated discharge response over long time (top), one month (bottom left) and five days (bottom right). Simulation a) is the reference simulation without preferential flow, simulations b-f) were parameterized with different dual permeability parameter sets (same as in Fig. 5). For all six cases, the effect of the input data resolution is small compared to the difference to measured discharge. The differences between the hydraographs modelled with the different parameterizations is clearly visible for both input data resolutions.

5

S3 Supplemental result data
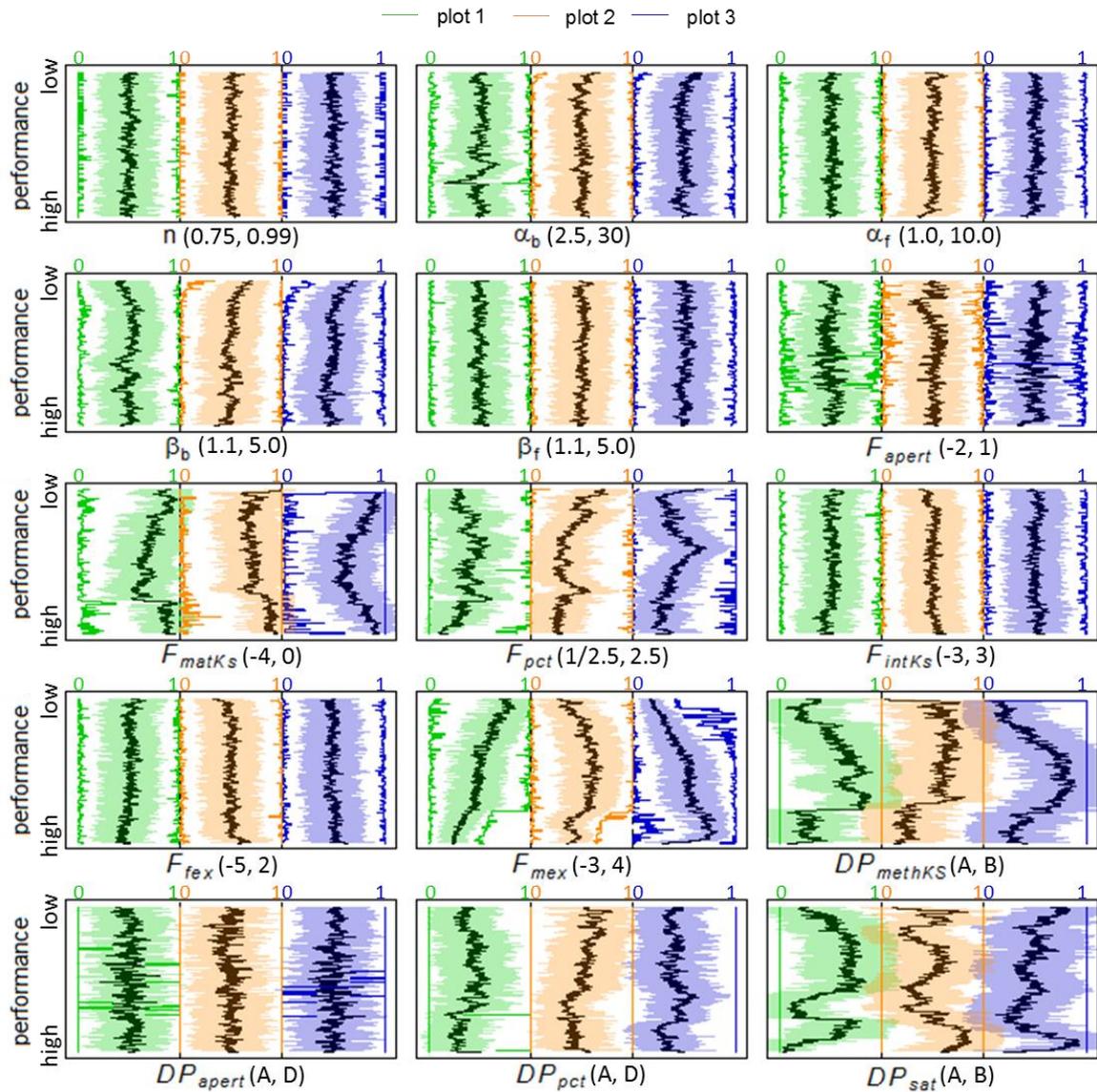
Fig. S2. Relations between the model performance for the irrigation plots and the preferential flow parameters, parameter modification factors, and predefined parameter depth profiles. The parameter values were normalized to the respective pre-constraint value ranges (in brackets) with 0 corresponding to the lowest constraint value or depth profile definition method type A, respectively. The plots show the average (black lines), standard deviation (colored shades) and the minimum and maximum (colored lines) of a moving window with a size of 1% of the MC simulations along the MC simulations sorted by their performance rank number (cf. Section 3.4). Note that the total number of MC simulations is 6000 for Plot 1 and Plot 3 and 8000 for Plot 2

6
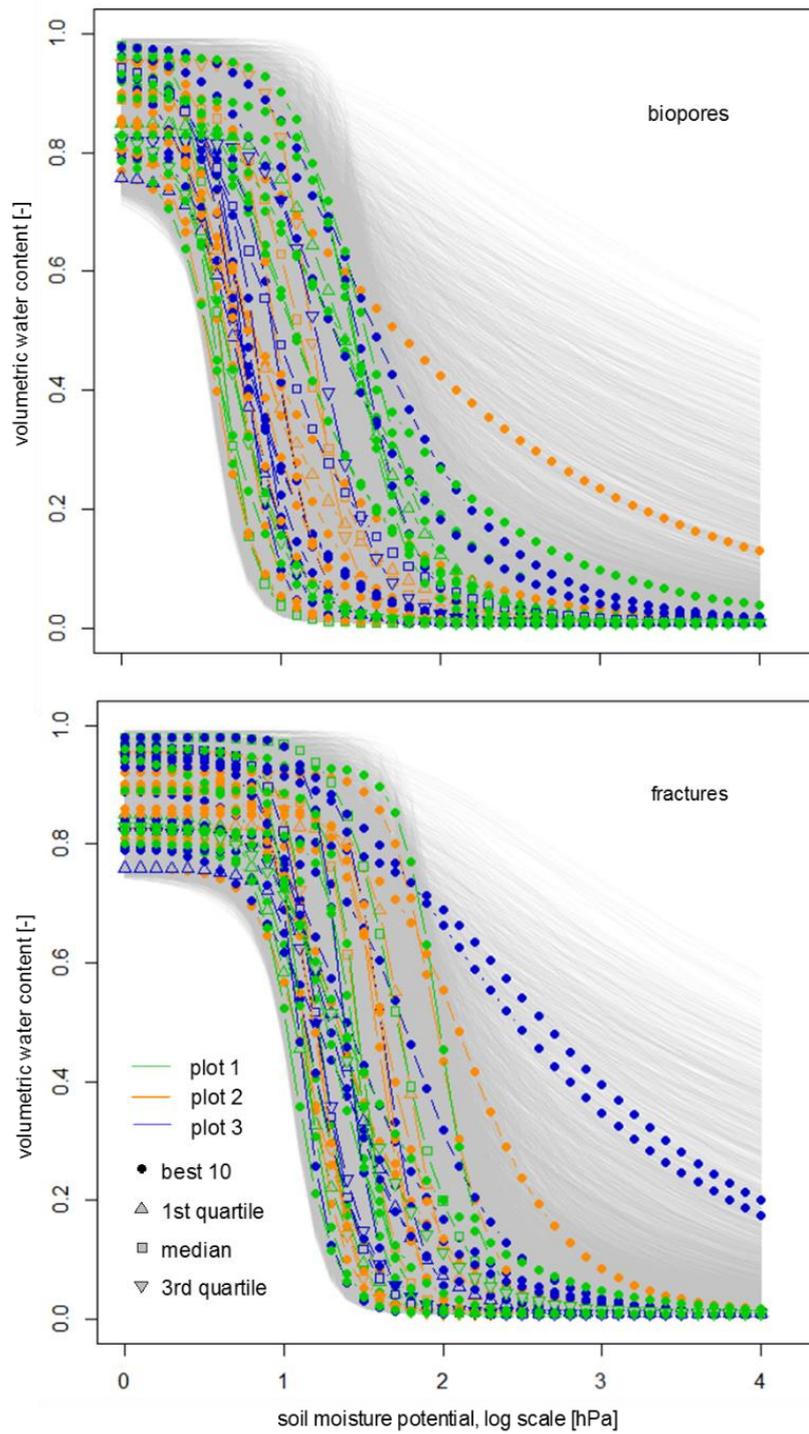
Fig. S3. Water retention curves as parameterized with the water retention parameters α, β, and *n* (cf. Table 2) for the two different macropore types: biopores (top) and fractures (bottom). The semitransparent grey lines depict the water retention curves of all 20,000 plot scale Monte Carlo parameterizations. The colored lines correspond to the 39 parameter sets selected for the catchment scale simulation. The colors of the lines indicate which irrigation plot was simulated with the parameter set at plot scale.The symbols correspond to the performance of the parameter sets at plot scale.

7

Table S3. Contingency table showing the allocation of the ten best + first quartile, median, and third quartile performing parameter sets of irrigation plot 1 (green), 2 (orange) and 3 (blue) to the different approaches applied for predefining depth profiles DP_{xxx} A-D (cf. Table 2, Section S1). Each cell refers to one specific method and irrigation plot; the number left of '+' indicates the number of runs from the ten best performing parameter sets, the number right of '+' indicates the number of runs from the first quartile, median, and third quartile performing parameter sets.

	[†]DP_{Ks}			[†]DP_{apert}			[‡]DP_{pct}			[‡‡]DP_{sat}		
	Plot 1	Plot 2	Plot 3	Plot 1	Plot 2	Plot 3	Plot 1	Plot 2	Plot 3	Plot 1	Plot 2	Plot 3
A	10+1	10+1	8+1	2+0	6+0	4+0	0+3	0+1	5+1	8+3	8+1	10+2
B	0+2	0+2	2+2	1+0	1+0	1+0	2+0	7+0	1+1	2+0	2+2	0+1
C	----	----	----	3+1	2+0	2+1	6+0	2+1	0+1	----	----	----
D	----	----	----	4+0	1+1	1+0	2+0	1+1	4+0	----	----	----

[†]DP_{Ks}: two methods for defining $K_s(z)$: A relied on pre-defined depth profiles $dp_{apert}(z)$ of macropore apertures, which were defined following four different pre-definition methods DP_{apert} A-D
B relied on maintaining effective hydraulic conductivities as $K_{s\text{-original}}(z)$, thus requiring adapted matrix conductivities $K_{s\text{-matrix}}(z)$

[‡]DP_{pct}: four methods for defining a depth profile $dp_{pct}(z)$: A and B relied on different interpretations of dye patterns, C assigned a global value, D relied on preliminary test simulations

[‡‡]DP_{sat}: two methods for defining a depth profile $dp_{sat}(z)$ of initial saturation: A was based on soil moisture observations, B was based on soil moisture simulations

8
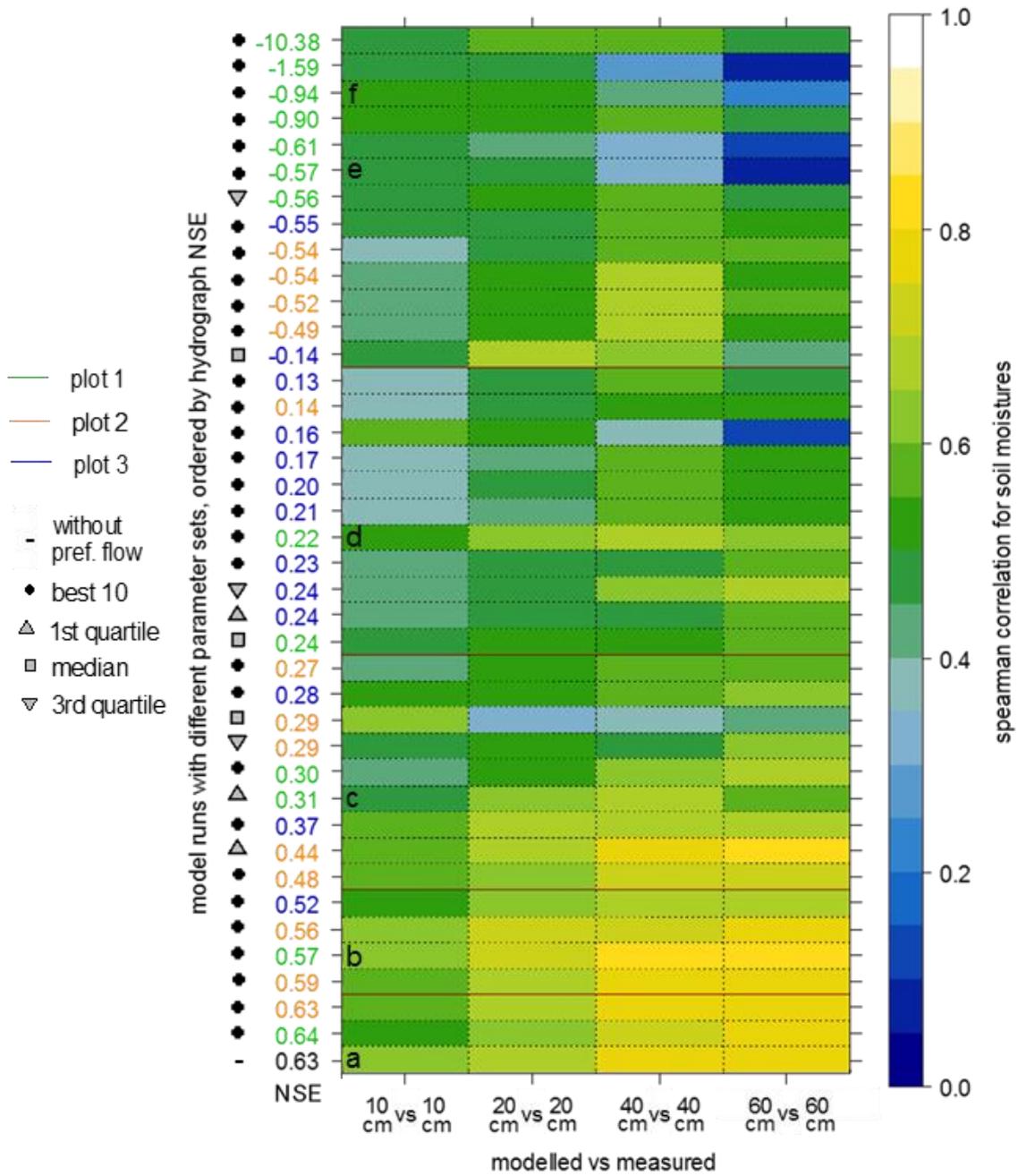
Fig. S4. Heatmap showing the correlations between observed and simulated soil moistures. Each line corresponds to one simulation, correlation is shown on a color scale (white = high correlation, blue = no correlation). The correlation values indicate moisture dynamic similarities between observed and simulated moisture for different depths. Lines a-f show the correlations for the simulated moisture time series shown in Fig. 7. The different simulations are sorted according to their hydrograph NSE (cf. Fig. 6). The colors of the NSE values indicate which irrigation plot was simulated with the parameter set at plot scale.The symbols correspond to the performance of the parameter sets at plot scale.
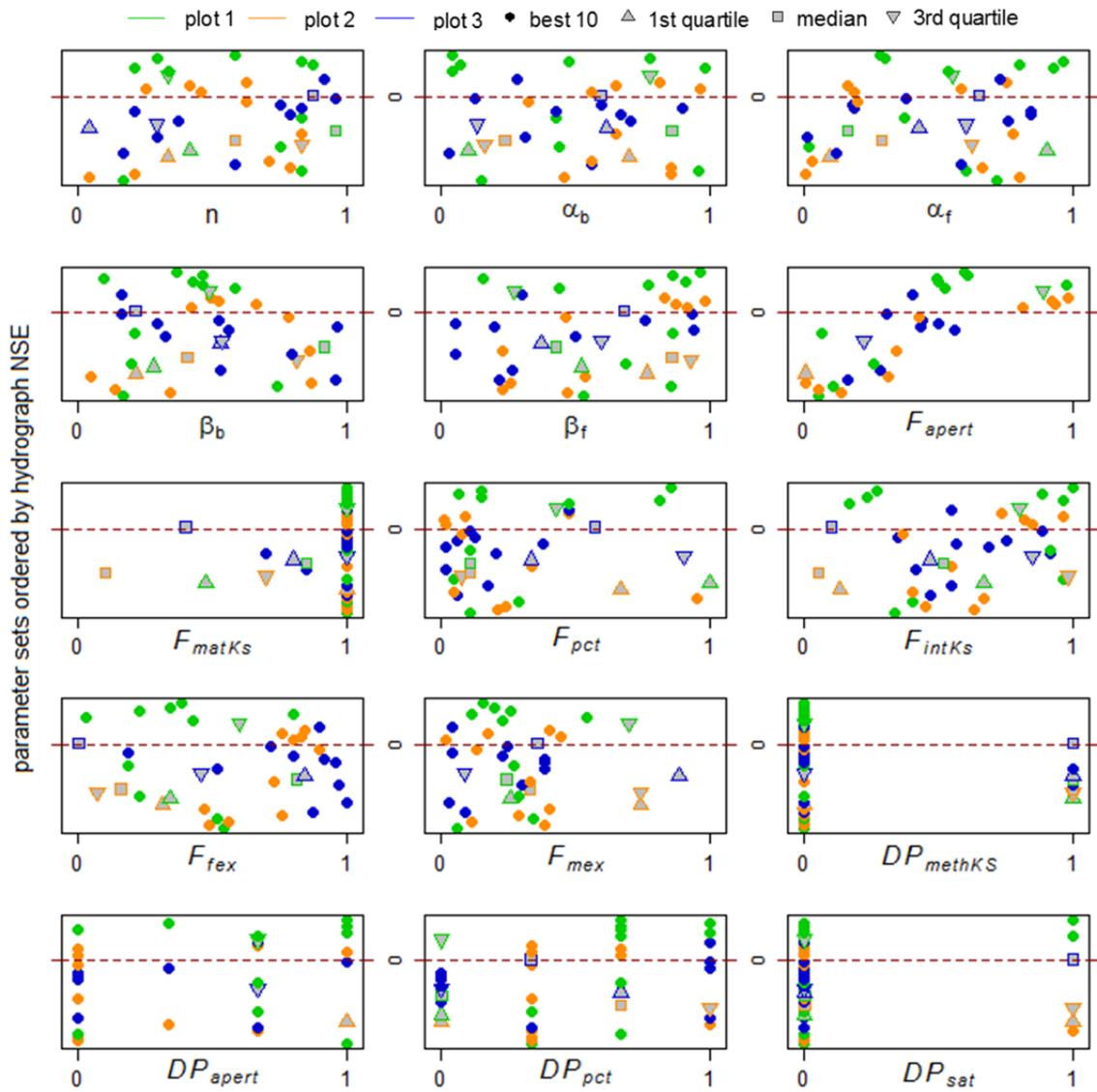
Fig. S5. Distribution of all modified parameters, parameter modification factors and predefined depth profiles within the catchment scale simulations. Parameter values were normalized to the respective constraint ranges (0 corresponding to the lowest constraint value or depth profile type A, respectively). Data points are sorted along the y-axis according to the hydrograph NSE of the parameter sets (cf. Fig. 6, Fig. 8). Data points above the red line correspond to parameter sets that resulted in a hydrograph NSE < 0, data points below the red line correspond to parameter sets that resulted in a hydrograph NSE > 0. The colors of the data points indicate which irrigation plot was simulated with the parameter set at plot scale.The symbols correspond to the performance of the parameter sets at plot scale.

10

References

Glaser, B., J. Klaus, S. Frei, J. Frentress, L. Pfister, and L. Hopp. 2016. On the value of surface saturated area dynamics mapped with thermal infrared imagery for modeling the hillslope-riparian-stream continuum. Water Resour. Res. 52: 1–26. doi: 10.1002/2015WR018414.

Rosenbom, A.E., R. Therrien, J.C. Refsgaard, K.H. Jensen, V. Ernstsen, and K.E.S. Klint. 2009. Numerical analysis of water and solute transport in variably-saturated fractured clayey till. J. Contam. Hydrol. 104: 137–152. doi: 10.1016/j.jconhyd.2008.09.001.

Wang, L., M.B. Cardenas, D.T. Slottke, R. a. Ketcham, and J.M. Sharp. 2015. Modification of the Local Cubic Law of fracture flow for weak inertia, tortuosity, and roughness. Water Resour. Res. 51: 1–17. doi: 10.1002/2014WR015815.

Witherspoon, P.A., J.S.Y. Wang, K. Iwai, and J.E. Gale. 1980. Validity of Cubic Law for fluid flow in a deformable rock fracture. Water Resour. Res. 16(6): 1016–1024. doi: 10.1029/WR016i006p01016.

Study 2: Technical note: Mapping surface-saturation dynamics with thermal infrared imagery

Authors: Barbara Glaser, Marta Antonelli, Marco Chini, Laurent Pfister, Julian Klaus

BG, MA, LP and JK designed and directed the study. BG and MA planned and carried out the field work. BG, MA and MC processed the TIR images. BG, MA, MC and JK discussed and interpreted the results. BG prepared the figures. BG prepared the manuscript with input from all co-authors. BG is the corresponding author.

Own contribution in %:

- Study concept and design:	65
- Field data acquisition:	50
- Image processing:	50
- Preparation of figures:	100
- Interpretation of the results:	75
- Preparation of the manuscript:	85

Hydrology and Earth System Sciences

Technical note: Mapping surface-saturation dynamics with thermal infrared imagery

Barbara Glaser[1,2]**, Marta Antonelli**[1,3]**, Marco Chini**[4]**, Laurent Pfister**[1,5]**, and Julian Klaus**[1]

[1]Catchment and Eco-Hydrology Research Group, Luxembourg Institute of Science and Technology,
5, avenue des Hauts-Fourneaux, Esch-sur-Alzette, 4362, Luxembourg
[2]Department of Hydrology, University of Bayreuth, Universitätsstr. 30, Bayreuth, 95447, Germany
[3]Hydrology and Quantitative Water Management Group, Wageningen University & Research, Droevendaalsesteeg 3a,
Building 100, 6708 PB, Wageningen, 6700, the Netherlands
[4]Remote Sensing and Ecohydrological Modelling Research Group, Luxembourg Institute of Science and Technology, 5,
avenue des Hauts-Fourneaux, Esch-sur-Alzette, 4362, Luxembourg
[5]Faculty of Science, Communication and Technology, University of Luxembourg, Maison du Savoir,
2 Avenue de l'Université, 4365 Esch-sur-Alzette, Luxembourg

Correspondence: Barbara Glaser (barbara.glaser@list.lu)

Abstract. Surface saturation can have a critical impact on runoff generation and water quality. Saturation patterns are dynamic, thus their potential control on discharge and water quality is also variable in time. In this study, we assess the practicability of applying thermal infrared (TIR) imagery for mapping surface-saturation dynamics. The advantages of TIR imagery compared to other surface-saturation mapping methods are its large spatial and temporal flexibility, its non-invasive character, and the fact that it allows for a rapid and intuitive visualization of surface-saturated areas. Based on an 18-month field campaign, we review and discuss the methodological principles, the conditions in which the method works best, and the problems that may occur. These considerations enable potential users to plan efficient TIR imagery-mapping campaigns and benefit from the full potential offered by TIR imagery, which we demonstrate with several application examples. In addition, we elaborate on image post-processing and test different methods for the generation of binary saturation maps from the TIR images. We test the methods on various images with different image characteristics. Results show that the best method, in addition to a manual image classification, is a statistical approach that combines the fitting of two pixel class distributions, adaptive thresholding, and region growing.

1 Introduction

The patterns and dynamics of surface-saturation areas have been on hydrological research agendas ever since the formulation of the variable source area (VSA) concept by Hewlett and Hibbert (1967). Surface saturation is relevant for runoff generation and for water quality, due to variable active and contributing areas (Ambroise, 2004) as well as critical source areas (e.g. Doppler et al., 2014; Frey et al., 2009; Heathwaite et al., 2005). Likewise, surface-saturation patterns and their dynamics are closely linked to groundwater–surface-water interactions (e.g. Frei et al., 2010; Latron and Gallart, 2007) and catchment storage characteristics and dynamics (e.g. Soulsby et al., 2016; Whiting and Godsey, 2016).

Despite the prominent role of saturated areas in hydrological processes research, mapping them remains a challenging exercise. The most straightforward mapping method consists of locating saturated areas by walking through the catchment. However, this simple but labour-intensive "squishy-boot" method (e.g. Blazkova et al., 2002; Creed et al., 2003; Latron and Gallart, 2007; Rinderer et al., 2012) is neither suitable for large areas nor for fine-scale spatial resolutions. Dunne et al. (1975) introduced topography, soil morphology, hydrometric measurements (soil moisture, water table level, base flow), and vegetation as useful indicators for delineating saturated areas. Today, it is still a valid research question

of how to best make use of these catchment characteristics to delineate saturated areas (e.g. Ali et al., 2014; Doppler et al., 2014; Grabs et al., 2009; Kulasova et al., 2014a, b). Hydrometric measurements offer the potential for monitoring the local temporal evolution (in increments ranging from minutes to months) of dynamic surface saturation. The analysis of topography, soil morphology, or vegetation allows lasting saturation patterns to be identified for large contiguous areas.

Remote sensing has proven to be well-suited for mapping temporal dynamic patterns of surface saturation over large areas. It is possible to extract flooded areas in the order of metres to kilometres from data acquired with satellite and airborne platforms, such as synthetic aperture radar (SAR) images (e.g. Matgen et al., 2006; Verhoest et al., 1998), or the normalized difference water index (NDWI) and the normalized difference vegetation index (NDVI; de Alwis et al., 2007; Mengistu and Spence, 2016). Observations at higher spatial resolutions (order of centimetres) require unmanned aerial vehicles (UAVs) or ground-based instruments. Due to various technical constraints, to date, SAR image acquisitions are rarely used for UAV-based applications or for ground-based applications that are not restricted to a fixed location (e.g. Li and Ling, 2015; Luzi, 2010). NDWI and NDVI are applicable at these scales (e.g. Orillo et al., 2017; Wahab et al., 2018), however, to the best of our knowledge, the necessary simultaneous acquisition of short-wave infrared and visible light (VIS) images has not yet been performed by UAVs or on the ground for mapping surface saturation.

Ishaq and Huff (1974) and Dunne et al. (1975) suggested the use of VIS or infrared photographs for mapping surface saturation. However, this suggestion has rarely been followed in the last 40 years (with Portmann, 1997, being a notable exception), despite VIS cameras having been deployed on the ground and mounted on UAVs, airborne platforms, or satellite platforms for a long time. Recently, Chabot and Bird (2013) and Spence and Mengistu (2016) successfully used VIS cameras mounted on UAVs for mapping surface water (a wetland of 128 ha and an intermittent stream surveyed via three transects of 2 km each). Silasari et al. (2017) mapped surface-saturated areas on an agricultural field (100 m × 15 m) using a VIS camera mounted on a weather station for high-frequency image acquisition.

Since the advent of affordable, handheld thermal infrared (TIR) cameras, TIR imagery features the same temporal and spatial flexibility as VIS imagery. In the context of this technical advancement, TIR imagery started to be used for analysing hydrological processes such as groundwater–surface-water interactions (e.g. Ala-aho et al., 2015; Briggs et al., 2016; Pfister et al., 2010; Schuetz and Weiler, 2011) or water flow paths, velocities, and mixing (e.g. Antonelli et al., 2017; Deitchman and Loheide, 2009; Schuetz et al., 2012). However, applications of TIR imagery for mapping surface saturation are rare. Two examples are from Pfister et al. (2010) and Glaser et al. (2016), who demonstrated the potential for TIR imagery to map surface saturation by carrying

out repeated TIR image acquisitions at small spatial scales (centimetres to metres) with handheld cameras.

One reason for the scarce number of studies that use TIR imagery for mapping surface saturation is certainly that few descriptions of the methodological advantages and challenges exist. However, there are several general guidelines and methodological descriptions for TIR imagery applications. These studies focus on one specific aspect of TIR imagery, such as co-registration (Turner et al., 2014; Weber et al., 2015) or on how to acquire correct surface water temperatures, which is the most common application of TIR imagery in hydrology (e.g. Dugdale, 2016; Handcock et al., 2006, 2012; Torgersen et al., 2001). Many of these recommendations can be directly applied for mapping surface saturation via TIR imagery (e.g. choice of sensor type). However, some recommendations are redundant (e.g. temperature corrections) or different (e.g. optimal time scheduling) for the application of TIR imagery for surface-saturation mapping.

Here, we go beyond the mere demonstration of the potential for TIR imagery to map saturated surface areas and address the related application-specific technical and methodological challenges. The novelty of this work is that we assimilate, within one study, fundamental principles, technical aspects, and methodological possibilities and challenges with an exclusive focus on the mapping of surface saturation. This includes all steps, from image acquisition to the generation of binary saturation maps. To do this, we (1) review relevant technical and methodological aspects from existing TIR imagery literature and (2) complement them with our expertise and results from an 18-month field campaign. The field campaign focused on the recurrent acquisition of panoramic images with a portable TIR camera in seven distinct riparian areas. The precautions and considerations that we describe in this technical note are also valid for surface-saturation mapping campaigns with permanently installed ground-based TIR cameras and TIR cameras mounted on UAVs and airborne or satellite platforms.

The paper is structured in two main parts. The first part (Sect. 2) focusses on the mapping approach itself and combines a literature review with examples of our own experience. The second part (Sect. 3) demonstrates the application of different pixel classification techniques for generating binary saturation maps from TIR images by applying and comparing them for different example images. A discussion and a conclusion section evaluate the key features of the paper and outline perspectives for future research and applications for TIR imagery in hydrological sciences.

2 Mapping surface saturation with TIR imagery: state of the art and examples

2.1 Fundamental principles

TIR cameras are used for measuring surface temperatures remotely (e.g. 100 µm penetration depth for water columns) within an area of interest. The cameras sense the intensity of thermal infrared radiation emitted by the objects the camera is pointed at. The surface temperature T (K) of the objects is then calculated from the sensed radiant intensity W (Wm^{-2}), based on Stefan–Boltzmann's law with the Stefan–Boltzmann constant, $\sigma = 5.67 \times 10^{-8}$ Wm^{-2} K^{-4}. This law can be formulated as

$$T = \sqrt[4]{(W/\sigma)}. \tag{1}$$

Considering radiometric corrections for material-specific emissivity ε, for reflections of radiation from the surroundings, and for atmospheric induced and attenuated radiation, the radiant intensity W is split into the emissions from the object (W_{obj}), from the ambient sources (W_{refl}), and from the atmosphere (W_{atm}).

$$W = \varepsilon \tau W_{obj} + (1 - \varepsilon) \tau W_{refl} + (1 - \tau) W_{atm}, \tag{2}$$

with τ being the transmittance of the atmosphere, which depends on the distance between the object and the camera sensor as well as on relative air humidity. Ultimately, values for the temperature of the ambient sources and the atmosphere, the targeted object's emissivity, the distance between object and camera, and the relative humidity are required for accurately estimating an object's surface temperature T.

Details on the principles of TIR imagery, TIR sensor types (i.e. wave length, sensitivity), and considerations for choosing the most appropriate camera and remote sensing platform for the desired acquisition (i.e. accuracy, resolution) are provided in the literature (cf. Dugdale, 2016; Handcock et al., 2012). For this study, we relied on two different hand-held TIR camera models: a FLIR B425 with a resolution of 320×240 pixels and an angle of view of 25° and a FLIR T640 with a resolution of 640×480 pixels and an angle of view of 45° (FLIR Systems, Wilsonville, USA). The wider angle of view of the FLIR T640 clearly facilitated the image acquisition in this study, while a pixel resolution lower than the resolutions of the two cameras would still have been sufficient for the identification of surface-saturation patterns.

We define surface saturation as water ponding or flowing on the ground surface (even if only present as a very thin layer). Mapping surface saturation with TIR imagery requires (1) a sufficient temperature contrast between surface water and the surrounding environment (e.g. dry soil, rock, vegetation) and (2) at least one pixel of the TIR image being known to correspond to surface water. When these two requirements are met, it is possible to visually identify the surface-saturation patterns in a TIR image. This is exemplified with a TIR image of a riparian-stream zone (Fig. 1). The substantial temperature contrast (requirement 1) allows us to differentiate between two TIR pixel groups, i.e. surface water pixels and surrounding environment pixels. With ground truth data at hand (here, VIS image – alternatives include stream-water temperature or knowing the location of the creek) for point 1 of Fig. 1 (requirement 2), the group of pixels with higher temperatures can be identified as surface water. The group of pixels with lower temperatures can be regarded as the non-saturated surrounding environment (cf. Fig. 1; point 2). With this classification in mind, the TIR image significantly amplifies the appearance of surface-saturated areas relative to a VIS image. Moreover, the TIR image reveals additional surface-saturated areas that are not clearly identifiable (cf. point 3; Fig. 1) or not visible (cf. area above point 6; Fig. 1) within a VIS image.

The example shows that the identification of surface saturation relies on temperature contrasts between surface water and the surrounding environment. Radiometric corrections of TIR images for obtaining correct temperature values are thus not necessary. However, interferences that affect temperature, such as shadow casts or reflections (cf. Dugdale, 2016; Handcock et al., 2012), cannot be disregarded, as they can influence the temperature contrast (see Sect. 2.2). In cases where the water temperature is too similar to the surrounding materials, saturated areas might be falsely identified as dry, whereas surrounding materials might be falsely identified as wet. In cases where non-uniform water temperatures occur, different water sources may be distinguished (cf. Fig. 1, where point 4 likely represents stream water, points 5 and 7 likely represent the exfiltration of warmer groundwater). However, a bimodal distribution of water temperatures (e.g. cold stream and warm exfiltrating groundwater or warm ponding water) can also lead to a misinterpretation of temperature contrasts to the surrounding environment (e.g. a surrounding material with a temperature that is in between the water temperatures might be identified as water).

For the above-mentioned reasons, it is important to evaluate the applicability of the TIR images for identifying the surface-saturated areas with some ground truth and validation data. For the validation, we relied on immediate visual verification during image acquisition as well as on VIS images. Another option is to install sensors that can verify the presence or absence of water on the ground surface locally, yet this is an experimental effort and only results in validation data for selective points. Validating the TIR images with other saturation mapping techniques is difficult, since most of these techniques implicitly include saturation in the upper soil layer, while the current use of TIR imagery excludes the soil. For example, saturated areas inferred via the squishy boot method account for areas where water is squeezed out of the soil when stepping on it, whereas such areas are not detected as saturated areas by the non-invasive TIR imagery.
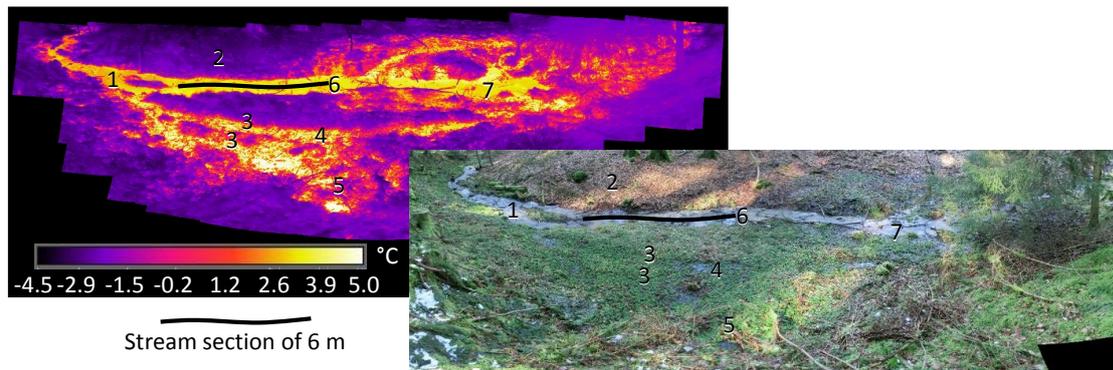
Figure 1. TIR image and VIS image of a riparian-stream zone. The temperature contrast between the water and the surrounding environment allows us to clearly differentiate between surface-saturated and dry areas in the TIR image. The numbers indicate identical locations in the TIR and VIS images and relate to dry areas (2), stream water (1, 4, 6), points of supposed groundwater exfiltration (5, 7: warmer water temperatures), and locations in which surface saturation is clearly visible in the TIR image but not in the VIS image (3, area above 6).

2.2   Image acquisition interferences

Impact of weather conditions

Weather conditions can interfere with TIR image acquisition (e.g. Dugdale, 2016; Handcock et al., 2012). The main problem stems from the similar temperatures of water and the surrounding environment, compromising an identification of surface saturation with TIR images (Fig. 2a). Water has a higher thermal capacity than most environmental materials, and the water surface temperature therefore generally aligns more slowly with the air temperature than the surface temperatures of surrounding materials. During our field campaign, it became clear that, particularly during day–night–day or seasonal transitions, this difference in thermal capacities induced a convergence of the surrounding environment's temperatures (which align to the air temperature) to the water temperature. Furthermore, the direct exposure of the study site to sunlight, combined with shadow casts, commonly distorted the temperature contrasts. Surrounding materials in the shade with temperatures different to the same surrounding materials in sunlight led to reduced temperature contrasts between these materials and the surface water (Fig. 2b). Once the direct sun exposure ceased, the different thermal capacities of different materials heated by the sun could still cause patches of warmer and colder temperatures. Rain and fog may also influence image quality due to water droplets falling between the TIR sensor and the ground, eventually blurring the images and causing uniform temperature signatures (Fig. 2c).

To avoid the acquisition of unusable TIR images, we advise to adapt the planning of field campaigns to the weather forecasts. The ideal situation is to work during dry weather with warm or cold air temperatures in order to ensure a clear difference between the temperature of the surrounding materials and the more temperate surface water temperatures.

Dugdale (2016) reported the time period from mid-afternoon to night-time as an optimal TIR image acquisition period for monitoring water surface temperatures. Based on our 18-month field campaign, we suggest that the optimal TIR image acquisition time for identifying surface-saturation patterns is early morning. At this time, there are no undesirable effects due to sunlight (shadows, warming-up), and there are generally high temperature contrasts between water surfaces and the surrounding environment. Cloudy conditions can also help to avoid the effect of direct sunlight. A site-specific analysis of the sun exposure throughout the day can help pinpoint the other times at which images can be taken in favourable conditions for a specific study site.

Camera position

Obstructions in the TIR camera's field of view are obviously problematic. Yet permanent view obstructions on the ground (e.g. tree trunks; Fig. 2d, point 6) proved to be useful ground reference points during our field campaign. Temporary view obstructions, such as growing vegetation (Fig. 2d), recent litter, and snow cover are a problem for repeated imaging campaigns. Cutting the vegetation during the growing season is an option for small study sites. Our experience is that the coverage of grasses and herbaceous plants with small leaves is normally low enough to permit the recording of the ground surface temperature, while the coverage of ferns or tree leaves is normally completely opaque. Snow cover usually hides surface saturation. Yet periods where the amount of snow is low are commonly unproblematic, since the saturated areas mainly stay uncovered due to a warmer water temperature and thus the fast melting of the snow.

Ideally, images are taken from above and at nadir to the study site. Oblique angles of view ($>30°$ of nadir) reduce the object's emissivity and thus distort the detected temperatures in the TIR images (Dugdale, 2016). The incorrect tem-
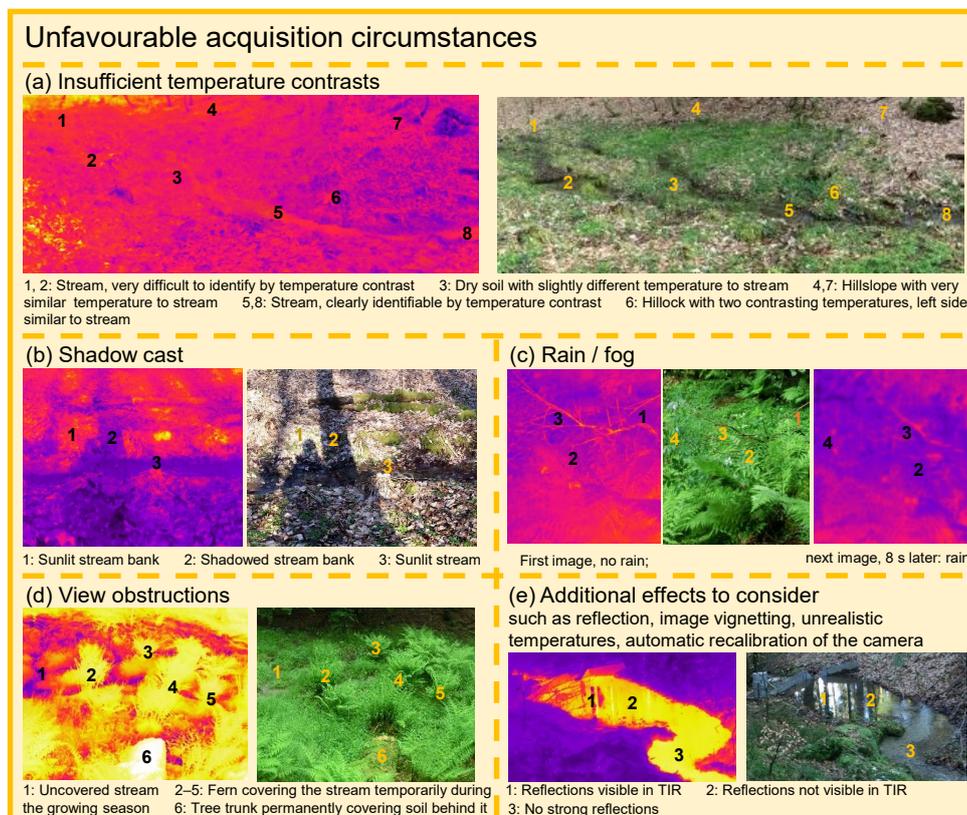
Figure 2. Example images showing how unfavourable image acquisition circumstances influence the usability of TIR imagery for the identification of surface saturation. The numbers indicate identical locations in the TIR and VIS images.

perature values are not critical as such for mapping surface-saturation patterns, but we observed that wide ranges of angles can result in distinct temperature distortions and thus reduced temperature contrasts within the images. In a similar way, varying distances between camera and ground surface for different positions within one image (e.g. top and bottom, left and right) not only provoke pixels with varying area equivalents but can also distort the temperature detection and thus temperature contrasts. Therefore, ground-based cameras should be positioned at locations that minimize the range of angles of view and the distances between camera and ground surface. In the event of a repeated image acquisition of a given area of interest, we took the pictures from the same position each time in order to facilitate subsequent image comparisons. For repeated image campaigns, it could be useful to install a structure that allows several images to be acquired by moving the camera to specific positions with fixed heights above the ground and fixed angles of view. This could simplify the post-processing and assemblage of the images into panoramic images (cf. Sect. 2.3).

Measurement artefacts during image acquisition

For determining surface saturation, the TIR images should cover an area known to be surface saturated (e.g. stream, visually obvious wet spots) in order to have a reference for water temperature (cf. Sect. 2.1). In addition, a VIS image should be acquired simultaneously to the TIR image for comparison. The TIR imagery parameters necessary for correcting and converting the radiation signal to temperature values (e.g. air temperature, humidity) do not need to correspond to the actual conditions, since only the temperature contrast, and not the correct temperature value, is required for defining saturated areas. Certainly, "wrong" temperatures influence the temperature contrast between the surroundings and the water, but this effect on the contrast can be negative or positive. If correct temperatures are targeted, radiometric corrections need to be applied during the image post-processing procedure. This allows, for example, for the consideration of different emissivities for different surface materials by using appropriate values for each individual image pixel (Aubry-Wake et al., 2015). However, in our experience, setting realistic parameter values during the image acquisition helped

the auto-focus process of the camera and prevented the observation of unrealistic surface temperatures. Nonetheless, in the event of clear skies or on cold winter days, we occasionally observed negative temperatures for flowing water. The explanations for these observations remain speculative. Potentially, a particularly strong reflection of the radiation from the surroundings and the sky in the water influenced the temperature detection. However, for the identification of surface-saturation patterns, such unrealistic negative temperatures do not pose a problem, since the temperature range stays correct (Antonelli et al., 2017).

Reflections of surrounding objects on the water surface (Fig. 2e) and image vignetting can occur during image acquisition and can compromise a further use of the TIR images. Vignetting is the falloff of radiation intensity towards the edges of the image, which is mainly generated by the geometry of the sensor optics (especially wide angle lenses; cf. Kelcey and Lucieer, 2012). As a consequence, the monitored temperature can change towards the edges of the picture (cf. aura effect in Antonelli et al., 2017). In this study, the image vignetting was unproblematic, especially where a panorama was built from several images (cf. Sect. 2.3). This is due to the fact that the effect of image vignetting only occurs at the edges of the pictures and it is of minor relevance in images with high temperature contrasts. Reflections of surrounding objects on the water surface limit the value of the images for saturation identifications in a similar way to shadows (cf. Fig. 2d and e). The difference with shadows is that reflections also occur with diffuse light, which makes it difficult to predict their occurrence and thus to avoid them.

2.3 Generation of TIR panorama images

We acquired the images used for the assemblage of a panoramic view in two different ways: (1) by taking single, overlapping images and (2) by taking a video of the area of interest. While both approaches deliver similar final results, videos are recorded faster than sequences of individual images. Independently from the chosen data format, we ensured that the saving format retained the temperature information as radiometric data for further image processing (see below and Fig. 3). Sun disappearance and appearance and automatic noise corrections by the camera (non-uniformity corrections; cf. Dugdale, 2016) can lead to considerable shifts in recorded temperatures from one image or video frame to another. Since correcting such temperature shifts is difficult (cf. Dugdale, 2016), we opted to control them by fixing the temperature–colour scale and restarting image acquisition if the colour (and thus temperature) of overlapping image parts changed.

We acquired the images and video frames in such a way that the area of interest formed the central part of a panorama. This allowed us to avoid image gaps and distortion effects at the borders of the area of interest. When possible, we ensured that the single pictures and video frames included over-

lapping parts with identifiable structures, such as the stream bank, tree stems, or stones, as natural reference points. For videos, it was essential to move the camera slowly enough to obtain sharp images and to use a low frame rate (e.g. 2 Hz) to keep the number of video frames reasonable (enough frames for obtaining area overlaps, but not too many frames showing the same area).

The generation of a panorama from overlapping TIR images or video frames acquired with a ground-based camera involves some challenges that specifically relate to TIR or ground-based images. This needs to be addressed in TIR-specific panorama generation and image processing steps, as presented briefly by Cardenas et al. (2014). Our approach consisted of transforming the acquired images and video frames containing the radiometric information (see above) into grey-scaled, standard-format images and videos (Fig. 3, step 1) in order to allow for the use of ordinary panorama assemblage software. We relied on grey-colour-scale images, linearly splitting the colour shades over the global temperature range of the acquired images and video frames, since this prevents the creation of artefacts by colour-mixing effects and allowed us to embed the temperature information in the generated panoramas. When the extreme temperature values of an image were not relevant for the identification of saturated areas, we truncated the global temperature range in favour of a better colour contrast and a finer temperature class width retained in the grey values (e.g. the retained temperature class width is 0.1 °C in case of a temperature range of 25.5 °C and an image with 255 grey values).

We employed Microsoft's Image Composite Editor (ICE) and the PTGui panorama software (New House Internet Services) to create panorama images (Fig. 3, step 2). ICE and PTGui allow for the creation of panoramas from single images (and from video frames for ICE) with an automatic mosaicking function (i.e. a function that geometrically transforms, aligns, and overlaps the single images). TIR images generally show less identifiable features and lower contrasts than VIS images (cf. Weber et al., 2015). Therefore, a (partial) failure of automatic mosaicking is not uncommon, and manual interactions with image alignment (i.e. defining control points for matching distinct points in overlapping images in PTGui) were frequently necessary for the TIR images taken during our 18-month field campaign.

In order to compare several panorama images of the same area, one needs to co-register the panoramas (Fig. 3; step 3). In principle, it is possible to geo-rectify the TIR images by allocating geographical coordinates to the images, which are derived from ground control points (cf. Keys et al., 2016; Silasari et al., 2017) or from a virtually projected elevation model (cf. Cardenas et al., 2014; Corripio, 2004; Härer et al., 2013). However, this can result in large gaps or strong interpolations and distortions in the images, due to view obstructions in the picture. Instead of this, therefore, we co-registered TIR panoramas of the same area against each other (cf. Cardenas et al., 2014; Glaser et al., 2016). More specif-
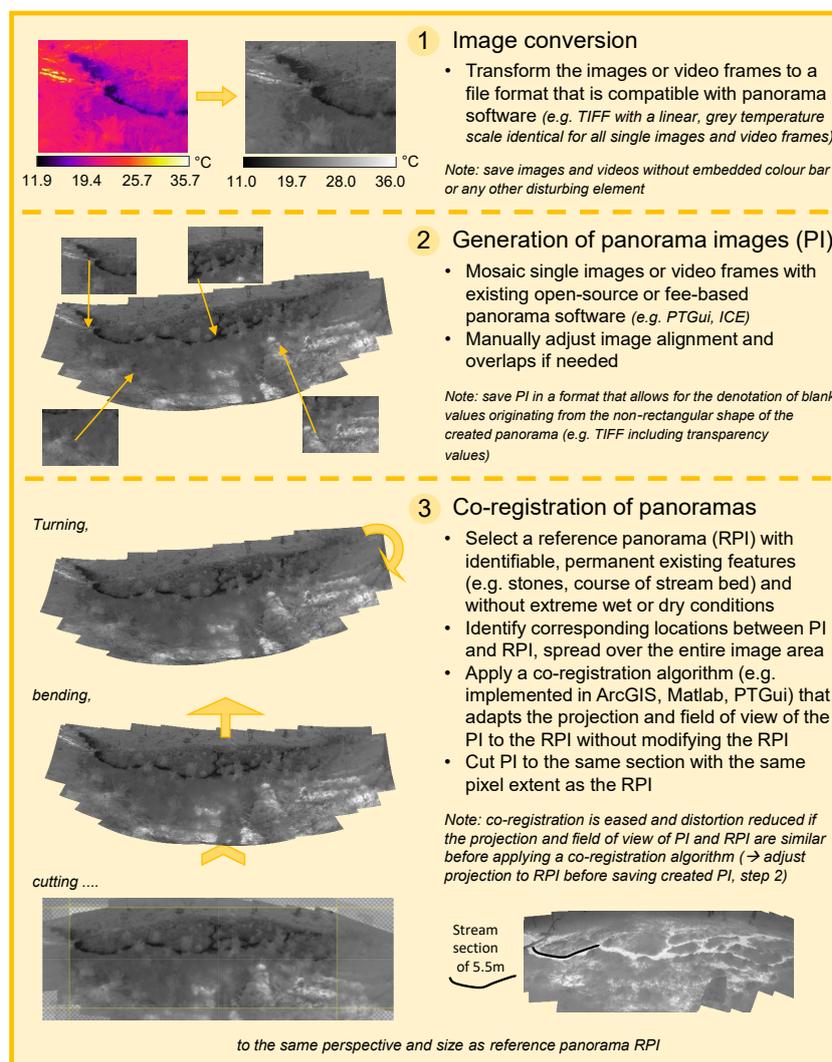
**Figure 3.** Workflow for processing single TIR images and video frames to co-registered panoramic images.

ically, we registered and cropped them to the dimensions of a reference TIR panorama of the area of interest (Fig. 3; step 3).

2.4 Application examples

In this section, we present three examples from our 18-month field campaign that demonstrate the potential for TIR imagery to analyse surface-saturation patterns and their dynamics. All images were taken in the Weierbach catchment – a forested, 42 ha headwater research catchment in western Luxembourg (Glaser et al., 2016; Klaus et al., 2015; Martínez-Carreras et al., 2016; Schwab et al., 2018). We avoided unfavourable environmental conditions for the image acquisitions (cf. 2.2, Fig. 2) by allowing a few days

of tolerance around the targeted biweekly or weekly recurrence frequency. Additionally, we cut ferns that obstructed the camera view during the summer months. The 364 acquired panorama images were divided into three groups classified as usable without restrictions (32.4 %), usable with some restrictions (small negative effects of low temperature contrasts or covering vegetation visible, 31.1 %), and unusable (36.5 %).

The usable panoramas captured the temporal evolution of surface saturation over the 18-month field campaign. This demonstrates the robustness of TIR imagery through the complete range of seasonal conditions (Fig. 4), including snow and growing vegetation as well as warm and cold water. The full extent of the added value provided by TIR imagery
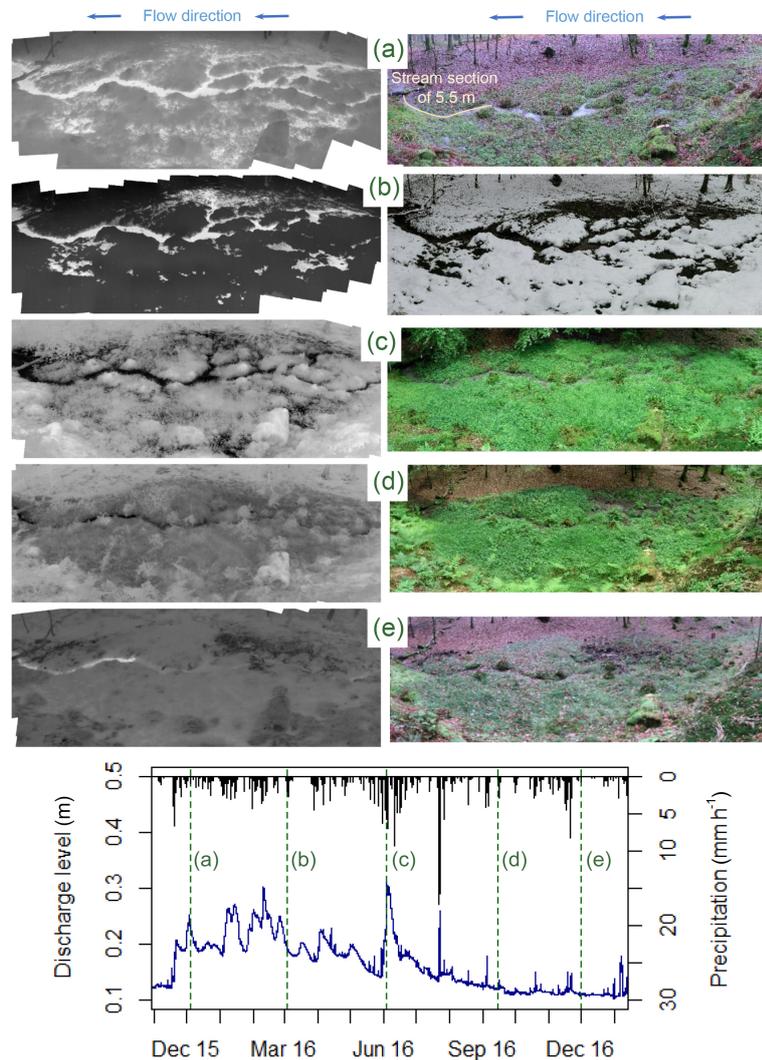
Figure 4. Time-lapse TIR and VIS panoramas, showing the variation of surface-saturation patterns with varying discharge levels under diverse seasonal conditions.

compared to VIS imagery was documented for cases with different seasonal conditions (Fig. 4), particularly for situations with less-pronounced differences in discharge levels (e.g. Fig. 4a–c). For example, the comparison of the VIS images of December 2015 and June 2016 (Fig. 4a vs. Fig. 4c) suggests wetter conditions for December 2015, while the two TIR images show similar saturation patterns for the two dates.

In addition to surface-saturation dynamics, the TIR images can also reveal distinct types of saturation patterns. For example, the orientation of saturated areas may change over a few metres from perpendicular (Fig. 5a, b) to parallel (Fig. 5c, d) to the adjacent stream. The extension of saturated areas along the left bank (Fig. 5c, d) appears to be created by

a parallel extension of the stream in a flat riparian zone that becomes an extended stream bed. The surface saturation oriented perpendicularly to the stream at the right bank (Fig. 5) appears to be generated from exfiltrating groundwater that flows downhill to the stream at the soil surface. Thus, the different directional extents of the saturated areas can indicate different processes underlying the surface-saturation formation.

Finally, the images allow us to identify the spatial heterogeneity of temporal saturation dynamics across different study sites. Figure 6 shows TIR images of the riparian zone of two different source areas with different degrees and dynamics of surface saturation. In area 1 (Fig. 6, panels a,c, and e), the pattern of saturation areas barely changed from Febru-

Hydrol. Earth Syst. Sci., 22, 5987–6003, 2018   www.hydrol-earth-syst-sci.net/22/5987/2018/

83
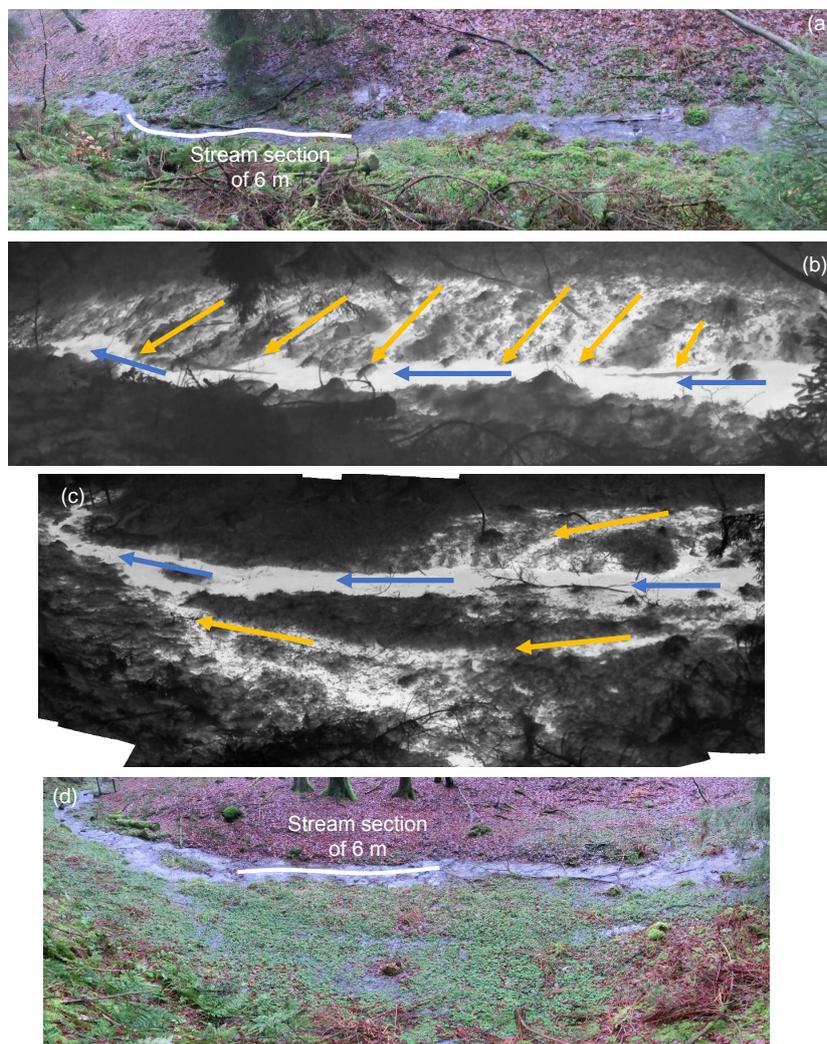
Figure 5. Comparison of different types of surface-saturation patterns. The yellow arrows indicate the orientation of the saturated areas towards the stream (blue arrows represent flow direction). The perpendicular direction **(a, b)** is likely caused by exfiltrating groundwater connecting to the stream, and the parallel direction **(c, d)** is likely caused by a parallel flow of the stream expanding into the riparian zone. The red ovals indicate where the two panorama images connect.

ary to April, while in area 2 (Fig. 6, panels b, d, and f) some locations had dried out (red circles). In December 2016, the riparian zones of both source areas were completely dry, and the stream started further downstream in comparison to the other observation dates (red arrows). This suggests that both source areas evolve from very wet to very dry conditions (during which surface saturation is mainly represented by spots with stable groundwater exfiltration) with distinctly different transition dynamics.

3   Quantification of saturation through pixel classification

3.1   Methods for generating binary saturation maps

The application examples described in Sect. 2.4 demonstrate the potential for TIR images to rapidly and intuitively visualize surface-saturated areas. However, the "raw data" images need to be transformed into binary saturation maps for further analyses based on quantitative values (e.g. saturation percentages). A common approach to making an image binary is histogram thresholding (e.g. Rosin, 2002). This allows a TIR image to be transformed into a binary satura-
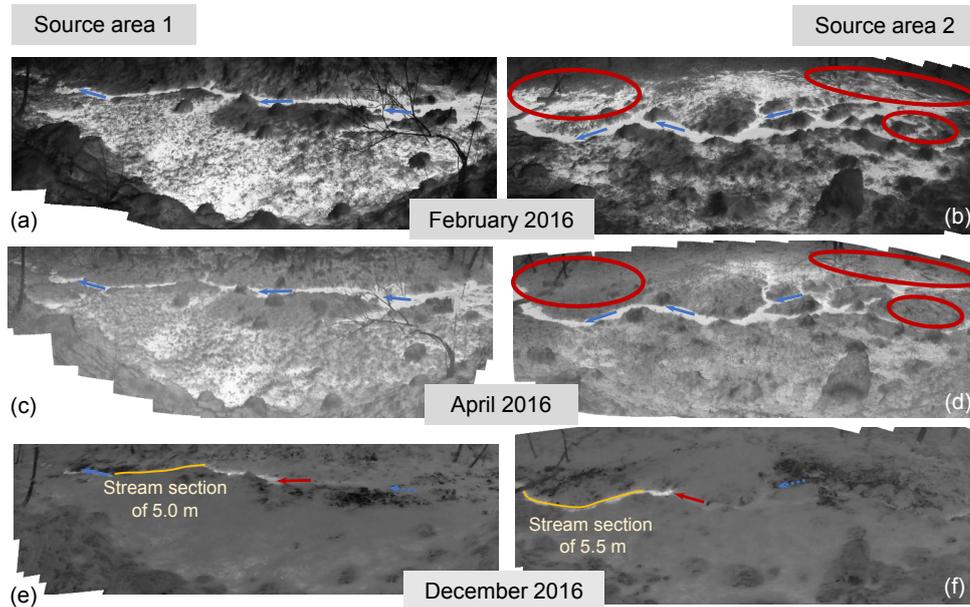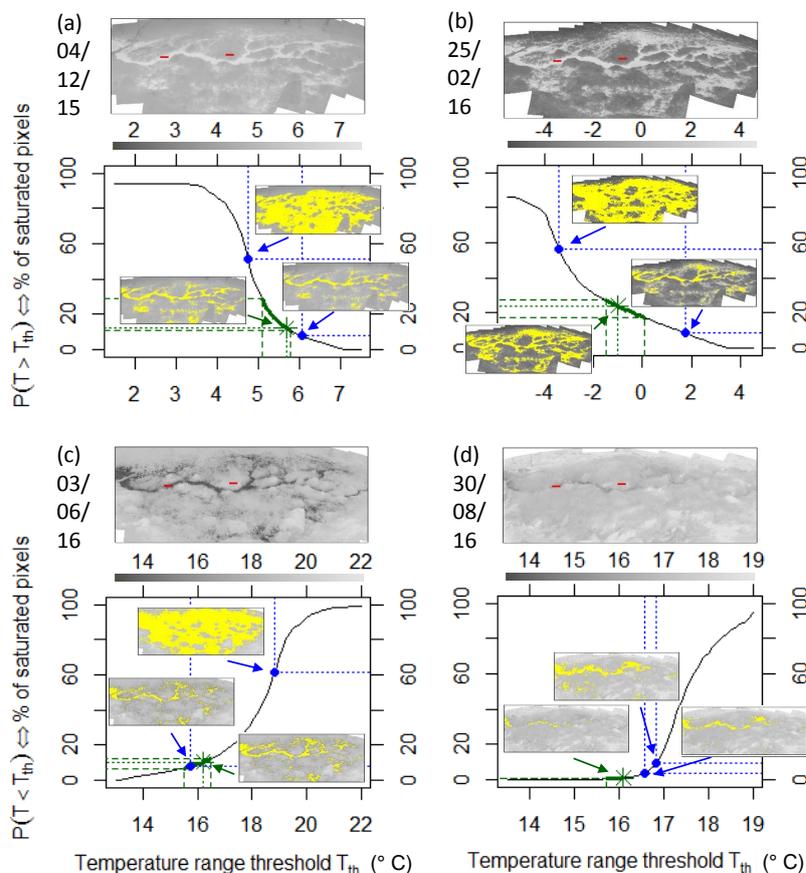
Figure 6. Transition of two source areas (**a**, **c**, **e** vs. **b**, **d**, **f**) from very wet (**a**, **b**) to very dry conditions (**e**, **f**). Surface saturation in source area 1 (**a**, **c**, **e**) barely changed between February and April 2016, whereas source area 2 is clearly drier at some locations (red ovals) in April 2016. In December 2016, both source areas were completely dry on each side of the stream (blue arrows represent flow direction), and the stream started further downstream (red arrows).

tion map by taking the temperature range of pixels that are known to be saturated (i.e. stream pixels) and defining all pixels in that image that fall into that temperature range as saturated (cf. Glaser et al., 2016; Pfister et al., 2010). Several thresholding algorithms can be found in the literature, each of which has its characteristic assumptions with respect to image content (Patra et al., 2011). Unsupervised approaches other than thresholding are also used for making an image binary, e.g. clustering (Li et al., 2015). Yet thresholding is the most rapid technique for achieving a binary classification of an image, even though the selection of an adequate threshold value represents a critical step and its choice strongly influences the classification outcome.

One possibility for selecting a threshold value for classifying surface saturation is to manually adapt the temperature range until the resulting saturation map matches best the visual assessment of the original TIR and – if possible – VIS image. A more objective and, for time-lapsed images, faster method consists of relying on the temperature of preselected pixels or a predefined mask for saturated and unsaturated parts in all images. Such pixels and masks can be selected based on a visual interpretation of the images or on information obtained from reference sensors in the field, indicating whether a location was wet or dry at the surface at the time of image acquisition.

Silasari et al. (2017) applied an automatic image classification for unimodal distributions based on a threshold pa-

rameter that needs to be calibrated to specific image conditions (in this case, the brightness of VIS images).This is only straightforward in cases where the temperature distribution between water and the surrounding environment is clearly bimodal. Chini et al. (2017) presented a parametric adaptive thresholding algorithm especially suited for images that do not show a clear bimodal distribution. The algorithm makes use of an automatic selection of image subsections with clear bimodal distributions, a hierarchical split-based approach (HSBA), and a subsequent parameterization of the distributions of the two pixel classes. Since the two decomposed distributions might still overlap to a certain extent, Chini et al. (2017) advise complementing the decomposed distribution information with contextual information of the image for the final generation of a binary image, instead of selecting a single threshold value between the two decomposed distributions. Several approaches are available in the literature for including contextual information in the classification of a single spectral image, such as mathematical morphology (Chini et al., 2009) or second-order textural parameters (Pacifici et al., 2009). Chini et al. (2017) suggested a region-growing algorithm where the seeds and the stopping criteria are constrained by the identified distribution of the class of interest (here, saturation).

B. Glaser et al.: Technical note: Mapping surface-saturation dynamics with thermal infrared imagery   **5997**

Figure 7. Example TIR images, with their cumulative saturation curves showing the percentage of pixels that have a higher **(a, b)** or lower **(c, d)** temperature than the temperature range threshold T_{th} and are thus defined as saturated (marked as yellow pixels in the inset TIR images). The green asterisks mark the temperature ranges that were manually chosen as optimum following a visual assessment of the images. Green dashed lines define the uncertainty of the optimum temperature ranges. The red rectangles in the TIR images depict the masks used for the identification of temperature ranges from a constantly wet **(a, c)** and constantly dry **(b, d)** area. The respective temperature ranges and saturation percentages are marked in blue. As a reference for the spatial dimension of the images, we refer the reader to the indicated stream section in Figs. 3 or 4.

3.2 Comparison of methods for generating binary saturation maps for TIR images

We applied three of the approaches described above to generate the binary saturation maps of our TIR image data set. Here, we present the results for four example images with differing conditions during image acquisition (e.g. very wet or dry conditions, water being the warmest or coldest material; Fig. 7). We evaluated the results of the three different approaches based on our observations from the field and the corresponding VIS image as ground truth.

First, we manually chose a temperature range of saturation for each image. By nature, this pixel classification approach creates results that are very close to ground truth. However, finding an unequivocal temperature range was not feasible, and the selection of the most plausible temperature

range (Fig. 7; dark-green asterisk) remained somewhat subjective. Furthermore, artefacts (such as pixels corresponding to vegetation covering the stream) induced some uncertainty in the pixel classification, eventually leading to discrepancies compared to visually identified saturation patterns. Consequently, a pixel classification based on this manual procedure remained tarnished by some uncertainties. The definition of an uncertainty range within which the temperature range can be considered plausible (Fig. 7; dark-green, dashed lines) was also subjective. Generally, the uncertainty range was small for images with low saturation and gradually increased with higher saturation (compare Fig. 7d–b). Accordingly, images with a large difference in percentages of saturated pixels (e.g. Fig. 7b vs. Fig. 7d) did not encounter an overlap of the uncertainty ranges. For some images, the uncertainty range was rather high (Fig. 7a), and a compar-

ison with other images with percentages of saturated pixels in the same range was thus problematic. In such cases, it is preferable that only one person defines the optimal temperature ranges and thus saturation patterns for all images that are intended to be compared in order to ensure consistency in the image interpretation.

Secondly, we performed an objective selection of the temperature range of saturation based on masks with known pixel classes. For this, we used two masks, one with 2000 pixels falling into an area that always stayed dry and one with 2000 pixels falling into an area where the stream was flowing all year (red rectangles; Fig. 7). Based on the mask, we selected the threshold for the temperature range as the 90th percentile and 10th percentile of the temperature of the stream mask pixels and dry mask pixels, respectively (i.e. 90 % of the pixels falling below the mask were defined as saturated and dry, respectively). By using the two different masks, we obtained two temperature ranges, resulting in two different saturation percentages for each image (Fig. 7; blue points). The identification of saturated areas based on the dry mask was clearly not constrained enough. The identification of saturated areas based on the stream mask sometimes approached the manual identification of saturation (Fig. 7a, c) but, in other cases, even exceeded it (Fig. 7d). The uncertainty range of saturation obtained with the two masks could be reduced by selecting a more extreme percentile for the temperature threshold definition. However, this increased the risk of obtaining a clearly incorrect value (cf. Fig. 7d), since the stream and dry mask can cover pixels of the wrong category (due to artefacts like vegetation covering the stream or due to distorted co-registered images, resulting in a shifted mask). A reduced mask size prevents such wrong pixels but also reduces the captured variability in temperature (in an extreme case, down to one temperature value), which in turn increases the risk of missing the warmest or coldest temperature of the wet or dry areas.

Finally, we tested the usability of the approach proposed by Chini et al. (2017), constraining a region-growing algorithm to (a) a bimodal distribution derived from the HSBA applied to the entire image, (b) a bimodal distribution derived from the HSBA where the selection of bimodal image subsections was constrained to image-specific manual predefinitions of temperature ranges of saturation, and (c) a bimodal distribution derived from preselected parts of the image that include clearly wet and dry areas. While in some cases the fully automatic image classification (point a) worked very well in comparison to the manual selection of a temperature range (cf. Fig. 8; 4 December 2015, 30 August 2016), for the other cases, saturation was mostly underestimated (cf. Fig. 8; 25 February 2016, 3 June 2016). The additional constraint with image-specific temperature ranges (point b) improved the matches overall with the manually defined saturation patterns, but the result was strongly influenced by the match of the given constraint range to the range that was defined as the optimum for the image. A constraint with a roughly esti-
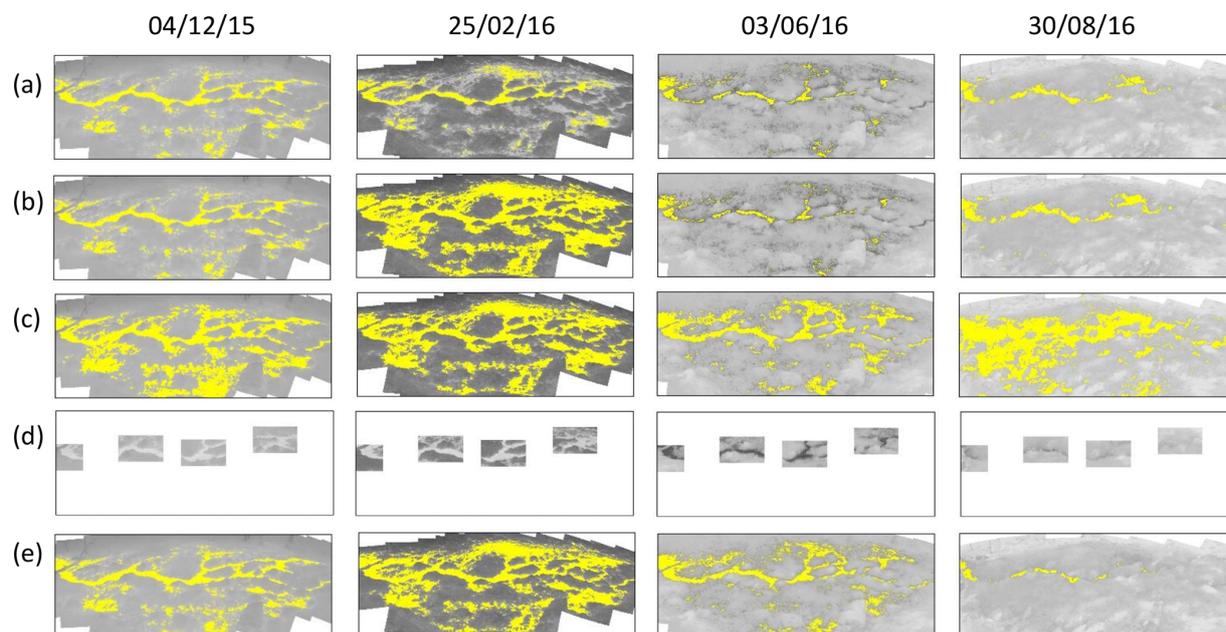
mated temperature for saturation worked less well than a constraint with the temperature range as selected in the detailed manual assessment described earlier in the section (cf. Fig. 7; green asterisks and lines). The classification based on preselected parts of the image (c) tended to result in higher saturation amounts. This improved the match for the cases that were underestimated with the fully automatic classification (point a – cf. Fig. 8; 25 February 2016, 3 June 2016), but this overestimated saturation for the cases where the fully automatic classification (point a) showed good results (cf. Fig. 8; 4 December 2015, 30 August 2016).

4   Discussion

4.1   Mapping surface saturation with TIR imagery

The main advantages of TIR imagery in comparison to other surface-saturation mapping methods are its non-invasive character and its large temporal and spatial flexibility (centimetres to kilometres, minutes to months). Another advantage is that TIR images allow a rapid and intuitive identification and analysis of the dynamics of surface-saturation patterns. The raw data images can be used without any additional processing to study surface-saturated areas, their evolution over time, and how and where they occur – ultimately contributing to a better mechanistic understanding of the hydrological processes prevailing in the studied area. The pure visual information provided by the images per se is also usable as soft data, e.g. for model validation (e.g. different types of extent compared to stream, Fig. 5; more and less stable saturation patterns, Fig. 6). VIS imagery offers similar advantages (Silasari et al., 2017), but commonly the saturated areas are not as clearly visible as with TIR imagery (cf. Figs. 1, 4). Moreover, VIS imagery is not usable during the night and cannot provide additional information about water sources and processes underlying the surface-saturation formation (cf. Figs. 1 and 5, groundwater inflow vs. stream water). Nevertheless, VIS imagery provides good complementary information to the TIR imagery and should always be considered as a ground truth information source.

In our study, unfavourable image acquisition conditions (cf. Sect. 2.2) caused 36.5 % of the acquired images to be unusable for further processing. High amounts of unusable images are a common problem in environmental imagery (e.g. cloud cover for satellite images, night-time for VIS images; de Alwis et al., 2007; Silasari et al., 2017). Flexibility in the scheduling of a field campaign is thus necessary for reducing the number of acquisitions during unfavourable conditions. A concern for the use of TIR imagery for mapping saturation patterns is that some saturated areas (e.g. warmed-up ponding water) might not be identified as saturated due to a temperature that is very different from the stream temperature. This relates to the fact that temperature is only used as an indicator for saturation. Compared to other saturation

B. Glaser et al.: Technical note: Mapping surface-saturation dynamics with thermal infrared imagery **5999**

Figure 8. Comparison of saturation maps (yellow represents saturation), generated with a region-growing process whose seeds and stopping criteria were automatically constrained to **(a)** bimodal distributions derived from the HBSA applied to the entire image, **(b)** bimodal distributions derived from the HBSA where the selection of bimodal image subsections was constrained to image-specific manual predefinitions of temperature ranges of saturation, and **(c)** bimodal distributions derived from preselected parts of the image (which include clearly wet and dry areas and are shown in **d**). The saturation maps generated with manually selected temperature ranges based on visual assessment (cf. Fig. 7; green asterisk) are shown for comparison **(e)**. As a reference for the spatial dimension of the images, we refer the reader to the indicated stream section in Fig. 3 or 4.

indicators, such as vegetation mapping or hydrometric measurements (cf. Dunne et al., 1975), we consider TIR imagery with the above-mentioned advantages as the better indirect mapping method. However, the only way to directly map surface saturation consists of walking through the area of interest (e.g. squishy boot method), which remains restricted to small areas or low mapping frequencies.

The amount of fieldwork for imagery mapping is generally reduced compared to other methods for mapping surface saturation (e.g. vegetation or soil mapping), allowing more frequent campaigns with higher spatial precision. Yet consistent with other imagery-mapping studies (e.g. Spence and Mengistu, 2016), the image post-processing in this study was time-consuming. Mosaicking and the co-registering of images is often considered particularly difficult for TIR images, since ground control points with a thermal signature are needed (Dugdale, 2016; Weber et al., 2015). Our experience showed that the images normally offered enough natural thermal ground control points (e.g. the stream bank) in cases where the temperature contrast between water and ambient materials was good enough for image usability. In combination with the post-processing workflow presented, the post-processing effort was reasonable. More automatized workflows like the one proposed by Turner et al. (2014) for

mosaicking UAV-acquired TIR images could also be adapted and applied.

The image acquisition considerations, post-processing steps, and application examples described focused on biweekly or weekly panoramic images of small areas, acquired with a portable TIR camera. A transfer of the TIR imagery technique to different temporal or spatial scales does not change the principles and possibilities of the technique, but it will require some additional scale- and platform-dependent considerations. For example, using permanently installed ground-based cameras for image acquisitions with high temporal frequencies might challenge technical aspects such as protection of the camera against environmental influences, an automatic triggering of image acquisition, and power supply. These aspects might also be relevant for TIR imagery acquisition at larger spatial scales, especially when using UAVs. Besides this, image acquisitions based on UAV or aeroplane overflights might, for example, require considerations of overflight regulations. Users of UAVs or aeroplanes should also be aware that saturation patterns within a forest might only – if at all – be mapped during the dormant season and that ground control points and ground truth data might be more difficult to obtain. Such challenges are partly addressed in existing literature (e.g. Vivoni et al., 2014; We-

ber et al., 2015), but others will need to be figured out by applying the TIR technique at such different scales.

4.2 Pixel classification methods

More challenging than TIR image mosaicking and co-registering was the generation of saturation maps from the TIR images. The different pixel classification methods tested all yielded somewhat different results compared to pixel classification based on manual visual assessment. Nevertheless, realizing an objective, automatic classification of saturated areas is not more challenging than for other surface-saturation mapping methods. Saturation maps created based on the squishy boot method or vegetation or soil mapping are subjective due to decisions made during the fieldwork. The supervised and unsupervised classification methods that are commonly used for creating saturation maps from remote sensing data (e.g. VIS images, NDVI or NDWI) also contain some uncertainty (Chabot and Bird, 2013; DeAlwis et al., 2007; Mengistu and Spence, 2016; Spence and Mengistu, 2016).

Moreover, the main problem for all of the tested saturation map generation methods (cf. Sect. 4) is that they are not applicable without being adapted to individual image conditions (very wet, very dry, water being the warmest or coldest material, slightly different fields of view). Other image processing methods for deriving saturation maps also do not fulfil this requirement; it is necessary to adapt the parameters (e.g. Silasari et al., 2017) or to perform a new supervision (with new classification pixels or masks) for the classification of images with different conditions (e.g. Chabot and Bird, 2013; Keys et al., 2016). At this stage, we consider a manual choice of temperature range for saturated pixels as the best approach for time-lapsed images with very variable conditions and slight perspective shifts, even though it is labour-intensive and somewhat subjective. For time-lapsed images with a fixed vantage point and for time spans with similar conditions (e.g. storm events), the automatable methods presented represent valuable options. In particular, the combination of an automatic decomposition of two pixel class distributions with a region-growing algorithm yielded objective saturation maps close to the manual saturation classification and visual assessment of the TIR images (Fig. 8). Small adaptations of the constraint for the decomposition of two pixel class distributions were sufficient for obtaining good results for the different image conditions (cf. Fig. 8a–c), and further developments of the method might even allow such adaptations to be performed in semi-automatic and automatic ways. More work on pixel classification might also include the application of machine-learning techniques or, especially for time-lapsed images, the analysis of the temperature signals of individual pixels over time. Another interesting option may consist of combining the TIR images with additional data (e.g. VIS images or NIR images), which will allow multi-spectral classification methods to be applied

(Chini et al., 2008) and contextual information to be integrated at the same time (Chini et al., 2014).

5 Summary and conclusions

This technical note presents recent work carried out in the Weierbach catchment, where we tested the potential for TIR imagery to map surface-saturation dynamics. To the best of our knowledge, this is the first comprehensive review and summary of the TIR imagery-related methodological principles and the required precautions and considerations for a successful application of TIR imagery for mapping surface saturation. We give advice for all steps, from image acquisition to processed saturation maps. The main requirement is a clear temperature contrast between water and the surrounding environments. Image acquisition during an 18-month campaign showed that the method works best during dry nights or dry early mornings and that images should be taken from well-chosen positions without obstructions in view towards the ground. The workflow presented for acquiring panoramic images is particularly suitable for small areas of interest (centimetres to metres) that are monitored with intermediate to low mapping frequencies (days to months). Moreover, the information contained in this technical note is also beneficial for applications at different temporal and spatial scales (fixed cameras for high-frequency images, drone and satellite images for larger spatial scales), considering that some adaption and further developments of the methodology might be necessary.

We demonstrated with three examples that TIR imagery is applicable throughout the year and can reveal spatially heterogeneous surface-saturation dynamics and distinct types of saturation patterns. The saturation patterns can also be used to identify different processes underlying the surface-saturation formation, such as groundwater exfiltration or stream expansion. The surface-saturation information visualized in the images can be used directly as soft data for characterizing field conditions, for analysing ongoing hydrologic processes, and for model validation.

The methods presented for obtaining binary, objective saturation maps from TIR images contain some uncertainties and are not automatable for data sets containing many images with varying characteristics (e.g. very wet or dry, water warmest or coldest material, slightly different fields of view). In such cases, a manual choice of the temperature range for saturated pixels is the most reliable approach. Yet for image subsets with similar conditions, the pixel classifications tested work well, and we think that the combination of an automatic decomposition of the image distribution in two pixel classes and a region-growing algorithm is a very promising option for obtaining objective, comparable saturation maps. In conclusion, we consider the TIR imagery a very powerful method for mapping surface saturation in terms of practicability and spatial and temporal flexibility, and we believe it

B. Glaser et al.: Technical note: Mapping surface-saturation dynamics with thermal infrared imagery   **6001**

can provide new insights into the role of saturated areas and subsequent spatial and temporal dynamics in rainfall–runoff transformation.

References

Ala-aho, P., Rossi, P. M., Isokangas, E., and Kløve, B.: Fully integrated surface–subsurface flow modelling of groundwater–lake interaction in an esker aquifer: Model verification with stable isotopes and airborne thermal imaging, J. Hydrol., 522, 391–406, https://doi.org/10.1016/j.jhydrol.2014.12.054, 2015.

Ali, G., Birkel, C., Tetzlaff, D., Soulsby, C., Mcdonnell, J. J., and Tarolli, P.: A comparison of wetness indices for the prediction of observed connected saturated areas under contrasting conditions, Earth Surf. Proc. Land., 39, 399–413, https://doi.org/10.1002/esp.3506, 2014.

de Alwis, D. A., Easton, Z. M., Dahlke, H. E., Philpot, W. D., and Steenhuis, T. S.: Unsupervised classification of saturated areas using a time series of remotely sensed images, Hydrol. Earth Syst. Sci., 11, 1609–1620, https://doi.org/10.5194/hess-11-1609-2007, 2007.

Ambroise, B.: Variable "active" versus "contributing" areas or periods: a necessary distinction, Hydrol. Process., 18, 1149–1155, https://doi.org/10.1002/hyp.5536, 2004.

Antonelli, M., Klaus, J., Smettem, K., Teuling, A. J., and Pfister, L.: Exploring streamwater mixing dynamics

via handheld thermal infrared imagery, Water, 9, 358, https://doi.org/10.3390/w9050358, 2017.

Aubry-Wake, C., Baraer, M., Mckenzie, J. M., Mark, B. G., Wigmore, O., Hellström, R. Å., Lautz, L., and Somers, L.: Measuring glacier surface temperatures with ground-based thermal infrared imaging, Geophys. Res. Lett., 42, 8489–8497, https://doi.org/10.1002/2015GL065321, 2015.

Blazkova, S., Beven, K. J., and Kulasova, A.: On constraining TOPMODEL hydrograph simulations using partial saturated area information, Hydrol. Process., 16, 441–458, https://doi.org/10.1002/hyp.331, 2002.

Briggs, M. A., Hare, D. K., Boutt, D. F., Davenport, G., and Lane, J. W.: Thermal infrared video details multiscale groundwater discharge to surface water through macropores and peat pipes, Hydrol. Process., 30, 2510–2511, https://doi.org/10.1002/hyp.10722, 2016.

Cardenas, M. B., Doering, M., Rivas, D. S., Galdeano, C., Neilson, B. T., and Robinson, C. T.: Analysis of the temperature dynamics of a proglacial river using time-lapse thermal imaging and energy balance modeling, J. Hydrol., 519, 1963–1973, https://doi.org/10.1016/j.jhydrol.2014.09.079, 2014.

Chabot, D. and Bird, D. M.: Small unmanned aircraft: precise and convenient new tools for surveying wetlands, J. Unmanned Veh. Syst., 1, 15–24, https://doi.org/10.1139/juvs-2013-0014, 2013.

Chini, M., Pacifici, F., Emery, W. J., Pierdicca, N., and Del Frate, F.: Comparing Statistical and Neural Network Methods Applied to Very High Resolution Satellite Images Showing Changes in Man-Made Structures at Rocky Flats, IEEE T. Geosci. Remote, 46, 1812–1821, https://doi.org/10.1109/TGRS.2008.916223, 2008.

Chini, M., Pierdicca, N., and Emery, W. J.: Exploiting SAR and VHR Optical Images to Quantify Damage Caused by the 2003 Bam Earthquake, IEEE T. Geosci. Remote, 47, 145–152, https://doi.org/10.1109/TGRS.2008.2002695, 2009.

Chini, M., Chiancone, A., and Stramondo, S.: Scale Object Selection (SOS) through a hierarchical segmentation by a multispectral per-pixel classification, Pattern Recognit. Lett., 49, 214–223, https://doi.org/10.1016/j.patrec.2014.07.012, 2014.

Chini, M., Hostache, R., Giustarini, L., and Matgen, P.: A hierarchical split-based approach for parametric thresholding of SAR images: Flood inundation as a test case, IEEE T. Geosci. Remote, 55, 6975–6988, https://doi.org/10.1109/TGRS.2017.2737664, 2017.

Corripio, J. G.: Snow surface albedo estimation using terrestrial photography, Int. J. Remote Sens., 25, 5705–5729, https://doi.org/10.1080/01431160410001709002, 2004.

Creed, I. F., Sanford, S. E., Beall, F. D., Molot, L. A., and Dillon, P. J.: Cryptic wetlands: integrating hidden wetlands in regression models of the export of dissolved organic carbon from forested landscapes, Hydrol. Process., 17, 3629–3648, https://doi.org/10.1002/hyp.1357, 2003.

de Alwis, D. A., Easton, Z. M., Dahlke, H. E., Philpot, W. D., and Steenhuis, T. S.: Unsupervised classification of saturated areas using a time series of remotely sensed images, Hydrol. Earth Syst. Sci., 11, 1609–1620, https://doi.org/10.5194/hess-11-1609-2007, 2007.

Deitchman, R. S. and Loheide, S. P.: Ground-based thermal imaging of groundwater flow processes at the seepage face, Geophys. Res. Lett., 36, L14401, https://doi.org/10.1029/2009GL038103, 2009.

www.hydrol-earth-syst-sci.net/22/5987/2018/   **Hydrol. Earth Syst. Sci., 22, 5987–6003, 2018**

90

Doppler, T., Honti, M., Zihlmann, U., Weisskopf, P., and Stamm, C.: Validating a spatially distributed hydrological model with soil morphology data, Hydrol. Earth Syst. Sci., 18, 3481–3498, https://doi.org/10.5194/hess-18-3481-2014, 2014.

Dugdale, S. J.: A practitioner's guide to thermal infrared remote sensing of rivers and streams: recent advances, precautions and considerations, WIREs Water, 3, 251–268, https://doi.org/10.1002/wat2.1135, 2016.

Dunne, T., Moore, T. R., and Taylor, C. H.: Recognition and prediction of runoff-producing zones in humid regions, Hydrol. Sci. Bull., 20, 305–327, 1975.

Frei, S., Lischeid, G., and Fleckenstein, J. H.: Effects of micro-topography on surface-subsurface exchange and runoff generation in a virtual riparian wetland – A modeling study, Adv. Water Resour., 33, 1388–1401, https://doi.org/10.1016/j.advwatres.2010.07.006, 2010.

Frey, M. P., Schneider, M. K., Dietzel, A., Reichert, P., and Stamm, C.: Predicting critical source areas for diffuse herbicide losses to surface waters: Role of connectivity and boundary conditions, J. Hydrol., 365, 23–36, https://doi.org/10.1016/j.jhydrol.2008.11.015, 2009.

Glaser, B., Klaus, J., Frei, S., Frentress, J., Pfister, L., and Hopp, L.: On the value of surface saturated area dynamics mapped with thermal infrared imagery for modeling the hillslope-riparian-stream continuum, Water Resour. Res., 52, 8317–8342, https://doi.org/10.1002/2015WR018414, 2016.

Grabs, T., Seibert, J., Bishop, K., and Laudon, H.: Modeling spatial patterns of saturated areas: A comparison of the topographic wetness index and a dynamic distributed model, J. Hydrol., 373, 15–23, https://doi.org/10.1016/j.jhydrol.2009.03.031, 2009.

Handcock, R. N., Gillespie, A. R., Cherkauer, K. A., Kay, J. E., Burges, S. J., and Kampf, S. K.: Accuracy and uncertainty of thermal-infrared remote sensing of stream temperatures at multiple spatial scales, Remote Sens. Environ., 100, 427–440, https://doi.org/10.1016/j.rse.2005.07.007, 2006.

Handcock, R. N., Torgersen, C. E., Cherkauer, K. A., Gillespie, A. R., Tockner, K., Faux, R. N., and Tan, J.: Thermal Infrared Remote Sensing of Water Temperature in Riverine Landscapes, in: Fluvial Remote Sensing for Science and Management, edited by: Carbonneau, P. E. and Piégay, H., 85–113, Wiley-Blackwell, Chichester, West Sussex, UK, 2012.

Härer, S., Bernhardt, M., Corripio, J. G., and Schulz, K.: PRACTISE – Photo Rectification And ClassificaTIon SoftwarE (V.1.0), Geosci. Model Dev., 6, 837–848, https://doi.org/10.5194/gmd-6-837-2013, 2013.

Heathwaite, A. L., Quinn, P. F., and Hewett, C. J. M.: Modelling and managing critical source areas of diffuse pollution from agricultural land using flow connectivity simulation, J. Hydrol., 304, 446–461, https://doi.org/10.1016/j.jhydrol.2004.07.043, 2005.

Hewlett, J. D. and Hibbert, A. R.: Factors affecting the response of small watershed to precipitation in humid areas, For. Hydrol., 275–279, available at: http://coweeta.ecology.uga.edu/publications/851.pdf (last access: 14 November 2018), 1967.

Ishaq, R. M. and Huff, D. D.: Application of remote sensing to the location of hydrodynamically active (source) areas, in: Proceedings of 9th international symposium on remote sensing of the environment, Environmental Research Institute of Michigan, Ann Arbor, Michigan, 653–666, 1974.

Kelcey, J. and Lucieer, A.: Sensor Correction of a 6-Band Multispectral Imaging Sensor for UAV Remote Sensing, Remote Sens., 4, 1462–1493, https://doi.org/10.3390/rs4051462, 2012.

Keys, T. A., Jones, C. N., Scott, D. T., and Chuquin, D.: A cost-effective image processing approach for analyzing the ecohydrology of river corridors, Limnol. Oceanogr.-Meth., 14, 359–369, https://doi.org/10.1002/lom3.10095, 2016.

Klaus, J., Wetzel, C. E., Martínez-Carreras, N., Ector, L., and Pfister, L.: A tracer to bridge the scales: on the value of diatoms for tracing fast flow path connectivity from headwaters to meso-scale catchments, Hydrol. Process., 29, 5275–5289, https://doi.org/10.1002/hyp.10628, 2015.

Kulasova, A., Beven, K. J., Blazkova, S. D., Rezacova, D., and Cajthaml, J.: Comparison of saturated areas mapping methods in the Jizera Mountains, Czech Republic, J. Hydrol. Hydromech., 62, 160–168, https://doi.org/10.2478/johh-2014-0002, 2014a.

Kulasova, A., Blazkova, S., Beven, K., Rezacova, D., and Cajthaml, J.: Vegetation pattern as an indicator of saturated areas in a Czech headwater catchment, Hydrol. Process., 28, 5297–5308, https://doi.org/10.1002/hyp.10239, 2014b.

Latron, J. and Gallart, F.: Seasonal dynamics of runoff-contributing areas in a small mediterranean research catchment (Vallcebre, Eastern Pyrenees), J. Hydrol., 335, 194–206, https://doi.org/10.1016/j.jhydrol.2006.11.012, 2007.

Li, C. J. and Ling, H.: Synthetic Aperture Radar Imaging Using a Small Consumer Drone, IEEE Int. Symp. Antennas Propagation, Vancouver, Canada, 685–686, https://doi.org/10.1109/APS.2015.7304729, 2015.

Li, H.-C., Celik, T., Longbotham, N., and Emery, W. J.: Gabor Feature Based Unsupervised Change Detection of Multitemporal SAR Images Based on Two-Level Clustering, IEEE Geosci. Remote S., 12, 2458–2462, https://doi.org/10.1109/LGRS.2015.2484220, 2015.

Luzi, G.: Ground based SAR interferometry: a novel tool for Geoscience, in: Geoscience and Remote Sensing New Achievements, edited by: Imperatore, P. and Riccio, D., InTechOpen, London, UK, 1–26, https://doi.org/10.5772/9090, 2010.

Martínez-Carreras, N., Hissler, C., Gourdol, L., Klaus, J., Juilleret, J., Iffly, J. F., and Pfister, L.: Storage controls on the generation of double peak hydrographs in a forested headwater catchment, J. Hydrol., 543, 255–269, https://doi.org/10.1016/j.jhydrol.2016.10.004, 2016.

Matgen, P., El Idrissi, A., Henry, J. B., Tholey, N., Hoffmann, L., de Fraipont, P., and Pfister, L.: Patterns of remotely sensed floodplain saturation and its use in runoff predictions, Hydrol. Process., 20, 1805–1825, https://doi.org/10.1002/hyp.5963, 2006.

Mengistu, S. G. and Spence, C.: Testing the ability of a semidistributed hydrological model to simulate contributing area, Water Resour. Res., 52, 4399–4415, https://doi.org/10.1002/2016WR018760, 2016.

Orillo, J. W., Bansil Jr., G., Bernardo, J. J., Dizon, C., Imperial, H., Macabenta, A. M., and Palima Jr., R.: Determination of Green Leaves Density Using Normalized Difference Vegetation Index via Image Processing of In-Field Drone-Captured Image, J. Telecommun. Electron. Comput. Eng., 9, 2–6, 1–5, 2017.

Pacifici, F., Chini, M., and Emery, W. J.: A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use

B. Glaser et al.: Technical note: Mapping surface-saturation dynamics with thermal infrared imagery 6003

classification, Remote Sens. Environ., 113, 1276–1292, https://doi.org/10.1016/j.rse.2009.02.014, 2009.

Patra, S., Ghosh, S., and Ghosh, A.: Histogram thresholding for unsupervised change detection of remote sensing images, Int. J. Remote Sens., 32, 6071–6089, https://doi.org/10.1080/01431161.2010.507793, 2011.

Pfister, L., McDonnell, J. J., Hissler, C., and Hoffmann, L.: Ground-based thermal imagery as a simple, practical tool for mapping saturated area connectivity and dynamics, Hydrol. Process., 24, 3123–3132, https://doi.org/10.1002/hyp.7840, 2010.

Portmann, F. T.: Hydrological runoff modelling by the use of remote sensing data with reference to the 1993–1994 and 1995 floods in the river Rhine catchment, Hydrol. Process., 11, 1377–1392, https://doi.org/10.1002/(sici)1099-1085(199708)11:10<1377::aid-hyp533>3.0.co;2-r, 1997.

Rinderer, M., Kollegger, A., Fischer, B. M. C., Stähli, M., and Seibert, J.: Sensing with boots and trousers – qualitative field observations of shallow soil moisture patterns, Hydrol. Process., 26, 4112–4120, https://doi.org/10.1002/hyp.9531, 2012.

Rosin, P. L.: Thresholding for Change Detection, Comput. Vis. Image Und., 86, 79–95, https://doi.org/10.1006/cviu.2002.0960, 2002.

Schuetz, T. and Weiler, M.: Quantification of localized groundwater inflow into streams using ground-based infrared thermography, Geophys. Res. Lett., 38, L03401, https://doi.org/10.1029/2010gl046198, 2011.

Schuetz, T., Weiler, M., Lange, J., and Stoelzle, M.: Two-dimensional assessment of solute transport in shallow waters with thermal imaging and heated water, Adv. Water Resour., 43, 67–75, https://doi.org/10.1016/j.advwatres.2012.03.013, 2012.

Schwab, M. P., Klaus, J., Pfister, L., and Weiler, M.: Diel fluctuations of viscosity-driven riparian inflow affect streamflow DOC concentration, Biogeosciences, 15, 2177–2188, https://doi.org/10.5194/bg-15-2177-2018, 2018.

Silasari, R., Parajka, J., Ressl, C., Strauss, P., and Blöschl, G.: Potential of time-lapse photography for identifying saturation area dynamics on agricultural hillslopes, Hydrol. Process., 31, 3610–3627, https://doi.org/10.1002/hyp.11272, 2017.

Soulsby, C., Bradford, J., Dick, J., McNamara, J. P., Geris, J., Lessels, J., Blumstock, M., and Tetzlaff, D.: Using geophysical surveys to test tracer-based storage estimates in headwater catchments, Hydrol. Process., 30, 4434–4445, https://doi.org/10.1002/hyp.10889, 2016.

Spence, C. and Mengistu, S. G.: Deployment of an unmanned aerial system to assist in mapping an intermittent stream, Hydrol. Process., 30, 493–500, https://doi.org/10.1002/hyp.10597, 2016.

Torgersen, C. E., Faux, R. N., McIntosh, B. A., Poage, N. J., and Norton, D. J.: Airborne thermal remote sensing for water temperature assessment in rivers and streams, Remote Sens. Environ., 76, 386–398, https://doi.org/10.1016/S0034-4257(01)00186-9, 2001.

Turner, D., Lucieer, A., Malenovský, Z., King, D. H., and Robinson, S. A.: Spatial co-registration of ultra-high resolution visible, multispectral and thermal images acquired with a micro-UAV over antarctic moss beds, Remote Sens., 6, 4003–4024, https://doi.org/10.3390/rs6054003, 2014.

Verhoest, N. E. C., Troch, P. A., Paniconi, C., and De Troch, F. P.: Mapping basin scale variable source areas from multitemporal remotely sensed observations of soil moisture behavior, Water Resour. Res., 34, 3235–3244, https://doi.org/10.1029/98wr02046, 1998.

Vivoni, E. R., Rango, A., Anderson, C. A., Pierini, N. A., Schreiner-McGraw, A. P., Saripalli, S., and Laliberte, A. S.: Ecohydrology with unmanned aerial vehicles, Ecosphere, 5, 130, https://doi.org/10.1890/ES14-00217.1, 2014.

Wahab, I., Hall, O., and Jirström, M.: Remote Sensing of Yields?: Application of UAV Imagery-Derived NDVI for Estimating Maize Vigor and Yields in Complex Farming Systems in Sub-Saharan Africa, Drones, 2, 28, https://doi.org/10.3390/drones2030028, 2018.

Weber, I., Jenal, A., Kneer, C., and Bongartz, J.: PANTIR-a dual camera setup for precise georeferencing and mosaicing of thermal aerial images, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. – ISPRS Arch., XL-3/W2, 269–272, https://doi.org/10.5194/isprsarchives-XL-3-W2-269-2015, 2015.

Whiting, J. A. and Godsey, S. E.: Discontinuous headwater stream networks with stable flowheads, Salmon River basin, Idaho, Hydrol. Process., 30, 2305–2316, https://doi.org/10.1002/hyp.10790, 2016.

Study 3: Intra-catchment variability of surface saturation − insights from long-term observations and simulations

Authors: Barbara Glaser, Marta Antonelli, Luisa Hopp, Julian Klaus

BG, LH and JK designed and directed the study. BG and MA planned and carried out the field work and processed the TIR images. Hydrometric and meteorologic data were provided by field engineers from LIST and ASTA. BG set up the simulation and processed the model output. BG, MA, LH and JK discussed and interpreted the results. BG prepared the figures and tables. BG prepared the manuscript with input from JK and LH. BG is the corresponding author.

Own contribution in %:

- Study concept and design: 80
- Field data acquisition and processing: 50
- Simulation: 100
- Data analysis, figures, and tables: 90
- Interpretation of the results: 75
- Preparation of the manuscript: 90

Intra-catchment variability of surface saturation – insights from long-term observations and simulations

Barbara Glaser[1,2], Marta Antonelli[3,1], Luisa Hopp[2], Julian Klaus[1]

5 [1]Catchment and Eco-Hydrology Research Group, Luxembourg Institute of Science and Technology, Esch/Alzette, 4362, Luxembourg
[2]Department of Hydrology, University of Bayreuth, Bayreuth, 95447, Germany
[3]Hydrology and Quantitative Water Management Group, Wageningen University & Research, Wageningen, 6700, The Netherlands

10

Correspondence to: Barbara Glaser (barbara.glaser@list.lu)

Abstract

The inundation of flood-prone areas varies in space and time and can have crucial impacts on runoff generation and water quality when the surface saturated areas become connected to the stream. In this study, we aimed to investigate and explain

15 the variability of surface saturation patterns and dynamics within a forested headwater catchment. On the one hand, we mapped surface saturation in seven distinct riparian areas of the Weierbach catchment (Luxembourg) with thermal infrared images, taken weekly to bi-weekly over a period of two years. On the other hand, we simulated the surface saturation generation in the catchment with the integrated surface subsurface hydrologic model HydroGeoSphere over the same period. Both the observations and simulations showed that the saturation dynamics were similar across the catchment, but that small differences

20 between the dynamics at different areas occurred. Moreover, the model reproduced the observed saturation patterns well for all seasonal and hydrologic conditions and at all investigated locations. Based on the observations and simulation results and the matches and mismatches between them, we concluded that the generation of surface saturation in the Weierbach catchment was largely controlled by exfiltration of groundwater into local depressions. However, we also illustrate that the entire variability of the patterns, dynamics and frequencies of surface saturation within the different riparian areas of the catchment

25 can only result from additional controlling factors to microtopography and groundwater exfiltration, such as differing hysteretic behaviour, differing subsurface structures, or additional water sources.

1 Introduction

It is critical for flood risk assessment to understand where and when water is standing or flowing on the ground surface outside of perennial surface water bodies. When such surface saturated areas connect to the stream via overland flow, they also become

30 crucial for runoff generation and water quality. In general, surface saturated areas arise from 1) water ponding on the surface due to exceedance of the infiltration capacity of unsaturated soil, 2) water ponding on impermeable surfaces or saturated soil, 3) water exfiltrating from the subsurface or, 4) stream water extending into the floodplain (e.g. Megahan and King, 1985). Over the past years and decades, various field studies mapped and analysed the spatial and temporal occurrence of surface saturation within different landscapes (e.g. Ambroise, 1986, 2016; Dunne et al., 1975; Gburek and Sharpley, 1998; Latron and

35 Gallart, 2007; Silasari et al., 2017; Tanaka et al., 1988). From the field studies it is well recognized that surface saturation varies in space and time and that its appearance is affected by structural (e.g. topography) and dynamic factors (e.g. precipitation intensity, antecedent moisture). Yet there is limited understanding on how surface saturation evolves spatially and temporally between and within landscapes and how the interplay of different controlling factors and generation processes controls the spatio-temporal variability of surface saturation.

1

Spatially distributed and dynamic hydrological models are potential tools for analysing the generation and development of surface saturation in space and time. Such models allow a detailed investigation of surface saturation at any desired location and time that goes far beyond the information that can be gained by any field observation. Several simulation studies systematically assessed the influence of static and dynamic factors on the temporal evolution, connectivity, and spatial

5   distribution of surface saturation by performing virtual experiments with hillslope models (Ogden and Watts, 2000; Reaney et al., 2014) or by testing a range of terrain indices for predicting time-integrated saturation patterns (Güntner et al., 2004). Other studies relied on dynamic distributed and semi-distributed simulations for analysing connectivity of surface saturation in relation to wetness conditions and catchment runoff (Mengistu and Spence, 2016; Qu and Duffy, 2007; Weill et al., 2013). Weill et al. (2013) and Partington et al. (2013) analysed the processes and water sources that generate surface saturation in a

10   wetland and a pre-alpine grassland headwater, respectively. Both studies applied a model belonging to the group of integrated surface-subsurface hydrologic models (ISSHMs, Sebben et al., 2013), which can simulate the interplay of different surface and subsurface processes of surface saturation generation (e.g. ponding of precipitation from the surface, exfiltration from the subsurface). Modelling studies that focus on a comprehensive spatio-temporal analysis of surface saturation dynamics within a landscape by evaluating the spatially distributed model outputs rather than aggregating the outputs are scarce (e.g. Nippgen

15   et al. (2015) for subsurface saturated areas)

When complementing field observations with simulations to analyse the generation and development of surface saturation in space and time, it is important to ensure that the model yields realistic results. Glaser et al. (2016) demonstrated for a small riparian area that a good match between modelled and observed discharge or soil moisture does not automatically imply a realistic simulation of saturation patterns. They concluded that a spatial validation of the dynamic saturation patterns itself is

20   crucial. However, only few of the existing modelling studies explicitly checked the realism of their simulated surface saturation with field observations before using them for further analyses. These studies focussed either on temporally integrated spatial patterns (Grabs et al., 2009; Güntner et al., 2004) or on temporal dynamics of overall catchment saturation (Birkel et al., 2010; Mengistu and Spence, 2016), but barely any study combined the observation and simulation of both surface saturation patterns and dynamics (Ali et al., 2014; Glaser et al., 2016). The lack of such studies is certainly explainable by the resources that are

25   necessary for obtaining appropriate field data. Today, we still lack a standard method to map surface saturation and the different existing methods such as the 'squishy boot' method, the usage of 'on-off' surface saturation sensors, the mapping of soil morphology or vegetation as surrogates, or the usage of remote sensing techniques (e.g. Dunne et al., 1975; Gburek and Sharpley, 1998; Güntner et al., 2004; Latron and Gallart, 2007; Mengistu and Spence, 2016; Silasari et al., 2017) all have their own advantages and disadvantages.

30   A relatively new and powerful method for mapping surface saturation is thermal infrared (TIR) imagery. TIR mapping relies on the difference between the surface temperature of water and other materials for identifying surface saturation. Previous work showed that recurrent mapping of surface saturation with high spatial resolution is possible with TIR imagery (Glaser et al., 2016; Pfister et al., 2010). Glaser et al. (2018) and Antonelli et al. (2019) applied TIR imagery mapping in the 42 ha forested Weierbach catchment in western Luxembourg and monitored the dynamics of surface saturation within several distinct

35   riparian areas along the Weierbach stream with a weekly to biweekly mapping frequency over several seasons.

In this study, we explore the intra-catchment variability of temporal and spatial characteristics of surface saturation (dynamics, frequencies, patterns) with a combination of field observation and modelling. We perform the study in the Weierbach catchment, where we can rely on existing TIR imagery data (Antonelli et al., 2019; cf. Glaser et al., 2018) and on previous modelling work for a 6 ha headwater of the catchment (Glaser et al., 2016, 2019) with the ISSHM HydroGeoSphere. Glaser

40   et al. (2016, 2019) simulated the 6 ha area of the catchment by accounting for a layering of the subsurface, while spatial heterogeneity was only represented by microtopography and by a different sequence of subsurface layers in the riparian zone compared to the hillslopes and plateau. Here, we extend the model setup to the entire 42 ha catchment without introducing additional heterogeneity and without performing a re-calibration. We simulate surface saturation in the catchment and contrast

2

the results with the observed saturation patterns from the TIR imagery, focussing on the long-term saturation dynamics over different seasons and wetness conditions (25 months with weekly to biweekly mapping resolution) and the spatial patterns of surface saturation occurrence and frequency at seven different riparian areas across the catchment. The two research objectives that we aim to address with this approach are:

5
1) How variable are surface saturation dynamics and patterns within a catchment and to what extent can we reproduce the variability of the saturation characteristics (dynamics, frequencies, patterns) with a rather homogenously set-up ISSHM?

2) What do we learn about the reasons for the intra-catchment variability of surface saturation characteristics from the matches and mismatches between simulation results and observations?

10 **2 Study site and data**

**2.1 Physiography, climate and hydrometry**

The Weierbach catchment is an intensively studied headwater catchment (42 ha) in western Luxembourg. About half of the catchment area is characterized by gentle slopes $<5°$, forming a plateau landscape unit (Martínez-Carreras et al., 2016). The rest of the catchment is characterized by hillslopes with slopes $> 5°$, forming a central V-shaped stream valley from north to

15 south and a V-shaped tributary valley in the east. A third, few metres long stream branch is situated in the west of the central stream valley. Riparian zones along the stream account for 1.2 % of the catchment area (Antonelli et al., 2019). Large parts of the catchment are forested with deciduous trees (mainly European beech and Sessile oaks), the south-east and some other small parts are forested with conifers (mainly Norway spruce and Douglas spruce). The riparian zones are free of tree canopy and covered with ferns, moss, and herbaceous plants. Soil developed from Pleistocene Periglacial Slope Deposits as shallow and

20 highly-permeable siltic, sceletic Cambisol with a depth ranging between 0.4 and 0.9 m (Gourdol et al., 2018; Juilleret et al., 2011; Moragues-Quiroga et al., 2017). Beneath the solum, a $0.5 – 1$ m thick basal layer with bedrock clasts oriented parallel to the slope overlies fractured Devonian slate and phyllites (Gourdol et al., 2018; Juilleret et al., 2011; Moragues-Quiroga et al., 2017; Scaini et al., 2017). In the riparian zones, soil and basal layer have been eroded and the fractured bedrock is overlain by shallow organic Leptosols (Glaser et al., 2016).

25 The climate is oceanic-continental without apparent seasonality in precipitation and with negligible amounts of snow (Carrer et al., 2019). Mean annual precipitation during the period from October 2013 to September 2017 was 955±53 mm. Mean annual discharge was 546±253 mm, with exceptionally dry conditions in the hydrological year 2017. During wet periods, discharge is characterized by double peak hydrographs with first peaks appearing as immediate response to precipitation and second pronounced peaks appearing 48h to 72h later (cf. Martínez-Carreras et al., 2016). During dry periods, only first

30 hydrograph peaks occur and the stream dries out intermittently starting from the source areas downstream.

Hydrological and meteorological data that were used in this study were measured from October 2013 to January 2018. Data from the period from October 2013 to September 2015 were used for spin-up simulations, data from the period from October 2015 to January 2018 were used to drive and validate the actual simulation (cf. Section 3). Discharge was measured with water pressure transducers (ISCO 4120 Flow Logger, 15 min logging intervals) at four v-notches, installed at the outlet of the

35 catchment (SW1, Fig.1) and upstream of the confluences of the three branches (SW2-SW4). Groundwater levels were continuously recorded every 15 minutes with pressure sensors (OTT CTD) in five piezometers installed in different landscape units (riparian zone, hillslope, plateau) of the catchment (Fig. 1, GW1-3, GW5, GW7). Soil moisture was continuously monitored (30 min logging intervals) with water content reflectometers (CS650, Campbell Scientific) installed horizontally in 10, 20, 40 and 60 cm depth at four different sites (Fig.1, SM3-5, SM7). At each site, two depth profiles were monitored. In

40 addition, soil moisture in 10 cm depth was monitored with water content reflectometers (CS616, Campbell Scientific, 30 min logging intervals) at five locations crosscutting the riparian zone of the stream source area of the middle stream branch (Fig. 1, SSM1-5).

3

Cumulative precipitation was recorded every 5 minutes with a tipping bucket raingauge (Young 52203, unheated, 1 m height) at an open area within the catchment (data gaps were filled by estimating a linear regression to data from a station approximately 4.5 km southward). Potential reference evapotranspiration was estimated based on measured air temperature, relative humidity, wind speed, and net radiation according to the FAO Penman-Monteith formulation (Allen et al., 1998). Air

5   temperature and relative humidity data were recorded next to the soil moisture profile SM5 (Fig. 1, HMP45C-LC, Campbell Scientific, 15 min logging intervals, 2 m height). Wind speed and radiation data were recorded approximately 4.5 km southward of the study site. Wind speed (Young Wind Monitor 05103, Vector A100R Anemometer) was recorded every 15 minutes in 3 m height and converted to wind speed in 2 m height (data gaps closed with data from a station approximately 11.5 km north-eastward) following the FAO guidelines (Allen et al., 1998). Net radiation was recorded every 15 minutes (Kipp

10   & Zonen NR Lite net radiometer) until May 2017. From June 2017 on (and for closing other data gaps), we used net radiation data recorded every 5 minutes close to Luxembourg Airport (~40 km southeast of the study site), as these measurements were highly correlated (linear regression with an intercept of 7.6 W m^{-2}, a slope of 0.92, R^2 = 0.81) with the measurements close to the study site in the years before.
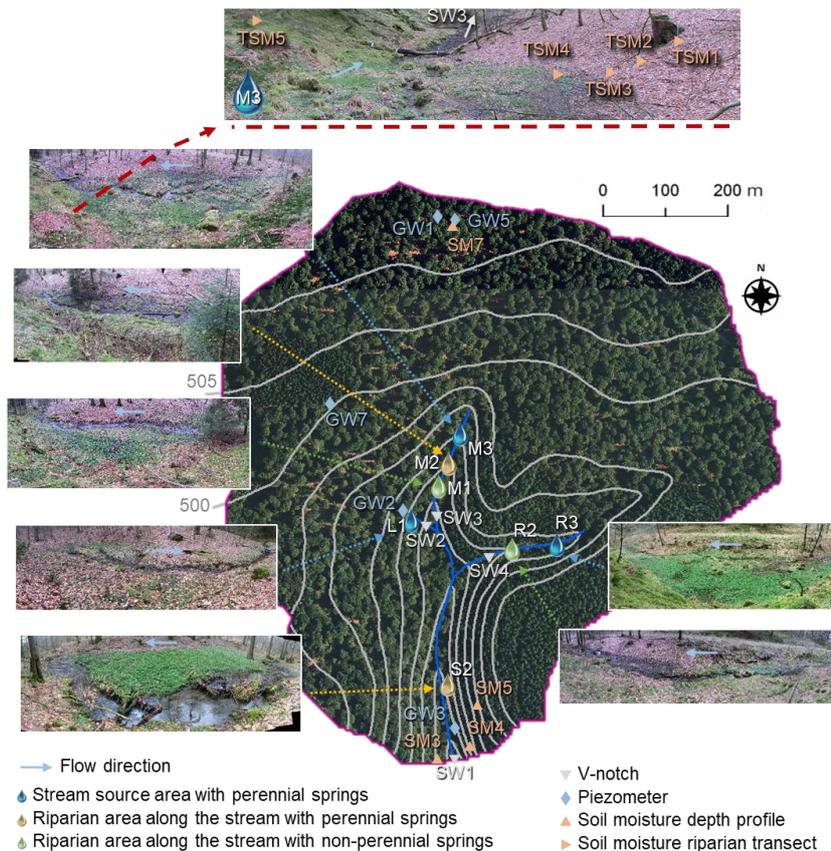
Figure 1: The Weierbach catchment with the locations of the installed v-notches, piezometers, soil moisture sensors and the seven investigated riparian areas.

15

4

Hydrology and
Earth System
Sciences
Discussions

Open Access

EGU

2.2 Surface saturation

Here, we define surfaces as saturated as soon as water is standing or flowing on the ground surface (Glaser et al., 2018). This involves water bodies such as lakes and streams, but excludes mere saturation in the topsoil. According to this definition, surface saturation in the Weierbach catchment generally only occurs in the streambed and the adjacent riparian zones. Other

5   areas that were occasionally observed to be surface saturated during very wet conditions or 'rain on snow' events are forest roads and the prolongation of the streambed above the source regions into the hillslopes. We focus in this study on seven distinct riparian areas in the catchment, which can be classified into three different categories (cf. Antonelli et al., 2019): i) stream source areas with perennial springs (L1, M3, R3, blue areas Fig. 1), ii) areas along the stream with perennial springs (M2, S2, yellow areas Fig. 1), and iii) areas along the stream with non-perennial springs (M1, R2, green areas Fig.1).

10   We mapped the surface saturation in these seven riparian areas weekly to biweekly from November 2015 to December 2017 with thermal infrared imagery (TIR). Details on the identification of surface saturation with TIR imagery and on the collected surface saturation dataset are presented and discussed in Glaser et al. (2018) and Antonelli et al. (2019). In brief, we created panoramic TIR images of the distinct areas and identified the locations of surface saturation (including the stream) within the images. To do this, each pixel in an image was assigned to be saturated or unsaturated based on the temperature range of

15   locations that were obvious to be saturated from field observations and visual images. In case the contrast between water temperature and temperature of surrounding materials was not sufficient for a reliable pixel classification, the images were excluded from the analysis. In case the pixel classification was affected by a poor temperature contrast or by pixels representing vegetation or snow cover in the images, the images were analysed but flagged as less reliable. Altogether, we obtained 291 binary panoramic images showing the temporal dynamics of surface saturation patterns in the seven studied riparian areas with

20   total numbers of images per site ranging between 34 (L1) and 48 (M2).

Time series of saturation were created for each area by accounting for the percentage of saturated pixels within the individual panoramic images. We normalized the saturation percentages to the maximum observed percentage of saturation in the distinct areas in order to allow a comparison of the saturation dynamics between the different riparian areas. For picturing the spatial surface saturation dynamics within a distinct riparian area, we created maps of saturation frequency. We counted for each area

25   how often the individual pixels of the panoramic TIR images were classified as saturated and normalized the resulting frequency numbers by the total number of TIR images analysed for that area.

The resulting maps of normalized saturation frequency rarely showed pixels that were always saturated (i.e. reaching a normalized frequency of 1). In reality, surface saturation was more persistent than indicated by the frequency maps. The reason for this artefact is that the perspective of the individual TIR panoramas was not 100% identical for all mapping instances and that vegetation sometimes covered parts of the saturated surface, especially during near dry conditions. We co-registered the

30   individual panoramas against a reference panorama for each area, but slight position shifts were inevitable. As a result, the images that were placed on top of each other did not always overlap exactly and the generated saturation frequency maps are blurred. Nonetheless, the maps of normalized saturation frequencies are very useful to understand at a glance where surface saturation occurs more and less frequent within an area and to be used for model validation.

35   ### 3 Catchment model

3.1 Model setup and parameterisation

We simulated the spatio-temporal dynamics of surface saturation across the Weierbach catchment with HydroGeoSphere (HGS, Aquanty Inc.). HGS is an integrated surface subsurface hydrological model and allows simultaneous simulation of transient surface and subsurface flow. Subsurface flow is simulated based on the 3D Richards equation. Surface flow is

40   simulated based on the diffusive-wave approximation of the 2D Saint Venant equation. Evapotranspiration is simulated with a comparatively simple approach, following the mechanistic concept of Kristensen and Jensen (1975). The equations are

<div align="center">5</div>

Hydrology and
Earth System
Sciences

Discussions

linearized implicitly using the Newton-Raphson approach and solved in an unstructured finite element grid. HGS has been used in the past for addressing diverse questions at various temporal and spatial scales (e.g. Ala-aho et al., 2015; Davison et al., 2018; Erler et al., 2019; Frei et al., 2010; Munz et al., 2017; Nasta et al., 2019; Partington et al., 2013; Schilling et al., 2017; Tang et al., 2018). It also has already been applied for a 6 ha headwater region of the Weierbach catchment (Glaser et

5   al., 2016, 2019). In this study, we applied the parameterization of Glaser et al. (2016) to the entire 42 ha catchment without performing an additional parameter calibration.

The catchment was spatially discretized into 42,274 triangular elements, using the mesh generator AlgoMesh (HydroAlgorithmics Pty Ltd). Edge lengths of the mesh elements ranged from > 30 m at the plateau to < 0.4 m for the seven analysed riparian zones and the streambed (Fig. 2). It was crucial to use such a fine mesh resolution in the riparian zone in

10   order to enable a comparable spatial detail as obtained with the TIR imagery for the surface saturation patterns. Vertically, the model grid comprised 5 m, which were divided into 14 layers with element depths ranging from 0.15 m for the top layers to 0.5 m for the bottom layers (Fig. 2).
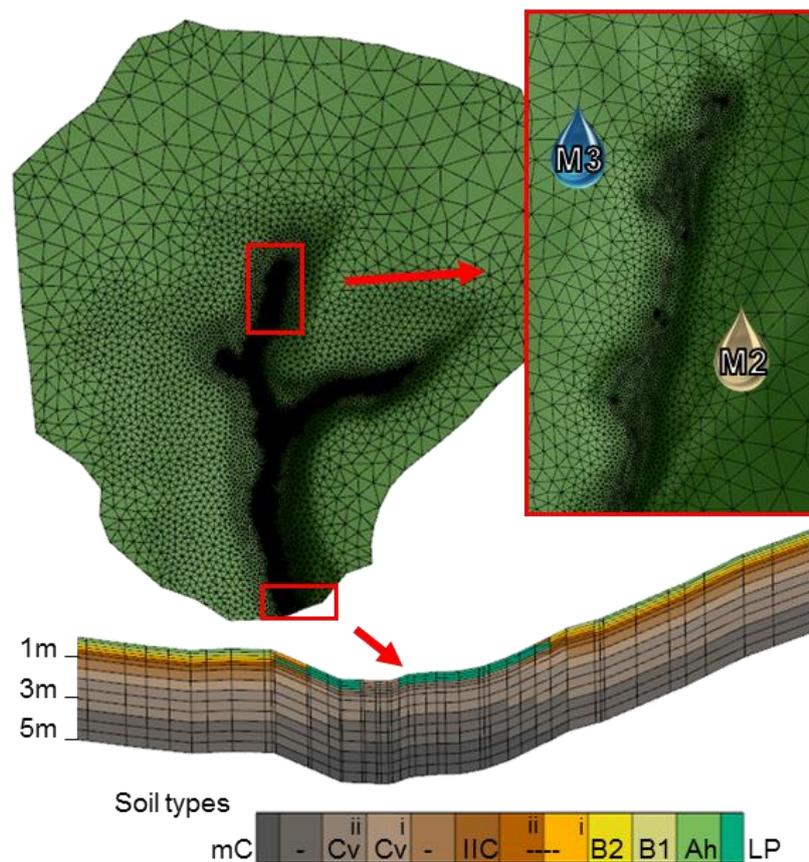
15   **Figure 2: Setup of the model mesh with a zoom on the fine horizontal resolution in the riparian areas and the streambed (inset on the right) and a vertical cross section through the stream valley and adjacent hillslopes (bottom) showing the vertical discretization and assignment of different soil properties (cf. Tab. 1). Ah = topsoil, B1 and B2 =subsoil, IIC = basal layer, Cv = fractured bedrock, Cm = fresh bedrock.**

6

Hydrology and
Earth System
Sciences
Open Access
Discussions

The subsurface was parameterized homogeneously with 10 different property layers, representing top- and subsoil (Ah, B1, B2), the basal layer (IIC), fractured and fresh bedrock (Cv, Cm), and layers of transition between subsoil, basal layer, and fractured bedrock (Fig. 2). We implemented spatial heterogeneity in the stream valleys, where soil and basal layer were eroded and the outcropping fractured bedrock was overlain with organic, stagnic Leptosol in the riparian zones (Fig. 2). We used the

5   Mualem-van Genuchten soil hydraulic functions for describing the saturation-pressure relation. The necessary soil hydraulic parameter values for the different property layers (porosity, residual saturation, van Genuchten α, van Genuchten β, saturated hydraulic conductivity, Tab. 1) were assigned according to Glaser et al. (2016). We only parameterised one additional layer for the fractured bedrock (Cv (ii)) in order to account for the adapted depth of 5 m in the catchment model compared to the depth of 3 m in the headwater model.

10

Table 1: Soil hydraulic parameters of the different soil property zones. Table adapted from Glaser et al. (2016)

Soil property zone	Residual saturation	van Genuchten parameter α [m^{-1}]	van Genuchten parameter β	Porosity	Saturated hydraulic conductivity [m d^{-1}]
Ah	0.12	6.6	1.46	0.74	1.71E+01
B1	0.10	22.1	1.42	0.61	1.71E+01
B2	0.10	22.1	1.42	0.45	4.59E+01
B2-IIC (i)	0.10	22.1	1.42	0.3	9.30E+02
B2-IIC (ii)	0.10	22.1	1.42	0.15	2.04E+03
IIC	0.02	6.0	1.50	0.20	8.40E+02
IIC-Cv	0.02	6.0	1.50	0.15	3.00E+00
Cv (i)	0.02	6.0	1.50	0.10	1.20E-02
Cv (ii)	0.02	6.0	1.50	0.07	1.20E-02
Cv-mC	0.02	6.0	1.50	0.05	9.00E-04
mC	0.02	6.0	1.50	0.01	2.40E-05
LP	0.10	22.1	1.42	0.61	7.80E+00

Surface and subsurface flow were coupled via a Darcy flux exchange through a thin coupling layer (10^{-4} m). We assumed different Manning's surface roughness values for the forested area ($1.24*10^{-6}$ d m$^{-1/3}$), the riparian zone ($9.41*10^{-7}$ d m$^{-1/3}$),

15   and the stream bed ($4.4*10^{-7}$ d m$^{-1/3}$) (cf. Glaser et al., 2016). Evapotranspiration properties (Tab. S1) were assigned individually for the deciduous forest, the coniferous forest in the southeast of the catchment, and the riparian zones including the streambed and values were based on the calibrated values of Glaser et al. (2016). The simulation was driven with daily sums of precipitation and reference evapotranspiration, which were treated as being spatially uniform. The outer edge of the surface domain was assigned as critical depth boundary, allowing water to leave the model domain via surface flow. Side and

20   bottom boundaries of the subsurface domain were no flow boundaries. A spin-up simulation drained the catchment from full saturation to steady state conditions (for 1 mm d^{-1} of precipitation, no evapotranspiration) and subsequently repeated the period from October 2013 to October 2015 three times for obtaining realistic initial conditions. The actual simulation spanned over the period from October 2015 to January 2018, the period where we mapped surface saturation with TIR imagery.

3.2 Assessment of model performance

25   We benchmarked the model against measured discharge, groundwater level, soil moisture, and surface saturation patterns and dynamics at various locations (Fig. 1). We calculated the Kling Gupta Efficiency (KGE) as a combined measure for correlation, bias, and relative variability (Gupta et al., 2009) between simulated and observed discharge. We also calculated KGEs for the simulated groundwater levels, but particularly evaluated the groundwater level dynamics rather than absolute values based on

Hydrology and
Earth System
Sciences
Discussions

Pearson correlation coefficients. Soil moisture was also evaluated based on its dynamics with Pearson correlation coefficients, while absolute values were only compared visually. Since simulated soil moisture was extracted from model nodes whose depths did not exactly correspond with the measurement depths, we interpolated depth-weighted average values from the model output for calculating the correlation with the observations in the respective depths. The interpolated model values of

5   volumetric water content were then correlated with the observations of water content, averaging the measurements of the twin depth profiles at the monitoring sites.

For comparing the simulation output with the surface saturation information obtained with the TIR images, it was necessary to convert the model output into a comparable format via several processing steps: First, we extracted the surface water depths in the surface domain of the model for noon of the days where TIR images were taken and analysed. Next, we transformed the

10   surface water depths into a binary saturation map of the entire catchment by classifying the surface domain cells as saturated if water depths were $>10^{-4}$ m. The depth of 10^{-4} m corresponds to the penetration depth of the used TIR camera for water columns and thus is the minimum depth that could be detected as pure water temperature signal with the camera. Finally, we projected the model output into jpeg images with the same perspective and extent of the TIR panoramic images by turning, bending, and cutting the modelled saturation maps according to each of the seven riparian areas individually. This model

15   output processing allowed us to perform the same calculations for the model output as for the TIR images, i.e. to create time series of normalized saturation and maps of normalized saturation frequencies for the seven riparian areas with comparable perspectives and extents. Since it was not possible to project the model output identically to the perspectives of the TIR images, we compared the saturation dynamics and patterns of the model images with the observations qualitatively (visually) only. A quantitative comparison would have been biased by differences in image distortions and total area extent.

20   Furthermore, we compared the simulated frequency of surface saturation with the simulated frequency of groundwater reaching the surface. To do this, we marked the surface cells below which the subsurface domain was fully saturated from the bottom to the surface as cells where groundwater reached the surface. This binary information was transformed into a frequency map analogous to the procedure for creating the surface saturation frequency maps, using the same output times.

4 Results

25   ### 4.1 Simulation of discharge, groundwater level and soil moisture

The model reproduced the seasonal dynamics of measured discharge very well (Fig. 3, Fig. S1). The best fit was obtained at the outlet (SW1) with a KGE of 0.74. Discharge at SW2, SW3, and SW4 was reproduced equally well with KGEs of 0.49, 0.48, and 0.47. Groundwater levels were captured well with the model at the locations close to the riparian zone (KGE=0.57, r=0.78 for GW2; KGE=0.64, r=0.84 for GW3). At hillslopes and plateau, simulated groundwater levels were similar to the

30   observed levels during the wet season, but during dry conditions the groundwater levels did not fall deep enough (Fig. 3, Fig. S1). This level discrepancy was reflected in low KGEs (0.30 for GW1, 0.21 for GW5, 0.02 for GW7). However, general dynamics of level increasing and decreasing were also captured at hillslopes and plateau (r = 0.66 for GW5, r = 0.62 for GW7, and r = 0.76 for GW1; note that the value for GW1 only includes data for wet periods, since the piezometer fell dry during summer months).

35   Simulated soil moisture generally showed a transition from higher to lower responsiveness from topsoil to subsoil layers consistent with the monitored soil moisture (Fig. 3, Fig. S1) and Pearson correlation coefficients indicated overall a good agreement between simulated and observed soil moisture dynamics (Tab. 1). As for the groundwater levels at the hillslopes and plateau, soil moisture observations showed a distinct decrease in water content during dry periods, which the simulation could not reproduce to the same extent. The observed water content in the riparian zone was always close to saturation (TSM4,

40   Fig. 3), while the simulation showed a decrease in water content during dry periods in the riparian zone. Yet the simulation also showed a spatial trend for more permanent soil saturation in the riparian zone (TSM4) and its vicinity (TSM3, Fig. S1)

8

Hydrology and
Earth System
Sciences
Discussions

than at the hillslopes and plateau. The simulated values of water content were similar to the observed values at some locations (e.g. TSM2, SM4, Fig. 3) and clearly differed at other locations (e.g. SM7, Fig. 3), but the matches and mismatches of the volumetric water content did not clearly depend on specific areas or landscape units. Moreover, we think that moisture dynamics and responsiveness are more informative for model performance than the absolute water content values, since also

5   the measured values of volumetric water content differed from each other within small distances (e.g. measurements of water content in 10 cm depth at profile SM7, Fig. 3).
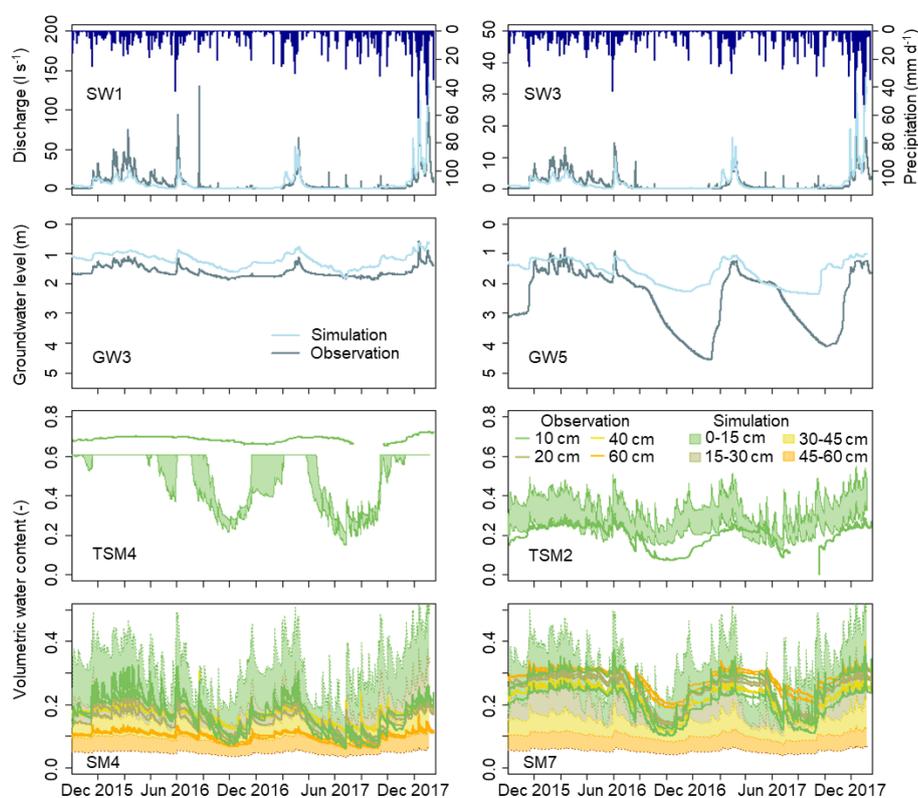

Figure 3: Simulated and observed time series of discharge, groundwater level below the surface, and volumetric water content.
10   **Colour bands indicate the possible span of simulated volumetric water contents in the depths between two model nodes. The time series of the observation locations (cf. Figure 1) that are not shown here, are shown in the supplemental material (Figure S1).**

Table 2: Coefficients of Pearson correlation between simulated and observed volumetric water content of the soil for the different measurement locations and depths (cf. Fig. 1).

	SM3	SM4	SM5	SM7	TSM1	TSM2	TSM3	TSM4	TSM5
10cm	0.54	0.75	0.70	0.59	0.60	0.62	0.67	0.30	0.85
20cm	0.67	0.82	0.76	0.62					
40cm	0.82	0.89	0.88	0.79					
60cm	0.85	0.92	0.91	0.82					

9

4.2. Dynamics of surface saturation

The observed dynamics of normalized surface saturation (Fig. 4, coloured lines) were similar for all seven investigated riparian areas and followed the seasonal trend of the catchment discharge. Yet some differences between the studied areas were discernible. For example, saturation was less persistent between February and April 2016 in the two areas without perennial

5   springs (M1, R2, Fig. 4) than in the other areas. Maximum saturation was reached in December 2017 at M1, R2 and S2, but between February and April 2016 at the other locations (Fig. 4). As for the observations, the simulated dynamics of normalized surface saturation (Fig 4, black lines) followed the general trend of the simulated discharge dynamic. The simulation showed a faster decrease and increase of the normalized saturation during dry periods than it was observed in most areas. However, simulated discharge also seemed to decrease and increase earlier than it was observed (c.f. Section 4.3). The simulated

10   saturation dynamics did not clearly differ between the different locations and thus behaved more synchronous than the observations (e.g. maximum simulated saturation in December 2017 in all areas). As a result, the match between simulated and observed dynamics of normalized saturation was better for some areas (e.g. M1, R2, Fig. 4) than for others (e.g. S2, L1, Fig. 4).

The dynamic changes of normalized simulated saturation matched the normalized observations generally well, despite of

15   under- and over-estimated amounts of minimum and maximum absolute saturation for all areas. The minimum amount of saturated pixels in the TIR panoramas ranged between 0.02 % at M3 and R3 and 3.38 % at S2, while the model did not simulate any surface saturation during the driest period (Fig. 4). In addition, simulated normalized saturation stayed longer close to the minimum than the observed saturation for several areas (L1, S2, M1). These results show that the model simulated a stronger dry-out than observed in the Weierbach. At the same time, the simulation overestimated maximum saturation in the riparian

20   zone (Fig. 4). The overestimation was not equally strong at the seven investigated areas and as a result, the distinction between areas showing higher or lower maximum saturation was not the same for observation and simulation (e.g. R3 showing one of the highest maximum saturation in the observation, but one of the lowest maximum saturation in the simulation compared to the other areas).

4.3 Discharge – surface saturation relationship

25   The Pearson correlation between normalized saturation and discharge at the outlet SW1 was > 0.65 for both the simulation and the observation in almost all riparian areas. L1 was the only exception with $r_{obs} = 0.54$ (Fig. 5). The simulated relationships between normalized saturation and discharge resembled the observed relationships in terms of value range and shape (Fig. 5), although the observation data scattered distinctly more than the simulation data. A power law relationship approximated the observed relationship between discharge and saturation for all seven areas, when data that were taken during rainfall or rising

30   discharge were excluded (cf. Antonelli et al., 2019). For some areas, the simulation matched the trend lines of the observation data closely (e.g. L1, M2). For other areas, the visual fit of the model output to the observation data was less good (e.g. S2, R3), but still described a similar trend.

Despite the common shape of a power law function, the saturation – discharge relationships were slightly different between the different areas, both for observation and simulation data. For example, the power law functions fitted to the observations

35   showed that saturation during high flow conditions ($> 5 \, l \, s^{-1}$) increased most strongly with discharge in the sources areas (especially M3 and R3). During low flow conditions ($< 1 \, l \, s^{-1}$), the source areas (L1, M3, R3) showed the lowest amount of normalized saturation and the least change relative to discharge compared to the other areas. In the simulated relationships, the increase in saturation for high discharge ($> 5 \, l \, s^{-1}$) was strongest for M3 and S2. The simulated relationship between discharge and surface saturation during low flow ($< 1 \, s^{-1}$) was similar for all areas in terms of slope, but differed in the amount

40   of normalized saturation, being highest for areas in the east stream branch (R2, R3), followed by the middle upstream branch (M1, M2, M3), and L1 and S2.

10

Hydrology and
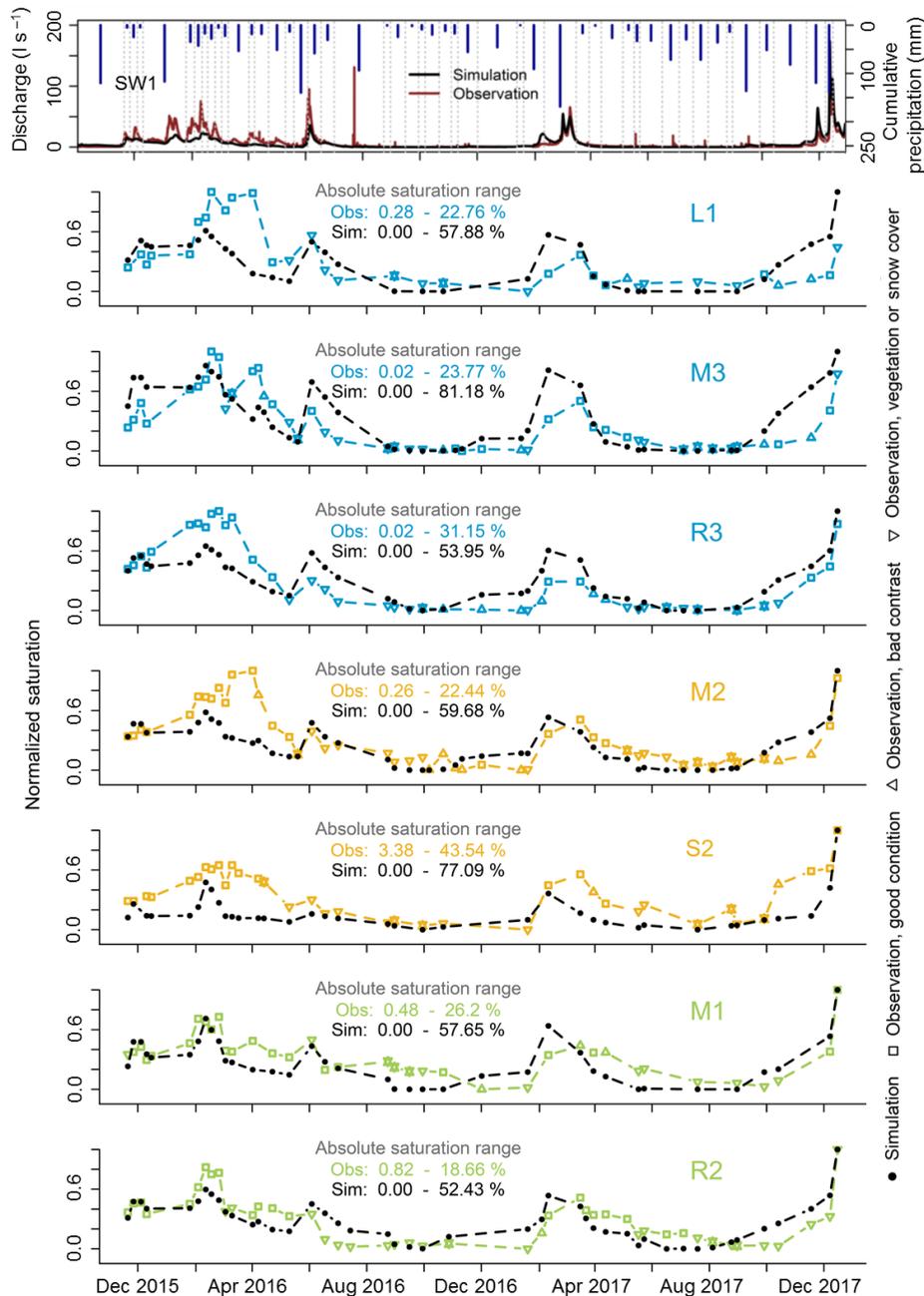Earth System
Sciences
Discussions
Open Access

Figure 4: Time series of observed and simulated surface saturation in the seven investigated riparian areas. Surface saturation is normalized to the minimum and maximum amount of saturation that was observed and simulated in the individual areas, respectively. Observations that were derived from TIR images with a poor temperature contrast or with influences of vegetation and snow cover are deemed less reliable. Cumulative precipitation between the measurement dates (grey dashed lines) and discharge at catchment outlet SW1 are shown in the top panel for facilitating the comparison to precipitation and flow conditions.

11
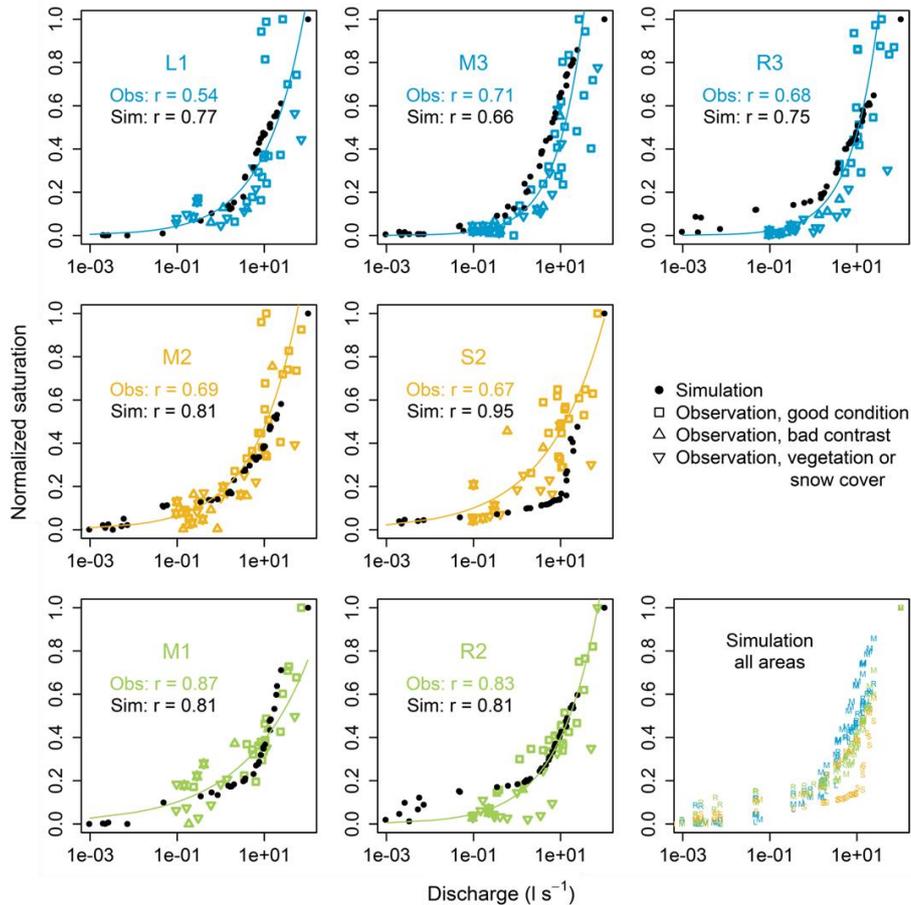
Figure 5: Observed and simulated relationships and Pearson correlations between normalized surface saturation and discharge at the catchment outlet SW1 for the seven investigated riparian areas. Observations that were derived from TIR images with a poor temperature contrast or with influences of vegetation and snow cover are deemed less reliable. Solid lines are power law curves
5 **fitted to the observation data, excluding data taken during rainfall or rising discharge. For facilitating the comparison between the seven areas, the panel on the bottom right contains the simulated data points from all seven areas and the area affiliation is indicated with the respective colour and letter.**

4.4 Spatial patterns of surface saturation

10   The realism of simulated patterns of surface saturation was evaluated for each riparian area by visually comparing the surface
saturation frequency maps obtained from the simulations and observations (Fig. 6). The model captured the location of the
stream and the locations that intermittently became surface saturated well for most of the seven investigated areas. For example,
both observation and simulation showed that only the right side of the stream became saturated in M1, that the riparian zone
of the right streamside in M2 became saturated only in the upstream part, and that saturation mainly developed on the left

15   streamside in R3, surrounding some permanently dry areas next to the stream (Fig. 6). The only area with a clear mismatch
between observed and simulated patterns of surface saturation was area L1, where surface saturation was simulated on the
opposite streamside and at a clearly wrong position along the stream (upstream vs downstream).

12

The simulated surface saturation also reflected the observed saturation frequencies well. The simulation reproduced the general picture of more frequent surface saturation in the streambed than at the streamsides, but - as for the saturation patterns - simulated and observed frequencies corresponded better in some areas (e.g. S2, Fig. 6) than in others (e.g. R3, Fig. 6). For example, the observed frequency of surface saturation in the streambed was generally lower in the source areas (L1, M3, R3) than in the mid- and downstream areas (M2, S2, M1, R2), while the simulated frequency of surface saturation in the streambed was more similar between the areas and particularly overestimated in L1 and R3.
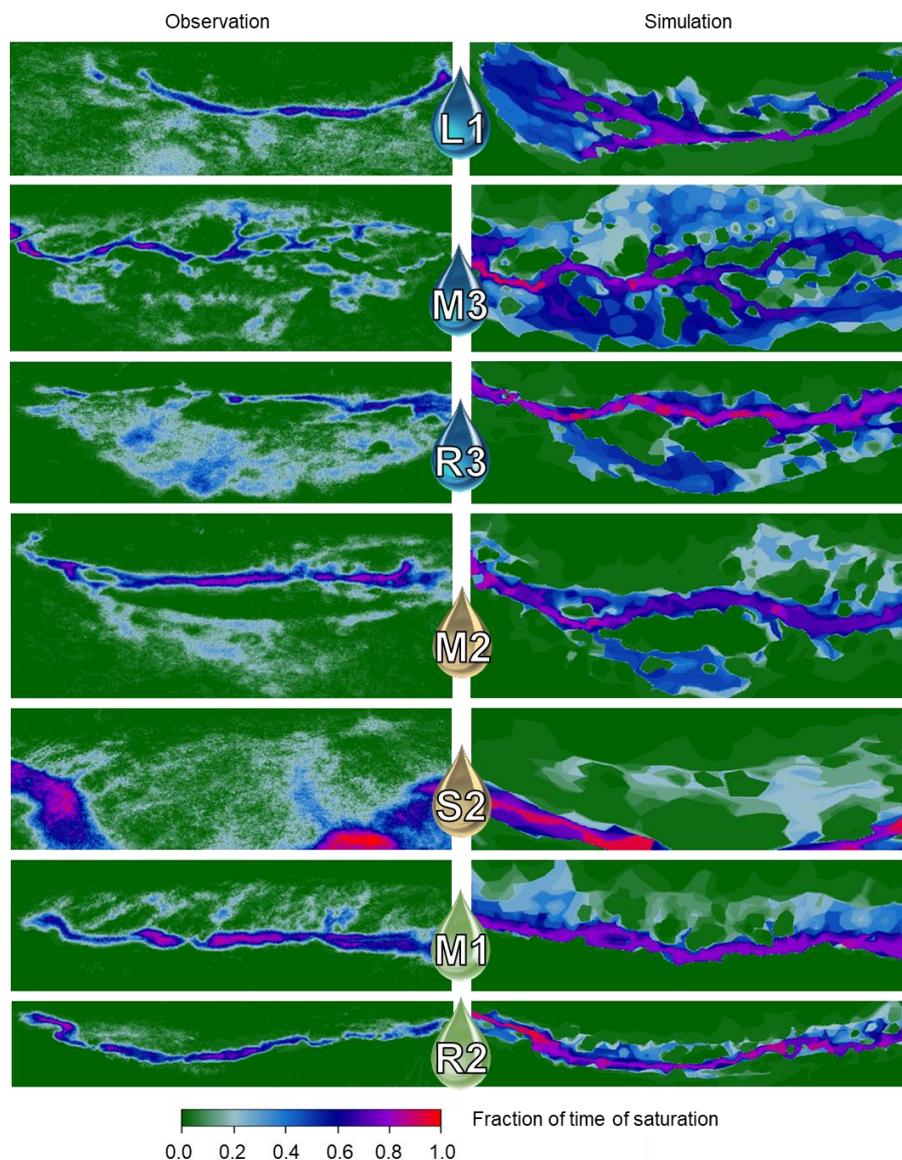
Figure 6: Observed (left) and simulated (right) frequencies of surface saturation in the seven investigated riparian areas. The maps were created by first counting how often the individual pixels were classified as saturated in the individual panoramic images and second normalizing the resulting frequency numbers by the total number of images analysed for the respective area.

13

4.5 Simulated patterns and dynamics of surface saturation versus groundwater reaching the surface at catchment scale

Simulated surface saturation generally occurred only in the streambed and adjacent riparian zones (Fig. 7a). During the wettest conditions of the study period (winter 2017/2018), surface saturation also occurred as prolongation of the eastern stream branch into the hillslope above the source area R3. This simulated occurrence behaviour of surface saturation across the catchment is

5   in accordance with field evidence, where we observed surface saturation outside of the valley bottom only during very wet conditions or rain on snow events (cf. Section 2.2). The simulated patterns of where and how frequently groundwater reached the ground surface (Fig. 7b) proved to be very similar to the surface saturation frequency map of the catchment (Fig. 7a). The only obvious difference occurred in the area above the source area of the eastern stream branch (R3), with a smaller extent of groundwater reaching the surface than extent of surface saturation.

10
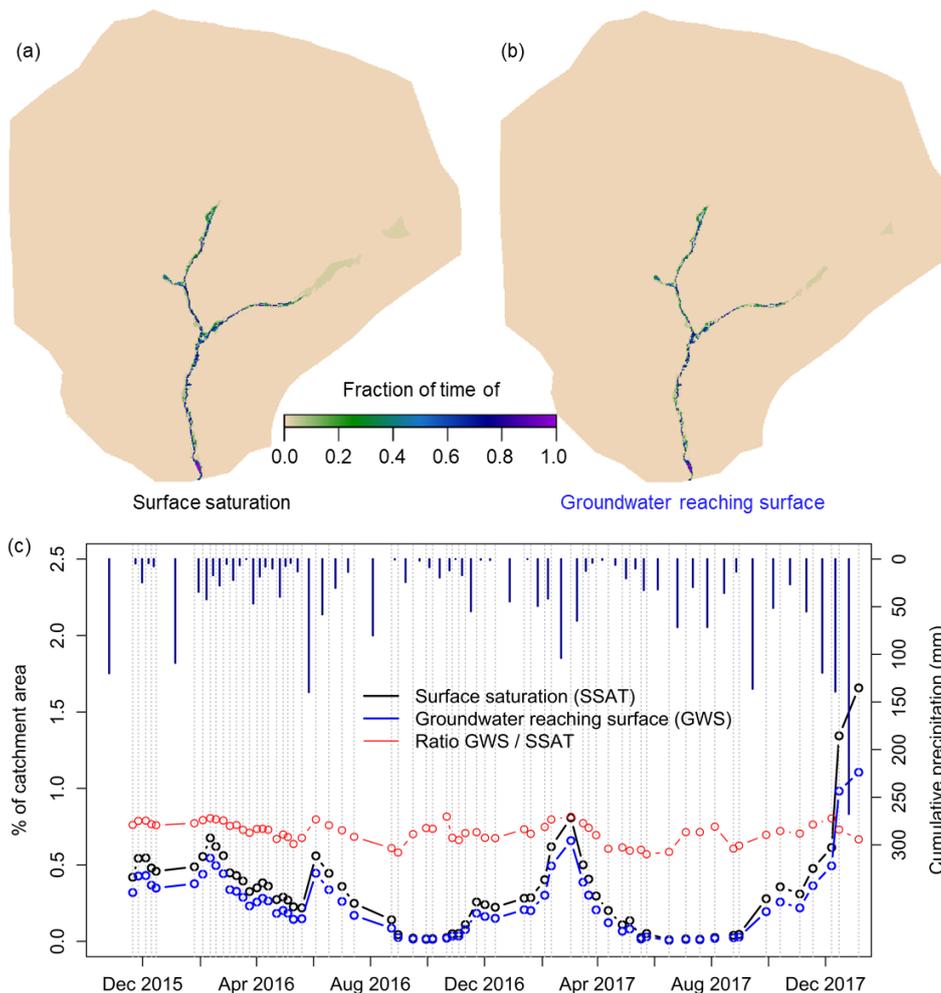
Figure 7: Simulated frequency maps (a, b) and time series of percentage (c) of surface saturation and groundwater reaching the surface in the Weierbach catchment. Precipitation is given as cumulative amounts between the observation dates (grey dashed lines).

14

Hydrology and
Earth System
Sciences

Open Access

EGU

Discussions

The time series of simulated percentage of catchment area with surface saturation and groundwater reaching the surface revealed that the area where groundwater reached the surface was always smaller in extent than the surface saturated area, even after dry condition (Fig. 7c). The biggest absolute difference between the areal extent of surface saturation and groundwater reaching the surface was simulated during winter 2017/2018 (1.66 % vs 1.1 % of catchment area), where the conditions were very wet with high discharge and high cumulative precipitation and where the difference in areal extent was also visible in the frequency maps (Fig. 7a and b). However, the ratio between the extent of groundwater reaching the surface and the extent of surface saturation was not exceptionally high during winter 2017/2018. Instead, the ratio scattered without a clear trend between 0.57 and 0.82 during the entire simulation period, apparently independent from the cumulative amount of precipitation or surface saturation.

5 Discussion

The aim of this study was to analyse the spatio-temporal variability of surface saturation within the Weierbach catchment, with a focus on the stream valleys and riparian zones. Even though simulated discharge, groundwater levels and soil moisture showed some discrepancies to observations in terms of absolute values, we would argue that the performance of the different time series at different locations was quite good for a model that was not calibrated and set up rather homogenously across the catchment. While the model had some problems to reproduce soil moisture and groundwater levels during the dry conditions at hillslopes and plateau, the simulated time series matched the observations especially well in the riparian zone and vicinity. This gives us confidence that the model setup was valid for evaluating and analysing the spatio-temporal dynamics of surface saturation and its intra-catchment variability.

5.1 Temporal dynamics of surface saturation

The model reproduced the observed long-term dynamics of surface saturation over different seasons and wetness conditions well. Our study goes beyond previous works that compared the simulation of surface saturation dynamics with observations (e.g. Ali et al., 2014; Birkel et al., 2010; Glaser et al., 2016; Mengistu and Spence, 2016) by relying on a longer study period and a higher number of observations in time. This allowed us to analyse and compare various hydrological conditions and the dynamic transition between them over all seasons with a frequent number of observations. Moreover, we accounted for spatial variability of saturated area dynamics within the catchment. Unlike the various quasi dynamic wetness indices presented in Ali et al. (2014), which could not satisfyingly reproduce the spatio-temporal variability of connected surface saturation observed in a catchment in the Scottish Highlands, our model reproduced the distributed dynamics of surface saturation well, without clear performance differences for different wetness conditions.

Simulations and observations showed both that the temporal dynamics of surface saturation were mostly consistent across the catchment. Moreover, our simulations showed that the spatio-temporal development of surface saturation was very similar to the spatio-temporal dynamics of groundwater reaching the surface (cf. Fig. 7). This suggests that the generation of surface saturation in the Weierbach catchment is largely driven by the synchronous exfiltration of groundwater in topographic depressions. Antonelli et al. (2019) drew consistent conclusions based on a statistical analysis of the observation data.

5.2 Relation between surface saturation and discharge

We found that the observed and simulated relationships between surface saturation and discharge resembled power law relationships (cf. Fig. 5). This is consistent with earlier studies that showed power law relationships between contiguous connected surface saturated areas and discharge (Mengistu and Spence, 2016; Weill et al., 2013). In contrast to these studies, we did not observe hysteretic loops in the relationship between saturation and high streamflow. Nonetheless, the scatter in the observed discharge – surface saturation relationships might indicate that the development of surface saturation in the

15

Hydrology and
Earth System
Sciences
Discussions

Weierbach catchment follows hysteretic loops, but that the hysteresis was not resolved with the available temporal resolution of the observations. For example, it is likely that surface saturation evolved in the riparian areas during high flow conditions and persisted on the ground surface during decreasing streamflow due to restricted infiltration capacities of the riparian soil (cf. Antonelli et al., 2019).

5 The lack of such a hysteretic process in the simulation could explain why the model showed the tendency for less persistent and faster contracting surface saturation. It may also explain why the simulated saturation dynamics differed less between the different investigated areas than the observed dynamics. It is likely that the observed saturation dynamics were not synchronous between the different areas due to a less persistent (and thus hysteretic) generation of surface saturation in the relatively narrow riparian areas without perennial springs (M1 and R2) compared to the wider riparian areas with perennial springs (cf.

10 observation of less persistent saturation in M1 and R2 during February and April 2016, Fig. 4). The model, instead, simulated a non-hysteretic saturation behaviour for all investigated riparian areas, which resulted in a better fit between simulated and observed dynamics in the areas M1 and R2 compared to the other areas.

At the same time, it might also be that the simulated relationship between saturation and discharge was correct in all riparian areas and that the scattering of the observation data did not result from hysteretic behaviour, but from uncertainties in the TIR

15 methodology. A good argument for a correct simulation of the discharge – surface saturation relationship is that not only simulated saturation but also simulated discharge seemed to be less persistent and to decrease and increase earlier than it was observed. In reality, the scatter of the observation data is likely related to both measurement uncertainties and hysteretic aspects and a future study with higher temporal resolution of field observations and corresponding simulation output could further analyse this.

20 Independently from the question on hysteretic loops, we found that the discharge – surface saturation relationships somewhat differed between the different areas. We could connect the main differences to different topographical and morphological features, yet we cannot decipher why the main controlling feature for the discharge – surface saturation relationship was different between observations (source areas vs non-source areas) and simulations (different stream branches, cf. Section 4.3). Nonetheless, our findings are in line with experimental studies that discussed that the relationships between baseflow discharge

25 and total extent of contributing saturated areas differ between catchments with different physiographic characteristics (e.g. Dunne et al., 1975; Latron and Gallart, 2007).

By comparing our model results to the double logarithmic plot presented by Latron and Gallart (2007) (Figure 8), we could identify similar shape varieties of the discharge – surface saturation relationship for the different areas studied within the Weierbach catchment as observed for the different catchments presented in Latron and Gallart (2007). We cannot compare our

30 results directly with the results shown in Latron and Gallart (2007), since we evaluated absolute discharge and normalized saturation, while they evaluated connected saturated areas in percentage of catchment area, but discharge normalized to the catchment area. In order to facilitate the comparison and to connect the two plots (Fig. 8a, 8b), we show the simulated relationship between discharge and surface saturation of the entire Weierbach catchment in both plots, once with normalized discharge and absolute saturation (Fig. 8a), and once with absolute discharge and normalized saturation (Fig. 8b). The shape

35 of the relationship for the entire Weierbach catchment was nearly linear, similar to the relationship observed in the Can Vila catchment investigated by Latron and Gallart (2007) (Fig. 8a). The relationships of the seven studied riparian areas differed from the catchment relationship and between each other (Fig. 8b). For example, area S2 and M1 showed a convex shape similar to the observations in the Vermont W2 catchment made by Dunne et al. (1975), area M3 showed a rather concave shape similar to the relationships found for a sub-catchment of the Asker basin (Myrabø, 1986) and the Strengbach catchment (Latron, 1990),

40 area M2 showed a rather linear shape similar to the Can Vila catchment studied by Latron and Gallart (2007). This clearly shows that differences in the relationship between surface saturation and discharge do not only occur between different catchments, but that they also occur as intra-catchment variability.

16

Hydrology and
Earth System
Sciences
Discussions
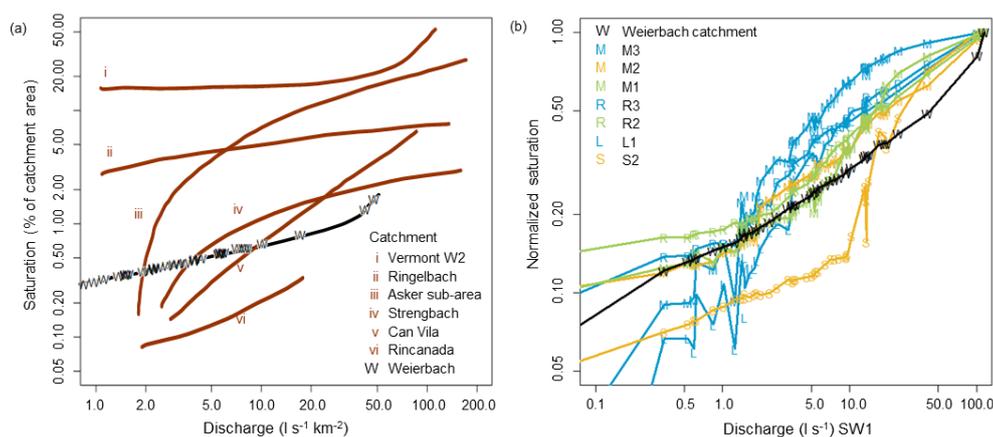
Figure 8: Simulated relationship between discharge and surface saturation of the entire Weierbach catchment (marked with W) in comparison to (a) the relationships observed in other catchments (Figure modified from Latron and Gallart (2007)) and (b) the relationships simulated for the seven investigated riparian areas within the catchment. The presented relationships of the other catchments were investigated by i) Dunne et al. (1975), ii) Ambroise (1986), iii) Myrabø (Myrabø, 1986), iv) Latron (Latron, 1990), v) Latron and Gallart (2007), and vi) Martinez-Fernandez et al. (2005). Area affiliation for the investigated riparian areas of the Weierbach catchment is indicated with the respective colour and letter (cf. Fig. 4-6).

5.3 Spatial patterns of surface saturation

The observed spatial patterns of surface saturation were reproduced with the simulations in great detail for most of the investigated areas. We attribute the successful simulation of the spatial patterns to microtopography (local topographical features with extents of centimetres to few metres) since i) microtopography described the main spatial variability between the seven investigated areas in the model setup and ii) we observed that small changes in the setup and resolution of the model mesh in the riparian zones changed some details of the simulated surface saturation patterns (Fig. S2, especially area M2, S2).

Therefore, we would like to stress that not only major topographic features of the catchment (e.g. hillslope shape, slope angle, valley width) but also its microtopography needs to be considered for identifying locations where surface saturation may occur. This may sound trivial and several studies have already pointed out the importance of microtopography for the simulation of different hydrological aspects such as hydraulic heads, hyporheic surface-subsurface water exchange, bank storage and overbank flooding, water quality of shallow groundwater systems and runoff generation (e.g. Aleina et al., 2015; Frei et al., 2010; Käser et al., 2014; Van der Ploeg et al., 2012; Tang et al., 2018). Still, microtopography is not often considered in the simulation of surface saturation patterns.

When microtopography is not resolved detailed enough, it is more likely that the simulated surface water extends over a large area instead of accumulating in topographic depressions and thus overrates the extent of surface saturation. In this context it is interesting to note that there are studies that simulated maximum extents of surface saturation up to 80 % of the study area (Qu and Duffy, 2007; Weill et al., 2013), while field observations have only reached maximum extents up to 25 % - 50 % of catchment area (Ali et al., 2014; Birkel et al., 2010; Dunne et al., 1975; Mengistu and Spence, 2016) and often show maximum extents around 10 % (Ambroise, 2016; Grabs et al., 2009; Güntner et al., 2004; Latron and Gallart, 2007; Tanaka et al., 1988). Microtopography might partly explain this discrepancy, even though the maximum extent of surface saturation certainly also depends on the climatic and physiographic conditions of the catchment and on the timing of the observations (e.g. baseflow conditions vs storm events) and there are some studies that simulated similar or less maximum extent of surface saturation

17

than observed without considering the microtopography (e.g. Ali et al., 2014; Birkel et al., 2010; Grabs et al., 2009; Güntner et al., 2004; Mengistu and Spence, 2016).

In our study, the simulated maximum extent of surface saturation was 1.6 % of catchment area, which is small compared to other simulation studies, but matches the observation that surface saturation commonly only occurs within the riparian zone

5 and streambed (extent of 1.2 %). Nonetheless, also our maximum saturation within the individual areas was overestimated compared to the observations (cf. Fig. 4). Besides the effect of microtopography, there are two other possible explanations for this. First, the largest simulated saturation occurred during winter 2017/2018, which is the same period where the model clearly overestimated discharge. This mismatch could partly explain the overestimation of saturation, assuming that the relationship between discharge and saturation was correctly captured with the model (cf. Section 5.2). Second, the overestimation of

10 absolute saturation could result from different perspectives and extensions of model output and TIR images (cf. section 3.2, Fig. 6). The TIR images included parts of the hillslopes around the riparian zones, which were not included to the same extent in the extracted model images. Since the hillslopes normally remained unsaturated, the maximum possible amount of saturated pixels in the TIR images was thus lower than in the model images, while the minimum possible amount of saturation was not affected. This could also explain why overestimation of total amounts of saturation was different between the different areas.

15 Despite the importance of microtopography, the model results showed that microtopography alone was not sufficient to capture the spatial patterns of surface saturation correctly. The simulated patterns of surface saturation clearly did not match the observed patterns equally well in all seven investigated areas (cf. Fig. 6), although the topographical information source and mesh resolution was consistent for the simulated riparian areas. This means that there are additional factors that control the spatial patterns of surface saturation that were not accounted for in the simulations. Such a factor could for example be the

20 structure of the subsurface, which was treated as being homogeneous between all investigated riparian areas in the simulations. In reality, the subsurface structure may locally differ to some degree, for example in the riparian area of the western stream branch (L1), where saturation was simulated at a clearly wrong side along the stream.

5.4 Frequency maps of surface saturation

The frequency maps of surface saturation combine information on when and where surface saturation occurs. We do not think

25 that the exfiltration of subsurface water into local depressions (cf. Section 5.1 and 5.2) can fully explain the spatial variability of saturation frequencies that was observed and simulated satisfactorily within the different riparian areas (Fig. 6). Instead, we suppose that the differences in saturation frequency were controlled by additional water sources than exfiltrating groundwater, such as stream water or direct precipitation, and that the contribution of these additional water sources to surface saturation varied in space and time. For example, the lower frequencies of surface saturation observations at the streamsides compared

30 to the streambed and the lower frequencies in the streambed of the source areas (L1, M3, R3) compared to the mid- and downstream areas (M2, S2, M1, R2) might reflect a lower and less frequent contribution of upstream water in these areas. The overestimation of simulated saturation frequencies in the streambed of R3 could thus indicate an overestimated upstream contribution due to simulating the stream extent too far upstream from the source area. Future work should analyse potential water sources and generation processes of surface saturation with a suitable model framework (cf. Partington et al., 2013;

35 Weill et al., 2013) in order to complement the interpretation of the observation data and to identify the mixture of different water sources of surface saturation (e.g. stream water, exfiltrating subsurface water, ponding precipitation), how the sources might vary in space and time, and how this might reflect in the surface saturation frequencies.

6 Summary and conclusions

We explored the intra-catchment variability of surface saturation in the Weierbach catchment with joint observations and

40 simulations. We showed that the model could reproduce the observed variability of the surface saturation characteristics

18

Hydrology and
Earth System
Sciences

Discussions

(dynamics, frequencies, patterns) with great detail, although the model setup was rather homogeneous and parameters were not calibrated at catchment scale. Our results demonstrated that a spatially distributed, physically-based, integrated hydrological model such as HGS is well-suited for reproducing and analysing the generation and development of surface saturation in space and time.

5 Based on the matches and mismatches between the simulation results and observations, we could identify some key factors controlling the surface saturation generation. The temporal occurrence of surface saturation was observed and simulated to be similar across the catchment, which we related – based on the simulation results – to a large influence of groundwater that reacts synchronous across the catchment. The spatial occurrence of the surface saturation differed between and within the seven investigated riparian areas, which we mainly could relate to the influence of microtopography. Furthermore, we

10 discussed that the full variability between the different areas and the mismatches between observations and simulation can only be explained with additional factors besides groundwater exfiltration and microtopography.

The spatially varying frequencies of surface saturation within the riparian areas indicated that there might be additional water sources than subsurface water that contribute to the generation of surface saturation. Since the model could reproduce the observed frequencies, the model can be used in a future study to analyse such a potential mixing of different water sources and

15 their variation in space and time. The observed differences between the investigated riparian areas with regard to the seasonal dynamics of saturation extension and contraction and the surface saturation – discharge relationship likely resulted from different morphological characteristics (width, existence of perennial springs) of the riparian areas. Although the model could not reproduce a varying hysteretic occurrence and persistence of surface saturation in the different investigated areas, also the simulation results demonstrated that the relationship between surface saturation and discharge can differ within a catchment

20 in the same manner as between catchments with different topographical and morphological conditions.

Data availability. Data underlying the study are property of the Luxembourg Institute of Science and Technology. They are available on request from the authors.

25 *Author contributions.* BG, LH and JK designed and directed the study. BG and MA planned and carried out the field work and processed the TIR images. BG set up the simulation and processed the model output. BG, MA, LH and JK discussed and interpreted the results. BG prepared the manuscript with contributions from JK and LH.

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

30

References

40 Ala-aho, P., Rossi, P. M., Isokangas, E. and Kløve, B.: Fully integrated surface–subsurface flow modelling of groundwater–lake interaction in an esker aquifer: Model verification with stable isotopes and airborne thermal imaging, J. Hydrol., 522, 391–406, doi:10.1016/j.jhydrol.2014.12.054, 2015.

Hydrology and
Earth System
Sciences
Discussions

Aleina, F. C., Runkle, B. R. K., Kleinen, T., Kutzbach, L., Schneider, J. and Brovkin, V.: Modeling micro-topographic controls on boreal peatland hydrology and methane fluxes, Bio, 12, 5689–5704, doi:10.5194/bg-12-5689-2015, 2015.

Ali, G., Birkel, C., Tetzlaff, D., Soulsby, C., Mcdonnell, J. J. and Tarolli, P.: A comparison of wetness indices for the prediction of observed connected saturated areas under contrasting conditions, Earth Surf. Process. Landforms, 39(3), 399–
5   413, doi:10.1002/esp.3506, 2014.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: Crop evapotranspiration (guidelines for computing crop water requirements), FAO Irrig. Drain. Pap., 56, 1998.

Ambroise, B.: Rôle hydrologique des surfaces saturées en eau dans le basin du Ringelbach à Soultzeren (Hautes-Vosges), France, in Recherches sur l'Environnement dans la Région, Actes du 1er Colloque Scientifique des Universités du Rhîn
10   Supérieur, edited by O. Rentz, J. Streith, and L. Ziliox, pp. 620–630, Université Louis Pasteur - Conseil de l'Europe, Strasbourg., 1986.

Ambroise, B.: Variable water-saturated areas and stream fl ow generation in the small Ringelbach catchment (Vosges Mountains , France): the master recession curve as an equilibrium curve for interactions between atmosphere , surface and ground waters, Hydrol. Process., 30, 3560–3577, doi:10.1002/hyp.10947, 2016.

15   Antonelli, M., Glaser, B., Teuling, R., Klaus, J. and Pfister, L.: Saturated areas through the lens: 1. Spatio-temporal variability of surface saturation documented through Thermal Infrared imagery, In preparation, 2019.

Birkel, C., Tetzlaff, D., Dunn, S. M. and Soulsby, C.: Towards a simple dynamic process conceptualization in rainfall – runoff models using multi-criteria calibration and tracers in temperate, upland catchments, Hydrol. Process., 24, 260–275, doi:10.1002/hyp.7478, 2010.

20   Carrer, G. E., Klaus, J. and Pfister, L.: Assessing the Catchment Storage Function Through a Dual-Storage Concept, Water Resour. Res., 55, 476–494, doi:10.1029/2018WR022856, 2019.

Davison, J. H., Hwang, H.-T., Sudicky, E. A., Mallia, D. V. and Lin, J. C.: Full Coupling Between the Atmosphere, Surface, and Subsurface for Integrated Hydrologic Simulation, J. Adv. Model. Earth Syst., 10, 43–53, doi:10.1002/2017MS001052, 2018.

25   Dunne, T., Moore, T. R. and Taylor, C. H.: Recognition and prediction of runoff-producing zones in humid regions, Hydrol. Sci. Bull., 20, 305–327, 1975.

Erler, A. R., Frey, S. K., Khader, O., Orgeville, M., Park, Y.-J., Hwang, H.-T., Lapen, D. R., Peltier, W. R. and Sudicky, E. A.: Simulating Climate Change Impacts on Surface Water Resources Within a Lake-Affected Region Using Regional Climate Projections, Water Resour. Res., 55, 130–155, doi:10.1029/2018WR024381, 2019.

30   Frei, S., Lischeid, G. and Fleckenstein, J. H.: Effects of micro-topography on surface-subsurface exchange and runoff generation in a virtual riparian wetland --- A modeling study, Adv. Water Resour., 33(11), 1388–1401, doi:10.1016/j.advwatres.2010.07.006, 2010.

Gburek, W. J. and Sharpley, A. N.: Hydrologic Controls on Phosphorus Loss from Upland Agricultural Watersheds, J. Environ. Qual., 27, 267–277, doi:10.2134/jeq1998.00472425002700020005x, 1998.

35   Glaser, B., Klaus, J., Frei, S., Frentress, J., Pfister, L. and Hopp, L.: On the value of surface saturated area dynamics mapped with thermal infrared imagery for modeling the hillslope-riparian-stream continuum, Water Resour. Res., 52, 8317–8342, doi:10.1002/2015WR018414, 2016.

Glaser, B., Antonelli, M., Chini, M., Pfister, L. and Klaus, J.: Technical note: Mapping surface-saturation dynamics with thermal infrared imagery, Hydrol. Earth Syst. Sci., 22(11), 5987–6003, doi:10.5194/hess-22-5987-2018, 2018.

40   Glaser, B., Jackisch, C., Hopp, L. and Klaus, J.: How meaningful are plot-scale observations and simulations of preferential flow for catchment models ?, Vadose Zo. J., doi:10.2136/vzj2018.08.0146, 2019.

Gourdol, L., Clément, R., Juilleret, J., Pfister, L. and Hissler, C.: Large-scale ERT surveys for investigating shallow regolith properties and architecture, Hydrol. Earth Syst. Sci. Discuss., 1–39, doi:10.5194/hess-2018-519, 2018.

20

Hydrology and
Earth System
Sciences
Discussions
Open Access

EGU

Grabs, T., Seibert, J., Bishop, K. and Laudon, H.: Modeling spatial patterns of saturated areas: A comparison of the topographic wetness index and a dynamic distributed model, J. Hydrol., 373(1-2), 15–23, doi:10.1016/j.jhydrol.2009.03.031, 2009.

Güntner, A., Seibert, J. and Uhlenbrook, S.: Modeling spatial patterns of saturated areas: An evaluation of different terrain
5 indices, Water Resour. Res., 40, W05114, doi:10.1029/2003WR002864, 2004.

Gupta, H. V, Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria : Implications for improving hydrological modelling, J. Hydrol., 377, 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Juilleret, J., Iffly, J. F., Pfister, L. and Hissler, C.: Remarkable Pleistocene periglacial slope deposits in Luxembourg
10 (Oesling): pedological implication and geosite potential, Bull. la Société des Nat. Luxemb., 112(1), 125–130 [online] Available from: http://www.snl.lu/publications/bulletin/SNL_2011_112_125_130.pdf, 2011.

Käser, D., Graf, T., Cochand, F., McLaren, R., Therrien, R. and Brunner, P.: Channel Representation in Physically Based Models Coupling Groundwater and Surface Water : Pitfalls and How to Avoid Them, Groundwater, 52(6), 827–836, doi:10.1111/gwat.12143, 2014.

15 Kristensen, K. J. and Jensen, S. E.: A model for estimating actual evapotranspiration from potential evapotranspiration, Nord. Hydrol., 6(3), 170–188, doi:10.2166/nh.1975.012, 1975.

Latron, J.: Caractérisation géomorphologique et hydrologique du bassin versant versant du Strengbach (Aubure), Université Louis Pasteur, Strasbourg I., 1990.

Latron, J. and Gallart, F.: Seasonal dynamics of runoff-contributing areas in a small mediterranean research catchment
20 (Vallcebre, Eastern Pyrenees), J. Hydrol., 335(1-2), 194–206, doi:10.1016/j.jhydrol.2006.11.012, 2007.

Martínez Fernández, J., Ceballos Barbancho, A., Morán Tejeda, C., Casado Ledesma, S. and Hernández Santana, V.: Procesos hidrológicos en una cuenca forestal del Sistema Central: Cuenca experimental de Rinconada, Cuad. Investig. Geográfica, 31, 7–25 [online] Available from: https://dialnet.unirioja.es/servlet/articulo?codigo=1975892, 2005.

Martínez-Carreras, N., Hissler, C., Gourdol, L., Klaus, J., Juilleret, J., Iffly, J. F. and Pfister, L.: Storage controls on the
25 generation of double peak hydrographs in a forested headwater catchment, J. Hydrol., 543, 255–269, doi:10.1016/j.jhydrol.2016.10.004, 2016.

Megahan, W. F. and King, P. N.: Identification of critical areas on forest lands for control of nonpoint sources of pollution, Environ. Manage., 9(1), 7–17, doi:10.1007/BF01871440, 1985.

Mengistu, S. G. and Spence, C.: Testing the ability of a semidistributed hydrological model to simulate contributing area,
30 Water Resour. Res., 52, 4399–4415, doi:10.1002/2016WR018760, 2016.

Moragues-Quiroga, C., Juilleret, J., Gourdol, L., Pelt, E., Perrone, T., Aubert, a., Morvan, G., Chabaux, F., Legout, a., Stille, P. and Hissler, C.: Genesis and evolution of regoliths: Evidence from trace and major elements and Sr-Nd-Pb-U isotopes, Catena, 149, 185–198, doi:10.1016/j.catena.2016.09.015, 2017.

Munz, M., Oswald, S. E. and Schmidt, C.: Coupled Long-Term Simulation of Reach-Scale Water and Heat Fluxes Across
35 the River-Groundwater Interface for Retrieving Hyporheic Residence Times and Temperature Dynamics, Water Resour. Res., (53), 8900–8924, doi:10.1002/2017WR020667, 2017.

Myrabø, S.: Runoff Studies in a Small Catchment, Nord. Hydrol., 17, 335–346, doi:10.2166/nh.1986.0025, 1986.

Nasta, P., Boaga, J., Deiana, R., Cassiani, G. and Romano, N.: Comparing ERT- and scaling-based approaches to parameterize soil hydraulic properties for spatially distributed model applications, Adv. Water Resour., 126, 155–167,
40 doi:10.1016/j.advwatres.2019.02.014, 2019.

Nippgen, F., McGlynn, B. L. and Emanuel, R. E.: Water Resources Research, Water Resour. Res., 51, 4550–4573, doi:10.1002/2014WR016719.Received, 2015.

Ogden, F. L. and Watts, B. a.: Saturated area formation on nonconvergent hillslope topography with shallow soils: A numerical investigation, Water Resour. Res., 36(7), 1795, doi:10.1029/2000WR900091, 2000.

21

Hydrology and
Earth System
Sciences

Open Access

Discussions

EGU

Partington, D., Brunner, P., Frei, S., Simmons, C. T., Werner, A. D., Therrien, R., Maier, H. R., Dandy, G. C. and Fleckenstein, J. H.: Interpreting streamflow generation mechanisms from integrated surface-subsurface flow models of a riparian wetland and catchment, Water Resour. Res., 49(9), 5501–5519, doi:10.1002/wrcr.20405, 2013.

5 Pfister, L., McDonnell, J. J., Hissler, C. and Hoffmann, L.: Ground-based thermal imagery as a simple, practical tool for mapping saturated area connectivity and dynamics, Hydrol. Process., 24(21), 3123–3132, doi:10.1002/hyp.7840, 2010.

Van der Ploeg, M. J., Appels, W. M., Cirkel, D. G., Oosterwoud, M. R., Witte, J.-P. M. and Van der Zee, S. E. A. T. M.: Microtopography as a Driving Mechanism for Ecohydrological Processes in Shallow Groundwater Systems, Vadose Zo. J., 11(3), doi:10.2136/vzj2011.0098, 2012.

Qu, Y. and Duffy, C. J.: A semidiscrete finite volume formulation for multiprocess watershed simulation, Water Resour.
10 Res., 43(8), 1–18, doi:10.1029/2006WR005752, 2007.

Reaney, S. M., Bracken, L. J. and Kirkby, M. J.: The importance of surface controls on overland flow connectivity in semi-arid environments: Results from a numerical experimental approach, Hydrol. Process., 28(4), 2116–2128, doi:10.1002/hyp.9769, 2014.

Scaini, A., Audebert, M., Hissler, C., Fenicia, F., Gourdol, L., Pfister, L. and Beven, K. J.: Velocity and celerity dynamics at
15 plot scale inferred from artificial tracing experiments and time-lapse ERT, J. Hydrol., 546, 28–43, doi:10.1016/j.jhydrol.2016.12.035, 2017.

Schilling, O. S., Gerber, C., Partington, D. J., Purtschert, R., Brennwald, M. S., Kipfer, R., Hunkeler, D. and Brunner, P.: Advancing Physically-Based Flow Simulations of Alluvial Systems Through Atmospheric Noble Gases and the Novel 37Ar Tracer Method, Water Resour. Res., 53, 10,465–10,490, doi:10.1002/2017WR020754, 2017.

20 Sebben, M. L., Werner, A. D., Liggett, J. E., Partington, D. and Simmons, C. T.: On the testing of fully integrated surface – subsurface hydrological models, Hydrol. Process., 27, 1276–1285, doi:10.1002/hyp.9630, 2013.

Silasari, R., Parajka, J., Ressl, C., Strauss, P. and Blöschl, G.: Potential of time-lapse photography for identifying saturation area dynamics on agricultural hillslopes, Hydrol. Process., 1–18, doi:10.1002/hyp.11272, 2017.

Tanaka, T., Yasuhara, M., Sakai, H. and Marui, A.: The Hachioji experimental basin study -- storm runoff processes and the
25 mechanism of its generation, J. Hydrol., 102, 139–164, 1988.

Tang, Q., Schilling, O. S., Kurtz, W., Brunner, P., Vereecken, H. and Hendricks Franssen, H.-J.: Simulating Flood-Induced Riverbed Transience Using Unmanned Aerial Vehicles , Physically Based Hydrological Modeling, and the Ensemble Kalman Filter, Water Resour. Res., (54), 9342–9363, doi:10.1029/2018WR023067, 2018.

Weill, S., Altissimo, M., Cassiani, G., Deiana, R., Marani, M. and Putti, M.: Saturated area dynamics and streamflow
30 generation from coupled surface-subsurface simulations and field observations, Adv. Water Resour., 59, 196–208, doi:10.1016/j.advwatres.2013.06.007, 2013.

Supplemental material

Table S1: Evapotranspiration parameter used in the model setup of this study, table adapted from Glaser et al. 2016. Root depth and interception storage parameters for the coniferous forest were changed compared to the calibrated parameter values of Glaser et al. 2016. The leaf area index was set to zero for all vegetation types, since this showed to reduce the calculation time during dry conditions substantially without a pronounced effect on the simulated discharge and surface saturation (while the partitioning between transpiration and evaporation changed).

|	Deciduous tree land	Coniferous tree land	Riparian zone + streambed
Root depth (m)	2	1.5	0.5
Root distribution function	Cubic	Cubic	Cubic
Evaporation depth (m)	0.2	0.2	0.2
Evaporation distribution function	Cubic	Cubic	Cubic
Transpiration fitting parameter C1	0.3	0.3	0.3
Transpiration fitting parameter C2	0.2	0.2	0.2
Transpiration fitting parameter C3	0.7	0.7	0.1
Canopy storage parameter (m)	5.00E-04	7.50E-04	1.00E-04
Initial interception storage (m)	5.00E-05	7.50E-05	1.00E-05
Wilting point saturation	0.165	0.165	0.165
Field capacity saturation	0.51	0.51	0.51
Oxic limit saturation	0.7	0.7	0.8
Anoxic limit saturation	0.9	0.9	0.98
Evaporation limiting saturation (min.)	0.1	0.1	0.1
Evaporation limiting saturation (max.)	0.5	0.5	0.5

1
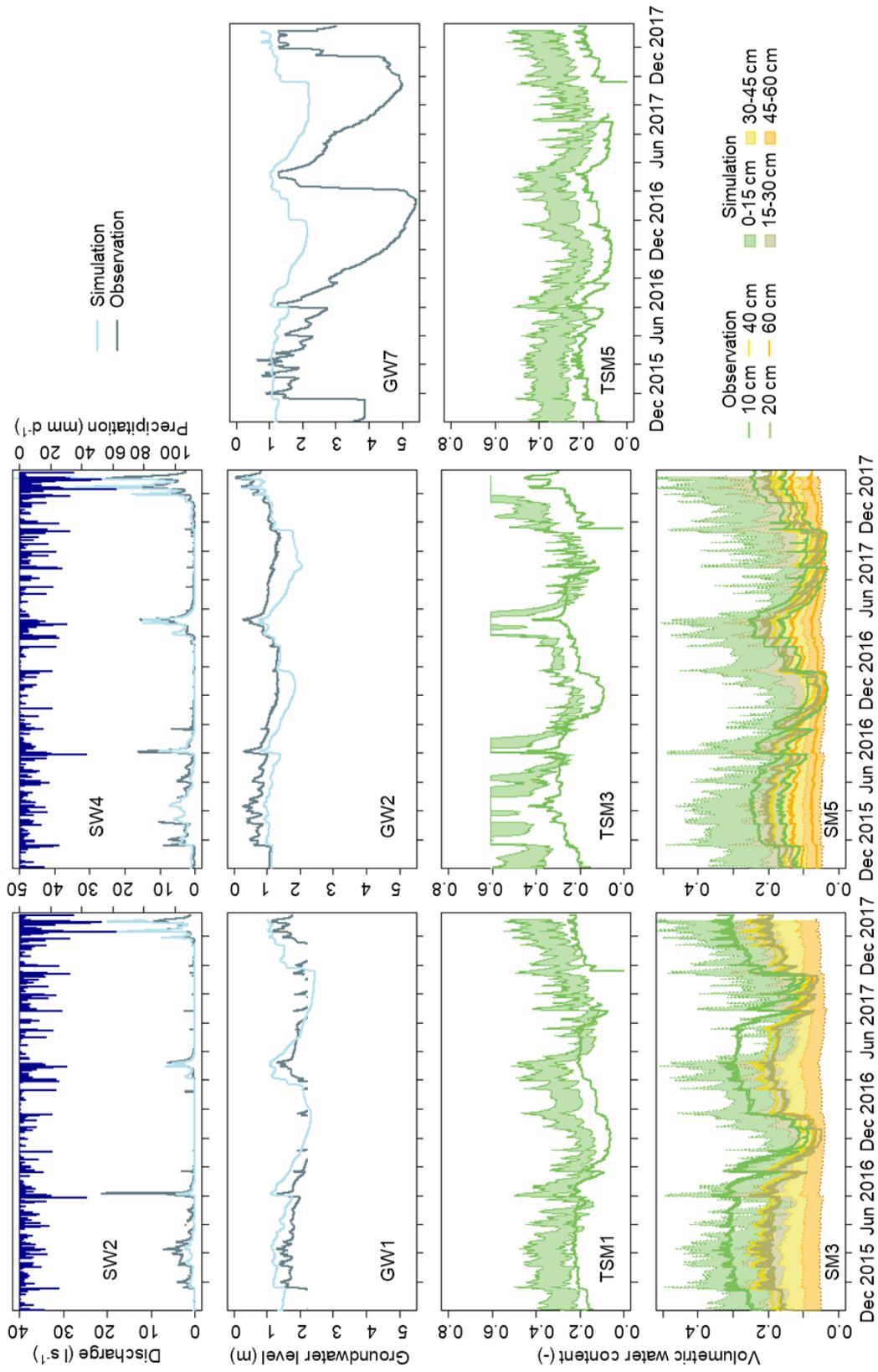
Figure S1: Simulated and observed time series of discharge, groundwater level below the surface, and volumetric water content. Colour bands indicate the possible span of simulated volumetric water contents in the depths between two model nodes. The time series of the observation locations (cf. Figure 1) that are not shown here, are shown in Figure 3.
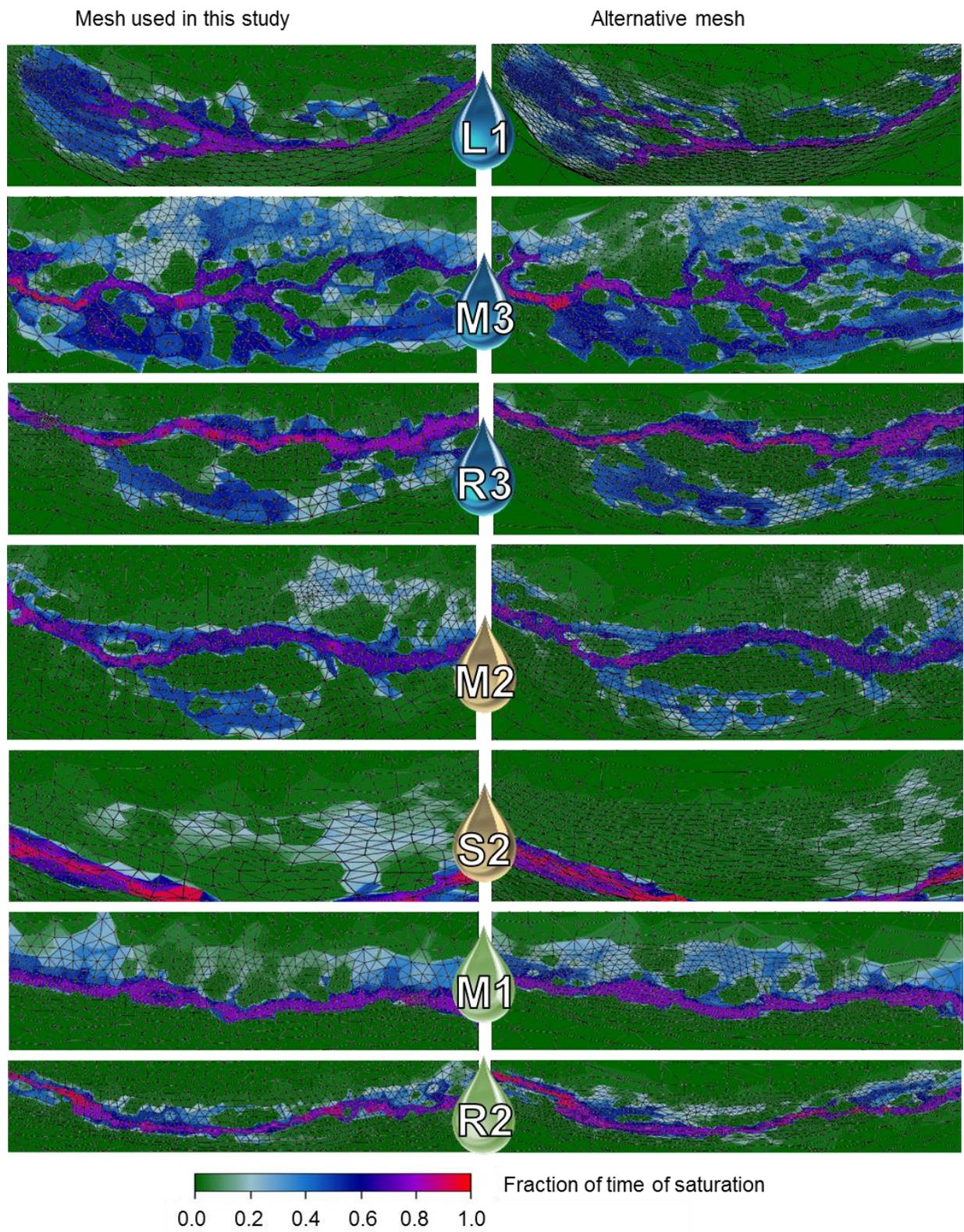
2

Figure S2: Comparison of simulated frequencies of surface saturation in the seven investigated riparian areas for two different model meshes. Left: mesh as used in this study; Right: alternative mesh with finer resolution in the riparian areas.

3

Study 4: Sources of surface water in space and time

Status: In preparation

Authors: Barbara Glaser, Dan Partington, Luisa Hopp, Philip Brunner, René Therrien, Julian
            Klaus

BG, LH, PB, RT and JK designed and directed the study. BG and DP set up the simulation. BG
processed the model output and prepared the figures and tables. BG interpreted the results and
prepared the manuscript with input from all co-authors.

Own contribution in % at the current stage:

- Study concept and design:	90
- Simulation:	90
- Data analysis, figures, and tables:	100
- Interpretation of the results:	90
- Preparation of the manuscript:	90

Sources of surface water in space and time

Barbara Glaser[1,2], Dan Partington[3], Luisa Hopp[2], Philip Brunner[4], René Therrien[5], and Julian Klaus[1]

[1]Catchment and Eco-Hydrology Research Group, Luxembourg Institute of Science and Technology, Esch/Alzette, Luxembourg.

[2]Department of Hydrology, University of Bayreuth, Bayreuth, Germany.

[3]National Centre of Groundwater Research and Training, School of the Environment, Flinders University, Adelaide, Australia.

[4]Centre for Hydrogeology and Geothermics, University of Neuchâtel, Neuchâtel, Switzerland.

[5]Department of Geology and Geological Engineering, Laval University, Quebec, Canada.

Abstract

Knowing the sources of surface water is important for understanding its hydrological, biogeochemical, and ecological impacts in riparian zones or floodplains and on runoff generation. We investigated which ultimate mechanisms delivered water to the surface (immediate delivery path) and from where the surface water originated (geographical sources) for 34 different locations within the riparian zone and streambed of a humid-temperate, forested, headwater catchment. We applied a comprehensively evaluated integrated surface subsurface model (HydroGeoSphere) in combination with a hydraulic mixing cell approach to identify the general variability of the water sources in space as well as their temporal variability for different wetness states and phases of wetting or drying. We found that water was homogeneously and consistently delivered to the surface of the riparian zone by return flow and mainly consisted of a mixture of water from the fractured bedrock, riparian soil and subsolum. Some local variations of the mixing ratio of the water from the different subsurface stores occurred, which might have impacts on water sampling campaigns or ecological and biogeochemical activity. The identified sources of surface water in the streambed indicated that streamflow was generated as well by return flow occurring and accumulating all along the stream and that riparian surface water was basically nothing else than stream water outside of the streambed. Finally, the simulated fractions of precipitation in surface water were generally small, while we observed shifts in the mixing ratio of the surface water with increasing wetness towards higher contributions from more distant subsurface stores.

1

1 Introduction

Floodplains and riparian zones are critical landscape units for a wide range of hydrological, biogeochemical, and ecological processes. A main characteristic of these ecohydrological interfaces (cf. Krause et al., 2017) is irregular flooding in space and time. The spatial and temporal variability of water availability creates specific microhabitats for flora and fauna (e.g. Mallik et al., 2001; Ramey & Richardson, 2017) and hot spots and moments of biogeochemical activity (e.g. Frei et al., 2012; Harms & Grimm, 2008; Singer et al., 2016), which in turn can influence water quantity and quality (e.g. Grabs et al., 2012; Williams & Scott, 2009). Moreover, it is relevant for flood risk assessment to understand where and when water occurs at the surface of floodplains and riparian zones and potentially contributes to streamflow generation via overland flow.

Several processes can induce the occurrence of standing and flowing water at the surface. Surface water can originate from exfiltration of subsurface water or from precipitation ponding due to infiltration or saturation excess and continued delivery of water from the subsurface or precipitation can cause an expansion of surface saturation via overland flow (e.g. Dunne & Black, 1970; Hewlett & Hibbert, 1967; Megahan & King, 1985). In addition, expansion of streamflow can induce the occurrence of surface saturation in riparian zones or floodplains, even though streamflow is basically return flow of subsurface water, infiltration excess, or saturation excess overland flow as well. Depending on the mechanism underlying the generation of surface saturation in riparian zones and floodplains, the origin and characteristics of surface water and thus its impact on ecology, water quality, runoff, and flood risk may differ. Therefore it is necessary to investigate and understand which processes dominate the generation of surface saturation, if and which different types of water contribute in what ratio to surface water, and how the sources of surface water vary in space and time. While it is one of the most prominent questions in catchment hydrology to ask where stream water is coming from (e.g. Hewlett & Hibbert, 1967; McDonnell, 2003), little is known on the sources of surface water beside the streambed apart from apparent inundation.

The investigation of the sources of surface water can focus on three different aspects (cf. Sklash & Farvolden, 1979): i) How old is the water, i.e. what are the temporal sources? ii) Where does the water come from, i.e. what are the geographical sources? iii) Which mechanism ultimately transferred the water to the surface, i.e. what are the immediate delivery paths? Plenty of studies have investigated one or several of these three aspects for sources of the stream hydrograph based on physical measurements of water fluxes, hydrological tracers and mixing analyses (cf. literature overviews given in Barthold & Woods, 2015; Cowie et al., 2017; Klaus & McDonnell, 2013). Several experimental studies demonstrated that the mixing of water from different subsurface stores and landscape units, the ratio of event and pre-event water, and the distribution of water age in the hydrograph change within and between runoff events and for different seasons and wetness conditions (e.g. Birkel et al., 2012; Cartwright & Morgenstern, 2018; Correa et al., 2017; Martínez-Carreras et al., 2015; McGlynn & Seibert, 2003). Others analyzed how the mixture of streamflow sources changes depending on the catchment size or other landscape characteristics (e.g. Correa et al., 2019; Cowie et al., 2017; Gordon et al., 2015; Kirchner, 2009; Laudon et al., 2007; McGlynn et al., 2004; Zhang et al., 2018). Yet we are not aware of any field study except of the study of Brown et al. (1999) that investigated the variability of sources of surface water along a short stream section or for inundated areas in the riparian zone or floodplains.

2

1   Comprehensive sampling of water sources in space and time is limited by its high labor
2   and cost requirements (cf. Correa et al., 2019). Sampling of surface water in the riparian zone or
3   floodplains is certainly further complicated by the intermittent occurrence and commonly low
4   depth of water. A way to avoid these problems is to investigate the spatial and temporal
5   variability of sources of surface water with hydrologic modelling, in particular with integrated
6   surface subsurface hydrologic modelling (ISSHM), where the occurrence of water at the surface
7   and exchange processes between the surface and subsurface are not defined a priori but develop
8   and change depending on the current conditions in the simulation (cf. e.g. Kollet et al., 2017;
9   Maxwell et al., 2014; Paniconi & Putti, 2015; Sebben et al., 2013). Various recent studies have
10 developed and applied particle tracking schemes to simulate transient transit and residence time
11 distributions, i.e. the temporal variability of the temporal sources of surface water (e.g. Engdahl
12 & Maxwell, 2015; Maxwell et al., 2019; Remondi et al., 2018; de Rooij et al., 2013; Yang et al.,
13 2018). Others applied solute transport simulations and particle tracking to identify temporal or
14 geographical sources of streamflow (e.g. Chow et al., 2016; Jones et al., 2006; Liggett &
15 Werner, 2014). Moreover, methods were developed to estimate and track the mixing ratio of
16 initial geographical and temporal water sources on a cell-by-cell basis throughout the simulation
17 (Partington et al., 2011, 2013; Sayama & McDonnell, 2009).

18   Nonetheless, we are only aware of two studies that made use of simulations to investigate
19 in detail the different sources of surface water across space, including surface water generated
20 outside of a streambed. Weill et al. (2013) analyzed the simulated exchange fluxes between the
21 surface and subsurface to decipher the mechanisms generating surface saturation across the
22 hillslopes and riparian zone of a small pre-alpine headwater catchment. Partington et al. (2013)
23 applied a hydraulic mixing cell approach to a riparian wetland to investigate the in-stream and
24 overland flow generation during a storm event. Both studies demonstrated that the dominating
25 processes for the generation of surface saturation varied locally, but they did not evaluate
26 comprehensively if their simulated surface water occurrence across space actually matched
27 reality and they mainly aimed to identify the contribution of overland flow to streamflow
28 generation and to assess the applicability and improvements of the previously developed
29 hydraulic mixing cell approach (Partington et al. 2011), respectively.

30   In this study, we investigate the spatial and temporal variability of sources of surface
31 water within the streambed and riparian zone with the integrated surface subsurface hydrologic
32 model HydroGeoSphere combined with a hydraulic mixing cell approach (Partington et al.,
33 2011, 2013). We perform the study in the 42 ha Weierbach catchment (Luxembourg), where
34 previous work has shown that it is possible to accurately simulate the dynamics, patterns, and
35 frequencies of surface saturation occurrence (Glaser et al., 2019) as observed with repeated
36 thermal infrared imagery mapping (Antonelli et al., 2019a; Glaser et al., 2018). These thermal
37 infrared imagery investigations focused on the spatio-temporal occurrence of surface saturation
38 within several distinct riparian areas along the stream and suggested that there are distinct
39 locations of subsurface water exfiltration that generate and maintain surface saturation in the
40 riparian zone, while additional surface water may originate from the stream extending into the
41 riparian zone or precipitation. In line with that, the previous simulation of the surface saturation
42 in the catchment (Glaser et al., 2019) suggested that the surface water was mainly the result of
43 groundwater exfiltration into microtopographic depressions, yet it remained open how additional
44 water sources influence and reflect the spatio-temporal occurrence of surface water. Here, we
45 complement and enhance the work of the previous studies by a detailed simulation analysis of
46 the sources of surface water in the riparian zone and streambed for three of the previously

3

investigated areas located at a upstream, midstream, and downstream section of the stream. Our research questions are:

1) Which mechanisms ultimately induce and maintain the occurrence of water at the surface, i.e. what are the immediate delivery paths of surface water?

2) Where does the surface water come from, i.e. what are the geographical sources of surface water?

In addition, each of the questions is subdivided into the two questions:

a) Is there a spatial variability for the immediate delivery paths / geographical sources in the riparian zone and along the stream?

b) Are the immediate delivery paths / geographical sources constant in time or do they vary with different initial wetness conditions and for phases of wetting and drying?

2 Study site and previous work

2.1 Weierbach catchment

The Weierbach catchment is a forested catchment (42 ha) in western Luxembourg. The stream network consists of three tributaries that merge into a main stream flowing through a steep, v-shaped stream valley (Fig. 1). The streamsides are bordered by a narrow, flat riparian zone that widens at several sections along the stream and in the source areas of the three tributaries (Fig. 1). The riparian zone comprises 1.2% of the catchment area and is vegetated with ferns, mosses and herbaceous plants. Hillslopes (slopes > 5°, 45 % of catchment area) and the plateau (slopes < 5°, 54 % of catchment area) are covered by deciduous forest (mainly European beech, Sessile oak) with some patches of coniferous trees and a coniferous forest (mainly Norway spruce and Douglas spruce) in the south-east of the catchment.

Catchment geology is dominated by Devonian slates and phyllites. Fractured bedrock appears at a depth of around 1.4 meter and fractures gradually close down to fresh bedrock in an average depth of 5 meter (e.g. Gourdol et al., 2018). A subsolum classified as regolithic saprock (cf. Juilleret et al., 2016) and a shallow Cambisol (average solum depth 0.5 m) overly the fractured bedrock at the plateau and hillslopes of the catchment. In the riparian zone, fractured bedrock is overlain by a shallow organic Leptosol (cf. Glaser et al., 2016). In the streambed, the fractured bedrock is directly exposed at the surface (cf. Fig. 1).

The climate is oceanic-continental with an average annual precipitation around 950 mm and average annual potential evapotranspiration around 590 mm (cf. Carrer et al., 2019; Pfister et al., 2017; data based on years 2006-2014). Precipitation is distributed rather uniformly over the year, while evapotranspiration and runoff (annual average around 480 mm) show clear seasonal variation (cf. Pfister et al., 2017). During dry conditions, streamflow intermittently ceases from the source areas downstream and rainfall-runoff reaction is characterized by sharp, short-lasting discharge peaks. During wet conditions, additional response comes from a broad, long-lasting second discharge peak that starts to appear few hours after the onset of precipitation. This second peak only occurs once a certain storage threshold of subsurface water is exceeded (e.g. Martínez-Carreras et al., 2016; Scaini et al., 2018), but then largely outweighs the volume of the first discharge peaks.

4
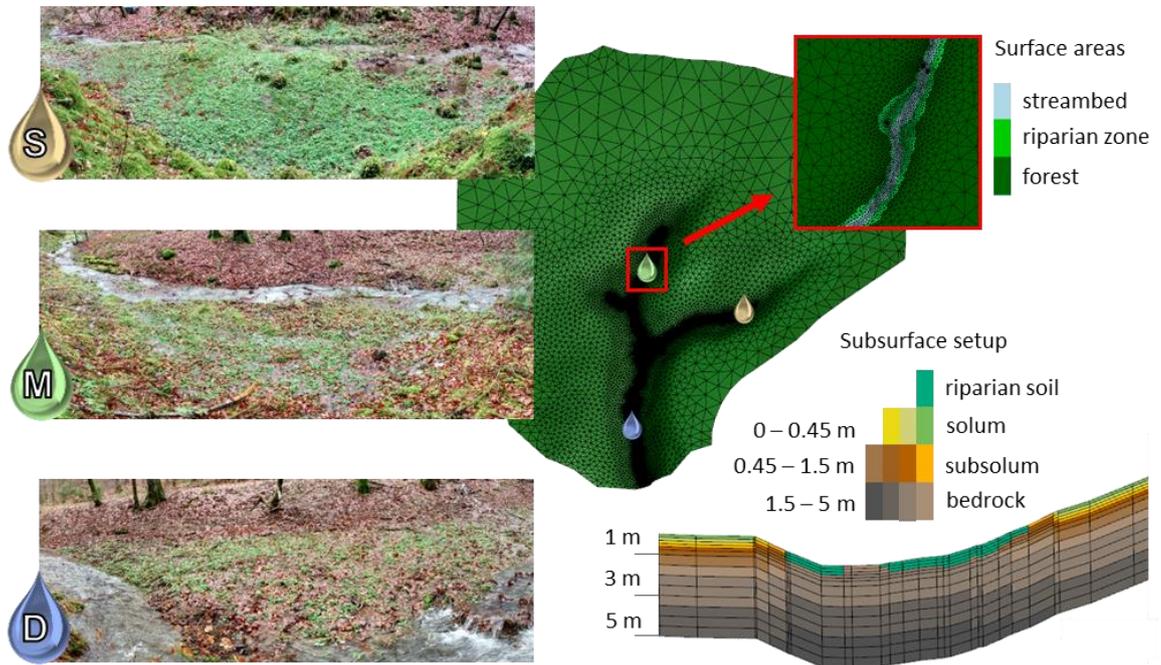
1
2 **Figure 1**. Riparian areas and model setup of the 42 ha Weierbach catchment in western Luxembourg. The three
3 photographs show the investigated riparian areas of the source area of the eastern tributary (S, 155 m^2), the
4 midstream section of the middle tributary (M, 169 m^2), and a downstream section of the main stream (D, 170 m^2).
5 The map of the catchment shows the stream valleys, the location of the three riparian areas, and the nested model
6 mesh with the delineation of the surface domain into forested area, riparian zone, and streambed. The profile at the
7 bottom right shows the vertical discretization of the subsurface model domain and the assigned parameterization
8 zones in and aside the riparian zone, corresponding to the fresh and fractured bedrock (slate and phyllites), the
9 subsolum (regolithic saprock), the solum of the hillslopes and plateau (Cambisol), and the riparian soil (Leptosol) of
10 the catchment. Figure freely based on Glaser et al. (2019).

11   **2.2 Weierbach model**

12   In this study we rely on the integrated surface subsurface hydrologic model
13 HydroGeoSphere (HGS, Aquanty Inc.) as previously implemented for the Weierbach catchment
14 (Glaser et al., 2019). The model consists of a 5 meter deep subsurface domain where transient
15 subsurface flow is simulated based on the 3D Richards equation and a surface domain in which
16 surface flow is simulated based on the diffusive-wave approximation of the 2D Saint-Venant
17 equation. Surface and subsurface flow are simulated simultaneously and exchange between the
18 two domains is simulated as Darcy flow through a very thin coupling layer. Actual
19 evapotranspiration is simulated based on potential evapotranspiration, actual water availability,
20 and some plant and soil characteristics (e.g. rooting depth, evaporation depth) following the
21 conceptual approach of Kristensen and Jensen (1975).

22   The surface domain is spatially discretized into 42,274 triangular elements with edge
23 lengths ranging from > 30 m at the plateau to < 0.4 m in the riparian zone and streambed (Fig.
24 1). The nested mesh was generated with the mesh generator AlgoMesh (HydroAlgorithmisc Pry
25 Ltd) and topographic information was assigned from a 0.1 m elevation raster that interpolated
26 and combined topographic information from 10 m contour lines of a topographic map for the
27 plateau and hillslopes and a high-resolution ground-based LiDAR DEM for the riparian zone and

5

streambed. The subsurface domain is discretized with fourteen layers of triangular prisms with element depths ranging from 0.15 m for the top layers to 0.5 m for the bottom layers (cf. Fig. 1). Sides and bottom of the subsurface domain are no flow boundaries. At the sides of the surface domain, a critical depth boundary allows water to leave the model domain.

The parameterization of the model subsurface distinguishes twelve different property layers representing the specific pedolithology of the catchment with solum, subsolum, fractured bedrock, and a different soil type in the riparian zone (cf. 2.1, Fig. 1). The subsurface is set up homogeneously across the hillslopes and plateau, whereas the upper property layers are removed in the riparian zone and streambed and replaced by riparian soil and lower property layers in dependency of the topography (cf. Fig. 1 and Glaser et al. (2016) for details on the implementation). Evapotranspiration parameters and Manning's surface roughness differ between the deciduous and coniferous forest, the riparian zone, and the streambed. Exact parameter values are given in Glaser et al. (2019). Here, it should be noted that the model parameters were not calibrated at catchment scale, but were transferred from a HGS model implementation of a 6 ha headwater area of the middle tributary (Glaser et al., 2016). The parameterization of this 6 ha headwater catchment was largely based on field experience and measurements (e.g. ERT profiles, soil profiles, hydraulic conductivity measurements) and literature values. Only some evapotranspiration parameters, porosity values and hydraulic conductivity were adapted in a manual calibration procedure at headwater scale (cf. Glaser et al., 2016).

Glaser et al. (2019) simulated the hydraulic states and fluxes in the Weierbach catchment from October 2015 to January 2018, driven with daily input data of precipitation recorded within the catchment and potential evapotranspiration (FAO reference evapotranspiration) calculated from meteorological data from nearby weather stations. A multi-data evaluation showed that simulated discharge, groundwater level, and soil moisture matched observation data well. Kling Gupta efficiency was 0.74 for the simulated discharge at the catchment outlet and ranged from 0.47 to 0.49 for the simulated discharge of the three tributaries. The seasonal dynamic of groundwater levels was captured at all five monitoring locations distributed across the catchment (Pearson correlation coefficients r ranging between r = 0.62 and r = 0.84) and groundwater level was captured particularly well for locations in and close to the riparian zone. Simulated water content showed some deficiencies during dry conditions, but captured the overall seasonal soil moisture dynamics at different locations and in different depths (r = 0.73 +- 0.15).

2.3 Previous investigations of surface saturation in the riparian zone

The riparian zone of the Weierbach catchment has been suggested to play an important role in runoff generation in a range of experimental and modelling studies (Fenicia et al., 2014; Glaser et al., 2016; Klaus et al., 2015; Martínez-Carreras et al., 2015; Martínez-Carreras et al., 2016; Schwab et al., 2018; Wrede et al., 2015), especially with regards to the occurrence of surface saturation and overland flow. A recent detailed characterization of the seasonal dynamics of surface saturation in seven distinct riparian areas of the catchment has revealed that the temporal development of surface saturation is largely synchronous across space, but that there is some spatial variability in the relationship between the local extent of surface saturation and catchment discharge (Antonelli et al., 2019a) and in the correlation between the local saturation extent and the contribution of the correspondent stream reach to catchment discharge (Antonelli et al., 2019b). The mapping of the surface saturation extent and dynamics for these studies has

6

1   been done with thermal infrared imagery (cf. Glaser et al., 2018; Pfister et al., 2010), which
2   implies that surface saturation was defined as all water standing or flowing at the surface,
3   including stream water but excluding 'mere' saturation in the top soil layer. Glaser et al. (2019)
4   used the same comprehensive data set of thermal infrared images taken weekly to biweekly in
5   the different riparian areas from October 2015 to January 2018 in comparison to the simulations
6   with the integrated surface subsurface hydrologic HGS model of the Weierbach catchment (cf.
7   Section 2.2.). They showed that the model could satisfactorily reproduce the observed dynamic
8   patterns of surface saturation for the different investigated riparian areas, including the spatial
9   patterns of surface saturation frequency (Fig. 2).
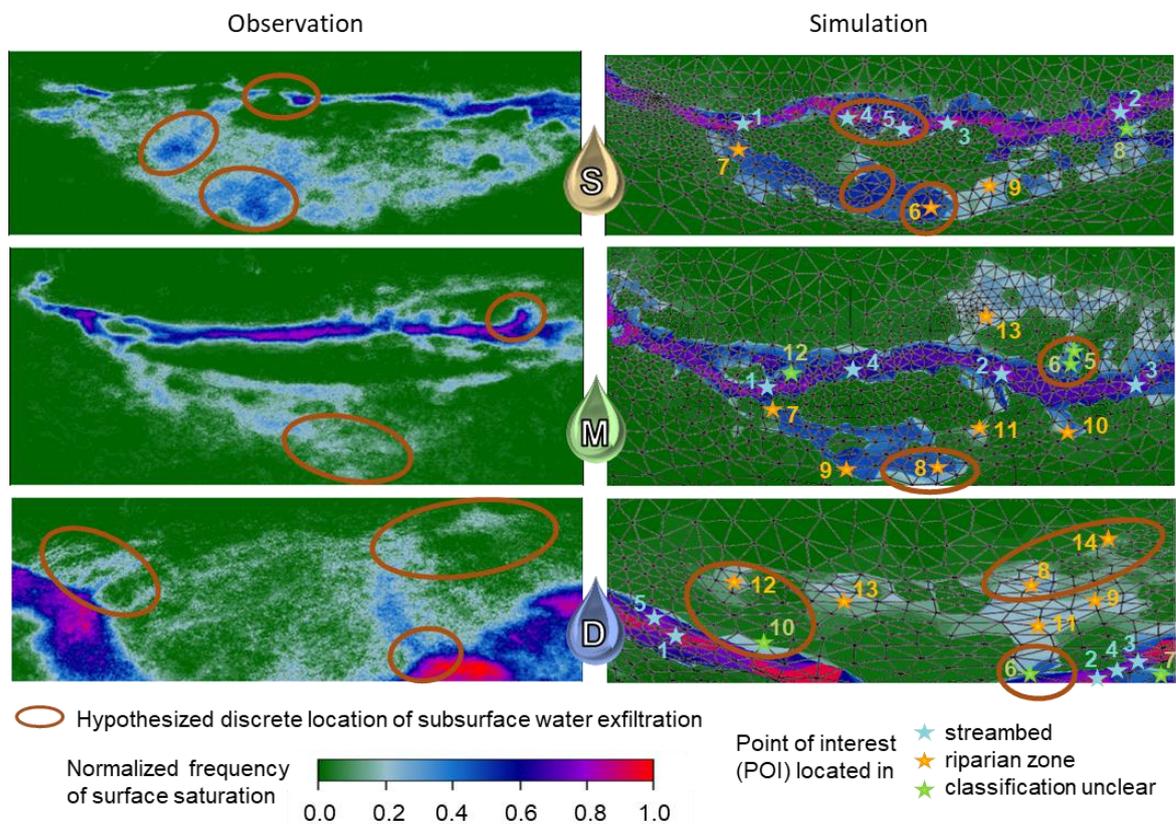
10
11 **Figure 2**. Patterns of observed (left) and simulated (right) frequency of surface saturation in the source area of the
12 eastern tributary (S), the midstream area at the middle tributary (M), and the downstream area at the main stream
13 (D). The observed frequencies indicate how often surface water was standing or flowing within the riparian zone
14 and streamed over the entity of 43 (area S), 48 (area M), and 37 (area D) thermal infrared image snapshots taken
15 and analyzed within the period from October 2015 to January 2018 (cf. Fig. 4, Glaser et al. (2019)). The simulated
16 frequencies indicate how often water was simulated in the surface domain (water depth $> 10^{-4}$m) during the same
17 moments in time (cf. Glaser et al., 2019). Locations that were hypothesized to be locations of discrete subsurface
18 water exfiltration based on the recurrent field observations of water temperature, saturation frequency and saturation
19 stability are labelled with brown circles. Points of interest (POIs) that were selected in the simulation for a detailed
20 analysis of the mixing of sources of surface water with the hydraulic mixing cells (cf. section 3) are colored
21 depending on their location in the riparian zone (green stars) or streamed (blue stars) and numbered according to
22 decreasing saturation frequency within the different areas. Figure adapted from Glaser et al. (2019).

7

Based on the way of functioning and the good performance of the model, Glaser et al (2019) concluded that the generation of surface saturation in the Weierbach catchment is largely driven by groundwater exfiltration into topographic depressions, which matches evidences from the thermal infrared field observations (Antonelli et al., 2019a; Glaser et al., 2018). However, both the field observations and the simulations suggested that exfiltration from the subsurface is not the exclusive source for water at the surface. The field observations suggested that the delivery of subsurface water to the surface was maintained at discrete exfiltration locations (Fig. 2). These discrete locations of subsurface water exfiltration were hypothesized based on the temperatures monitored with the thermal infrared images, the identified surface saturation frequencies, and the observed stability of the surface saturation over time. Surface saturation at other locations was not related exclusively to exfiltration of subsurface water, but to additional immediate delivery paths of surface water such as overland flow from the discrete exfiltration locations through the riparian area, streamflow extending into the riparian zone, or precipitation. Furthermore, Glaser et al. (2019) suggested that the observed and similarly simulated variability of surface saturation frequency within the investigated areas reflects a mixture of different water sources and immediate delivery paths of surface saturation. As example, they pointed out that the generally higher saturation frequencies in the streambed than in the riparian zone (cf. Fig. 2) might result from maintenance of surface water delivery in the streambed by upstream water contribution. Here, we applied the hydraulic mixing cell approach with the Weierbach catchment model as described in the following section to eventually clarify, specify, and quantify the previously assumed mixing of different geographical sources and immediate delivery paths of surface water.

3 Application of hydraulic mixing cells to identify surface water sources

3.1 The hydraulic mixing cell approach

The hydraulic mixing cell (HMC) approach is a modified mixing cell approach developed for integrated surface subsurface hydrologic models by Partington et al. (2011, 2013). The HMC approach enables to track the mixing of predefined initial water sources at any location and at any time based on information from the hydraulic flow solution. Each model cell of the surface and subsurface domain is assigned to a source area and water that was initially stored within the cell is tracked as water from the respective source area throughout the simulation. Water newly entering the model system during the simulation via precipitation is assigned as precipitation throughout the simulation. Thus, the water keeps the initial source assignment when passing through cells of different source areas. However, water originating from different source areas mixes within a model cell and the mixing ratio is tracked and adapted throughout the simulation.

The fractions of different water sources in a model cell are calculated based on the simulated hydraulic fluxes into and out of the model cell, the fluid volume in the cell, and the assumption that water mixes within the cell following the 'modified mixing rule', i.e. mixing follows a regime that ranges between perfect mixing and piston flow (cf. Campana & Simpson, 1984). In order to avoid numerical instability and numerical dispersion while ensuring computational efficiency, the possibility to calculate mixing ratios for sub-time-steps of the flow solution and several stability criteria were introduced (cf. Partington et al., 2013). These stability criteria are checked at each time step and for each model cell individually and in case one

8

criterion is not met, the affected cell is excluded from the mixing calculation of the current time step. Instead, the fractions of water sources in the cell are reset and the cell is assigned a so-called reset fraction of 1. This reset fraction is tracked in the further course of the simulation as well and indicates an unknown origin of water.

3.2 HMC simulation in the Weierbach catchment

We analyzed the sources of surface water with the HMC approach for 34 distinct points of interest (POIs) in the riparian zone and streambed of the Weierbach catchment. The POIs were distributed in three riparian areas, comprising a source area (S), midstream (M), and downstream (D) section of the stream with extents of 155 m^2 (S), 169 m^2 (M), and 170 m^2 (D) (cf. Fig. 2). As POIs, we selected mesh nodes of the surface domain of the HGS model (cf. section 2.2) that covered a large range of different saturation frequencies (cf. order number of the POIs, Fig. 2) as well as nodes that were located in the streambed (blue labelled POIs, Fig. 2), in the riparian zone (yellow labelled POIs, Fig. 2), at locations that could not unambiguously be assigned as streambed or riparian zone (green labelled POIs, Fig. 2), and at locations with assumed distinct exfiltration of subsurface water (POIs within brown circles, Fig. 2).

For the identification of different sources of surface water with the HMC approach, we distinguished between incoming precipitation and seven different initial source areas. The initial source areas corresponded to the zones of the model parameterization (cf. section 2.2) and separated the surface domain into forest, riparian, and streambed areas and the subsurface domain into fractured bedrock, subsolum, solum of the hillslopes and plateau, and riparian soil (cf. Fig. 1, Fig. 3).

We ran the HMCs enabled HGS model from October 2015 to January 2018 with hourly meteorological forcing data. In order to avoid that the signal of initially stored water was fully replaced by new incoming precipitation and that mass balance errors and reset fractions could accumulate, we split the HMC simulation of the 28 months into 64 consecutive but individual simulation periods and re-initialized the fractions of water sources at the start of each of these individual sub-periods of simulation (cf. Fig. 3, grey lines). We chose the sub-periods according to the model output analysis of Glaser et al. (2019), who analyzed the simulated surface saturation patterns for the times where TIR images were available (cf. Fig. 2, Fig. 4). This allowed us to rely on previous simulation output to initialize the hydraulic conditions of the consecutive sub-periods and thus to considerably reduce the wall-clock time of the simulation by running the individual simulation periods in parallel.

3.3 Processing of the HMC output

We extracted the time series of simulated fractions of water sources for each of the 39 selected POIs (cf. Fig. 2). For time periods where no discharge was simulated at a POI, the time series of water fractions naturally feature gaps (cf. Fig. 3). In case the mass balance error of the simulated water fractions was > 5 % (i.e. the sum of the individual fractions was < 0.95 or > 1.05) or a water fraction was simulated with a negative contribution > 1 % of simulated discharge, we excluded the affected time step from any further investigations by artificially introducing a gap into the time series of the affected POI.

In order to identify the immediate delivery paths of surface water, i.e. the mechanisms that ultimately induce and maintain the occurrence of water at the surface, we extracted the

9

1 fractions of water sources after the first time step of the simulation from each of the 64 sub-
2 period simulations (cf. Fig. 3). Since water of an initial source area is not mixed with water from
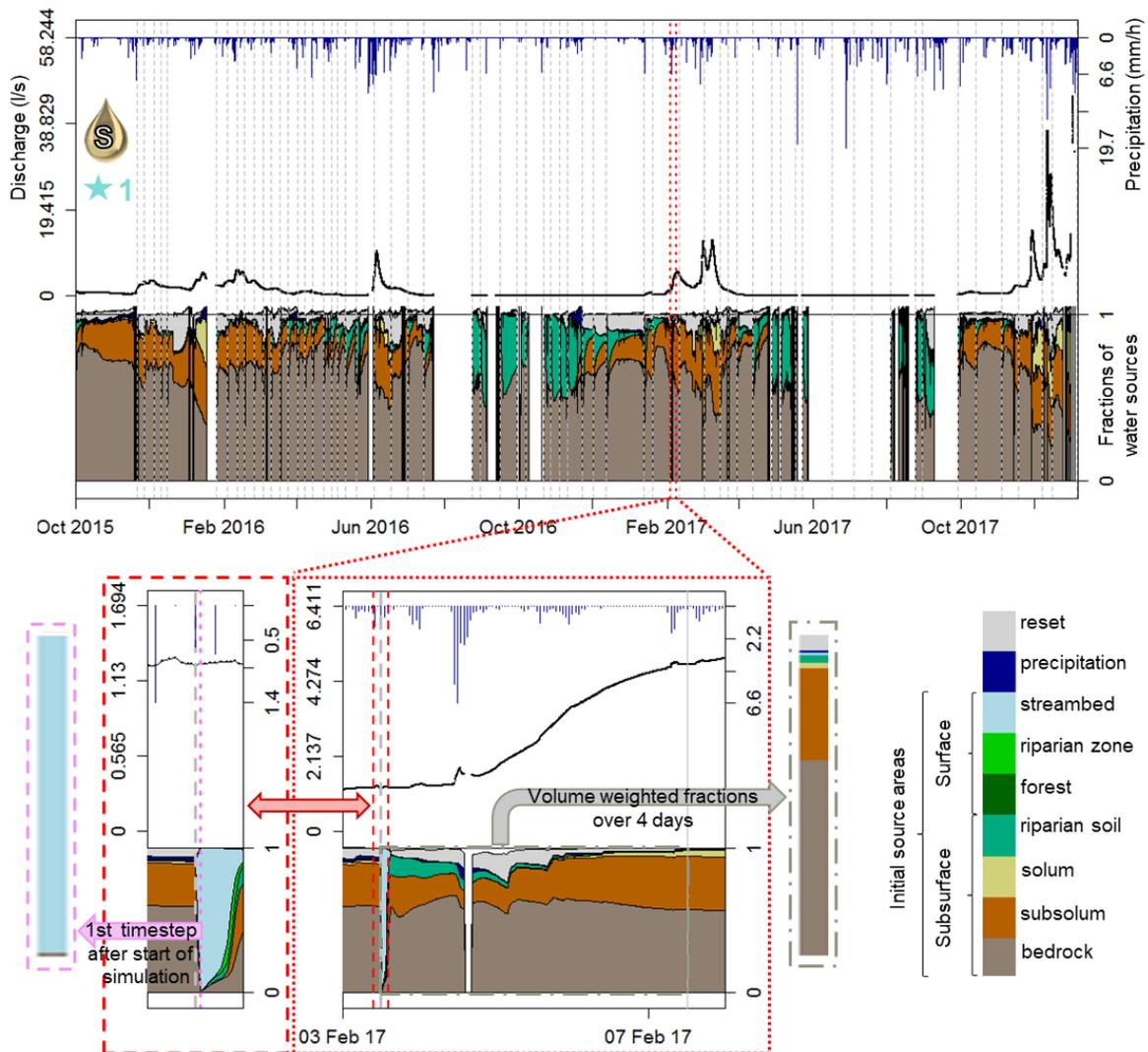
3
4 **Figure 3**. Output and processing of the HMC simulation, shown for POI 1 located in the streambed of the source
5 area S of the eastern tributary (cf. Fig. 1 and 2). Fractions of water initially originating from the different source
6 areas defined in the surface and subsurface model domain (cf. Fig. 1), fractions of precipitation, and reset fractions
7 are presented as stacked time series. Above that, time series of hourly precipitation and discharge simulated at the
8 POI are shown above. Gaps in the time series of fractions of water sources indicate either that the POI was inactive
9 (simulated discharge is zero) or that the time points were excluded from the analysis due to high mass balance errors
10 (no simulated discharge shown). Grey dotted lines indicate the start and end times of the 64 sub-periods of
11 simulation where fractions of water sources were reset to their initial source areas. For the identification of the
12 immediate delivery paths, the fractions of water sources were extracted after the first time step of the sub-period
13 simulations (bottom left). For the identification of geographical water sources, the volume of the different water
14 sources was summed up individually for the first four days of a sub-period of simulation and normalized with the
15 total water volume that was discharged at the POI during the four days to volume weighted fractions of water
16 sources (bottom right).

10

other initial source areas within the first time step, the fractions of water sources after the first time step depict from where water entered into a model cell. Consequently, it can be inferred after the first time step, if the immediate delivery of water to a POI happened via exfiltration from the subsurface, overland flow, or direct precipitation. Based on the assigned initial source areas (cf. above, Fig. 3), overland flow could be further distinguished between streamflow and overland flow from the forest and riparian zone and exfiltration from the subsurface could be further distinguished into exfiltration from the fractured bedrock, subsolum, solum, and riparian soil.

In order to identify the geographical sources of surface water, i.e. the source areas where water was initially stored before reaching a POI, we calculated the volume weighted fractions of water sources over the first four days of each of the 64 sub-period simulations (cf. Fig. 3). We chose to analyze the first four days of a simulation period as compromise between the facts that water initially stored in adjacent source areas of a POI was displaced by precipitation or water from more distant sources areas, but that water originating from distant source areas needed some time to reach a POI after the start of the simulation. The assigned initial source areas allowed us to interpret the geographical sources of surface water regarding different pedolithological water stores as well as regarding different landscape units. We distinguished contributions of precipitation, upstream water, riparian soil and riparian surface water, water from soils and surface from the hillslopes and plateau, and water from the subsolum and fractured bedrock. Water from the subsolum and fractured bedrock could indicate contributions from deep subsurface stores from the hillslopes and plateau as well as contributions from subsurface stores directly underlying the streambed and riparian soil (fractured bedrock) and the edges of the riparian zone (subsolum), respectively (cf. Fig. 1).

The analysis of the spatial variability of the immediate delivery paths and geographical sources of surface water considered the variability along the stream by separating between the source area S, midstream M, and downstream D area (cf. Fig. 1). Furthermore, the variability within one area was considered by separating between POIs in the streambed, the riparian zone, and at their interface (cf. blue, yellow and green stars in Fig. 2). In total, we identified the immediate delivery paths and geographical sources at the 39 POIs for 64 different time periods. In order to obtain a picture of the average spatial variability of the mixing of water sources, we determined for each of the 39 POIs the arithmetic mean of the 64 identified mixing ratios of the immediate delivery paths and geographical sources.

The temporal variability of the mixing of the immediate delivery paths and geographical sources was analyzed for eight selected periods in time, representing different initial wetness states and phases of wetting and drying. To distinguish different initial wetness states, we grouped the 64 sub-periods of simulation into the four categories i) dry, ii) intermediate, iii) wet, and iv) very wet initial conditions. Since discharge dynamics showed to be similar at all POIs and the catchment outlet (data not shown), the classification was based on discharge Q at the catchment outlet at the start of the simulations: i) Q < 0.5 l/s, ii) 0.5 l/s < Q < 5 l/s, iii) Q > 5 l/s but not all POIs active, iv) Q > 5 l/s and all POIs active. Within each of the four categories, we visually selected one period in time where the simulated catchment discharge was rising within the first four simulation days that were analyzed to identify the geographical sources of surface water and one period in time where the simulated discharge was falling within the first four simulation days (Fig. 4, Table 1). The discharge at the catchment outlet showed to be well correlated to the extent of surface saturation (cf. Fig. 4, Table 1, and Glaser et al. (2019)), thus

11

1 we assumed that the selected time periods represented phases of wetting with increasing
2 discharge and expanding surface saturation across the entire catchment and phases of drying with
3 decreasing discharge and contracting surface saturation across the entire catchment.
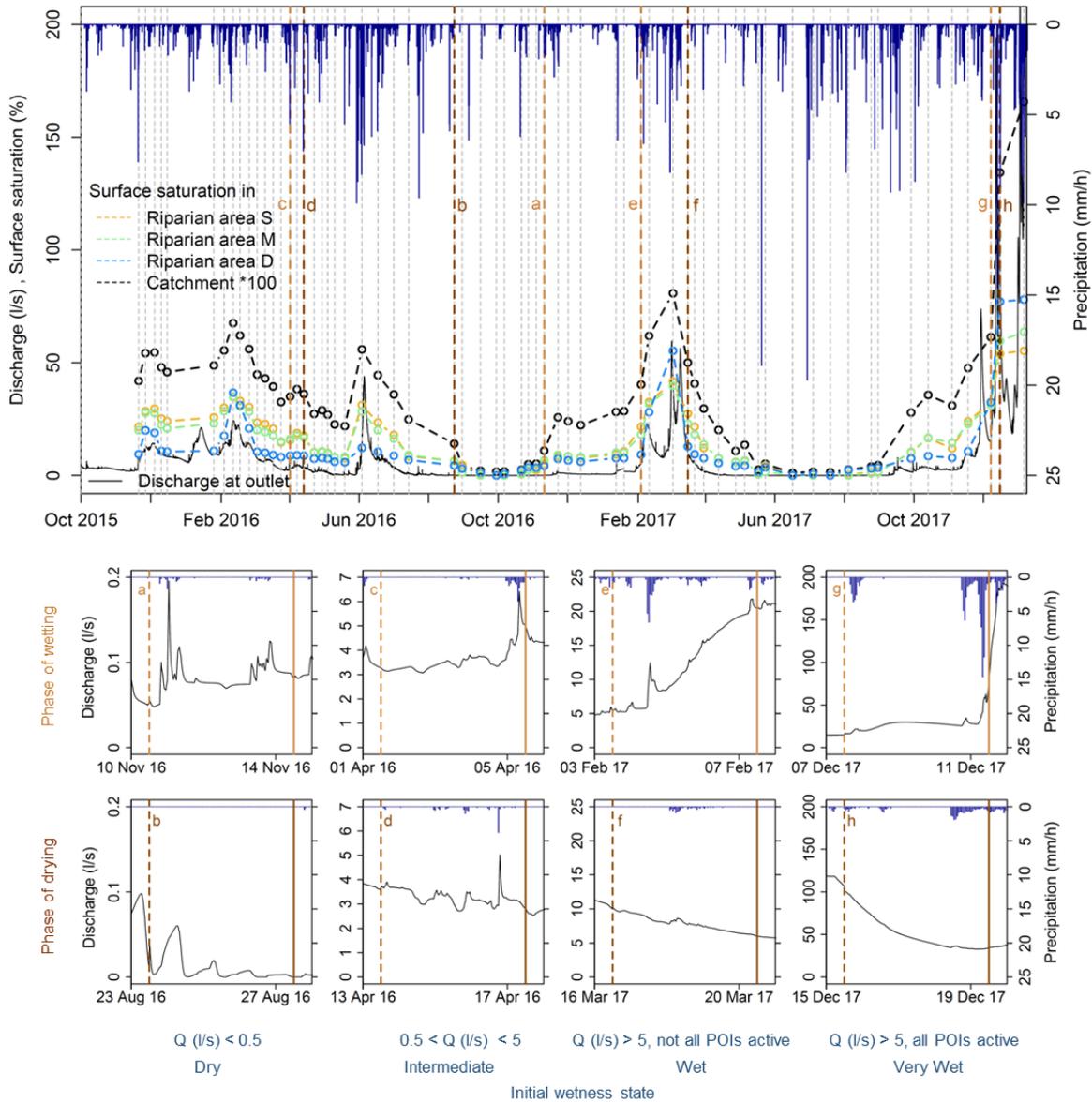
4
5 **Figure 4**. Simulated discharge at the catchment outlet and areal percentage of surface saturation simulated in the
6 entire catchment and in the three investigated areas comprising a source area (S), midstream (M) and downstream
7 (D) section of the stream (cf. Fig. 1 and 2). The areal percentage of surface saturation (water depth > 10^{-4} m) was
8 estimated by Glaser et al. (2019) for the times of thermal infrared imagery mapping (cf. Fig. 2) and corresponds to
9 the extent of surface saturation at the beginning of the 64 sub-periods of HMC simulation (grey dotted lines, cf. Fig.
10 3). Brown dashed and solid lines a-h mark the beginning and end of the eight time periods selected to represent
11 phases of wetting (light brown) and drying (dark brown) starting from different initial wetness states as classified by
12 the simulated discharge Q at the catchment outlet (cf. Table 1).

12

Table 1. Quantitative hydrologic characteristics for the eight selected time periods (cf. Fig. 4) representing dry (a,b), intermediate (c,d), wet (e,f), and very wet (g,h) initial wetness states and phases of wetting (a,c,e,g) and drying (b,d,f,h). Q_t is the simulated discharge (l/s) at the catchment outlet at the start and end of the selected periods of four days, $P_{t=0}$ is the precipitation rate (mm/h) during the first time step of the selected simulation periods, and P_{cum} is the cumulative precipitation (mm) over the four day periods. Areal percentage of surface saturation was estimated for the entire catchment and the three distinct investigated areas (cf. Fig. 1 and 2) at the start of the selected periods in time only (cf. Glaser et al., 2019).

Time period	Start date	$Q_{t=0}$	$Q_{t=4days}$	$P_{t=0}$	P_{cum}	Surface saturation (%) at start date in			
						Catchment	Area S	Area M	Area D
a	10/11/2016	0.05	0.08	0.0	7.7	0.11	5.1	6.75	4.15
b	23/08/2016	0.01	0	0.0	0.0	0.14	6.43	6.33	4.42
c	01/04/2016	3.27	5.05	0.0	14.1	0.35	15.64	16.21	8.76
d	13/04/2016	3.55	2.84	0.0	10.5	0.36	17.59	16.76	8.77
e	03/02/2017	5.27	20.44	0.6	40.2	0.4	21.59	18.1	9.22
f	16/03/2017	9.99	5.99	0.0	6.8	0.5	27.44	22.98	12.86
g	07/12/2017	14.94	78.64	0.0	82.9	0.61	32.49	31.14	32.41
h	15/12/2017	106.72	34.18	0.3	24.6	1.34	53.95	59.68	77.09

4 Results
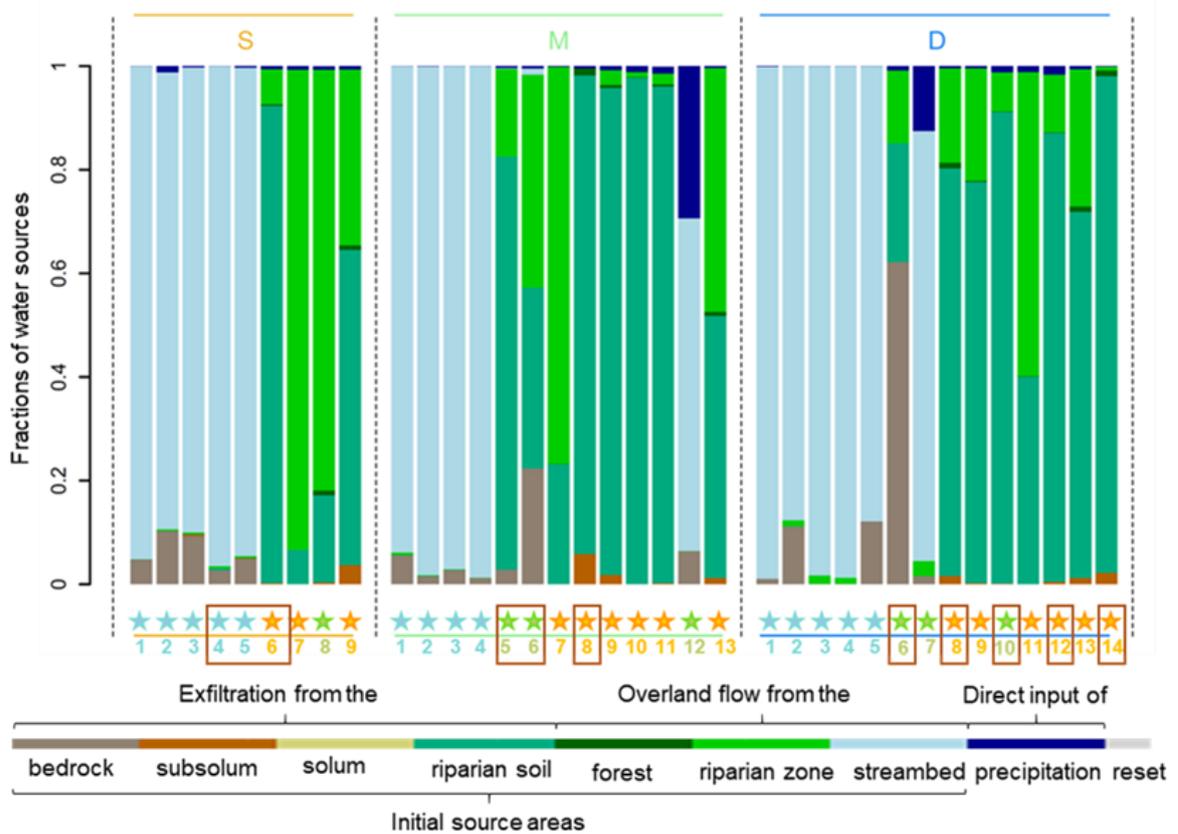
4.1 Immediate delivery paths of surface water

4.1.1 Spatial variability

The immediate delivery path of surface water clearly differed for the streambed and the riparian zone, regardless where the studied area was located along the stream (Fig. 5). Surface water in the streambed was essentially delivered by overland flow, i.e. streamflow (blue POIs, Fig. 5), while surface water in the riparian zone was largely fed by water exfiltration from the riparian soil (yellow POIs, Fig. 5). In addition, fractions of riparian overland flow were found at most locations in the riparian zone, but the relevance of the fractions varied locally. In the riparian zone of the upstream area S, one location (POI 6) received surface water nearly exclusively from the riparian soil, while the other POIs showed a high (POI 9) or even dominant (POI 7) fraction of overland flow. This is consistent with the drawn assumptions from field observations that distinct locations of groundwater exfiltration exist (cf. section 2.3, Fig. 2). However, in area M and D and in the streambed of area S, POIs positioned at locations that we assumed to be predominantly fed by groundwater exfiltration (brown labelled POIs in Fig. 5) did not show distinctly different immediate delivery paths of surface water than other locations in the riparian zone and streambed, respectively.

A considerable direct contribution of precipitation was only found for two locations that could not be clearly assigned to the streambed or the riparian zone based on their location and saturation frequency (Fig. 5, green POIs): POI 12 in area M and POI 7 in area D. The position of these POIs in the streambed suggests that they are 'island POIs' that exclusively receive water from direct precipitation when stream water level is low. Once the stream level rises, they are flooded and the contribution of precipitation becomes minor compared to the streamflow. The other POIs that could not be clearly assigned to the streambed or riparian zone showed similar immediate delivery paths as found for the locations in the riparian zone, with the exception of

13

POI 6 in area D, where the dominant immediate delivery path was exfiltration from the fractured
bedrock.

Figure 5. Spatial variability of immediate delivery paths of surface water. The bars represent the mixing ratios of
sources for surface water at the 34 POIs located in the streambed (blue stars), in the riparian zone (yellow stars), and
at unclassifiable positions (green stars) in the three areas S, M and D along the stream (cf. Fig. 2). Brown labelled
POIs indicate locations with assumed distinct exfiltration of subsurface water (cf. Fig. 2). The fractions of water
sources are the arithmetic mean of the fractions of water sources extracted after the first time step of each of the 64
sub-period simulations (cf. Fig. 3). The water sources that contributed to a POI during the first time step of the
simulations represent the immediate delivery paths of surface water and can be distinguished between precipitation,
exfiltration from different subsurface source areas, and streamflow or overland flow from the riparian zone or forest.

4.1.2 Temporal variability

The mixing ratio of the immediate delivery paths was at most locations similar for the
eight selected times representing different initial wetness states and phases of drying and wetting
(Fig. 6). The largest change in the mixing ratio of the immediate delivery paths at streambed
locations occurred during periods with dry initial conditions (Fig. 6a and b, blue POIs). The
fraction of groundwater exfiltration (i.e. water from the bedrock) was considerably higher during
these time periods than during periods with wetter initial conditions (Fig. 6c - h), especially
during the wetting phase (Fig. 6a). Nonetheless, streamflow was still the predominant immediate
delivery path of surface water at all locations in the streambed except of one (POI 5 in area D).
The mixing ratio of the immediate delivery paths at locations in the riparian zone and at most
unclassified locations stayed stable once the delivery of water to the surface was activated (Fig.

14
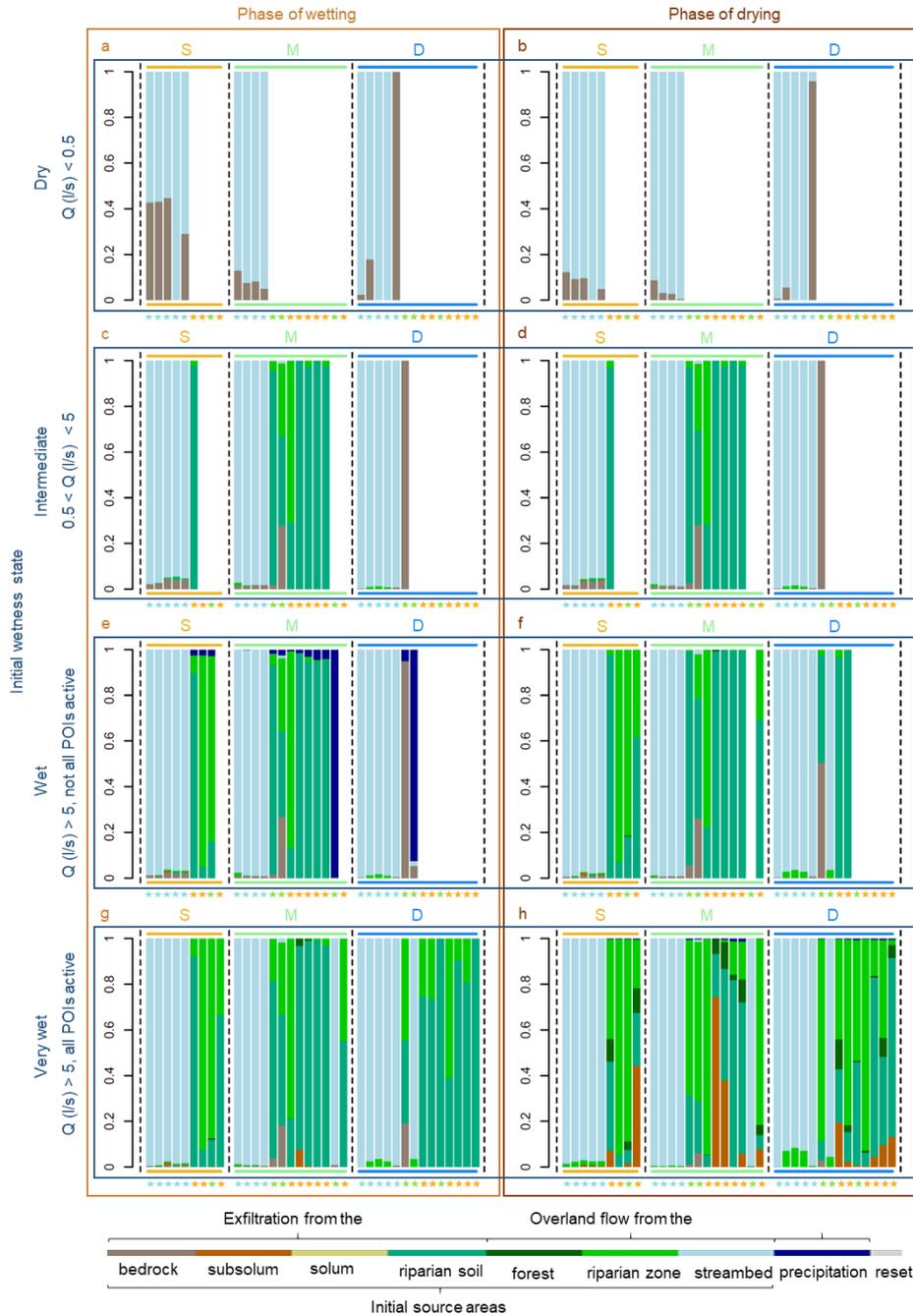
1
2 **Figure 6.** Temporal variability of immediate delivery paths of surface water depending on the initial wetness state
3 and phases of wetting or drying. The bars represent the mixing ratios of sources for surface water at the 34 POIs
4 located in the streambed (blue stars), in the riparian zone (yellow stars), and at unclassifiable positions (green stars)
5 in the three areas S, M and D along the stream (cf. Fig. 2). The fractions of water sources are the fractions of water
6 sources extracted after the first time step (cf. Fig. 3) for each of the eight selected time periods (a-h, Fig. 4, Table 1).
7 The water sources that contributed to a POI during the first time step of the simulations represent the immediate
8 delivery paths of surface water and can be distinguished between precipitation, exfiltration from different subsurface
9 source areas, and streamflow or overland flow from the riparian zone or forest.

6, yellow and green POIs). Only during drying after very wet initial conditions (Fig. 6h), the fraction of overland flow increased relative to the exfiltration of subsurface water. Moreover, overland flow and exfiltrating subsurface water partly originated from areas that are located further away from the stream (i.e. subsolum and forest) during that drying period.

A clear change of the predominating immediate delivery path with changing wetness conditions only occurred at the three POIs that have been pointed out before to be characterized by different immediate delivery paths than in the streambed or riparian zone (section 4.1.1). At POI 6 in area D, the immediate delivery of surface water clearly shifted from exfiltration from the bedrock to exfiltration from riparian soil to overland flow from the riparian zone with increasing initial wetness. The 'island POIs' (POI 12 in area M, POI 7 in area D) were dry during dry and intermediate wetness conditions (Fig. 6a-d), inundated by streamflow during very wet conditions (Fig. 6g and h), and nearly exclusively watered by direct precipitation in case of the wetting phase starting from wet initial conditions (Fig. 6e).

Precipitation occurred at the beginning of the simulation period in two of the eight selected time periods (time period e and h, cf .Tab. 1). Since we identified the immediate delivery path of surface water by extracting the simulated water sources after the first time step (cf. section 3.3), direct precipitation could technically only appear as immediate delivery path at these two of the eight selected times. For the wet initial conditions (Fig. 6e), direct precipitation contributed by a small fraction to the generation of surface saturation in the riparian zone. However, contributions of direct precipitation were negligible compared to the contribution from other immediate delivery paths in the streambed and during the drying phase starting from very wet conditions (Fig. 6h).
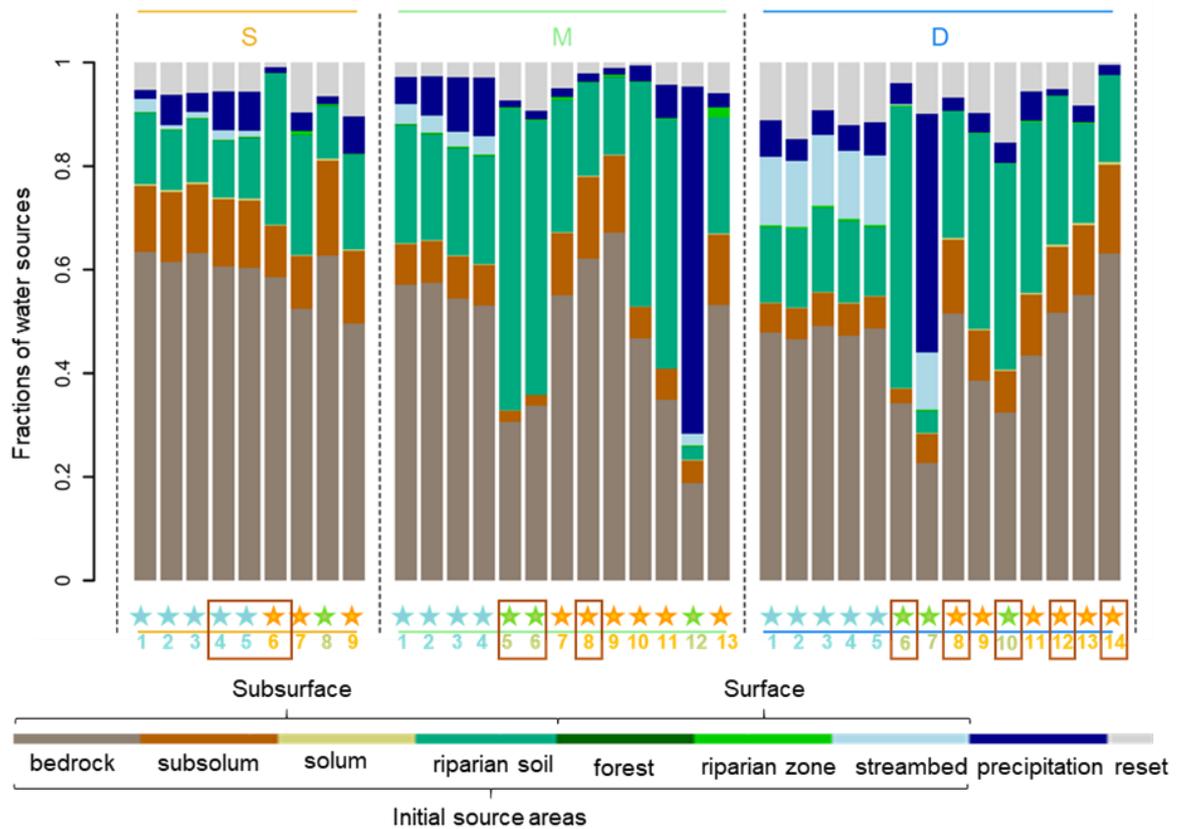
4.2 Geographical sources of surface water

4.2.1 Spatial variability

The most apparent spatial variability for the geographical sources of surface water was that upstream water was exclusively found at streambed locations and at the 'island POIs' (POI 12 in area M, POI 7 in area D, cf. section 4.1), with increasing fractions downstream (Fig. 7). Apart from that, the mixing of water from different geographical sources was quite homogeneous across the catchment with a large fraction of water that was initially stored in the fractured bedrock, followed by water that was initially stored in the riparian soil, and some contributions of water initially stored in the subsolum. The fraction of subsolum water was similar to the fraction of riparian soil water in area S, whereas the fraction of riparian soil water was clearly higher than the fraction of subsolum water in area M and D. Furthermore, the fraction of riparian soil water tended to be smaller in the streambed (blue POIs, Fig. 7) than in the riparian zone and at unclassified positions (yellow and green POIs, Fig. 7). Some distinct local variation of the fractions of riparian soil water and water from the fractured bedrock occurred within the riparian zone and for the unclassified locations.

Water from newly incoming precipitation (cf. Section 3.3.) generally only made up a small fraction of surface water (Fig. 7). Exceptions were again the two 'island POIs' (POI 12 in area M, POI 7 in area D). In addition, the fraction of precipitation seemed to be slightly higher in the streambed than in the riparian zone, especially in area M. Yet the reset fraction and thus the fraction of unknown water origin (cf. section 3.1) was at many locations higher than the fraction of precipitation. This makes it uncertain to interpret the spatial distribution and the contribution

16

1 of precipitation and other geographical water sources with small or non-existing fractions in the
2 mixing of surface water.

3
4 **Figure 7**. Spatial variability of geographical sources of surface water. The bars represent the mixing ratios of
5 sources for surface water at the 34 POIs located in the streambed (blue stars), in the riparian zone (yellow stars), and
6 at unclassifiable positions (green stars) in the three areas S, M and D along the stream (cf. Fig. 2). Brown labelled
7 POIs indicate locations with assumed distinct exfiltration of subsurface water (cf. Fig. 2). The fractions of water
8 sources are the arithmetic mean of the volume weighted fractions of the first four days (cf. Fig. 3) of each of the 64
9 sub-period simulations. The water sources that contributed to a POI over the four days represent the geographical
10 sources of surface water and can be distinguished between precipitation, upstream water, riparian soil and riparian
11 surface water, water from soils and surface from the hillslopes and plateau (solum and forest), and deeper water
12 from the subsolum and fractured bedrock (being exposed to the surface in the streambed).

13   **4.2.2 Temporal variability**

14   The most apparent difference in the mixing ratio of the geographical sources of surface
15 water between phases of wetting and drying was a varying fraction of precipitation (Fig. 8).
16 While the fraction of precipitation was nearly negligible during drying phases, it was small but
17 considerable during wetting phases. Exceptions were the wetting and drying phase starting from
18 intermediate initial wetness conditions (Fig. 8c and d), showing similar fractions of precipitation
19 in the surface water. The cumulative amount of precipitation was similar for these two time
20 periods as well (cf. Table 1), while it differed between the other wetting and drying phases. This
21 suggests that the varying fractions of precipitation in surface water rather reflected the amount of
22 precipitation in relation to discharge than the distinction between wetting and drying phases. The

17
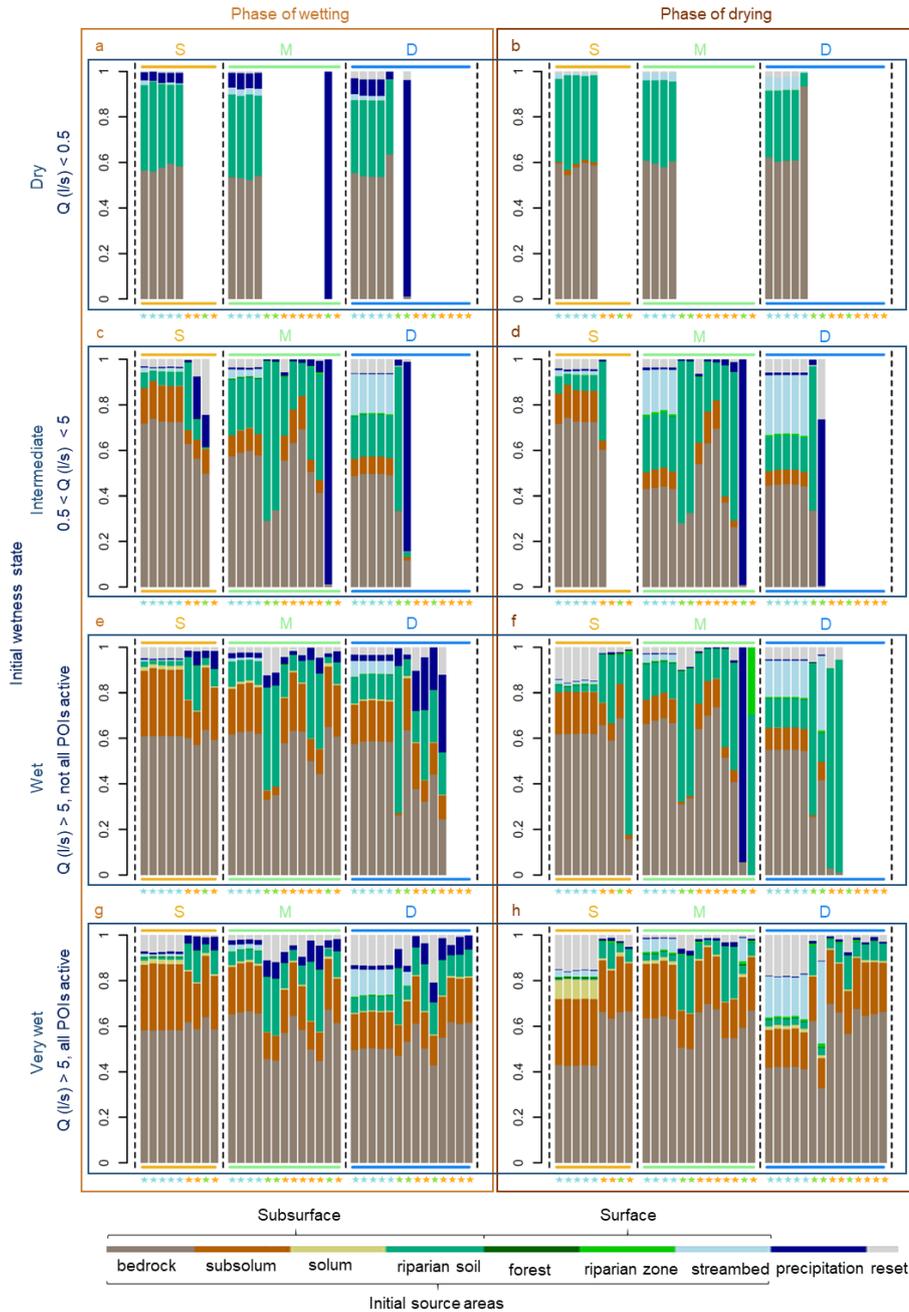
Figure 8. Temporal variability of geographical sources of surface water depending on the initial wetness state and phases of wetting or drying. The bars represent the mixing ratios of sources for surface water at the 34 POIs located in the streambed (blue stars), in the riparian zone (yellow stars), and at unclassifiable positions (green stars) in the three areas S, M and D along the stream (cf. Fig. 2). The fractions of water sources are the volume weighted fractions of the first four days (cf. Fig. 3) for each of the eight selected time periods (a-h, Fig. 4, Table 1). The water sources that contributed to a POI over the four days represent the geographical sources of surface water and can be distinguished between precipitation, upstream water, riparian soil and riparian surface water, water from soils and surface from the hillslopes and plateau (solum and forest), and deeper water from the subsolum and fractured bedrock (being exposed to the surface in the streambed).

18

fractions of the other geographical water sources differed in some areas or for some initial wetness states between phases of wetting and drying. However - as for the immediate delivery paths (cf. section 4.1) - we did not find any indication for systematically and clearly different sources of surface water between phases of wetting and drying.

A distinct variability of the mixing ratio of the geographical sources of surface water was found between the different initial wetness states at streambed locations (Fig. 8, blue POIs). For dry initial conditions (Fig. 8a and b), the water mainly originated from the riparian soil and fractured bedrock. For intermediate initial wetness conditions (Fig. 8c and d), the fraction of riparian soil water decreased and instead, considerable fractions of upstream water and water from the subsolum appeared. With further wetting of the initial conditions (Fig. 8e-h), the mixing ratio between riparian soil water and subsolum water tended to further shift towards a higher contribution of subsolum water. In addition, considerable fractions of solum water appeared, especially in area S (Fig. 8e-h).

The mixing ratio of the geographical sources of surface water in the riparian zone and at unclassified positions (Fig. 8, yellow and green POIs) tended as well to shift with increasing wetness from high fractions of water from the fractured bedrock and riparian soil to smaller fractions of riparian soil water in favour of higher fractions of subsolum water. Only at the two 'island POIs' (POI 12 in area M, POI 7 in area D), surface water was nearly exclusively precipitation during phases with dry and intermediate wetness conditions (Fig. 8a-d). For wetter conditions (Fig. 8e-h), the 'island POIs' did then show a similar mixture of geographical sources of surface water as found in the streambed.

5 Discussions

5.1 Assumptions and limitations of the hydraulic mixing cell approach

Premise for a meaningful application of the hydraulic mixing cell approach is that the simulations appropriately reflect reality. The performance of our model previously has been intensively evaluated and it has been proved that the spatial and temporal occurrence of the analyzed surface saturation was consistent with field observations (cf. section 2, Glaser et al. (2019)). This model assessment goes far beyond other studies that relied on simulations for analyzing the generation of surface saturation across space and evaluated model performance only against stream discharge (Partington et al., 2013; Weill et al., 2013). In principal, the simulation results could be further evaluated by measurements of geochemistry or isotope data for an identification of the water sources. Yet a main reason for the application of the hydraulic mixing cell simulations was to complement the process investigation where field measurements are difficult to perform, which is the case for a spatial identification of water sources along and sideways the streambed (cf. section 1).

The principal of the HMC approach is to delineate the mixing of water according to defined initial source areas in the surface and subsurface, thus it basically indicates if and where water was stored at the start of the simulation. Tracking the flow path of water through different source areas is currently only possible for overland flow, but not for flow through the subsurface. Nonetheless, we identified the immediate delivery paths and the geographical sources of surface water by investigating the mixing of the initial water sources i) after the first time step and ii) as volume weighted fractions over four days. It is likely that the identified mixing ratios of the sources would look different when applying different time intervals. However, since our study

19

focused on the comparison of the water sources between different locations and times, it was mainly important to apply the same analysis for all locations and sub-simulations. Precipitation was considered both in the analysis of the immediate delivery paths and the geographical sources of surface water. Actually, it mainly gives information about the temporal aspect of the water sources, indicating water that newly entered the system and was not already stored as 'old' water in the surface or subsurface at the start of the simulation.

The main challenge for the application of the HMC approach to identify the sources of surface water in space and time was to deal with the reset fractions, i.e. the fractions of water with unknown origin (cf. Section 3.1). The reset fractions ensure to avoid mass balance errors in the mixing calculation due to numerical instability or numerical dispersion, yet they still introduce uncertainties in the mixing ratios. Partington et al. (2013) suggested to accept reset fractions < 1 % and pointed out that the numerical stability of the HMC approach is particularly critical for model cells with small water storage volume relative to water flow. Accordingly, highly fluctuating water occurrence and shallow water depths in the riparian zone and streambed of the Weierbach catchment were problematic for the HMC simulations. We managed to limit the reset fractions and mass balance errors by using hourly input data, splitting the simulation periods in short sub-periods, and relaxing some numerical stability criteria, but we could not minimize the reset fractions to values < 1 % for stable simulations with reasonable calculation times. At various locations within the model domain our adaptations were not sufficient to obtain useable simulation results (data not shown) and further adaptations of the HMC code itself might help to overcome these current limitations. Nonetheless, for the locations that we investigated in this study, the times where we obtained high reset fractions or where we had to remove the simulation output from the analysis due to mass balance problems were acceptable. If water that was assigned with reset fractions was all originating from the same initial source, e.g. precipitation, the mixing ratio of the water sources might be largely distorted, especially for periods with high reset fractions (cf. Fig. 8f-h). However, it is unlikely that the reset fraction represents only one initial water source and we think that the obtained information about the mixture of the water sources is reliable for the investigation of the general relevance of different water sources and their spatial and temporal variability.

A limitation of our HMC analysis is that we did not systematically investigate the contribution of precipitation to surface water. The investigated time periods did not explicitly include periods with varying precipitation amounts for similar wetness states or periods with similar precipitation amounts for varying initial wetness conditions and phases of wetting or drying. Thus, we cannot clarify with this study if the fraction of precipitation in surface water actually varies for different initial wetness states and phases of wetting and drying or if the observed variation (Fig. 8) only reflects different ratios between the amount of precipitation and discharge volume (cf. Section 4.2.2, Fig. 8). However, a comparison of the contribution of precipitation between different locations is possible and since the two 'island POIs' (POI 12 in area M, POI 7 in area D, cf. section 4.1) often showed dominant contributions of precipitation, it is certain that the small fraction of precipitation observed at all other locations is not only an artefact of the characteristics of the selected time periods. Moreover, it needs to be kept in mind that we analyzed the fraction of the immediate delivery paths and geographical sources of surface water. Since the discharge volume differed for different POIs and different times, it is possible that the volume of a contributing source was constant in time or space, but that the fraction of the source relative to other sources changed. For example, the clear decrease of the relative contribution of exfiltration from the fractured bedrock into the streambed with increasing

20

wetness (cf. Fig. 6) does not mean that the exfiltration ceased. Instead, it is likely the exfiltration persisted but that the volume and thus relevance of water delivered via streamflow increased.

5.2 Sources of surface water in the Weierbach catchment in space and time

5.2.1 How is surface water generated and sourced in the riparian zone?

Previous studies in the Weierbach catchment suggested that the generation of surface saturation in the riparian zone is largely driven by groundwater exfiltration (Antonelli et al., 2019a; Glaser et al., 2019). The hydraulic mixing cell analysis could confirm these early suggestions. The identified high contribution of water that directly exfiltrated from the subsurface (Fig.5, Fig. 6) and mainly originated from riparian soil and fractured bedrock (Fig. 7, Fig. 8) suggests that surface saturation in the riparian zone was generated in situ by return flow. Accumulation of overland flow from neighboring locations in the riparian zone only exceeded the immediate delivery of water via return flow at some locations, especially close to the stream, which implies that return flow occurred widely across the riparian zone as dominant immediate delivery path of surface water. This does not confirm suggestions of previous work in the Weierbach catchment (Antonelli et al., 2019a; Glaser et al., 2018, 2019) that the delivery of surface water to the riparian zone is maintained by discrete locations of groundwater exfiltration (cf. Fig. 2) from where the surface water is distributed across the riparian zone via overland flow.

Moreover, we only found a small contribution of precipitation to the surface saturation in the riparian zone. This suggests that the mechanisms of saturation and infiltration excess as well as precipitation falling onto saturated areas generally played a minor role in inducing the occurrence of standing or flowing water at the surface of the riparian zone. This differs from what has been identified from ISSHM simulations for pre-alpine hillslopes (Weill et al., 2013) and riparian wetlands (Partington et al., 2013) and is only partly in line with the variable source area concept, where the generation of surface saturation and overland flow in the riparian zone of forested humid catchments is mainly related to a rising groundwater level and return flow, but also to continued precipitation on saturated areas (Dunne & Black, 1970; Hewlett & Hibbert, 1967; Megahan & King, 1985). We cannot entirely exclude a relevant influence of precipitation on the extension of surface saturation during rain events (cf. Fig.8), but we can exclude that such an influence was long-lasting or dominant at any location except of the 'island POIs' (cf. discussion in Section 5.1).

The observed temporal consistency of the mixing ratios of the immediate delivery paths of surface water (Fig. 6) rather suggested that surface saturation was consistently induced and maintained by the same process once the delivery of water was activated, independently if the surface saturation extent extended or contracted. The increased delivery of riparian surface water via overland flow and the shift of the mixing ratio of the geographical sources to higher fractions of subsolum and even solum water with increasing initial wetness (cf. Fig. 6, Fig. 8) is likely the result of generally more extended surface saturation across the riparian zone, which activates overland flow from the edges of the riparian zone where return flow from the underlying subsolum created surface saturation. This corresponds to the perceptual model described in previous work on the development of surface saturation in the riparian zone of the Weierbach catchment (Antonelli et al., 2019a). However, contrary to this previous work in the Weierbach catchment, our results did not support the suggestion that return flow and extension of streamflow into the riparian zone generated surface saturation during wetting phases, while the

21

surface saturation locally only persisted or was maintained by overland from neighboring areas during drying phases.

Finally, also the spatial variability of the immediate delivery paths and geographical sources of surface water was small across the riparian zone of the Weierbach. Based on the often discussed hot spots of biogeochemical and ecological activity in riparian zones and floodplains (e.g. Frei et al., 2012; Grabs et al., 2012; Harms & Grimm, 2008; Krause et al., 2017; Singer et al., 2016), the results from studies simulating the generation of surface saturation for pre-alpine hillslopes (Weill et al., 2013) and riparian wetlands (Partington et al., 2013), and the high heterogeneity of the occurrence and frequency of surface saturation in the riparian zone of the Weierbach catchment (cf. Fig. 2, Glaser et al., 2019), we would have expected a higher variability of the water sources. Instead, it seems that microtopography, which has been previously discussed as main explanatory factor for the patchiness of surface saturation occurrence in the Weierbach catchment (Glaser et al., 2019) as well as for hot spots of biogeochemical and microbiological activity in riparian zones (Frei et al., 2012; Harms & Grimm, 2008), is indeed the key factor to induce heterogeneous groundwater fluctuations, soil saturation, and flooding frequencies and magnitudes and thus hot spots of biogeochemical and ecological activity. However, we performed for the first time a detailed analysis of the immediate delivery paths and geographical sources of surface water across a riparian zone and we expect that the spatial variability of the water sources and thus their impact on the spatial variability of biogeochemical and ecological processes might be larger for riparian areas or floodplains of larger size and in other geographical settings than humid, temperate forests.

Moreover, although the spatial variability was small, we observed some spatial differences in the importance of overland flow (Fig. 5) and the contribution of riparian soil water relative to water from the fractured bedrock and subsolum (Fig. 7). This small variability might already be enough to create microhabitats due to varying water quantity and quality, thus nutrient availability and in consequence to generate hotspots of biogeochemical activity (cf. Krause et al., 2017; Ramey & Richardson, 2017). Furthermore, the spatial variability of the geographical sources might be critical when sampling riparian surface water, especially since the spatial difference of the water sources was not related to any visually apparent criteria such as saturation frequency, saturation persistence, locations of obvious groundwater exfiltration, or distance to the stream. Thus, when for example sampling riparian water as end member for a mixing analysis, the result of the hydrograph separation might change depending on the chosen sampling location. This highlights that it is important to consider a possible spatial variability of the water sources in riparian zones and floodplains and HMC simulations might actually be useful to support and inform the planning of field work beforehand of starting extensive sampling campaigns.

5.2.2 How is surface water generated and sourced in the streambed?

In addition to the successful identification of immediate delivery paths and geographical sources of surface water in the riparian zone, our HMC analysis allowed the investigation of the processes for surface saturation generation in the streambed and thus streamflow generation. We found a generally small fraction of precipitation in streamflow (Fig. 7, Fig. 8), which is consistent with results from previous tracer studies in the Weierbach catchment (Martínez-Carreras et al., 2015; Wrede et al., 2015) that showed that old water stored in the subsurface is the main contributor to streamflow. Yet the previous studies showed higher contributions of

22

precipitation to runoff events than what we simulated, especially for the characteristic sharp, short-lasting first discharge peaks (cf. Section 2.1), where precipitation represented up to 60% of the hydrograph (Martínez-Carreras et al., 2015). This discrepancy probably results from the fact that we did not specifically analyze the water sources for individual events and especially did not distinguish the first peak events from the second peak events. A follow-up study might use the HMC simulations to explicitly analyze the water sources of the hydrograph in the Weierbach catchment with a focus on distinct rain events (cf. discussion in Section 5.1) and the first and second discharge peaks in comparison and as complement to the field observations. Nonetheless, our analysis included periods with rain and discharge events (cf. Fig. 4) and we found that subsurface water was always dominating in the streamflow, independently from the initial wetness state or phases of wetting or drying (cf. Fig. 8). Thus, the model reproduced the phenomenon of high contributions of pre-event water in event runoff that has been observed in catchments worldwide (e.g. Cartwright & Morgenstern, 2018; Correa et al., 2019; Laudon et al., 2007) and is known as the 'old water paradox' (cf. Kirchner, 2003). This indicates that precipitation-induced pressure waves can explain and simulate the fast release of previously stored subsurface water to the stream as immediate response to precipitation.

The identified similarity of the geographical sources of surface water in the streambed and riparian zone (cf. Fig.7) furthermore suggests that streamflow was generated in the same way as the surface saturation in the riparian zone, i.e. by direct return flow of subsurface water. Based on the facts that direct exfiltration from the underlying bedrock appeared as a relevant immediate delivery path of surface water to a POI during dry conditions (Fig. 6) and that the contribution of upstream water increased downstream (Fig.8), we suppose that streamflow accumulated from return flow occurring all along the streambed. With increasing wetness, return flow increased and extended the active stream length (cf. Antonelli et al., 2019b), which in turn increased the accumulation and thus fraction of upstream water at the investigated POIs (cf. Fig. 8). This matches the variable source area concept, where Hewlett and Hibbert (1967) described that streamflow in intermittent and ephemeral tributaries is likely to be generated and expanded by an expansion of subsurface saturation and consequently return flow. Here, we highlight that the same processes occur and generate streamflow in downstream areas in addition to a delivery of water via streamflow from upstream. Moreover, we suggest that an extent and increase of streamflow results from an extent and increase of the volume of return flow which is triggered by increasing subsurface saturation as well as the propagation of pressure waves induced by precipitation (cf. above).

With the exception of an increase of upstream water contribution in downstream direction, the sources of stream water where rather homogeneous in space (Fig. 7). This is consistent with findings from McGlynn et al. (2004) who showed that runoff water in the Maimai catchment consistently originated from headwater riparian areas, independently from the catchment size, and was transferred downstream through the channel network. Moreover, various studies have discussed accordingly that the sources of stream water mainly vary with varying physiographic settings, landscape features, or climatic conditions (Brown et al., 1999; Cowie et al., 2017; Gordon et al., 2015; Laudon et al., 2007), whereas mere catchment size and thus the location along the stream within a catchment with consistent characteristics – such as the Weierbach catchment – has a less significant influence on the hydrograph sources. Some studies found a spatially varying contribution of young water depending on catchment size or elevation (Brown et al., 1999; James & Roulet, 2009; Zhang et al., 2018), which we cannot confirm or contradict with certainty from our simulations due to the similar size of the simulated fractions of

23

precipitation and reset fractions (cf. Fig. 7, discussion in Section 5.1). Correa et al. (2019) identified a more pronounced and faster connectivity of hillslope soil water to the stream for upper than lower sub-catchments, which matches our identified tendency for lower fractions of subsolum water compared to riparian soil water in downstream direction (Fig. 7).

Temporally, the mixing ratio of the different subsurface water sources in the stream shifted with increasing wetness towards higher contributions from geographical sources from further uphill (i.e. riparian soil -> subsolum -> solum, cf. Fig. 8), which indicates that more distant geographical source areas got activated and connected to the stream with increasing wetness. In line with that, previous hydrograph separations in the Weierbach catchment revealed a more dominant contribution of riparian soil water during events in spring and summer than for events in autumn and winter, where the wetness state of the catchment was higher (Martínez-Carreras et al., 2015). Moreover, our findings match findings from tracer studies in other catchments that analyzed the temporal variation of the sources of stream water. For example, McGlynn and McDonnell (2003) showed as well an increasing contribution of hillslope water in comparison to riparian water during events with large discharge volume and related this finding to an initiation of hillslope runoff with increasing wetness. Also Correa et al. (2017) highlighted the importance of riparian soil water in streamflow and revealed a direct connection of water flow from the hillslopes to the stream during wet periods only. Yet Correa et al. (2017) found a slightly smaller fraction of riparian soil water during the drier than the wetter season. This relates to the circumstance that they found especially high contributions of spring water from the bedrock during the drier season, whereas the spring water became less relevant compared to shallow interflow through the soil for high flow conditions. Similarly, Cartwright and Morgenstern (2018) found indications that the activation of streamflow during events was related to an activation of water flow from soils or regolith in addition to groundwater baseflow. In line with both of these studies, our simulations showed a relevant delivery of water to the stream by direct exfiltration from the underlying fractured bedrock during dry conditions (Fig. 6) and increased fractions of subsolum and solum water in streamflow with increasing wetness (cf. Fig. 8), which indicates that return flow from the underlying fractured bedrock persisted throughout time but was increased and extended by water flow from upstream and uphill source areas and upper soil layers once these areas and layers were wetted up and activated.

Overland flow from the riparian zone showed to be a negligible immediate delivery path for streamflow generation (Fig. 5). This, together with the absence of considerable fractions of water originating from initial riparian surface saturation (Fig. 7) and the low fractions of new precipitation water (cf. above) in stream water, suggests minor contributions of overland flow from the riparian zone to streamflow generation. This is in contrast to previous studies in the Weierbach catchment that assumed saturation excess overland flow from the riparian zone to play a relevant role in streamflow generation, especially for the first discharge peaks (Fenicia et al., 2014; Glaser et al., 2016; Klaus et al., 2015; Martínez-Carreras et al., 2015; Martínez-Carreras et al., 2016; Schwab et al., 2018; Wrede et al., 2015). Yet, as discussed above, we did not explicitly analyze the generation of the first and second peaks of the hydrograph of the catchment and thus cannot fully exclude a high contribution of overland flow to parts of the hydrograph. Moreover, it is likely that water delivered to the surface of the riparian zone nonetheless passes through the riparian zone via overland flow (cf. Fig. 5) and eventually enters the stream. Since such a behavior could not be inferred from a difference between the geographical sources of the surface water in the riparian zone and streambed (Fig. 7), we suggest that this is a fast process and that overland flow in the riparian zone is basically intermittent

24

streamflow outside of the streambed. On one side this implies that the consideration of riparian overland flow as distinct component in streamflow generation is redundant for the Weierbach catchment. On the other side, a distinct consideration of riparian overland flow can still be justified by the visually apparent separation between the streambed and riparian zone and lower water volumes and surface saturation frequencies in the riparian zone than in the streambed. In that sense, further analyses would be needed to quantify which fraction of the water in the stream actually passed through the riparian zone as overland flow and to quantify the active contribution of the riparian surface saturation to streamflow in comparison to the actual extent of surface saturation (as e.g. done in Partington et al., 2013).

6 Conclusions

We investigated the sources of surface water and their variability in space and time using a hydraulic mixing cell approach in a comprehensively validated HydroGeoSphere model of the Weierbach catchment. The hydraulic mixing cell analysis allowed us to specify and quantify the different geographical sources and the flow mechanisms (i.e. immediate delivery paths) that generated surface saturation in the riparian zone and surface water in the streambed of three small study areas along a stream. Further development of the hydraulic mixing cell code might help to reduce uncertainties introduced by numerical instability of the mixing calculations when applying the code for intermittently flooded areas with small water volume. Nonetheless, we obtained useful information going far beyond the standard model output and we think that this or similar simulation approaches should be used much more often than currently done as complement to field data, either to help to interpret and understand the field data or to plan experiments beforehand.

Our main findings regarding the spatial and temporal variability of the immediate delivery paths and geographical sources of surface water in the riparian zone and streambed of the Weierbach catchment are as follows:

- Contribution of precipitation and thus new water to surface water was generally small
- Surface water was mainly delivered by exfiltration from the subsurface in the riparian zone and by streamflow in the streambed
- Once the delivery of surface water was activated, the immediate delivery path stayed rather constant, independently from the initial wetness state or phases of drying or wetting. Exception was a clear shift of the ratio of predominant sources towards streamflow and overland flow in the streambed and riparian zone during dry and very wet conditions, respectively.
- The geographical sources of surface water were similar among the three studied areas and the riparian zone and streambed, representing mainly a mixture of water from different subsurface stores.
- The mixing ratio of geographical sources remained similar for phases of wetting or drying, but changed depending on the initial wetness state, i.e. water from upslope source areas became more relevant with increasing wetness.

Regarding the generation of surface saturation in the riparian zone of the investigated Weierbach catchment, these findings implied that surface saturation was generated and maintained by return flow across the riparian zone. Nonetheless, we identified some spatial

25

variations in the delivery and mixing of surface water within the three small investigated areas (155 -170 m^2) that could have an effect on microhabitats and hotspots of biogeochemical activity or might be relevant for water sampling campaigns. In larger riparian zones or floodplains, this spatial variability of water sources and thus possible impacts on ecology, water quality, runoff, and flood risk may be larger. Therefore, we think more studies similar to ours are needed, especially for riparian zones or floodplains with different sizes or different landscape characteristics and climatic conditions.

Regarding the streamflow generation in the Weierbach catchment, we concluded from the results of the hydraulic mixing cell simulations that streamflow was mainly generated by return flow into the streambed occurring all along the stream. With increasing wetness, the return flow originated from more extended source areas, reaching up to the hillslopes and its upper soil layers during very wet conditions. We suggest that precipitation plays a minor immediate contribution to streamflow generation, but that it triggers an increase and extent of return flow by inducing pressure waves. Future analyses with the hydraulic mixing cells might focus explicitly on an analysis of the streamflow sources during single events to clarify the role of precipitation in inducing hydrograph peaks. Similarly, future analyses could clarify if the identified minor contribution of riparian overland flow to streamflow generation applies throughout single events and if and how much of riparian surface saturation eventually connects to the stream. Yet we discussed that a distinction between riparian overland flow and streamflow was mainly justified by a visual spatial delineation and discharge volume. In the end, the generation mechanism and water sources proved to be the same for riparian overland flow and streamflow, meaning that overland flow was basically nothing else than intermittent streamflow outside of the streambed.

Acknowledgments, Samples, and Data

References

Antonelli, M., Glaser, B., Teuling, A. J., Klaus, J., & Pfister, L. (2019a). Saturated areas through the lens: 1. Spatio-temporal variability of surface saturation documented through Thermal Infrared imagery. *Hydrological Processes*, *in review*, (HYP-19-0453).

Antonelli, M., Glaser, B., Teuling, A. J., Klaus, J., & Pfister, L. (2019b). Saturated areas through the lens: 2. Spatio-temporal variability of streamflow generation and its relationship with surface saturation. *Hydrological Processes*, *in review*, (HYP-19-0455).

Barthold, F. K., & Woods, R. A. (2015). Stormflow generation: A meta-analysis of field evidence from small, forested catchments. *Water Resources Research*, *51*, 3730–3753. https://doi.org/10.1002/2014WR016221

Birkel, C., Soulsby, C., Tetzlaff, D., Dunn, S., & Spezia, L. (2012). High-frequency storm event isotope sampling reveals time-variant transit time distributions and influence of diurnal cycles. *Hydrological Processes*, *26*(2), 308–316. https://doi.org/10.1002/hyp.8210

Brown, V. A., McDonnell, J. J., Burns, D. A., & Kendall, C. (1999). The role of event water, a rapid shallow flow component, and catchment size in summer stormflow. *Journal of Hydrology*, *217*(3–4), 171–190. https://doi.org/10.1016/S0022-1694(98)00247-9

Campana, M. E., & Simpson, E. S. (1984). Groundwater residence times and recharge rates using a discrete-state compartment model and 14C data. *Journal of Hydrology*, *72*, 171–185. https://doi.org/10.1016/0022-1694(84)90190-2

Carrer, G. E., Klaus, J., & Pfister, L. (2019). Assessing the Catchment Storage Function Through a Dual-Storage

Concept. *Water Resources Research*, *55*(1), 476–494. https://doi.org/10.1029/2018WR022856

Cartwright, I., & Morgenstern, U. (2018). Using tritium and other geochemical tracers to address the "old water paradox" in headwater catchments. *Journal of Hydrology*, *563*, 13–21. https://doi.org/10.1016/j.jhydrol.2018.05.060

Chow, R., Frind, M. E., Frind, E. O., Jones, J. P., Sousa, M. R., Rudolph, D. L., et al. (2016). Delineating baseflow contribution areas for streams – A model and methods comparison. *Journal of Contaminant Hydrology*, *195*, 11–22. https://doi.org/10.1016/j.jconhyd.2016.11.001

Correa, A., Windhorst, D., Tetzlaff, D., Crespo, P., Célleri, R., Feyen, J., & Breuer, L. (2017). Temporal dynamics in dominant runoff sources and flow paths in the Andean Páramo. *Water Resources Research*, *53*(7), 5998–6017. https://doi.org/10.1002/2016WR020187

Correa, A., Breuer, L., Crespo, P., Célleri, R., Feyen, J., Birkel, C., et al. (2019). Spatially distributed hydro-chemical data with temporally high-resolution is needed to adequately assess the hydrological functioning of headwater catchments. *Science of The Total Environment*, *651*, 1613–1626. https://doi.org/10.1016/j.scitotenv.2018.09.189

Cowie, R. M., Knowles, J. F., Dailey, K. R., Williams, M. W., Mills, T. J., & Molotch, N. P. (2017). Sources of streamflow along a headwater catchment elevational gradient. *Journal of Hydrology*, *549*, 163–178. https://doi.org/10.1016/j.jhydrol.2017.03.044

Dunne, T., & Black, R. D. (1970). Partial area contributions to storm runoff in a small New England watershed. *Water Resources Research*, *6*, 1296–1311. Retrieved from http://soilandwater.bee.cornell.edu/Research/VSA/papers/DunneWRR70.pdf

Engdahl, N. B., & Maxwell, R. M. (2015). Quantifying changes in age distributions and the hydrologic balance of a highmountain watershed from climate induced variations in recharge. *In Review.*, *522*, 152–162. https://doi.org/10.1016/j.jhydrol.2014.12.032

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

Frei, S., Knorr, K. H., Peiffer, S., & Fleckenstein, J. H. (2012). Surface micro-topography causes hot spots of biogeochemical activity in wetland systems: A virtual modeling experiment. *Journal of Geophysical Research: Biogeosciences*, *117*, G00N12. https://doi.org/10.1029/2012JG002012

Glaser, B., Klaus, J., Frei, S., Frentress, 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

Glaser, B., Antonelli, M., Chini, M., Pfister, L., & Klaus, J. (2018). Technical note: Mapping surface-saturation dynamics with thermal infrared imagery. *Hydrology and Earth System Sciences*, *22*(11), 5987–6003. https://doi.org/10.5194/hess-22-5987-2018

Glaser, B., Antonelli, M., Hopp, L., & Klaus, J. (2019). Intra-catchment variability of surface saturation &ndash; insights from longterm observations and simulations. *Hydrology and Earth System Sciences Discussions*, (May), 1–22. https://doi.org/10.5194/hess-2019-203

Gordon, R. P., Lautz, L. K., McKenzie, J. M., Mark, B. G., Chavez, D., & Baraer, M. (2015). Sources and pathways of stream generation in tropical proglacial valleys of the Cordillera Blanca, Peru. *Journal of Hydrology*, *522*, 628–644. https://doi.org/10.1016/j.jhydrol.2015.01.013

Gourdol, L., Clément, R., Juilleret, J., Pfister, L., & Hissler, C. (2018). Large-scale ERT surveys for investigating shallow regolith properties and architecture. *Hydrology and Earth System Sciences Discussions*. https://doi.org/10.5194/hess-2018-519

Grabs, T., Bishop, K., Laudon, H., Lyon, S. W., & Seibert, J. (2012). Riparian zone hydrology and soil water total organic carbon (TOC): implications for spatial variability and upscaling of lateral riparian TOC exports. *Biogeosciences*, *9*(10), 3901–3916. https://doi.org/10.5194/bg-9-3901-2012

Harms, T. K., & Grimm, N. B. (2008). Hot spots and hot moments of carbon and nitrogen dynamics in a semiarid riparian zone. *Journal of Geophysical Research: Biogeosciences*, *113*, G01020. https://doi.org/10.1029/2007JG000588

Hewlett, J. D., & Hibbert, A. R. (1967). Factors affecting the response of small watersheds to precipitation in humid areas. In W. E. Sopper & H. W. Lull (Eds.), *International Symposium on Forest Hydrology* (pp. 275–290). Oxford: Pergamon Press. Retrieved from http://coweeta.ecology.uga.edu/publications/851.pdf

James, A. L., & Roulet, N. T. (2009). Antecedent moisture conditions and catchment morphology as controls on spatial patterns of runoff generation in small forest catchments. *Journal of Hydrology*, *377*(3–4), 351–366. https://doi.org/10.1016/j.jhydrol.2009.08.039

27

Jones, J. P., Sudicky, E. A., Brookfield, A. E., & Park, Y.-J. (2006). An assessment of the tracer-based approach to quantifying groundwater contributions to streamflow. *Water Resources Research*, *42*(2), n/a-n/a. https://doi.org/10.1029/2005WR004130

Juilleret, J., Dondeyne, S., Vancampenhout, K., Deckers, J., & Hissler, C. (2016). Mind the gap: A classification system for integrating the subsolum into soil surveys. *Geoderma*, *264*, 332–339. https://doi.org/10.1016/j.geoderma.2015.08.031

Kirchner, J. W. (2003). A double paradox in catchment hydrology and geochemistry. *Hydrological Processes*, *17*, 871–874. https://doi.org/10.1002/hyp.5108

Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research*, *45*(2), 1–34. https://doi.org/10.1029/2008WR006912

Klaus, J., & McDonnell, J. J. (2013). Hydrograph separation using stable isotopes: Review and evaluation. *Journal of Hydrology*, *505*, 47–64. https://doi.org/10.1016/j.jhydrol.2013.09.006

Klaus, J., Wetzel, C. E., Martínez-Carreras, N., Ector, L., & Pfister, L. (2015). A tracer to bridge the scales: on the value of diatoms for tracing fast flow path connectivity from headwaters to meso-scale catchments. *Hydrological Processes*, *29*(25), 5275–5289. https://doi.org/10.1002/hyp.10628

Kollet, S., Sulis, M., Maxwell, R. M., Paniconi, C., Putti, M., Bertoldi, G., et al. (2017). The integrated hydrologic model intercomparison project, IH-MIP2: A second set of benchmark results to diagnose integrated hydrology and feedbacks. *Water Resources Research*, *53*, 867–890. https://doi.org/10.1002/2016WR019191

Krause, S., Lewandowski, J., Grimm, N. B., Hannah, D. M., Pinay, G., McDonald, K., et al. (2017). Ecohydrological interfaces as hot spots of ecosystem processes. *Water Resources Research*, *53*(8), 6359–6376. https://doi.org/10.1002/2016WR019516

Kristensen, K. J., & Jensen, S. E. (1975). A model for estimating actual evapotranspiration from potential evapotranspiration. *Nordic Hydrology*, *6*, 170–188. https://doi.org/10.2166/nh.1975.012

Laudon, H., Sjöblom, V., Buffam, I., Seibert, J., & Mörth, M. (2007). The role of catchment scale and landscape characteristics for runoff generation of boreal streams. *Journal of Hydrology*, *344*(3–4), 198–209. https://doi.org/10.1016/j.jhydrol.2007.07.010

Liggett, J., & Werner, A. (2014). Fully integrated modeling of surface-subsurface solute transport and the effect of dispersion in tracer hydrograph separation. *Water Resources …*, 1–16. https://doi.org/10.1002/2013WR015040.Received

Mallik, A. U., Lamb, E. G., & Rasid, H. (2001). Vegetation zonation among the microhabitats in a lacustrine environment: analysis and application of belowground species trait patterns. *Ecological Engineering*, *18*(2), 135–146. https://doi.org/10.1016/S0925-8574(01)00069-6

Martínez-Carreras, N., Wetzel, C. E., Frentress, J., Ector, L., McDonnell, J. J., Hoffmann, L., & Pfister, L. (2015). Hydrological connectivity inferred from diatom transport through the riparian-stream system. *Hydrology and Earth System Sciences*, *19*, 3133–3151. https://doi.org/10.5194/hess-19-3133-2015

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

Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J.-O., Ferguson, I. M., Ivanov, V., et al. (2014). Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks. *Water Resources Research*, *50*(2), 1531–1549. https://doi.org/10.1002/2013WR013725

Maxwell, R. M., Condon, L. E., Danesh-Yazdi, M., & Bearup, L. A. (2019). Exploring source water mixing and transient residence time distributions of outflow and evapotranspiration with an integrated hydrologic model and Lagrangian particle tracking approach. *Ecohydrology*, *12*(1), e2042. https://doi.org/10.1002/eco.2042

McDonnell, J. J. (2003). Where does water go when it rains? Moving beyond the variable source area concept of rainfall-runoff response. *Hydrological Processes*, *17*, 1869–1875. https://doi.org/10.1002/hyp.5132

McGlynn, B. L., & McDonnell, J. J. (2003). Quantifying the relative contributions of riparian and hillslope zones to catchment runoff. *Water Resources Research*, *39*(11). https://doi.org/10.1029/2003wr002091

McGlynn, B. L., & Seibert, J. (2003). Distributed assessment of contributing area and riparian buffering along stream networks. *Water Resources Research*, *39*(4), 1082. https://doi.org/10.1029/2002WR001521

McGlynn, B. L., McDonnell, J. J., Seibert, J., & Kendall, C. (2004). Scale effects on headwater catchment runoff timing, flow sources, and groundwater-streamflow relations. *Water Resources Research*, *40*(7), 1–14. https://doi.org/10.1029/2003WR002494

Megahan, W. F., & King, P. N. (1985). Identification of critical areas on forest lands for control of nonpoint sources of pollution. *Environmental Management*, *9*, 7–17. https://doi.org/10.1007/BF01871440

28

Paniconi, C., & Putti, M. (2015). Physically based modeling in catchment hydrology at 50: Survey and outlook. *Water Resources Research*, *51*, 7090–7129. https://doi.org/10.1002/2015WR017780

Partington, D., Brunner, P., Simmons, C. T., Therrien, R., Werner, a. D., Dandy, G. C., & Maier, H. R. (2011). A hydraulic mixing-cell method to quantify the groundwater component of streamflow within spatially distributed fully integrated surface water-groundwater flow models. *Environmental Modelling and Software*, *26*, 886–898. https://doi.org/10.1016/j.envsoft.2011.02.007

Partington, D., Brunner, P., Frei, S., Simmons, C. T., Werner, A. D., Therrien, R., et al. (2013). Interpreting streamflow generation mechanisms from integrated surface-subsurface flow models of a riparian wetland and catchment. *Water Resources Research*, *49*, 5501–5519. https://doi.org/10.1002/wrcr.20405

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

Pfister, L., Martínez-Carreras, N., Hissler, C., Klaus, J., Carrer, G. E., Stewart, M. K., & McDonnell, J. J. (2017). Bedrock geology controls on catchment storage, mixing, and release: A comparative analysis of 16 nested catchments. *Hydrological Processes*, *31*, 1828–1845. https://doi.org/10.1002/hyp.11134

Ramey, T. L., & Richardson, J. S. (2017). Terrestrial Invertebrates in the Riparian Zone: Mechanisms Underlying Their Unique Diversity. *BioScience*, *67*(9), 808–819. https://doi.org/10.1093/biosci/bix078

Remondi, F., Kirchner, J. W., Burlando, P., & Fatichi, S. (2018). Water Flux Tracking With a Distributed Hydrological Model to Quantify Controls on the Spatiotemporal Variability of Transit Time Distributions. *Water Resources Research*, *54*(4), 3081–3099. https://doi.org/10.1002/2017WR021689

de Rooij, R., Graham, W., & Maxwell, R. M. (2013). A particle-tracking scheme for simulating pathlines in coupled surface-subsurface flows. *Advances in Water Resources*, *52*, 7–18. https://doi.org/10.1016/j.advwatres.2012.07.022

Sayama, T., & McDonnell, J. J. (2009). A new time-space accounting scheme to predict stream water residence time and hydrograph source components at the watershed scale. *Water Resources Research*, *45*(7). https://doi.org/10.1029/2008WR007549

Scaini, A., Hissler, C., Fenicia, F., Juilleret, J., Iffly, J. F., Pfister, L., & Beven, K. (2018). Hillslope response to sprinkling and natural rainfall using velocity and celerity estimates in a slate-bedrock catchment. *Journal of Hydrology*, *558*, 366–379. https://doi.org/10.1016/j.jhydrol.2017.12.011

Schwab, M. P., Klaus, J., Pfister, L., & Weiler, M. (2018). Diel fluctuations of viscosity-driven riparian inflow affect streamflow DOC concentration. *Biogeosciences*, *15*, 2177–2188. https://doi.org/10.5194/bg-15-2177-2018

Sebben, M. L., Werner, A. D., Liggett, J. E., Partington, D., & Simmons, C. T. (2013). On the testing of fully integrated surface – subsurface hydrological models. *Hydrological Processes*, *27*, 1276–1285. https://doi.org/10.1002/hyp.9630

Singer, M. B., Harrison, L. R., Donovan, P. M., Blum, J. D., & Marvin-DiPasquale, M. (2016). Hydrologic indicators of hot spots and hot moments of mercury methylation potential along river corridors. *Science of The Total Environment*, *568*(March), 697–711. https://doi.org/10.1016/j.scitotenv.2016.03.005

Sklash, M. G., & Farvolden, R. N. (1979). The role of groundwater in storm runoff. *Journal of Hydrology*, *43*, 45–65. https://doi.org/10.1016/0022-1694(79)90164-1

Weill, S., Altissimo, M., Cassiani, G., Deiana, R., Marani, M., & Putti, M. (2013). Saturated area dynamics and streamflow generation from coupled surface-subsurface simulations and field observations. *Advances in Water Resources*, *59*, 196–208. https://doi.org/10.1016/j.advwatres.2013.06.007

Williams, D. G., & Scott, R. L. (2009). Vegetation-hydrology interactions: Dynamics of riparian plant water use. In J. C. Stromberg & B. J. Tellman (Eds.), *Ecology and Conservation of the San Pedro River* (pp. 37–56). Tucson: University of Arizona Press.

Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., et al. (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

Yang, J., Heidbüchel, I., Musolff, A., Reinstorf, F., & Fleckenstein, J. H. (2018). Exploring the Dynamics of Transit Times and Subsurface Mixing in a Small Agricultural Catchment. *Water Resources Research*, *54*(3), 2317–2335. https://doi.org/10.1002/2017WR021896

Zhang, Q., Knowles, J. F., Barnes, R. T., Cowie, R. M., Rock, N., & Williams, M. W. (2018). Surface and subsurface water contributions to streamflow from a mesoscale watershed in complex mountain terrain. *Hydrological Processes*, *32*(7), 954–967. https://doi.org/10.1002/hyp.11469

29

List of publications

First-author manuscripts in ISI-listed journals

Glaser, B., Antonelli, M., Hopp, L., Klaus, J. 2019: Intra-catchment variability of surface saturation – insights from long-term observations and simulations. UNDER REVIEW IN Hydrology and Earth System Sciences Discussion. DOI: 10.5194/hess-2019-203

Glaser, B., Jackisch, C., Hopp, L., Klaus, J. 2019: How meaningful are plot-scale observations and simulations of preferential flow for catchment models? PUBLISHED IN Vadose Zone Journal, Volume 18: 180146. DOI: 10.2136/vzj2018.08.0146

Glaser, B., Antonelli, M., Chini, M., Pfister, L., Klaus, J. 2018: Technical note: Mapping surface-saturation dynamics with thermal infrared imagery. PUBLISHED IN Hydrology and Earth System Sciences, Volume 22: 5987-6003. DOI: 10.5194/hess-22-5987-2018

Glaser, B., Klaus, J., Frei, S., Frentress, 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. PUBLISHED IN Water Ressources Research, Volume 52: 8317-8342. DOI: 10.1002/2015WR018414

Co-authored manuscripts in ISI-listed journals

Antonelli, M., Glaser, B., Teuling, A. J., Klaus, J., Pfister, L. 2019: Saturated areas through the lens: 2. Spatio-temporal variability of streamflow generation and its relationship with surface saturation. UNDER REVIEW IN Hydrological Processes. HYP-19-0455

Antonelli, M., Glaser, B., Teuling, A. J., Klaus, J., Pfister, L. 2019: Saturated areas through the lens: 1. Spatio-temporal variability of surface saturation documented through Thermal Infrared imagery. UNDER REVIEW IN Hydrological Processes. HYP-19-0453

Hopp, L., Glaser, B., Klaus, J., Schramm, T. 2019: Calibration of a three dimensional dual-permeability model at the catchment scale using DREAM. UNDER REVIEW IN Hydrological Processes. HYP-19-0345

(Eidesstattliche) Versicherungen und Erklärungen

(§ 8 Satz 2 Nr. 3 PromO Fakultät)

Hiermit versichere ich eidesstattlich, dass ich die Arbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe (vgl. Art. 64 Abs. 1 Satz 6 BayHSchG).

(§ 8 Satz 2 Nr. 3 PromO Fakultät)

Hiermit erkläre ich, dass ich die Dissertation nicht bereits zur Erlangung eines akademischen Grades eingereicht habe und dass ich nicht bereits diese oder eine gleichartige Doktorprüfung endgültig nicht bestanden habe.

(§ 8 Satz 2 Nr. 4 PromO Fakultät)

Hiermit erkläre ich, dass ich Hilfe von gewerblichen Promotionsberatern bzw. -vermittlern oder ähnlichen Dienstleistern weder bisher in Anspruch genommen habe noch künftig in Anspruch nehmen werde.

(§ 8 Satz 2 Nr. 7 PromO Fakultät)

Hiermit erkläre ich mein Einverständnis, dass die elektronische Fassung der Dissertation unter Wahrung meiner Urheberrechte und des Datenschutzes einer gesonderten Überprüfung unterzogen werden kann.

(§ 8 Satz 2 Nr. 8 PromO Fakultät)

Hiermit erkläre ich mein Einverständnis, dass bei Verdacht wissenschaftlichen Fehlverhaltens Ermittlungen durch universitätsinterne Organe der wissenschaftlichen Selbstkontrolle stattfinden können.

_____   _____
Ort, Datum   Barbara Glaser