Improving the conceptualization of the streamflow Travel Time Distribution and its estimation from isotopic tracers

Zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS
(Dr.-Ing.)

von der KIT-Fakultät für
Bauingenieur-, Geo- und Umweltwissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Nicolas Rodriguez
aus Mont Saint Martin (France)

Tag der mündlichen Prüfung: 04.02.2020

Referent: Prof. Dr.-Ing. Erwin Zehe
Korreferent: Prof. Dr. Andrea Rinaldo
Korreferent: Prof. Dr. Laurent Pfister

Karlsruhe (2020)

Declaration of Authorship

Eidesstattliche Versicherung gemäß § 6 Abs. 1 Ziff. 4 der Promotionsordnung des Karlsruher Instituts für Technologie für die Fakultät für Bauingenieur-, Geo- und Umweltwissenschaften:

1. Bei der eingereichten Dissertation zu dem Thema "Improving the conceptualization of the streamflow Travel Time Distribution and its estimation from isotopic tracers" handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Karlsruhe, im Oktober 2019

Abstract

Water travelling through hydrological landscapes (such as catchments) can take a variety of times before reaching a neighboring stream. In other words, water in streamflow has a distribution of ages. Travel time distributions (TTDs) characterize the water ages in catchment outflows and are fundamental descriptions of how catchments store and release water that entered days, months, and years ago as precipitation. TTDs are essential for a better management of water resources because they give insights about water flow paths and velocity in hydrological systems, and they are key to gear hydrological models towards water quality simulations. Streamflow TTDs have been estimated from hydrological tracers like stable isotopes of oxygen (O) and hydrogen (H) for more than 50 years. Yet, the detailed characteristics of the TTDs such as their shape, statistics, and their time variability are not fully deciphered, and the factors controlling these characteristics are not clearly identified. In addition, no general agreement exists regarding the estimation of unsteady streamflow TTDs from isotopic tracers or about their use to model the transport of tracers. This incomplete understanding partly stems from the prevalence of simple steady-state TTD models deduced from low resolution tracer data in the last decades. Most streamflow travel time studies relied on bi-weekly or monthly stable isotope data generally sampled over less than two years. Accordingly, the time variance of streamflow TTDs mostly remained unacknowledged in the analytical TTD models that were tested against the tracer time series. The overarching objective of this thesis is to understand what controls the shape and time variability of the streamflow TTD, and how it can be determined from tracer data.

This is achieved with a combination of theoretical investigations, experimental (field and laboratory) work, and modeling efforts to go beyond previous limitations. Isotopic tracers (^{2}H, ^{18}O, and ^{3}H) were collected at high frequency (sub-daily) in precipitation and streamflow of experimental catchments in Luxembourg over 2 years. Improved parameterizations of streamflow TTDs and novel analytical models were proposed based on theoretical and experimental grounds, notably using the acquired isotopic data set in Luxembourg. A method to leverage water age information from a dual-isotopic approach (using both ^{2}H and ^{3}H) was suggested to clarify an emerging misunderstanding about the limitations of the stable isotopes of O and H compared to ^{3}H. The work in this thesis shows that streamflow TTDs can have more varied shapes and more intricate variability than hypothesized in studies over the last decade. A Mediterranean climate can for example cause complex patterns of water ages released to the stream during transitions from summer to winter and vice versa. Superimposed streamflow generation processes associated either with contrasting flow paths or contrasting water velocities may generate multimodal TTDs containing several age peaks. It was also shown that only multiple tracers used jointly such as ^{2}H and ^{3}H can help to decipher the multiple peaks in the TTDs and their long tails associated with old water. Accurate streamflow TTDs will soon be a vital concept for water policy makers in their efforts to curb water quality degradation. The more efficient and more accurate determination of unsteady

streamflow TTDs from isotopic tracers, and their improved parameterization proposed in this thesis pave the way for a holistic understanding of water flow paths and water quality in catchments.

Zusammenfassung

Wasser, das durch hydrologische Landschaften (wie Einzugsgebiete) fließt, benötigt unterschiedlich lange, um den nächsten Fluss zu erreichen. In anderen Worten: das Wasser im Fluss besteht aus einer Verteilung von Wasser verschiedenen Alters. Transitzeitverteilungen (engl.: Travel Time Distributions (TTDs)) charakterisieren die Transitzeiten des Wassers bis zum Abfluss aus einem Einzugsgebiet und beschreiben, wie Einzugsgebiete Wasser speichern und abgeben, das vor Tagen, Monaten und Jahren als Niederschlag gefallen ist. Kenntnisse über Transitzeitverteilungen sind entscheidend für ein besseres Wasserressourcenmanagement, weil sie Einblicke in Fließwege und –geschwindigkeiten in hydrologischen Systemen gewähren. Außerdem sind sie entscheidend bei der Steuerung hydrologischer Modelle zur Simulation von Wasserqualität. Transitzeitverteilungen in Flüssen werden über hydrologische Tracer, wie die stabilen Isotope von Sauerstoff (O) und Wasserstoff (H), abgeschätzt. Detaillierte Eigenschaften der Transitzeitverteilungen, wie deren Form, Statistik, und zeitliche Variabilität, sind nicht vollkommen entschlüsselt und die Faktoren, die diese Eigenschaften beeinflussen, sind nicht klar identifiziert. Darüber hinaus gibt es keine Rahmenvereinbarungen für die Schätzung instationärer Transitzeitverteilungen sowie deren Anwendung bei der Modellierung des Transports von Isotopentracern. Diese Einschränkungen haben ihren Ursprung zum Teil in der weit verbreiteten Verwendung einfacher stationärer Transitzeitverteilungsmodelle, die aus niedrig aufgelösten Tracerdaten der letzten Jahrzehnte abgeleitet wurden. Die meisten Transitzeit-Studien basieren auf zweiwöchentlichen oder monatlichen Daten stabiler Isotope, die meist über einen Zeitraum von weniger als zwei Jahren aufgenommen wurden. Dementsprechend wurde die zeitliche Varianz der Transitzeitverteilung in den meisten analytischen Transitzeitverteilungsmodellen, mit denen die Tracer Zeitreihen verglichen wurden, nicht berücksichtigt. Das alles überspannende Ziel dieser Arbeit ist es, zu verstehen, was die Form und zeitliche Variabilität von Transitzeitverteilungen beeinflusst und wie sie aus Tracerdaten abgeleitet werden können.

Dafür werden theoretische Untersuchungen, experimentelle Arbeit (Feld- und Laborarbeit), sowie Modellierung auf eine Weise kombiniert, die weit über bisherige Arbeiten hinausgeht. In luxemburgischen Untersuchungsgebieten wurden über zwei Jahre Isotopentracer (^{2}H, ^{18}O, ^{3}H) im Niederschlag und Abfluss in hoher Auflösung (mehrere Messungen pro Tag) gemessen. Basierend auf den theoretischen und experimentellen Grundlagen, insbesondere unter Berücksichtigung des in Luxemburg erhobenen Isotopendatensatzes, wurden verbesserte Parametrisierungen von Transitzeitverteilungen und neue analytische Modelle vorgeschlagen. Es wurde eine Methode zur Einschätzung von Informationen über das Alter des Wassers über einen „dual-isotopischen"-Ansatz (unter Verwendung von ^{2}H und ^{3}H) vorgeschlagen, um aufkommende Missverständnisse über die Grenzen der stabilen Isotope von O und H im Vergleich zu ^{3}H aufzuklären. Die Arbeit dieser Dissertation zeigt, dass die Transitverteilungen des Wassers vielfältigere Formen und komplexere Variabilitäten haben können, als in Studien des letzten Jahrzehnts angenommen wurde. Im mediterranen Klima kann es beispielsweise bei übergängen zwischen Sommer

und Winter zu komplexen Mustern des Wasseralters kommen. überlagerte Abflussbildungsprozesse unterschiedlicher Fließwege oder -geschwindigkeiten können multimodale Transitzeitverteilungen generieren, die verschiedene Altersspitzen enthalten. Darüber hinaus wird gezeigt, dass nur Tracer, die in Kombination verwendet werden, so wie ^{2}H und ^{3}H, helfen können, um multiple Spitzen der Transitzeitverteilungen und deren langen Auslaufkurven, die mit altem Wasser assoziiert werden, zu entschlüsseln. Akkurate Transitzeitverteilungen des Abflusses werden bald zu einem wesentlichen Konzept für die Bemühungen von Entscheidungsträgern im Wassermanagement werden, eine Minderung der Wasserqualität einzudämmen. Die effizientere und präzisere Bestimmung von instationären Transitzeitverteilungen des Abflusses über Isotopentracer und ihre in dieser Arbeit vorgeschlagene verbesserte Parametrisierung ebnen den Weg für ein ganzheitliches Verständnis von Wasserfließwegen und Wasserqualität in Einzugsgebieten.

Acknowledgements

Numerous people need to be thanked for helping me to achieve this PhD work:

Julian Klaus for his devoted supervision, his example as a scientist, and for creating a stimulating research environment. Laurent Pfister for managing the project and the research team well, creating one of the best places to do a PhD in the world. Erwin Zehe for his supervision and support, and for the intense, thought-provoking discussions about physics and math. The CAT group at LIST, for a dynamic, exciting, and friendly environment. The IWG group at KIT for the pleasant working environment through winter 2017–2018. Marta Antonelli, Barbara Glaser, and other former PhD students for their amazing friendship and for all the great moments shared together in the office and after work. Jérôme Juilleret pour ses encouragements: "Allez allez !". Jeff Iffly, Jérôme, Gwenael Carrer, Barbara, and Michael Schwab for help in the field. François Barnich and Gwen for their help with the isotopes in the lab. Paolo Benettin for exciting discussions on travel times and SAS functions and for allowing me to visit EPFL in winter 2016–2017. The state of Baden-Württemberg for the well-maintained and well-documented computational infrastructures that I could easily use. Fabrizio Fenicia, Gerrit Schoups, and Uwe Ehret for help with Matlab scripts. Benjamin Renard, Dmitri Kavetski, and Sun Chun for advice on Bayesian Inference and numerics. Uwe Ehret for fruitful discussions and logistical help during scientific events. Judith Meyer and Barbara for translating the abstract to German. Paolo Benettin, Jeff McDonnell, and Chris for their useful comments on the initial version of the thesis. I sincerely thank: the doctoral referee and co-referees Prof. Dr.-Ing. Erwin Zehe, Prof. Dr. Andrea Rinaldo, and Prof. Dr. Laurent Pfister for reading and evaluating my thesis, the other oral examination members Prof. Eiff, Prof. Uhlmann, and Prof. Horn for their assessment and positive feedback, and finally the chair Prof. Dr.-Ing. Dr. h.c. mult. Franz Nestmann for making my doctoral defense possible. Et finalement je remercie Maman, Papa, et Christophe pour leur soutien pendant ces 4 dernières années.

Contents

List of Figures

List of Tables

Introduction

Streams and rivers are ultimately made of water that originated as precipitation. A droplet of water starting as precipitation and travelling to the stream can take countless pathways and flow at various speeds, i.e. it can take more or less time and have different ages when leaving the catchment via streamflow. The central concept of this thesis is the distribution of water ages in the stream. Water ages are an important concept in hydrology and a subject undergoing intense research.

General motivation on the use of water ages in hydrology

Population growth and increasing human activities are threatening water quality worldwide (e.g. Zal et al. 2018). Hydrologists are needed now more than ever to understand how hydrological systems (essentially catchments) determine the quality of water they store and release. The role of hydrologists has traditionally been to estimate water quantity changes (e.g. floods and droughts). As a result, computational hydrology has often focused on estimating how fast a water perturbation (typically a storm) propagates through a catchment and how large the river response is. Hydrological models have thus been often validated against observed hydrographs, meaning that these models were accepted solely based on their ability to simulate water "celerity" (McDonnell and Beven 2014). Water quality management on the other hand requires an understanding of water "velocity" (McDonnell and Beven 2014). Concretely, water velocity allows to know for example how fast a contaminant plume in groundwater reaches a drinking well (Oudijk and Schmitt 2000), or the location of the contaminant source (Solomon et al. 1995), or how much contaminant is naturally degraded during its transport (Farlin et al. 2013). More useful hydrological models representing accurately water celerity and water velocity are therefore needed for better hydrological predictions and improved water management.

Water age is directly related to water velocity and water quality (Aggarwal 2002). Confronting water ages calculated by hydrological models against those deduced from hydrological tracer data can overcome the challenge of simulating water velocity and water celerity in these models. In the last two decades, conservative hydrological tracers measured in catchment outlets have thus been increasingly used to validate hydrological models against tracer-derived water ages (Birkel and Soulsby 2015; Fenicia et al. 2008; Son and Sivapalan 2007; Vaché and McDonnell 2006) and to gear these models towards water quality simulations. In catchments, water quality is based on where it flows (i.e. its flow paths), how much time it spends in contact with the surrounding media (i.e. its residence time), and how long it takes to reach an outlet (i.e. its travel time). This is particularly true e.g. for weathering products such as silica, sodium, and calcium (Benettin et al. 2015a; Maher 2011; Rinaldo and Marani 1987), and for nitrate and phosphorus (Darracq et al. 2008; Ehrhardt et al. 2018; Erostate et al. 2018; van der Velde et al. 2012). Thus, is now clear that

the distribution of water ages in catchment outlets has for long remained the missing component in hydrological models to simulate water quality (Ameli et al. 2017; Hrachowitz et al. 2016, 2015).

Definitions and uses of the water ages distributions

Different distributions of water ages exist in catchments and have been used ambiguously. Only recent studies started clearly distinguishing the distributions (Benettin et al. 2015c; Rigon et al. 2016). The residence time distribution (RTD) describes the ages of water still inside the catchment. It can be useful to indicate the time it took for water to recharge a given system, giving an idea about water renewability that can be easily communicated to non-specialists (Broad 2005; Kazemi et al. 2005). The travel time (or transit time) distribution (TTD) describes the water ages in a given outlet (e.g. streamflow Q and evapotranspiration ET). Besides their fundamental link to water quality, TTDs have gained more attention in hydrological research since ∼1950 (see Małoszewski and Zuber 1982; McGuire and McDonnell 2006; Stewart et al. 2010; Turner and Barnes 1998) for many reasons:

- TTDs are to conservative tracer fluxes what the Instantaneous Unit Hydrograph is to water fluxes, i.e. TTDs are the transfer function between tracer inputs (typically concentrations in precipitation) and tracer outputs (typically concentrations in the stream) (Heidbüchel et al. 2012). As such, TTDs can be efficient lumped models of transport (Rinaldo and Marani 1987), with potentially only about 1–4 parameters (Leray et al. 2016; Małoszewski and Zuber 1982; McGuire and McDonnell 2006).

- TTDs can summarize complex 3-dimensional flow paths in a single function of water age, enabling intuitive interpretations and visual representations of transport processes (e.g. Sayama and McDonnell 2009; van Huijgevoort et al. 2016).

- TTDs are excellent tools to extract information about catchments functions of storage and release from tracer input-output relationships. This is evident in the fact that TTDs are a common concept for various methods accessing this information from the tracers using different mathematical backgrounds (Cirpka et al. 2007; Kirchner 2019; Kirchner et al. 2000; Massoudieh et al. 2012; McGuire and McDonnell 2006; Turner and Macpherson 1990a). Catchment storage is a very difficult information to access. Yet, it can been simply deduced from TTDs because of the fundamental link between storage, and the mean travel time and the total amount of water flowing through a catchment (Birkel et al. 2015; McGuire and McDonnell 2006; Pfister et al. 2017a; Soulsby et al. 2009).

- Travel time measures have been shown to correlate with various catchment characteristics (Broxton et al. 2009; Capell et al. 2012; Dunn et al. 2007; Hrachowitz et al. 2009a, 2010) making them a potentially meaningful criteria to use in a catchment classification system (Wagener et al. 2007).

- Travel times are now also used to understand water fluxes between hydrological compartments (Sprenger et al. 2019), in particular the complex soil-plant-atmosphere water interactions (Cain et al. 2019; Evaristo et al. 2019; Knighton et al. 2019; Sprenger et al. 2018).

- The research interest in TTDs is wider and is growing faster nowadays due to the potential of travel times for connecting several research communities of the critical zone (i.e. the permeable layer from the tops of the trees to the bottom of the groundwater) (Hrachowitz et al. 2016; Sprenger et al. 2019).

Unlike TTDs, the catchment RTD cannot be simply deduced from hydrological tracers because it is currently impossible to take tracer samples that represent all the water stored in catchments. Similarly, the TTD of ET is difficult to estimate because the absence of a single outlet means that representative tracer samples cannot be easily gathered. Thus, the streamflow TTD has been more often investigated, and it is the focus of this thesis.

Research needs on streamflow travel times tackled in this thesis

Research on streamflow travel times is needed because of the still incomplete understanding of the characteristics of the TTD (e.g. shape, statistics, variability) (Kirchner et al. 2000; Leray et al. 2016; McDonnell et al. 2010; McGuire and McDonnell 2006; Rinaldo et al. 2015) and of the factors (e.g. climate, catchment characteristics) controlling them (Dunn et al. 2007; Hale and McDonnell 2016; Heidbüchel et al. 2013; Hrachowitz et al. 2010; McGuire et al. 2005; Remondi et al. 2019; Tetzlaff et al. 2009; von Freyberg et al. 2018). Moreover, there is no consensus on how to infer TTDs from hydrological tracers nor on how to model the tracers with TTDs (McDonnell et al. 2010; Stewart et al. 2010). These research needs have been even more pronounced since the seminal work of Botter et al. (2011) and van der Velde et al. (2012) showing that streamflow TTDs are transient and can take a multitude of shapes at different times. For instance, considering a catchment, determining its streamflow TTD will require answering the following questions, for which no general solution is available: What is the expected shape of the TTD? Does it vary strongly with time? If yes, how does it vary? What climatic or geographic factors influence this variability? What tracers should be used and how should they be sampled to test hypotheses about the TTD (e.g. shape, controls on its variability)? How can these hypotheses be tested in a modelling framework taking into account unsteady flow conditions?

To answer this, three connected research avenues about streamflow travel times are explored in this thesis: theoretical investigations on the mathematical properties (e.g. expected shape and time variance) of the distributions [chapters 1, 2 & 3]; experimental work to choose where, when, how, and what tracers to sample and measure [chapters 2 & 4]; and modelling work to make the best joint use of the tracers and models to infer the TTDs [all chapters]. The overarching question that my thesis addresses is: "What controls the shape and time variability of the streamflow TTD and how can it be determined from tracer data?"

Modeling streamflow travel time variability using catchment storage

One of the main steps forward in catchment travel time theory in the past decade has been the acknowledgement of the time variance of TTDs. This has resulted in a novel equation expressing the co-evolution of the distributions of water ages in storage (RTD) and in outflows (TTDs) (Botter et al. 2011; van der Velde et al. 2012). This breakthrough was motivated by observations of stream tracer concentrations

that could not be explained by steady-state TTDs (Hrachowitz et al. 2009b; McGuire et al. 2007; van der Velde et al. 2010), and by the long-standing need to consider the time variance of hydrological fluxes in travel time theory (Botter et al. 2010; Duffy 2010; Rinaldo et al. 2011). StorAge Selection (SAS) functions were then proposed to formally link TTDs and the RTD in catchments (Botter et al. 2011; Rinaldo et al. 2015). Subsequent studies showed a clear dependence of travel times on catchment wetness often represented through the modelled catchment storage (Benettin et al. 2015a; Davies et al. 2013; Heidbüchel et al. 2012; Hrachowitz et al. 2013; Klaus et al. 2015a; McMillan et al. 2012b; Soulsby et al. 2015a; van der Velde et al. 2012). It was suggested that in general, a larger fraction of the younger stored water is released during wetter catchment conditions, a concept called "Inverse Storage Effect" and formally defined using SAS functions (ISE, Benettin et al. 2017b; Harman 2015; Kim et al. 2016; Pangle et al. 2017; Wilusz et al. 2017). This concept was proposed and initially tested in sites with low seasonality of precipitation and evapotranspiration and thus small storage changes through the year. The ISE has gained rapid momentum because it allows modeling the streamflow TTD by simply parameterizing it with catchment modeled storage. However, it is unlikely that the variability of the streamflow TTD is explained only by storage variations in all catchments. A clear knowledge gap that needs to be addressed is therefore the validity of the ISE in catchments with different hydrometeorological conditions than those where the ISE was initially proposed and tested.

In the first chapter of this thesis, I present a study of the relationship between catchment storage and travel times using SAS functions in a Mediterranean climate, where storage changes are large because of the pronounced seasonality of hydrological forcings. I examined the validity of the ISE in a well-investigated catchment (WS10) of the H.J. Andrews forest in Oregon and discussed the consequences of strong climate seasonality for the parameterization of the streamflow SAS function hence for the modelling of the streamflow TTD in general. For this, I developed a conceptual model of WS10 and I calibrated it against hydrographs and ^{18}O data. The streamflow TTD and SAS function were directly derived from the calibrated model. I then juxtaposed the shapes (and variability) of the streamflow SAS function and the previous findings on streamflow generation processes to show the relevance of SAS functions for efficiently summarizing time-varying transport processes to the stream in a lumped manner.

Expanding travel time theory towards water age multimodality

One of the most difficult aspects of modelling the transport of tracers or solutes with travel times is matching a theoretical TTD with the transport processes in the catchment (McGuire and McDonnell 2006). Considerable efforts are thus devoted in travel time research to investigate the theoretical shapes of TTDs and relate them to mechanisms of transport such as advection and dispersion in the subsurface, groundwater level dynamics, and preferential flow (e.g. Davies et al. 2013; Kirchner et al. 2001; Lindgren et al. 2004; Pangle et al. 2017). Most travel time studies nevertheless assume a simple parametric form for the streamflow TTD and calibrate its parameters using tracer data (McGuire and McDonnell 2006; Stewart et al. 2010). Nonparametric approaches have been more rare (e.g. Cirpka et al. 2007; Kirchner 2019; Massoudieh et al. 2012). The probability distribution functions commonly used for the TTDs were initially proposed for groundwater transport (Małoszewski and

Zuber 1982) and represent various assumptions about the flow field and the boundary conditions that are rarely true in catchments (Leray et al. 2016). Another analytical form, the gamma distribution, was suggested for TTDs in order to explain the spectral slopes of stream chemistry deduced from recent high-frequency sampling campaigns (Aubert et al. 2013; Godsey et al. 2010; Kirchner et al. 2001; Kirchner and Neal 2013; Kirchner et al. 2000). These simple and often unimodal models of TTD used to simulate tracers observations in the stream are however not consistent with the various studies (e.g. Cirpka et al. 2007; Eberts et al. 2012; Goderniaux et al. 2013; Haggerty et al. 2002; Long and Putnam 2009; Stolp et al. 2010) suggesting that true TTDs are more likely to have several peaks associated with different lengths of transport and/or different transport velocities (see other examples in Leray et al. 2016, pp. 78–80). Recent studies showed more complex shapes of TTDs or SAS functions (e.g. with many peaks) from catchment conceptual models calibrated to tracer data (Benettin et al. 2015b, 2013b; Hrachowitz et al. 2013; McMillan et al. 2012b; Rodriguez et al. 2018; van der Velde et al. 2015) or from tracking particles in physically-based distributed catchment models (Danesh-Yazdi et al. 2018; Davies et al. 2013; Kaandorp et al. 2018; Maxwell et al. 2018; Yang et al. 2018). Further research is vitally needed to understand the origin of water age multimodality and to propose ways to account for it in theoretical TTD and SAS functions models. Moreover, we still ignore the consequences of neglecting peaks in the TTDs or SAS functions used for water chemistry simulations. This is particularly concerning since TTDs are seen as the link between hydrology and biogeochemistry.

Thus, in the second and third chapters of this thesis, I present two studies addressing these issues related to representing at the catchment scale the relevant processes for solute transport in a lumped manner with TTDs (or SAS functions). The first study describes the theoretical development of composite SAS functions for an experimental catchment in Luxembourg (Weierbach). Composite SAS functions are written as a weighted sum of probability distribution functions. The following study explores the origins of water age multimodality, and shows two additional examples about the usefulness of composite SAS functions for tracer and especially for water quality simulations in complex hydrological systems.

A dual-tracer experimental approach to determine the entire streamflow TTD

Determining TTDs requires experimental work to collect and analyse tracer samples. Isotopic tracers have become the cornerstone to estimate TTDs in catchments. Stable isotopes of O and H (^{18}O and ^{2}H) and tritium (^{3}H) have been the most commonly used tracers because they are part of the water molecule itself and are thus considered more ideal than other tracers (Kendall and McDonnell 1998). Nowadays, ^{18}O and ^{2}H are much more used than ^{3}H because technological progress made their sampling easier (e.g. Berman et al. 2009; Herbstritt et al. 2012; Koehler and Wassenaar 2011; Pangle et al. 2013) and analyses cheaper (Gupta et al. 2009; Lis et al. 2008). Stewart et al. (2010) showed that mean travel times (MTT) estimated from stable isotopes and tritium differ and argued that this is because only tritium is able to detect old water signatures in the stream thanks to its radioactive decay. The term "truncation" of TTD tails with stable isotopes was then coined by Stewart et al. (2012) to warn the research community that tritium is strictly required to infer the long tails of TTDs and avoid a vision of TTDs generally biased towards young water. However, a complete understanding of the underlying causes for such differences

between MTTs deduced from (^{18}O, ^{2}H) and from ^{3}H has been hindered by limitations in tracer observations (e.g. baseflow sampling only), incoherent mathematical frameworks, and incorrect modelling assumptions (e.g. steady state). Furthermore, the true information contents of stable isotopes and tritium regarding stream water ages are not well known and require further investigation.

In the last chapter of this thesis, I present study relying on the experimental work in Luxembourg and describing how useful both stable isotopes and tritium are to estimate the whole streamflow TTD. I collected samples for two years in three Luxembourgish experimental catchments and analysed them for stable isotopes and tritium in the laboratory. In this study I first tested the hypothesis of truncation of streamflow TTD tails in the Weierbach catchment (one of the three experimental catchments) by comparing TTDs obtained using only stable isotope measurements and TTDs obtained using only tritium measurements. I deduced catchment storage from the TTDs to test if the potential truncation of TTDs yields underestimates of storage. I also calculated the age information contents of the isotopic tracers to provide guidelines on their sampling for future experimental studies aiming at deciphering the entire TTDs with both tracers.

Chapter 1

Modeling streamflow travel time variability using catchment storage

An edited version of this paper was published by Wiley[1] as:

N. B. Rodriguez, K. J. McGuire, and J. Klaus (2018). "Time-Varying Storage-Water Age Relationships in a Catchment With a Mediterranean Climate". In: *Water Resources Research* 54.6, pp. 3988–4008. DOI: 10.1029/2017wr021964

Key points

- Storage-water age relationships generally correspond to the inverse storage effect in a catchment exposed to a Mediterranean climate.
- The marked seasonal climate at this study site caused short-term deviations from the inverse storage effect.
- Streamflow SAS functions were a consistent lumped description of rainfall-runoff processes

Abstract

Recent studies on the relationships between catchment storage and water ages using Travel Time Distributions (TTDs), Residence Time Distributions (RTDs), and StorAge Selection (SAS) functions have led to the hypothesis that streamflow preferentially mobilizes younger water when catchment storage is high. This so-called "Inverse Storage Effect" (ISE) needs further evaluation in more catchments with diverse climates and physiographical features. In this work, we assessed the validity of the ISE in WS10 (H. J. Andrews forest, Oregon, USA), a forested headwater catchment in a Mediterranean climate. A conceptual model of the catchment, developed based on experimental observations of water flow paths in WS10, was calibrated to streamflow and $\delta^{18}O$ in streamflow. Based on the calibrated model results, we determined RTDs, and streamflow TTDs and SAS functions by assuming that the soil reservoir and the groundwater reservoir act as well-mixed systems. The streamflow SAS functions and travel time dynamics medskip that the ISE generally applies in WS10. Yet, during transitions from dry summer periods to wet winter periods and vice versa, the marked seasonal climate caused rapid and strong storage variations in the catchment, which led to deviations from the ISE. The seasonality of streamflow travel times in WS10 is the result of the seasonal contributions of younger water from the hillslopes added to the rather constant groundwater contributions of older water. The streamflow SAS functions were able to capture the relative importance of contrasting flow paths in the soils and in the bedrock highlighted by previous studies in WS10.

1.1 Introduction

The chemical composition of water in catchment storage and discharge is largely governed by the contact times between water and catchment material, i.e., by the age of water. The distributions of water ages in catchment storage (residence time distributions, RTDs) and in catchment outflows (travel time distributions, TTDs) have been used previously to describe how catchments store and release water and solutes (Benettin et al. 2015a, 2017b; Godsey et al. 2010; Hrachowitz et al. 2016; Kirchner et al. 2000; Maher 2010, 2011; McGuire and McDonnell 2006; Rinaldo and Marani 1987; van der Velde et al. 2012). A recent study gathered stable isotope tracer data from 254 catchments and suggested that, on a global scale, about a third of the streamflow volume consists of young water from recent precipitation (Jasechko et al. 2016). In contrast, the remaining two-thirds of the streamflow volume comes from much older water mobilized from catchment storage (Jasechko et al. 2016). Streamflow TTDs and catchment RTDs appear to be very different, which shows that streamflow is not a uniform mixture of all the water stored in a catchment, but represents only a fraction of the storage at any time. This perception is supported by detailed investigations of tracer transport in catchments using lumped conceptual models (e.g., Birkel et al. 2011b; Fenicia et al. 2010; Hrachowitz et al. 2013; Hrachowitz et al. 2015; McMillan et al. 2012b). In these studies, mixing coefficients were used within conceptual control volumes to represent how incoming rainfall mixes with the stored water and is eventually released by outflows. This suggested that a complete mixing of the entire catchment volume is inadequate to simulate tracer concentrations of outflows in many catchments. Furthermore, these studies medskip that catchment TTDs and RTDs are highly time-variant.

Other recent work used StorAge Selection (SAS) functions, defined as the ratio of TTD and RTD, for relating water ages in storage and outflows without assuming stationarity (Benettin et al. 2015c; Botter 2012; Botter et al. 2011; Rinaldo et al. 2015; van der Velde et al. 2012). SAS functions are a lumped characterization of the mixing of various water ages in storage when they are released as discharge or evapotranspiration. Thus, SAS functions provide a theoretically consistent lumped description of water and solute transport in catchments, in which TTDs and RTDs are allowed to differ from each other and vary individually in time (Rinaldo et al. 2015). Therefore, SAS functions are seen as a promising tool to better understand the complexity of streamflow generation processes, calculate time-varying TTDs and RTDs, and simulate water and solute transport in catchments (Benettin et al. 2015a, 2017b; Harman 2015; van der Velde et al. 2015).

Studies with SAS functions in catchments with strong oceanic climates, e.g., in the Upper Hafren catchment in Wales (Benettin et al. 2015b; Harman 2015), and in the Bruntland Burn catchment in Scotland (Benettin et al. 2017b), supported the emerging perception that younger water stored in the catchments is preferentially mobilized to generate streamflow when catchment storage is high. This was termed "Inverse Storage Effect" (ISE) by Harman (2015). The ISE was also shown to be relevant in a laboratory-scale experiment, where breakthrough curves of irrigation water tagged with different tracers on a sloping lysimeter were used to examine transport processes (Kim et al. 2016). SAS functions, calculated from tracking virtual particles with a physically based model of the same sloping lysimeter, medskip a similar dependence on the system wetness (Pangle et al. 2017) to what was found in the Upper Hafren and the Bruntland Burn catchments. Other studies have shown that streamflow travel times tend to decrease for wetter catchment states, which points to an ISE-like behavior (Botter et al. 2010; Heidbüchel et al. 2012; Hrachowitz et al. 2013; Hrachowitz et al. 2015; Klaus et al. 2015a; Rinaldo et al. 2011; Segura et al. 2012; Soulsby et al. 2015a; van der Velde et al. 2015). Until now, the ISE phenomenon has been observed mostly in systems with limited seasonality for precipitation, evapotranspiration, and streamflow, and where evapotranspiration is rather low. Calculating time-varying streamflow SAS functions and travel times in catchments with diverse climates and physiographical features is essential for determining the generality of the ISE.

Physically based particle tracking approaches recently estimated nonparametric SAS functions and linked them to flow paths on small scale (Pangle et al. 2017) and in virtual media (Danesh-Yazdi et al. 2018). Yet, catchment scale approaches for tracer transport and travel time estimates commonly rely on lumped representations based on parameterized SAS functions. At the catchment scale, SAS functions are usually assumed to be distinct functions of one-to-three parameters calibrated with tracer data (e.g., gamma, beta, and power law functions) (Rinaldo et al. 2015). To be meaningful, this assumption of the shape of SAS functions and their parameters needs to be compared to observed runoff generation and solute transport processes in catchments. Establishing a clearer link between time-varying SAS functions and travel times at the catchment scale, and observed physical processes will help hydrologists develop and justify lumped modeling approaches based on travel times. This is particularly important since lumped catchment models that employ time-varying travel times are seen as a unifying modeling framework for the currently separate catchment hydrology and water quality scientific communities (Hrachowitz et al. 2016). Well-researched experimental catchments provide the opportunity to link travel time theory to hydrological processes.

In this study, we designed and calibrated a conceptual catchment flow and transport model that allowed us to solve the Master Equation (Botter et al. 2011) and derive time-varying RTDs, and streamflow TTDs and SAS functions in the well-studied catchment WS10 in the H. J. Andrews Experimental Forest in Oregon, USA. We assumed that the modeled soil reservoir and groundwater reservoir act as well-mixed systems. The novelty of this work lies in the pronounced seasonality of the Mediterranean climate of the study site allowing the direct examination of the ISE. Moreover, the large number of previous studies on rainfall-runoff in WS10 allows a benchmarking of streamflow SAS functions against physical processes in the catchment. The present study aims to shed light on the following research questions:

1. Is the inverse storage effect a characteristic feature in the Mediterranean climate?

2. Are time-varying streamflow SAS functions and travel times consistent with previous detailed investigations of flow paths in an experimental catchment?

1.2 Materials and Methods

1.2.1 Study site

This study is carried out in WS10, a forested headwater catchment (10.2 ha) located in the H. J. Andrews forest (HJA) in the central Western Cascades of Oregon, USA (44.28°N, 122.258°W) (Figure 1.1). The climate is Mediterranean, with a strong contrast between the dry summers and the wet winters (Greenland 1994). Annual precipitation averages 2,200 mm (1990–2003), with 80% of precipitation falling predominantly as rainfall between October and April. Minor snow accumulation ($<$ 30 cm) is not uncommon in winter at WS10, but it rarely lasts longer than a few weeks, and on average melts within a few days (Harr and McCorison 1979; Sollins and McCorison 1981). During the isotopic sampling for this study, no major snow accumulation was observed in WS10 (McGuire and McDonnell 2010; McGuire et al. 2007). We therefore neglected snow processes for the purpose of this study, consistent with previous studies using the same data set (Klaus et al. 2015a; McGuire et al. 2005; McGuire et al. 2007). The mean annual runoff ratio in WS10 is 0.60 over February 2001–2003. Evapotranspiration varies seasonally between less than 1 mm/d in winter and 5 mm/d in summer (Barnard et al. 2010; McGuire et al. 2007). This creates a distinct seasonality in catchment streamflow with summer low flows commonly below 0.01 mm/h and peak flows that can reach 6 mm/h during winter storms of moderate intensity.

WS10 mostly consists of short ($<$ 200 m) and steep (30–45°) hillslopes, with elevations ranging from 473 to 680 m.a.s.l. The catchment is covered by Douglas-fir trees (Pseudotsuga menziesii), which regenerated naturally following a clearcut in 1975. Soil depth varies from 1.5 to 4.2 m, with an average of 3 m (Sayama and McDonnell 2009). Soils are identified as gravelly clay loams at the surface, as gravelly silty clay loams or clay loams in lower layers, and as gravelly clays or clay loams in the subsoils (Harr 1977). Soil total porosity is 60% on average, and varies little with depth. The soils in WS10 are highly conductive (saturated hydraulic conductivities well above 300 cm/h at the surface) due to their aggregated structure and their high drainable porosity, which in average declines with depth from 23% at the surface to 5% at 1.5 m depth (Dyrness 1969; Harr 1977). No overland flow has been

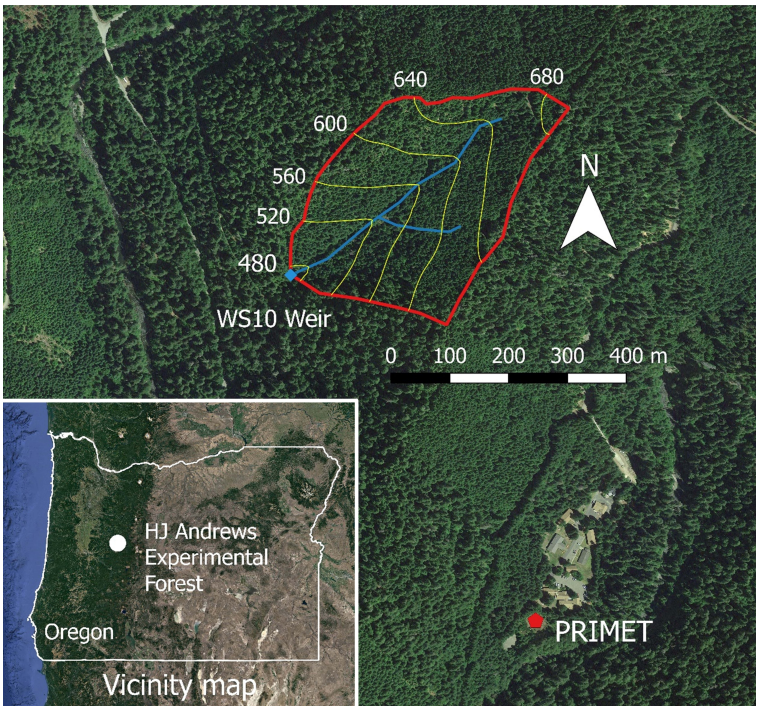

FIGURE 1.1: Map of WS10 in the HJA, showing the location of the PRIMET meteorological station nearby, and the location of HJA in Oregon, USA. The outlet weir is located at coordinates 44.21698°N and 122.26118°W. Yellow lines indicate the elevation in meters above sea level. Background map data: Google.

observed in WS10 (McGuire et al. 2005). Moreover, preferential flow strongly influences runoff generation, as revealed by the rapid transport of tracers applied on one of the WS10 hillslopes (McGuire and McDonnell 2010; McGuire et al. 2007). WS10 is underlain by volcanic bedrock including andesitic and dacitic tuff and coarse breccia (Swanson and James 1967), which is considered mostly impermeable except in weathered and highly fractured areas that can rapidly transfer groundwater reaching the regolith-bedrock interface laterally (Gabrielli et al. 2012; van Verseveld et al. 2017).Poorly permeable, 1–7 m thick partially weathered saprolite underlies the soils (Harr and McCorison 1979).Unweathered bedrock is found at 0.4–0.6 m depth on the stream-hillslope interface and at 3–8 m depth at the catchment divide. Debris flows occur periodically in WS10 (e.g., 1986, 1996), and maintain a narrow and steep stream channel with apparent bedrock on 60% of its length upstream from the outlet.

McGuire and McDonnell (2010) summarized the current understanding of water flow paths and streamflow travel times in WS10 in a perceptual model. Different streamflow generation mechanisms are activated sequentially depending on the wetness state: dry state in summer (June to October), transition states in fall (October to December) and in spring (April to June), and wet state in winter (December to April). During dry periods, groundwater above the soil-bedrock interface provides base flow with an age of 1–2 years. During wet-up periods in fall, water from rainfall events moves vertically in the soils as unsaturated flow, recharging gradually the deeper soils in the hillslopes, as suggested by stable isotope patterns at different soil depths (Brooks et al. 2010). Water from rainfall events can also reach the stream in less than 2 days by preferential flow. Vertical preferential flow occurs in the hillslope soils while lateral preferential flow occurs in the fractured bedrock at the soil-bedrock interface, as derived from tracer experiments (McGuire et al. 2007;

van Verseveld et al. 2017), and hydrogeological observations (Gabrielli et al. 2012). After the wet-up period, soil saturation expands upslope (Harr 1977) and mobilizes the soil water recharged over the last month, creating lateral saturated flow. This may cause a seasonality of streamflow travel times in WS10 with the onset of connectivity between the hillslopes and the stream. McGuire and McDonnell (2010) medskip that the connection and disconnection of hillslopes from the stream creates hysteresis in the hillslope-catchment discharge relationship. Hillslopes provide water that is contrastingly younger than the groundwater from the bedrock (Gabrielli et al. 2012).

1.2.2 Data set

Precipitation and air temperature were recorded at the PRIMET meteorological station located less than 1 km away from the WS10 gauge, at a similar elevation (436 m.a.s.l.). Potential evapotranspiration (PET) was calculated with Thornthwaite's method (Thornthwaite 1948), based on PRIMET air temperature measured 150 cm above ground. Streamflow data were converted with a rating curve from water levels measured at 15 min intervals (Rothacher 2016). We used a 2 year time series of ^{18}O in precipitation and catchment streamflow spanning from January 2001 to February 2003 and described by McGuire et al. (2005) (Figure 1.2). Precipitation was sampled as weekly bulk samples at the PRIMET station. Additionally, sequential rainfall samples (4.4 mm increments, i.e., on average, one sample every 20 h) were collected in WS10 during fall and winter 2002 McGuire and McDonnell (2010). We interpolated an hourly ^{18}O signature in precipitation by assuming that the precipitation $\delta^{18}O$ value between two consecutive samples had the value of the last sample. Streamflow was manually sampled at weekly intervals, including periods (a few days to a week) of subhourly to subdaily automatic sampling. When several stream samples were taken in less than 1 h, a flow-weighted value was calculated for the hourly interval. The ^{18}O analysis procedure is described in McGuire et al. (2005).

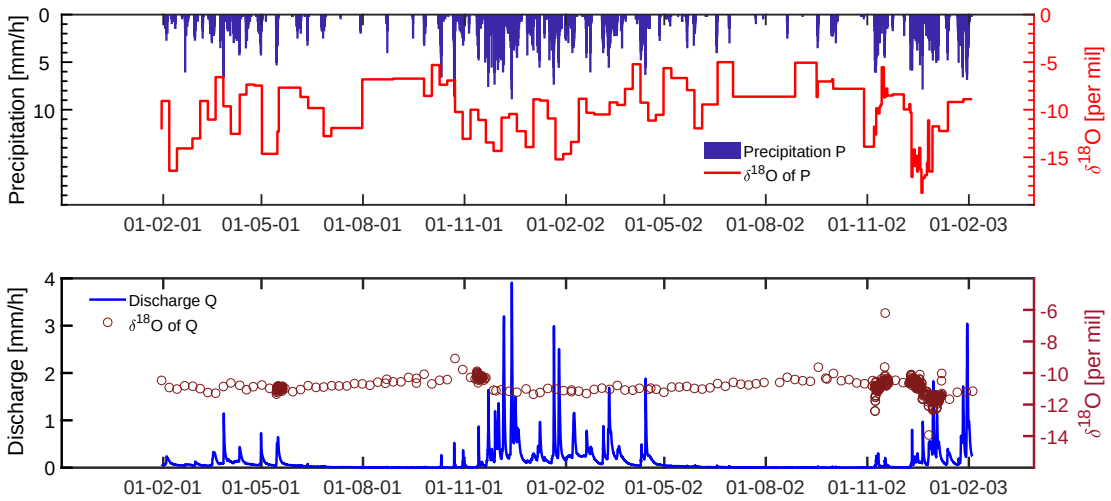

FIGURE 1.2: (top) Precipitation rates and precipitation $\delta^{18}O$ values used for this study. (bottom) Discharge in WS10 outlet and streamflow $\delta^{18}O$ values used for this study.

1.2.3 Catchment model

A conceptual catchment model was employed to simulate water storage and water fluxes in WS10 (Figure 1.3), as well as the transport of ^{18}O (or any other tracer) (Table 1.2). A total of nine parameters were needed (Table 1.1) and determined from calibration to observed data (see section 1.4 for details on calibration). We want to stress that the main purpose of the model was to provide physically meaningful estimates of the water storage indifferent compartments (i.e., soil water and groundwater) and of the fluxes from and between them. The model results were used to derive time-varying RTDs, and streamflow TTDs and SAS functions (see section for 1.5 details on how these were determined). We used the successful model realizations as the best approximation of the catchment and its internal flow paths. This approach was previously described and used for example by Benettin et al. (2015b), who successfully applied a similar model structure to the Upper Hafren catchment in Plynlimon, Wales. In this study, we justified the validity of the approach and of the reliability of its estimates of water storage and fluxes by comparing the model-simulated streamflow and $\delta^{18}O$ in streamflow to the observed data in WS10.

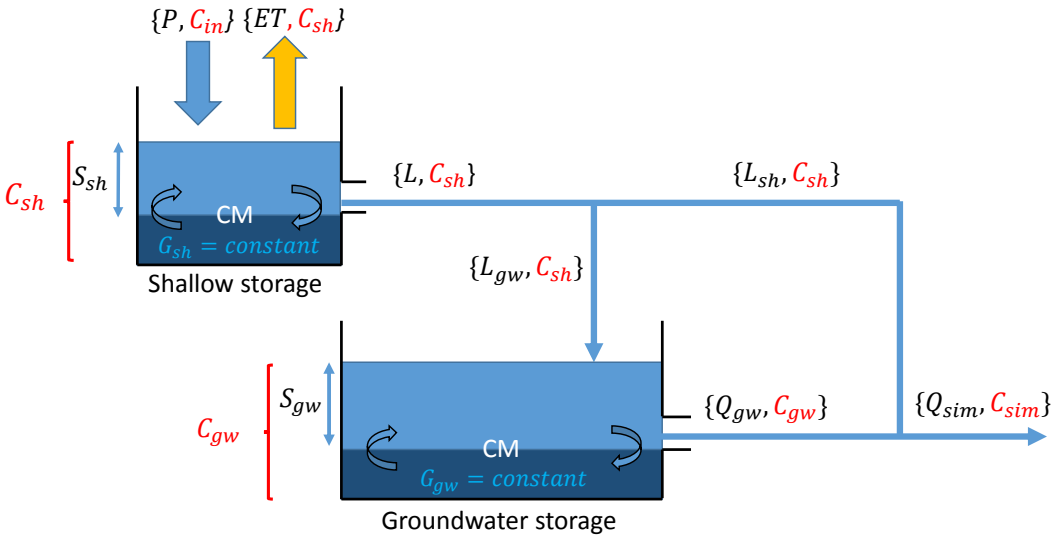

FIGURE 1.3: Conceptual catchment model employed for WS10 in this study. The inflows and outflows of each reservoir are indicated in pairs flux, concentration, representing the water fluxes and their associated tracer concentration in red font (i.e., $\delta^{18}O$ values). The passive storage volumes are pictured in dark blue. CM stands for Complete Mixing (see section 1.2.3).

The model structure was chosen because it is consistent with the experimental observations of the water flow paths in WS10 summarized by McGuire and McDonnell (2010). These flow paths consist of groundwater flow in the bedrock, and saturated and unsaturated flow in the hillslope soils during the wet season. The model uses two reservoirs connected in a series (one after the other) and in parallel (the first reservoir can bypass the second) (Figure 1.3). The upper reservoir represents the storage of water in the soils and subsoils in the hillslopes (called "shallow storage"), which is able to generate a rapid response to precipitation during wetter periods. The lower reservoir represents the storage of water in the weathered and unweathered bedrock(called "groundwater storage"), which provides base flow and remains active during the prolonged dry periods in summer.

We distinguished two forms of water storage in each reservoir. Active storage can freely drain by gravity and is directly available to generate discharge. In the

TABLE 1.1: Model Parameters.

Name	Symbol	Unit	Description
ET magnifier	Ψ	(–)	Coefficient increasing the magnitude of ET
Reference partition coefficient	λ_0	(–)	Maximum fraction of streamflow made of direct contributions from the shallow storage reservoir
Discharge coefficient	μ_Q	mm/h	Magnitude coefficient in the storage-discharge relationships
Shallow storage exponent	β_{sh}	(–)	Shape coefficient of the shallow reservoir storage-discharge relationship
Groundwater storage exponent	β_{gw}	(–)	Shape coefficient of the groundwater reservoir storage-discharge relationship
Shallow storage capacity	D_{sh}	mm	Capacity of the shallow storage reservoir
Groundwater storage capacity[a]	D_{gw}	mm	Capacity of the groundwater storage reservoir
Shallow passive storage	G_{sh}	mm	Amount of water contributing only to mixing with the shallow storage reservoir
Groundwater passive storage	G_{gw}	mm	Amount of water contributing only to mixing with the groundwater storage reservoir

[a]This parameter was treated differently in the calibration. See section 1.4 for details.

TABLE 1.2: Hydrological and Transport Equations Used in the Model.

State variable	Units	Balance Equation[a]	Constitutive functions[b]
S_{sh}	mm	$\frac{dS_{sh}}{dt} = P(t) - ET(t) - L(t)$	$L(t) = \mu_Q\, s(t)^{\beta_{sh}}$ $L_{sh} = \lambda_0\, s(t)\, L(t)$ $L_{gw} = (1 - \lambda_0\, s(t))\, L(t)$
S_{gw}	mm	$\frac{dS_{gw}}{dt} = L_{gw}(t) - Q_{gw}(t)$	$Q_{gw}(t) = \mu_Q \left(\frac{S_{gw}(t)}{D_{gw}}\right)^{\beta_{gw}}$
M_{sh}	mm‰	$\frac{dM_{sh}}{dt} = P(t)\, C_{in}(t) - (ET(t) + L(t))\, C_{sh}(t)$	$C_{sh}(t) = \frac{M_{sh}(t)}{S_{sh}(t)+G_{sh}}$
M_{gw}	mm‰	$\frac{dM_{gw}}{dt} = L_{gw}(t)\, C_{sh}(t) - Q_{gw}(t)\, C_{gw}(t)$	$C_{gw}(t) = \frac{M_{gw}(t)}{S_{gw}(t)+G_{gw}}$

[a]See Figure 1.3 for a visual interpretation of the balance equations. See section 1.3 for details on the equations and on the terms used in the equations.
[b] $s(t) = S_{sh}(t)/(n \times D_{sh})$, see section 1.3.

shallow storage reservoir, active storage is the state variable S_{sh}, while this variable is called S_{gw} in the groundwater storage reservoir (Figure 1.3). Passive storage (parameters G_{sh} and G_{sh}, Table 1.1 and Figure 1.3) does not modify the hydrological response of the reservoir but causes a dampening of the tracer response by mixing. It is essentially as if the passive storage was exchanging water molecules with the active storage (and incoming precipitation) at an infinite rate. We assumed constant passive storage, as is common in hydrological transport studies employing conceptual models (Benettin et al. 2015a; Fenicia et al. 2010; Hrachowitz et al. 2013). Thus, the hydrological storage-discharge equations only involve the active storage, while ^{18}O tracer transport equations use both the active and the passive storage (see Table 1.2 for equation details).

In this study, we used the Complete Mixing (CM) assumption that considers that each modeled reservoir acts as a fully-mixed system. This means that for each reservoir, all outflows have a $\delta^{18}O$ concentration equal to the mean reservoir $\delta^{18}O$, which involves the total reservoir storage (active + passive). The CM assumption implies that the distributions of water ages in the reservoir and each of its outflows are the same, which is known as Random Sampling (RS) (see section 1.5). The simulated tracer response of the catchment was governed by the relative contributions of the shallow and the groundwater storage to streamflow. Although each reservoir is a fully mixed system, the overall system is not fully mixed, which is consistent with previous findings for WS10 (Klaus et al. 2015a). This is because each reservoir generally has a different $\delta^{18}O$, a different RTD, and contributes in varying degrees to

streamflow (Benettin et al. 2015b).

1.3 Hydrological and Transport Equations

Simulations of water and tracer mass used four state variables. The first two state variables were shallow reservoir active storage S_{sh} and shallow reservoir δ^{18}O mass M_{sh}. The latter was taken as the product of total shallow storage ($S_{sh} + G_{sh}$) and δ^{18}O concentration in the shallow storage C_{sh}. The last two state variables were groundwater reservoir active storage S_{gw} and groundwater reservoir δ^{18}O mass M_{gw}. The latter was taken as the product of the total groundwater storage ($S_{gw} + G_{gw}$) and δ^{18}O concentration in the groundwater storage C_{gw}. These four state variables changed according to their balance equation involving model forcings: measured precipitation rates P and calculated evapotranspiration rates ET; and constitutive flux functions with their parameters (Tables 1.1 and 1.2). Some constitutive functions used the variable $s(t)$ that we defined as the active shallow storage normalized by the maximum value it can take, as in Benettin et al. (2015b) (equation 1.1):

$$s(t) = \frac{S_{sh}(t)}{n \times D_{sh}} \tag{1.1}$$

where D_{sh} is shallow storage capacity (mm), a calibrated parameter, and n is the model soil drainable porosity, a fixed parameter. This is consistent with our description of the model reservoirs, where active storage is at most the volume of water that can freely drain by gravity. We used a constant n value equal to the upper limit of drainable porosity in WS10, which is $n = 23\%$ in the upper soil (Harr 1977). Actual ET was calculated from PET in equation 1.2:

$$ET(t) = \Psi \frac{s(t)\,(1+m)}{s(t)+m}\, PET(t) \tag{1.2}$$

where Ψ is a calibrated parameter that enhances the seasonal amplitudes of ET and allows a good agreement between ET and the long-term average difference $P - Q$ (1980–2003). $m = 0.5$ is a fixed smoothing parameter used to ensure numerical stability of the solution by quickly reducing ET as water availability in soils (represented by $s(t)$) decreases (c.f. Fenicia et al. 2016). Discharge from each reservoir was formulated as a power function of the reservoir storage normalized by the maximum value it can take (Table 1.2): $Q(t) = \mu_Q \left(\frac{S(t)}{\max S}\right)^{\beta}$ where β is a dimensionless shape coefficient. For the shallow storage, the normalized storage $s(t)$ (Equation 1.1) was used, whereas for the groundwater storage we divided the active storage S_{gw} by calibrated parameter D_{gw} representing the groundwater active storage capacity in mm of water (Table 1.1). The magnitude coefficient μ_Q is an upper bound of the discharge from the reservoirs, with the same units as the observed streamflow (mm/h). μ_Q acts as a scaling factor for the storage-discharge relationships, depending on the average value of the ratio $\left(\frac{S(t)}{\max S}\right)^{\beta}$. We used μ_Q for the shallow and the groundwater storages to limit the number of parameters (Benettin et al. 2015b). The normalization of the storage-discharge relationship by the maximum storage has a number of advantages compared to the classical nonlinear storage-discharge formulation. It leads to more intuitive parameter units (mm and mm/h only), reduces the trade-off effects between shape and magnitude coefficients in the calibration, and it helps estimating a priori around what values the storages will fluctuate (Kirchner

2016b). Discharge from the shallow storage contributes directly to streamflow (L_{sh}) and recharges the groundwater reservoir (L_{gw}) depending on $s(t)$. At the catchment outlet, the simulated discharge was calculated as (equation 1.3):

$$Q_{sim}(t) = L_{sh}(t) + Q_{gw}(t) \tag{1.3}$$

and the simulated streamflow δ^{18}O value was calculated as (equation 1.4):

$$C_{sim}(t) = \frac{L_{sh}(t)\, C_{sh}(t) + Q_{gw}(t)\, C_{gw}(t)}{Q_{sim}(t)} \tag{1.4}$$

1.4 Numerical Schemes and Calibration Procedure

Each mass balance equation (Table 1.2) was solved numerically using two different schemes and different time steps depending on the period of interest (P_1 or P_2).

P_1 is a warm-up period of 21 years (1980–2001). Storage values were initialized at the start of P_1 as 50% of the maximum capacity of the shallow storage and 80% of the maximum capacity of the groundwater storage. Both reservoirs were assigned the average streamflow ^{18}O signature (over 2001–2003) $\delta^{18}O = -10.9‰$. We used the precipitation amounts and PET measured at PRIMET and looped back the 2 year record of δ^{18}O data from 1980 to 2001 to run the model for 21 years. The aim of this warm-up was to remove the impact of the initialization of storage values, the impact of the initial tracer masses, and the influence of the initial residence time distributions on the simulations during period P_2. To reduce the computation time and the memory requirements of the warm-up period, we used daily time steps and an Implicit Euler scheme to solve the mass balance equations without divergence from the solution.

P_2 is the period from January 2001 to February 2003, which we used to calibrate model parameters and interpret the water age results. We employed an Explicit Euler scheme to solve mass balance equations efficiently at hourly time steps. The hourly time steps were sufficiently small so that the solution converged. The hourly time steps also increased the precision of the numerical solution compared to the warm-up period (which has daily time steps), since the numerical errors are in these two numerical schemes directly proportional to the time step size.

The hydrological and transport parameters (Table 1.1) were calibrated using a single objective (but multicriteria) Monte Carlo approach, which consists of the following sequence of operations, repeated in a loop:

1. Sample all calibration parameters simultaneously from their prior distributions (Table 1.3).
2. Run the model with the selected parameter set over periods P_1 and P_2.
3. Evaluate model performance over period P_2 with the selected parameter set, by comparing the simulations with the observations. This is done for multiple performance criteria simultaneously, and these are summed into a single objective function.

We repeated this procedure 200,000 times to sample parameter sets. Initial calibration tests showed a correlation between the β_{gw} and D_{gw} parameters. To remove the existing trade-off effects between the groundwater storage exponent β_{gw} and groundwater storage capacity D_{gw}, a storage-discharge relationship was deduced

TABLE 1.3: Prior Distributions of the Calibrated Parameters.

Parameter	Unit	Prior distribution[a]
Ψ	(–)	wbl(1.6, 5)
λ_0	(–)	unif(0, 1)
μ_Q	mm/h	logn(2.43, 0.8)
β_{sh}	(–)	unif(0, 100)
β_{gw}	(–)	logn(4.62, 0.7)
D_{sh}	mm	unif(0, 3000)
D_{gw}	mm	[b]
G_{sh}	mm	unif(0, 3000)
G_{gw}	mm	unif(0, 5000)

[a] wbl(a, b): Weibull distribution with scale parameter a and shape parameter b. unif(min, max): Uniform distribution between min and max. logn(m, r): lognormal distribution with parameters m and r.

[b] This parameter was treated differently, see section 1.4.

from multiple recession segments with a Master Recession Curve (MRC) analysis, using the matching-strip algorithm of Lamb and Beven (1997). The groundwater reservoir storage-discharge relationship could then be expressed in the model parameter space as the functional relationship $D_{gw} = 30\,\beta_{gw}$. This method was described in detail and successfully used in various catchments in the Alzette River basin in Luxembourg (Fenicia et al. 2006). Thus, D_{gw} was not treated like the other eight free calibration parameters, but was deduced from the values of β_{gw} every time a new parameter set was selected.

Parameter prior distributions (Table 1.3) were used in order to make sure that the Monte Carlo simulation sampled certain parameters more often in certain ranges of interest. The Weibull distribution was used for parameter Ψ to allow a negative skew (i.e., emphasis on lower values). The Lognormal distribution was chosen for parameters μ_Q and β_{gw} to allow a positive skew (i.e., emphasis on higher values). The parameters of the distributions were chosen so that the mode of the distributions was located at values that we judged more appropriate for WS10 and so that the variance around the mode was not too restrictive for the Monte Carlo approach. Choosing $a = 1.6$ (value of the mode) and $b = 5$ (measure of the spread) in the Weibull distribution of the parameter Ψ (Table 1.3) constrained Ψ to values allowing a good agreement between simulated ET and the observed average difference $P - Q$ (1980–2003). Using Ψ was necessary to close the water balance, because average PET estimates were found to be 120 mm/yr below the average difference $P - Q$ (over 1980–2003) (see also Waichler et al. 2005). This choice of the prior for w also prevented ET rates from exceeding 5 mm/d, which is the upper limit for WS10 (Barnard et al. 2010). Choosing $\mu_Q = 2.43$ and $\sigma = 0.8$ in the Lognormal distribution of the parameter μ_Q allowed a good agreement between observed and simulated maximum streamflow (around 6 mm/h for the period 1980–2003), by placing the mode of the distribution at $e^{\mu-\sigma^2} \simeq 6$ and guaranteeing a sufficient positive skew. Finally, for the parameter β_{gw}, choosing $\mu = 4.62$ and $\sigma = 0.7$ in the Lognormal distribution placed the mode at $e^{\mu-\sigma^2} \simeq 62$. This yielded groundwater storage values consistent with the catchment storage in WS10 estimated from its time invariant streamflow Mean Travel Time (MTT) of 1.2 years (McGuire et al. 2005), since we used the relationship

$D_{gw} = 30\,\beta_{gw}$ and assumed that $MTT = D_{gw}/Q_{avg}$, where Q_{avg} is observed average annual streamflow (1980–2003).

Model performance Φ was evaluated for the 200,000 parameter sets over period P_2 simultaneously for the hourly streamflow data (observed Q and simulated Q_{sim}) and for the observed δ^{18}O and simulated δ^{18}O (C_{sim}) at the catchment outlet with (equation 1.5):

$$\Phi = NSE_Q + NSE_{\log Q} + V_Q + NSE_{\delta^{18}O} + V_{\delta^{18}O} \tag{1.5}$$

where NSE stands for Nash-Sutcliffe Efficiency (Nash and Sutcliffe 1970) and V stands for Volumetric Efficiency (Criss and Winston 2008). We chose those performance metrics to emphasize various parts of the simulated flows Q and δ^{18}O (correspondence in peaks reflected in NSE, recessions reflected in $\text{NSE}_{\log}$, and long-term trends reflected in V). This means that all our results and conclusions drawn from the use of the model are conditional on the simultaneous fit of the model to discharge and δ^{18}O.

The 200,000 parameter sets were ranked by decreasing Φ. We kept the top-ranked 1,000 parameter sets as behavioral (i.e., the best 0.5% parameters) and discarded the others. Parameter histograms were generated from the behavioral parameter sets. For travel time calculations, we selected the calibrated parameter set consisting of the most common occurring value for each parameter from the histograms (c.f. Benettin et al. 2015b). We also run a posteriori simulation with this identified parameter set to confirm that the associated model performance Φ was among the 1,000 best, as this specific combination was not sampled within the 200,000 parameter sets.

1.5 Residence Time Distributions, Streamflow Travel Time Distributions and StorAge Selection Functions

We calculated RTDs, and streamflow TTDs and SAS functions by tracking water ages in the calibrated catchment model. The approach consists of solving the Master Equation (ME) (equation 1.6) numerically for each model reservoir R, yielding its residence time distribution, p_R, at every moment (Botter et al. 2011):

$$\frac{\partial}{\partial t}(S_R(t)\,p_R(T,t)) = -S_R(t)\,\frac{\partial p_R(T,t)}{\partial T} - \sum_{k=1}^{n} Q_k(t)\,p_k(T,t) \tag{1.6}$$

where T is water age since it entered the catchment surface as precipitation, S_R is the total storage of the reservoir (active + passive), and p_k is the travel time distribution of the outflow Q_k from the reservoir R (in this case discharge and/or ET fluxes). Water storage in each reservoir and fluxes between reservoirs were necessary to solve the equation (equation 1.6). These quantities were provided by the simulation results from the calibrated conceptual model.

The RS scheme implies that outflows from a given reservoir draw water ages proportionally to their relative presence in storage, which corresponds to SAS functions ω_k equal to 1 across all ages T and all times t for each outflow Q_k. This means that the probability that a water parcel has the age T in outflow k at time t is the probability that a water parcel in the corresponding storage has the age T at time t (equation 1.7):

$$p_k(T,t) = \omega_k(T,t)\, p_R(T,t) = 1 \times p_R(T,t) \tag{1.7}$$

Boundary conditions for the ME were defined by precipitation inputs with $T = 0$ in storage at every time t (equation 1.8):

$$p_R(0,t) = \frac{P(t)}{S_R(t)} \tag{1.8}$$

Initial conditions for the ME were specified as an exponential distribution of water ages in storage at $t = 0$ (equation 1.9), with a mean $T_0 = 1.2$ years (McGuire et al. 2005):

$$p_R(T,0) = \frac{1}{T_0} \exp\left(-\frac{T}{T_0}\right) \tag{1.9}$$

The effect of these initial RTDs in storage was removed by the 21 year warm-up period P_1 (1980–2001). This method was shown to have little impact on travel time estimates in the following period P_2 used for the interpretations of the results (c.f. Hrachowitz et al. 2011).

The overall streamflow TTD, p_Q was calculated as a flow-weighted TTD of the different contributions of the reservoirs to streamflow (equation 1.10):

$$p_Q(T,t) = \frac{L_{sh}(t)\, p_{sh}(T,t) + Q_{gw}(t)\, p_{gw}(T,t)}{Q_{sim}(t)} \tag{1.10}$$

The overall catchment RTD, p_S was calculated as a storage-weighted RTD including all forms of storage (equation 1.11):

$$p_S(T,t) = \frac{(S_{sh}(t) + G_{sh})\, p_{sh}(T,t) + (S_{gw}(t) + G_{gw})\, p_{gw}(T,t)}{S_{sh}(t) + S_{gw}(t) + G_{sh} + G_{gw}} \tag{1.11}$$

The overall streamflow SAS function ω_Q was calculated as the ratio of the overall streamflow TTD and the overall RTD (equation 1.12):

$$\omega_Q(T,t) = \frac{p_Q(T,t)}{p_S(T,t)} \tag{1.12}$$

Note that ω_Q represents the contributions of both reservoirs to streamflow, and is different from the individual SAS functions for each reservoir ω_k, which are all equal to 1. The dynamics of the streamflow SAS function ω_Q were therefore mostly governed by the relative contributions of the shallow and the groundwater reservoirs with their respective RTDs.

1.6 Results

1.6.1 Simulated Q, ET, δ^{18}O, and Parameter Distributions

The calibrated model successfully reproduced streamflow seasonality, alternating between the high and the low flows (Figure 1.4) with $NSE = 0.85$ over the period P_2 (Table 1.4). The modeled ET in WS10 averaged 576 mm/yr (2001–2003), consistent with past ET estimates using a monthly Thornthwaite method with the same climate station (Rothacher et al. 1967). ET varied strongly across seasons. During the summer, ET peaked at around 5 mm/d, which is consistent with estimates using the Penman-Monteith method (Monteith 1981) over a summer period in WS10

(Barnard et al. 2010). In summer, ET depleted the water available in the soils for generating streamflow. In winter, modeled ET decreased to about 1 mm/d, consistent with estimates by (McGuire et al. 2007). Streamflow $\delta^{18}O$ was well reproduced by the model (Figure 1.4) with $NSE = 0.51$ (Table 1.4). However, the calibrated model partly failed to reproduce the first $\delta^{18}O$ peaks in the wet-ups of October 2001 and 2002. Furthermore, simulated and observed $\delta^{18}O$ deviated through low flows in the summer (when streamflow was near-zero and little rainfall-runoff response occurred). Generally, behavioral parameter ranges were rather wide (spanning 1–2 orders of magnitude, Figure 1.5 and Table 1.5), except for the parameters Ψ, λ_0, μ_Q, and β_{sh}. The behavioral distributions for parameters Ψ, μ_Q, and β_{gw} were rather similar to the prior distributions, which indicates that the prior distributions were decisive in the calibration. This introduced a certain degree of subjectivity.

TABLE 1.4: Model Performance Metrics.

Variable	NSE[a]	VE[b]
Q	**0.85** (0.74–0.89)	**0.67** (0.57–0.7)
$\log(Q)$	**0.73** (0.60–0.77)	
$\delta^{18}O$	**0.51** (0.31–0.57)	**1.02** (0.99–1.03)

Note. Bold values: calibrated model; ranges of the behavioral solutions in parentheses.
[a] Nash-Sutcliffe Efficiency (Nash and Sutcliffe 1970).
[b] Volumetric Efficiency (Criss and Winston 2008).

TABLE 1.5: Calibrated Parameter Values and Parameter Ranges of the Behavioral Solutions.

Parameter	Unit	Calibrated value	Behavioral range[a]
Ψ	(–)	1.6	0.88–2.1
λ_0	(–)	0.9	0.47–0.99
μ_Q	mm/h	9	5.8–60
β_{sh}	(–)	18	11–34
β_{gw}	(–)	60	22–2.2×10^2
D_{sh}	mm	1,500	1.1×10^3 to 2.9×10^3
D_{gw}	mm	1,800[b]	6.6×10^2 to 6.6×10^3
G_{sh}	mm	550	1.9×10^2 to 1.6×10^3
G_{gw}	mm	0[c]	1.6×10^2 to 4.9×10^3

[a] 2.5–97.5 percentile range of the distributions in Figure 1.5.
[b] This parameter was treated differently, see section 1.4.
[c] Parameter contrained to 0, see section 1.7.1.

1.6.2 Dependence of the Streamflow SAS Functions on Catchment Storage

We used the dynamic storage metric S_d as a proxy of catchment wetness (equation 1.13):

$$S_d(t) = S_{sh}(t) + S_{gw}(t) - \min(S_{sh} + S_{gw}) \tag{1.13}$$

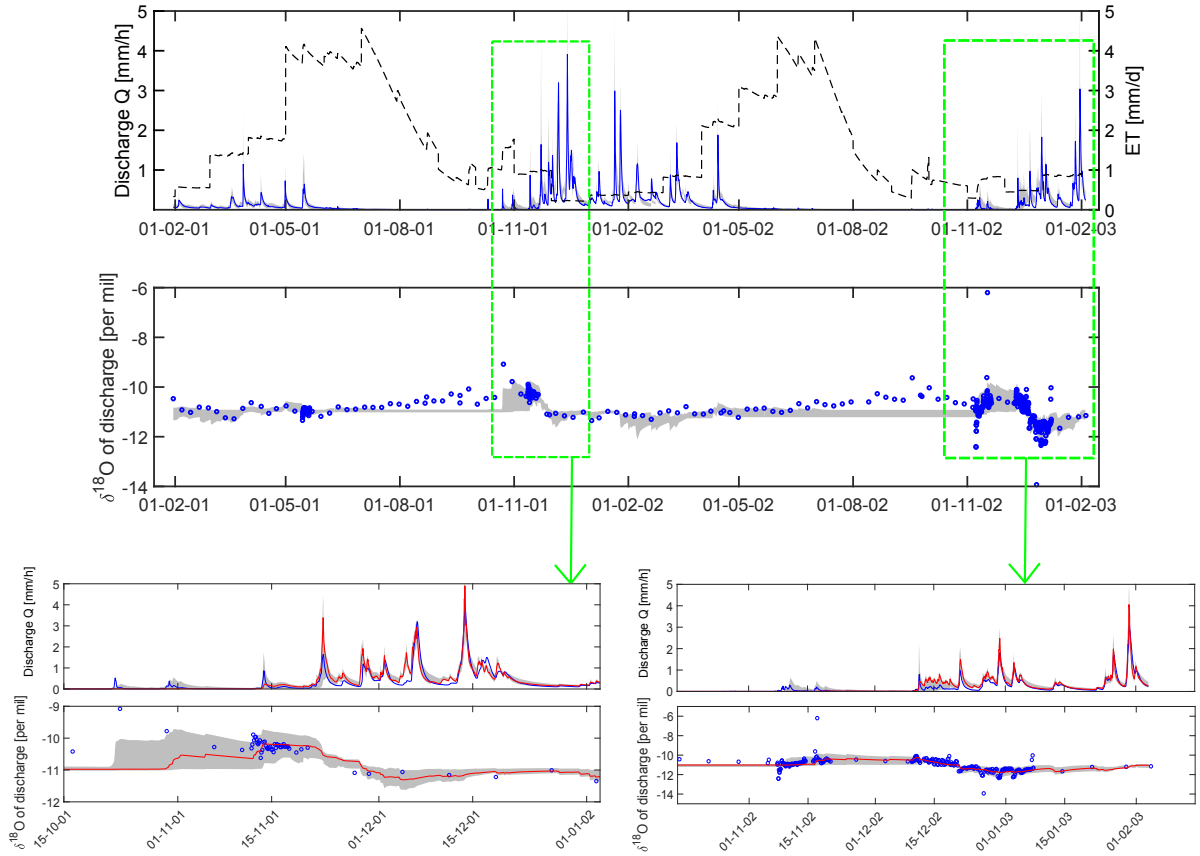

FIGURE 1.4: (a, c, and e) Simulation envelopes of the behavioral solutions (gray bands), for streamflow and (b, d, and f) its δ^{18}O values, (a and b) for the whole simulation period, including simulated ET (dashed blacked line), and for wet-up periods of (c and d) 2001 and (e and f) 2002, including the calibrated model simulation (red line). Observations are in blue.

This metric ranged from 0 mm, when the active storage (Ssh + Sgw) was at its minimum (around September), to almost 400 mm, when the active storage reached its maximum (around January). Distinct patterns of streamflow SAS functions could be identified, depending on the value of S_d (Figure 1.6). During dry periods, when S_d approached 0 (red curves), streamflow preferentially selected older water. Younger water was preferentially selected during wet periods (blue curves), when S_d approached its maximum (about 400 mm). This is consistent with the definition of the ISE. However, during transition periods, when S_d took intermediate values (e.g., 175–275 mm), streamflow SAS functions did not vary regularly with S_d, as they alternated rapidly between a preference for younger and a preference for older water (Figure 1.6). This indicates that during transition periods, the ISE was not always valid. The streamflow SAS functions in WS10 also displayed noticeable plateaus. For example, the top purple curves (Figure 1.6) are flat between $P_S = 0$ and $P_S = 0.2$, which shows that the youngest 20% of the water particles stored in the catchment were equally mobilized to generate discharge at that time. Such plateaus can be followed by sharp vertical variations as seen for P_S slightly above 0.2 (Figure 1.6, purple curves). This shows that the oldest 80% of the water particles were significantly less mobilized to generate discharge compared to the youngest 20% at that time, suggesting a strong variability of flow paths.

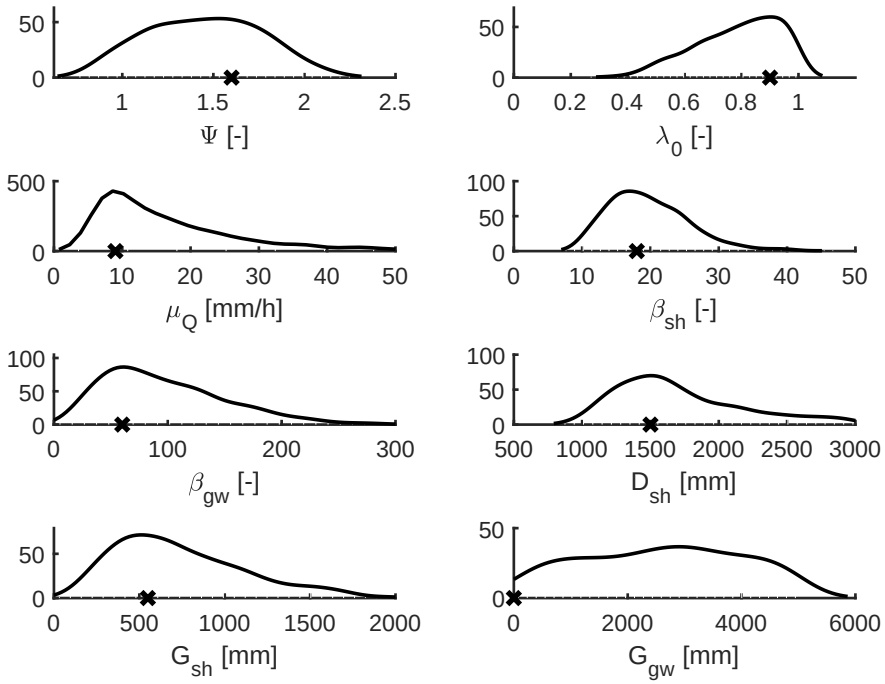

FIGURE 1.5: Histograms of the 1,000 best parameter sets (smoothed by a normal kernel distribution), and calibrated parameter values (crosses). Units in y axis are nondimensional numbers per parameter value.

1.6.3 Inverse Storage Effect (ISE) in WS10

Let t_m be the time when the catchment storage S reaches its minimum S_{min} and t_M the time when the catchment storage reaches its maximum S_{MAX}. The sign of the following metric χ generally defines the increase or decrease in the preference of streamflow for younger water, for increasing storage values (equation 1.14):

$$\chi = \frac{\omega_Q(P_S = 0, t_M) - \omega_Q(P_S = 0, t_m)}{S_{MAX} - S_{min}} \tag{1.14}$$

Note that in our case, we could compute χ using the dynamic storage metric S_d instead of the absolute storage S, since their time variation was the same. Water age T or age-ranked storage S_T could be used as an alternative variable for ω_Q in equation 1.14. $\chi > 0$ indicates an ISE overall, while $\chi < 0$ indicates a Direct Storage Effect (DSE) overall. Contrary to the ISE, the DSE is by definition a decrease of young water fractions in discharge for increasing storage (Harman, 2015). In WS10,the ISE is generally the dominating process since $\chi = \frac{2.4}{370\ mm} = 0.065\ mm > 0$.

During transition periods in WS10 (approximately for $175\ mm < S_d < 275\ mm$), the streamflow SAS functions varied strongly while S_d varied little, which prevented the use of a metric similar to χ (e.g., denominator close to 0 when variations of S_d were close to 0). Thus, we needed to compare the variations of ω_Q for the youngest water(i.e., $d\omega_Q(0,t)$) and the variations of S_d (i.e., $dS_d(t)$) in a time-varying way. We compared variations of ω_Q and S_d at weekly intervals, i.e., we compared $\omega_Q(0, t + 1\ week) - \omega_Q(0,t)$ and $S_d(t + 1\ week) - S_d(t)$ (Figure 1.7).It became apparent that the response of WS10 was mostly in agreement with the ISE (Figure 1.7). However, there were periods during which $dS_d > 0$ and $d\omega_Q < 0$, i.e., the DSE seemed to apply. A

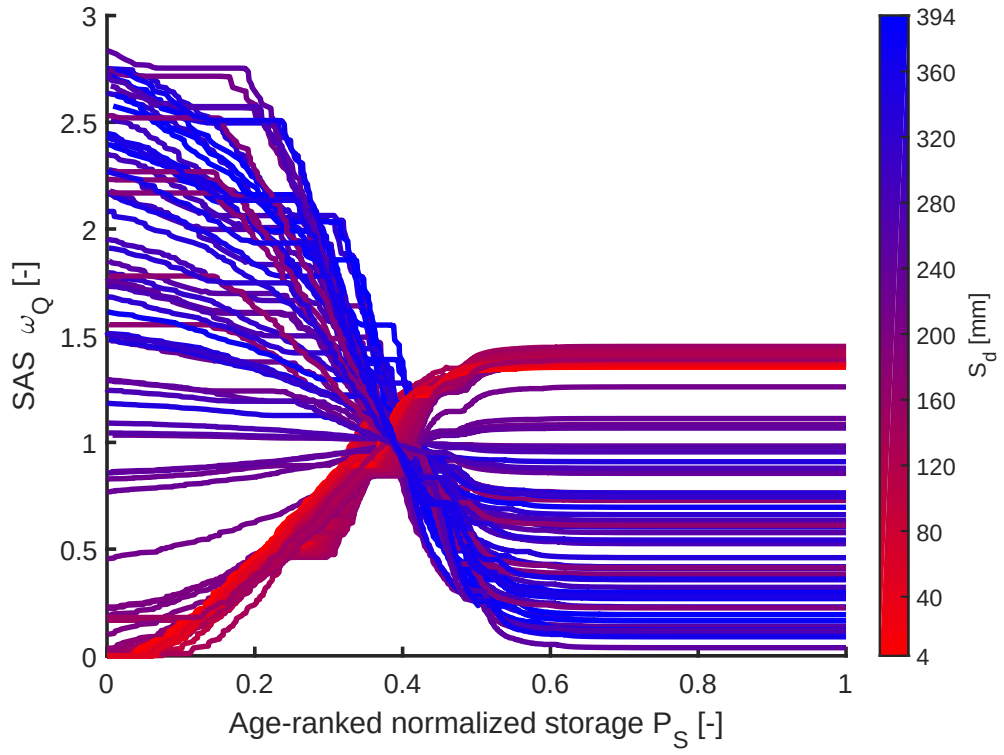

FIGURE 1.6: Streamflow SAS functions $\omega_Q(T,t)$ for different dynamic storage values $S_d(t)$. Each curve represents a different time t, associated with a color from red to blue, depending on the value $S_d(t)$. We used a change of variables $T \rightarrow P_S(T,t)$ for the SAS functions $\omega_Q(P_S(T,t),t)$, where P_S is the cumulative residence time distribution (i.e., age-ranked normalized storage). Younger ages in storage are toward the left part of the x axis, and a higher amplitude of ω_Q indicates a stronger preference of streamflow for a given age in storage.

particularly important number of points lay on the $d\omega_Q = 0$ axis for $dS_d < 0$ (Figure 1.7), which shows that weekly variations in $S_d(t)$ and weekly variations of $\omega_Q(0,t)$ were independent in these moments.

1.6.4 Streamflow Travel Time Dynamics

The streamflow median travel time (MdTT) in WS10 showed a pronounced seasonality, ranging from less than 100 days during wetter periods to more than 1,300 days during drier periods (Figure 1.7). Similar to the streamflow SAS functions, the MdTT appeared not to depend only on the value of S_d. When $S_d < 175$ mm, the MdTT was high and it increased as S_d decreased. Conversely, when $S_d > 275$ mm, the MdTT was low and it decreased as S_d increased. Yet, for values of S_d close to these two thresholds, switches between the low and high values of the MdTT occurred. Over the entire range of S_d, MdTT followed a counter-clockwise hysteresis loop. Many MdTT values corresponded to a given S_d, which indicates that the streamflow travel times in streamflow depend on the whole history of storage states. Note that a clockwise hysteresis loop exists for $S_d < 175$ mm only because water is aging steadily. Smaller counterclockwise hysteresis loops are visible for $S_d > 275$ mm and correspond to differences between the rising limb and the falling limb of runoff events (Figure 1.8).

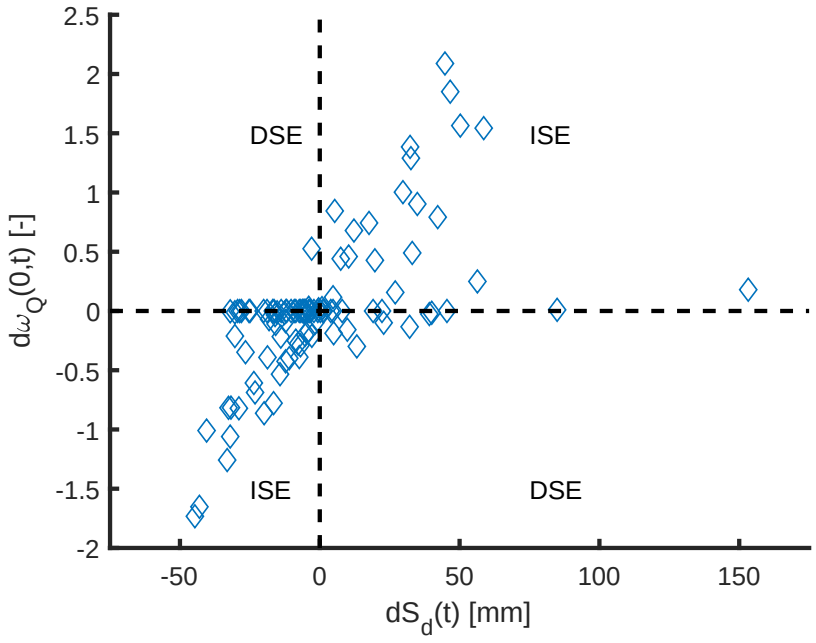

FIGURE 1.7: Variations of the streamflow SAS functions ω_Q at age 0 against variations of S_d in time (here we used weekly intervals to compute these variations).When dS_d and $d\omega_Q$ have the same sign, the catchment response corresponds to the ISE, otherwise it corresponds to the DSE. For practical calculations, since ω_Q is not defined at all ages for all t, we selected at time t the value of ω_Q at the youngest age T available (in our case up to $T =$ a few days).

Water age fractions in streamflow (Figure 1.9) help understand the observed MdTT patterns. Streamflow volume was dominated by younger water discharged during runoff events. At the beginning of winter 2001, the catchment passed through several states and periods that could be identified as follows (Figure 1.9):

1. Period 1 (6–27 October 2001), dry catchment state: streamflow is very low, rather constant, and 75% of the water is over 2 years old, which explains the very long MdTT of over 3 years. Water is aging steadily, as indicated by the small and constant increase in the MdTT.

2. Period 2 (27 October to 27 November 2001), wet-up: as soon as the stream starts reacting to precipitation events, even when the streamflow response is so low that it does not show up on the graph (Figure 1.9), the event water fraction (age < 2 days, pink color) increases rapidly (up to 12%). Young preevent water (2–20 days, yellow color) starts to be mobilized and ranges from 10% to 15% of the total flow. This causes the MdTT to drop significantly to under 2 years. Streamflow increases sharply and consists mostly of water younger than 1 year old.

3. Period 3 (27 November to 20 December 2001), wet state: through the wettest period of winter, the event water fraction slightly decreases (peaking around 10%), whereas young preevent water (2–20 days) becomes more relevant (up to 35% contribution). This causes the MdTT to decrease to less than 6 months.

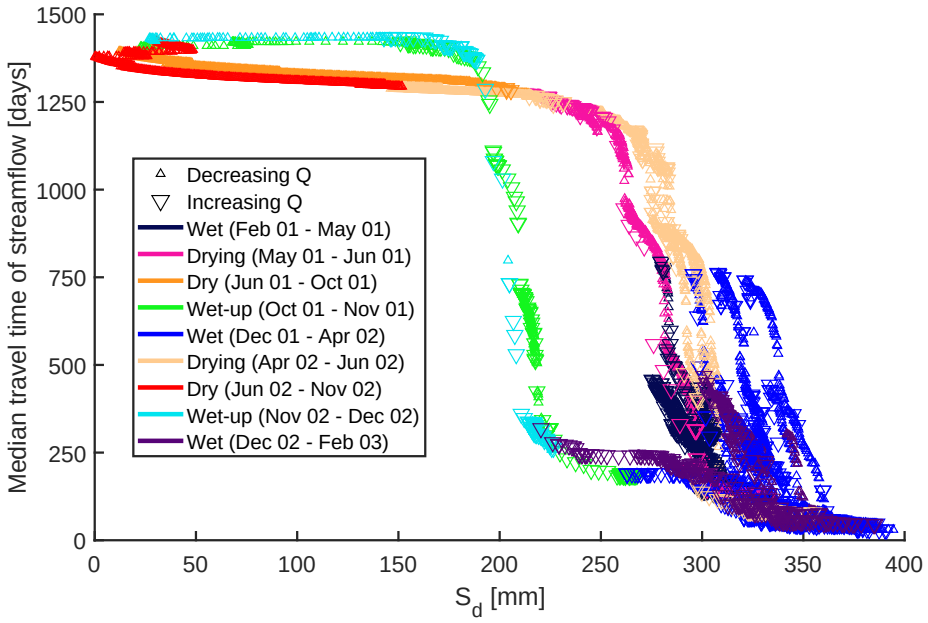

FIGURE 1.8: Streamflow Median Travel Time (MdTT) plotted against S_d. The colors of the triangles indicate various hydrological periods that we selected manually: wet-up (when the first fall rainfall-runoff events occur after a long dry period in summer), wet period (when winter larger rainfall events create large runoff events), dry-out (when discharge starts decreasing in spring while ET starts increasing), dry period (when discharge is minimum and PET is maximum in summer). Smaller triangles pointing upward indicate decreases in discharge Q (these are pointing upward because they are generally associated with increases in the MdTT), while bigger triangles pointing downward indicate increases in discharge Q. The main hysteresis loop runs counter-clockwise.

Streamflow is then at its maximum and mostly composed of water younger than 6 months.

4. Period 4 (20 December 2001 to 4 January 2002), recession: event water is not found in streamflow as recession proceeds and rainfall events cease, and young preevent water (2–20 days) contributes to less than 10% of the streamflow. About 50% of streamflow is older than 2 years. Consequently, the MdTT increases to more than 2 years. The system is still in a transitional phase, where event water can still reach the stream rapidly and causes a noticeable drop in the MdTT during precipitation events.

1.7 Discussion

1.7.1 Model Parameterization and Performance

The deviation between simulated and observed $\delta^{18}O$ in streamflow during the low-flow period of 2001 and 2002 (Figure 1.4) suggests an evaporative enrichment of stream water, which was not implemented in the model. During these periods, discharge was close to zero ($Q \ll 1$ L/s) and the stream was mainly flowing over a

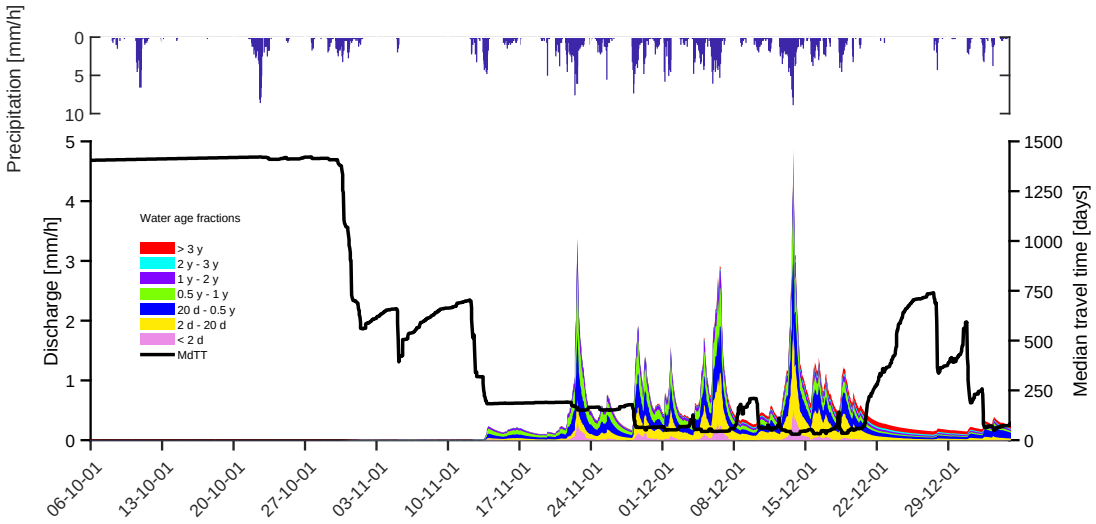

FIGURE 1.9: Simulated hydrographs over the beginning of winter 2001, showing the different fractions of streamflow ages in time (age brackets associated with contrasting colors) and the streamflow median travel time MdTT (black curve) in time. The top part of the plot shows precipitation rates. Note that the youngest water fraction is placed at the bottom of the hydrograph to make it more visible.

bedrock channel without much storage, and likely experienced evaporation. Without additional ^{2}H data, it is unfortunately not possible to plot the stream samples against the Local Meteoric Water Line (LMWL) and confirm the interpretation that water was fractionated by evaporation.

Within this work, we assumed RS for all outflows in the soil and groundwater reservoirs (Benettin et al. 2015b). Yet, Brooks et al. (2010) found that plants may be able to access water in the soils that does not contribute to discharge. The RS assumption we employed here was nevertheless sufficient to satisfactorily simulate ^{18}O tracer dynamics in the WS10 streamflow (NSE > 0.5 for streamflow $\delta^{18}O$), especially after combining the two storages in series-parallel. The related simplicity of representing transport in this way is a major advantage when solving equations numerically. An additional advantage is the parsimonious parameterization, which limits parameter uncertainties. More complex approaches could parameterize the interactions between the active storage and passive storage in the model. Recent studies have used mixing coefficients in conceptual models and concluded that in some cases, the use of partial mixing (PM) could result in better performances in simulating hydrological tracer transport compared to the use of CM (Fenicia et al. 2010; Hrachowitz et al. 2013; McMillan et al. 2012b). Yet, in some cases, this improvement came at the cost of more uncertainty due to decreased parameter identifiability (Fenicia et al. 2010; Hrachowitz et al. 2013). Alternatively, the constant SAS functions used in each reservoir in our study could be changed to varying functions of residence times. One could use, for example, time-variant gamma (Harman 2015), beta (van der Velde et al. 2012; van der Velde et al. 2015), or power law distributions (Benettin et al. 2017b) with respect to reservoir residence times. The parameters of these distributions could be made storage-dependent. The main difficulty in applying such approaches would be finding the minimum additional state variables or parameters required to capture the true hydrological and chemical processes in this

catchment (Haraldsson and Sverdrup 2013) without increasing uncertainties (Fenicia et al. 2010; Page et al. 2007) or attempting to match the data too closely with too complex and unrealistic parameterizations (Kirchner 2006). Future work may explore the importance of mixing assumptions in WS10, as done at other sites (van der Velde et al. 2015).

The parameter ranges in this study were similar to other studies using conceptual models calibrated to hydrographs and tracer data (Benettin et al. 2015a,b; Birkel and Soulsby 2016; Birkel et al. 2011a, 2015; Fenicia et al. 2010; Soulsby et al. 2015b). However, some other catchment transport studies (e.g., Birkel et al. 2010b; Hrachowitz et al. 2013; Hrachowitz et al. 2015) found narrower parameter ranges. A meaningful comparison of parameter uncertainty and ranges is yet challenging due to different processes, parameterizations, and calibration methods. Streamflow δ^{18}O in WS10 is one of the most damped tracer record among a range of well-studied catchments (Tetzlaff et al. 2009) and thus provides less information about transport processes. Most δ^{18}O samples in WS10 were weekly samples, which may hide more pronounced variations that could be used to identify transport processes (Birkel et al. 2010a; Dunn et al. 2008; Hrachowitz et al. 2011; Kirchner et al. 2004). The damped and low frequency tracer record in WS10 was partly responsible for the parameter uncertainties. These uncertainties were already discussed by McGuire and McDonnell (2006) a decade ago when applying a steady state convolution approach to derive catchment MTT in WS10.

The prior distributions had a noticeable weight in shaping the behavioral parameter distributions (Figure 1.5). These prior distributions were nevertheless necessary to simulate streamflow values in agreement with discharge measurements (governed by the parameter μ_Q), ET values corresponding to the previous estimates (parameter Ψ), and groundwater storage consistent with a previous MTT estimate (parameter β_{gw}) (see section 1.4). The parameter G_{gw} representing passive groundwater storage was not identifiable (Figure 1.5), thus we constrained it to 0 mm. The uncertainty of G_{gw} results from the inability of ^{18}O to distinguish the potential contributions of a groundwater volume with residence times over $\sim$2 years, since these contributions barely show variation in δ^{18}O. Constraining old water fractions of the streamflow TTDs requires additional tracers, such as tritium (Stewart et al. 2012). Uncertainty remains in the contributions and the age of old water in WS10. Similar results were found in the Hubbard Brook catchment, NH, USA (using stable isotopes) and in the Upper Hafren catchment, Plynlimon catchment, UK (using chloride) (Benettin et al. 2015a,b). Future work should assess the uncertainties of time-varying travel times and SAS functions to all model parameters using various water age metrics, such as the Median Travel Time (Benettin et al. 2017b) in a rigorous uncertainty framework.

1.7.2 Time-Varying Relationships Between Storage and Water Ages

Our results (Figures 1.6 and 1.8) suggest that streamflow SAS functions and travel times in WS10 mostly depend on catchment wetness, which is controlled by the climate seasonality at HJA. Age selection patterns generally followed the ISE (Figure 1.7) where the fractions of younger water in discharge increase with increasing storage. This study therefore supports the emerging idea that the ISE may be true for most catchments (Benettin et al. 2017b; Harman 2015; Kim et al. 2016; Pangle et al. 2017; Wilusz et al. 2017). The ISE is also consistent with findings from previous streamflow travel time studies in WS10 (Klaus et al. 2015a; McGuire et al. 2007;

Sayama and McDonnell 2009) that have not formally studied age selection patterns. However, our study revealed that during transitions from dry to wetter states in fall, and during transitions from wet to drier states in spring, the ISE did not strictly apply in WS10. During these transitions, the changes in the streamflow SAS functions were not linearly related to the changes of catchment storage. In part, we also found a decrease in the fraction of young water for increasing catchment storage. This may point to a DSE-like behavior in a few instances. The marked seasonality of the Mediterranean climate at HJA made it possible to detect this, resulting in rapid and strong storage variations across seasons. However, uncertainty was highest during these transition periods (Figure 1.4). Hydrological systems are known to be more uncertain during transitions of certain state variables like soil moisture around threshold values (Zehe et al. 2007). Deviations of travel time dynamics from the ISE have been reported at other sites (Klaus et al. 2013; Legout et al. 2007). For example, Klaus et al. (2013) found higher fractions of preevent water (i.e., "old" water) leaving the soil matrix after a certain moisture threshold was exceeded.

Parameterizing SAS functions with respect to catchment storage has recently been demonstrated using one-to-one relationships with age-ranked storage (Benettin et al. 2015b, 2017b; Harman 2015). In WS10, streamflow SAS functions seemed to depend on storage values with a many-to-many relationship (Figures 1.6 and 1.7). During transition periods, different storage values could result in a similar age selection pattern. Reciprocally, different age selection patterns could result from the same storage value. In WS10, catchment mixing, i.e., the variability of flow paths integrated at the catchment scale, thus seemed to depend rather on the history of storage states (i.e., hysteretic), and on the spatial distribution of water storage (e.g., in the hillslopes). This is consistent with the observation of hysteresis in hillslope-catchment discharge relationships in a previous study in WS10 (McGuire and McDonnell 2010). Contributing areas, where water flows through the subsurface down to the stream, can extend far upstream during the rising limb of the hydrograph, and remain in the higher elevations during the falling limb, creating storage-discharge hysteresis (Nippgen et al. 2015). Overall, our findings suggest that storage effects on water age selection (ISE, DSE) may need to be regarded as a time-varying phenomenon in catchments exposed to a marked seasonal climate. In future work, the parameterizations of SAS functions should ideally integrate this behavior.

1.7.3 Confronting Time-Varying Streamflow Travel Times and SAS Functions to Previous Flow Path Investigations in WS10

The streamflow travel time estimates in this study were similar to the results from previous studies in WS10. We found an average MTT (flow-weighted) of 1.36 years over the simulated period (2001–2003). Previous studies reported a MTT of 1.26 $\pm$ 0.29 years (McGuire et al. 2005), 0.47 years (Sayama and McDonnell 2009), and 1.1 years (Klaus et al. 2015a), even though different approaches were used. The relatively low MTT of Sayama and McDonnell (2009) may be attributed to the fact that their model did not rely on tracer data but only hydrographs to constrain the predictions. The ^{18}O time series contains more information about the velocity of water particles than streamflow data alone, which is more informative about the celerity of subsurface water pressure waves. Incorporating tracer data into the models is known to affect the estimation of parameters and thus water velocities and travel times (McDonnell and Beven 2014; McGuire et al. 2007; Scaini et al. 2017; Vaché and McDonnell 2006; van Verseveld et al. 2017). This clearly shows that estimated travel times that are not benchmarked against tracer data must be taken with caution (c.f.

Danesh-Yazdi et al. 2018). Nevertheless, the streamflow travel time dynamics followed climate seasonality (Figure 1.9), showing a clear climate control on the MTT in WS10, consistent with previous studies (Klaus et al. 2015a; Sayama and McDonnell 2009) This is consistent with findings from other sites (Heidbüchel et al. 2013; Hrachowitz et al. 2009a,b, 2010; van der Velde et al. 2012).

We used the streamflow median travel time (MdTT) in our analysis (Figures 1.8 and 1.9), which is less sensitive to the uncertainties induced by old water fractions (see section 1.7.1). The MdTT ranged from 0.08 to 3.9 years during the 2001–2003 period (Figure 1.9) with an average (flow-weighted) value of 0.28 years. Klaus et al. (2015a) reported an average MdTT of 0.9 years over the same period. One reason for this difference relates to the way tracer data is used in each approach. The method of Klaus et al. (2015a) gave more weight to the base flow samples to constrain simulations than our approach. These samples are associated with longer travel times. Our approach was able to better capture short-term variations during events that commonly have much younger age compositions than base flow periods. Another reason for this difference in the MdTT could be that we distinguished two main sources of streamflow (i.e., reservoirs) with drastically different TTDs. The most active reservoir was the one simulating the contributions of soil water from the hillslopes, with a high turnover rate. This reservoir provided mostly younger water (ages mostly below 1 year), especially during rainfall-runoff events (Figure 1.9) and gave the average MdTT a rather low value of 0.28 years. This is consistent with the estimated average MdTT of 0.25 years in the WS10 experimental hillslope (McGuire et al. 2007; van Verseveld et al. 2017). It is also consistent with the conclusion that soil depth (along with climate, as explained above) acts as a control on the MTT in WS10 (Sayama and McDonnell 2009), which was also found at other sites (Capell et al. 2012; Heidbüchel et al. 2013; Hrachowitz et al. 2009a, 2010). In contrast to the shallow reservoir, the groundwater reservoir in this study had a low turnover rate. The variations of modeled groundwater storage were only around 5% over 1 year, consistent with the near absence of groundwater level changes in some wells in WS10 (Gabrielli et al. 2012). The groundwater reservoir provided mostly older water fractions (ages well above 1 year) to the stream (Figure 1.9). These groundwater contributions seem to be water that has flowed through tortuous pathways in the bedrock (Gabrielli et al. 2012). Most streamflow is derived from precipitation that fell a few months earlier, but the comparatively small volumetric contributions of water that is several years old (turquoise and red water age fractions in Figure 9) can substantially raise the average MdTT and especially the average MTT in WS10, as suggested by (Gabrielli et al. 2012). Thus, the strong seasonality of streamflow travel times in WS10 is the result of seasonal hillslope contributions added to comparatively constant groundwater contributions. The hillslope connection and disconnection were driven by the precipitation regime.

Deep seepage loss was identified in WS10 and the surrounding catchments WS01, WS02, and WS03 (Gabrielli et al. 2012; Graham et al. 2010; Waichler et al. 2005). In our model conceptualization, this water was lost by ET active only in the shallow storage, which contains mostly younger water (ages < 1 year). This means that potentially more young water left the catchment model via ET than in reality, whereas the model should instead lose old groundwater via deep seepage. This may result in a bias toward older ages in the model compared to reality. Yet, Waichler et al. (2005) suggested that ET estimates in the lower part of the Lookout Creek at HJA are too low to close the water balance and need to be adjusted to account for the evaporation of the precipitation intercepted by humus and wood logs. More work in WS10

is needed to understand what water evaporates from the canopy and the soils, what water plants take up and eventually transpire (Barnard et al. 2010; Bond et al. 2002; Brooks et al. 2010), and how this may impact streamflow TTDs.

Our simulations with a conceptual model yielded streamflow travel time dynamics that were consistent with the previous understanding of how water ages are mobilized from storage in WS10 (Klaus et al. 2015a; McGuire and McDonnell 2010). This understanding was summarized in a perceptual model of the catchment (McGuire and McDonnell 2010) consisting of three water age reservoirs: (1) event water (1–2 days old), (2) shallow reservoir (10–25 days old), and (3) deep reservoir (1–2 years old) associated with flow paths in (1) preferential networks, (2) hillslope soils, and (3) bedrock. Here we went beyond this perception, by showing that the relative contributions of these age reservoirs are highly time-variant and can be represented by contrasting streamflow SAS functions for the dry and wet seasons (Figure 1.6) in agreement with ISE (Figure 1.7). However, the ISE may not apply during transition periods, since different and competing transport processes create a hysteretic catchment response (Figure 1.8, McGuire and McDonnell (2010)). Hysteresis in storage-travel time response was also observed in three Scottish catchments (Hrachowitz et al. 2013). Hysteresis in catchment response indicates that significant changes occur in the source of the water mobilized to generate streamflow. This may be important to explain the rapid chemical response of catchments during runoff events in contrast with the relative absence of response in conservative tracers (Godsey et al. 2010; Kirchner 2003). The transport processes influencing the WS10 response during wet-up periods could on the one hand be the redistribution of soil water from wetter to drier parts. On the other hand, preferential flow takes place in the unsaturated zone (McGuire and McDonnell 2010; McGuire et al. 2007) and in fractures below the soil-bedrock interface (Gabrielli et al. 2012), potentially acting as bypass mechanisms. During the spring dry-out, a generally upward movement of the water, due to plant water uptake and evaporation, competes with downward unsaturated flow in soils. Conceptualizing the mixing of soil water at a smaller scale (e.g., pore scale) as in McGuire et al. (2007) and van Verseveld et al. (2017) to understand catchment scale transport effects (represented by SAS functions) paves the way for future research in WS10.

1.8 Conclusion

Our catchment conceptual model, developed based on experimental observations of water flow paths in WS10 (McGuire and McDonnell 2010), was successful in modeling streamflow and ^{18}O transport. It allowed us to track water ages in catchment storage and outflows, and derive time-varying RTDs, and streamflow TTDs and SAS functions by solving the Master Equation. The streamflow SAS functions showed that the ISE dominates the age selection patterns in WS10 where the Mediterranean climate is highly seasonal. The catchment preferentially releases younger water in the winter after wetting up in late fall, while older water is preferentially released during dryer periods in the summer. The pronounced seasonality of streamflow travel times in WS10 is related to seasonal contributions of younger water from the hillslopes during wet periods. The streamflow SAS functions were a consistent lumped description of this mobilization of water from catchment storage, supported by the previous experimental and modeling work in WS10. During transition periods from dry states to wet states and vice versa, strong storage variations caused by the marked climate seasonality caused deviations from the ISE. These deviations

occurred when changes in catchment storage were not linearly related to changes in streamflow SAS functions. These deviations also showed a DSE-like behavior over short periods of time. During transition periods, the catchment response was likely influenced by preferential flow and soil water redistribution caused by the drying effects of evapotranspiration. The streamflow SAS functions and travel time dynamics furthermore revealed that storage-travel time hysteresis exists in WS10. Overall, these findings suggest that future work with SAS functions needs to consider time variance and hysteresis in the relationships between catchment storage and water ages for catchments exposed to seasonal climates. The modeling results in this study also suggest that more work is needed in characterizing the soil water interactions in WS10. Such interactions could be implemented in lumped modeling approaches, for instance by using mixing coefficients or different SAS functions than in our study.

Acknowledgements

We gratefully acknowledge funding for this work, provided by the Luxembourg National Research Fund(FNR) in the framework of the STORE-AGE project (FNR/CORE/C14/SR/8353440/STORE-AGE). Precipitation, discharge, air temperature, and GIS data were provided by the HJ Andrews Experimental Forest and Long Term Ecological Research program, administered cooperatively by the USDA Forest Service Pacific Northwest Research Station, Oregon State University, and the Willamette National Forest. This material is based upon work supported by the National Science Foundation under Grant No.DEB-1440409. The data is freely available by contacting the corresponding author by email, and from the H. J. Andrews Experimental Forest research program (http://andrewsforest.oregonstate.edu/). We would like to thank Fabrizio Fenicia for providing the MRCtool MATLAB script, and Laurent Pfister for his constructive comments on the manuscript. Finally, we would like to thank the reviewer Markus Hrachowitz, one anonymous reviewer, and the editor for their helpful, critical, and supportive comments that helped us improve the manuscript.

Chapter 2

Expanding travel time theory towards water age multimodality: Part 1

An edited version of this paper was published by Wiley[1] as:

N. B. Rodriguez and J. Klaus (2019). "Catchment Travel Times From Composite StorAge Selection Functions Representing the Superposition of Streamflow Generation Processes". In: *Water Resources Research* 55.11, pp. 9292–9314. DOI: 10.1029/2019WR024973

Key points

- We introduced composite StorAge Selection functions defined as a weighted sum of probability distribution functions (i.e. components).
- A composite streamflow StorAge Selection function was necessary to simulate the complex high resolution δ^2H dynamics of the stream.
- The components of this StorAge Selection function confirmed the hypothesized superposition of streamflow generation processes.

Abstract

Catchment travel times integrate the multitude of hydrological flow processes and provide insights into catchment functioning. StorAge Selection (SAS) functions describe how residence times of water in storage are related to travel times of water in catchment outflows. As such, SAS functions are useful to summarize transport processes in catchments and are ideal to simulate catchment outflows and the concentrations of various solutes and tracers. Recent studies suggested that using one probability distribution function (pdf) for SAS functions may not account for all transport processes in various hydrological systems. In this study we introduced a composite streamflow SAS function defined as a weighted sum of several pdfs (called components), using a uniform pdf for the youngest water and two gamma pdfs for the oldest water. The novel parameterization of this SAS function is nonlinear with respect to catchment storage and to a proxy of storage variations. This composite streamflow SAS function was needed to obtain realistic simulations of the complex high resolution (sub-daily) stream deuterium (δ^2H) dynamics measured for 2 years in the Weierbach, a forested headwater catchment in Luxembourg, whereas various SAS functions using only one component (i.e. pdf) failed. The three components of the composite streamflow SAS function confirmed the superposition of streamflow generation mechanisms with contrasting travel times suggested by previous experimental and modeling studies. Our work suggests that in some catchments, composite SAS functions may offer a more realistic perspective on transport processes derived from high resolution tracer data than simpler SAS functions.

Plain Language Summary

The age of stream water is the time it takes rainfall to reach the stream. Stream water age depends on how water flows through natural landscapes. It provides essential information because it represents the time water spends in contact with soils and rocks, which influences stream chemistry. Recent studies showed that the current mathematical models describing stream water ages cannot completely explain the observed variations of certain hydrological tracers in the stream. These tracers are the water itself, or solutes transported with water. In this study we propose an adaptation of the models describing stream water ages that better corresponds to such observations in a small stream located in a forest in Luxembourg. We suggest that considering that the streamflow is derived from a multitude of superimposed flow mechanisms is crucial for a better description of stream water ages and for better interpretations of the hydrological tracers measured in the streams.

2.1 Introduction

Distributions of water ages are essential descriptors of the hydrological state and functioning of a catchment. The backwards residence time distribution (RTD in the rest of the manuscript) characterizes the fractions of water ages in storage, whereas backwards travel time distributions (TTDs in the rest of the manuscript) characterize the fractions of water ages in outflows (Benettin et al. 2015c; Rigon et al. 2016). Backward distributions (called here RTD and TTDs) are based on the concept of age (time spent since entry). The complementary forward distributions defined using life expectancy (time remaining before exit) are not considered here, but they were exhaustively described by Benettin et al. (2015c) and Rigon et al. (2016). The age of

water is affected by the range of hydrological processes taking place, which makes TTDs integrated descriptors of the transport processes to outlets. The streamflow TTD has thus been used as a lumped model of tracer transport from precipitation to the stream in many catchments (Hrachowitz et al. 2016; McGuire and McDonnell 2006), usually by assuming steady-state flow conditions and the equality between the streamflow TTD and the catchment RTD. These assumptions initially created a mismatch between streamflow generation processes and the streamflow TTD, because streamflow is not at steady-state, and because all the water in storage may not necessarily contribute to streamflow at all times (Berry et al. 2018; Hrachowitz et al. 2013). This mismatch has been a hindrance to summarizing streamflow generation processes using a TTD. Formulating a streamflow TTD in a way that goes beyond these assumptions and that accounts for all transport processes has nevertheless been a challenging task (McGuire and McDonnell 2006), especially because the tracers used to estimate TTDs (mostly ^{18}O and ^{2}H) have been coarsely measured in the last two decades (e.g. weekly to monthly). Advances in stable isotope sampling techniques for precipitation, streamflow, and other water sources (Berman et al. 2009; Koehler and Wassenaar 2011; Munksgaard et al. 2011; Pangle et al. 2013) revealed high variability and complexity of the dynamics of ^{18}O and ^{2}H in catchments. This has been valuable to shed light on the time variance of streamflow generation processes (Kirchner et al. 2004; Klaus and McDonnell 2013), and was followed by some initial steps to account for time variance in the TTDs and RTDs (e.g. Heidbüchel et al. 2012; Hrachowitz et al. 2009b; Tetzlaff et al. 2014). These observations also fostered the use of more complex TTDs and RTDs that are not described by analytical probability distribution functions (pdfs) (Botter et al. 2010). In recent years, new methods to calculate time-variant TTDs and RTDs emerged to better describe transport processes in catchments. For several catchments, water ages were tracked in lumped hydrological models including tracers (Birkel et al. 2015; Hrachowitz et al. 2013; McMillan et al. 2012b; Rodriguez et al. 2018). Yet, only a limited number of study sites have high resolution (e.g. sub-daily) tracer records spanning a few years available for travel time research.

In parallel to improvements in tracer observations, theoretical progress has enabled the calculation of shape-free and time-varying TTDs and RTDs based on an age-labeled water mass conservation equation (Botter 2012; Botter et al. 2011). The closure of this equation can be achieved by using StorAge Selection (SAS) functions. As a result, SAS functions (Harman 2015; Rinaldo et al. 2015) have become a focus of travel time research. SAS functions link the RTD to TTDs and represent the preference of catchment outflows to mobilize certain water ages from storage, which can be viewed as catchment scale mixing from combinations of various flow paths (Benettin et al. 2013a). SAS functions have thus been increasingly used to summarize the transport processes in various hydrological control volumes, such as the soil-atmosphere interface (Queloz et al. 2015b), laboratory hillslopes (Pangle et al. 2017) or virtual hillslopes (Danesh-Yazdi et al. 2018), and various catchments (e.g., Benettin et al. 2017b; van der Velde et al. 2012; van der Velde et al. 2015). Following arguments of Rinaldo et al. (2015), more and more studies have used a transport formulation based directly on SAS functions for a single control volume encompassing the whole catchment rather than calculated from hydrological models (Benettin and Bertuzzo 2018; Benettin et al. 2017b; Harman 2015; Kim et al. 2016; Klaus et al. 2015a). This formulation allows a conceptualization of catchment internal flow paths and the resulting mixing dynamics by only specifying the SAS functions of the catchment outflows. This approach also yields directly the time-varying RTD and TTDs

as a model output, even when the SAS functions are at steady-state. Approaches to estimate non-parametric TTDs from tracer data have also been proposed as an alternative to fitting analytical SAS functions to tracer data (Kirchner 2019; Turner et al. 1987; Turner and Macpherson 1990a). However, such approaches do not distinguish the influence of external forcing dynamics from the influence of internal changes in flow paths on travel times (Kim et al. 2016), nor account for all the terms (e.g. ET) of the catchment age-labeled water budget. Therefore, as it was the case for the TTDs and the RTDs a decade ago, current research on the SAS functions often relies on fitting the parameters of their pdfs (e.g. beta distribution, gamma distribution, power-law function) to tracer data. Recent research has shown that considering time variance in the parameterization of the streamflow SAS function accounts for the time variance of hydrological processes and results in more accurate tracer transport simulations than at steady-state (Benettin et al. 2017b; Harman et al. 2016; Rodriguez et al. 2018). Furthermore, several studies suggested that the shape of the streamflow SAS function is challenging to parameterize with e.g. storage, because of catchment subsurface heterogeneity and nonlinear relationships between storage and streamflow (Danesh-Yazdi et al. 2018; Remondi et al. 2018; Rodriguez et al. 2018; van der Velde et al. 2015; Yang et al. 2018). Thus, an improved understanding of the shape of the streamflow SAS function and of its parameterization is a critical step towards a better link between streamflow generation processes and travel times.

Most studies that determined a time-varying streamflow TTD have found that using one pdf for the streamflow SAS function (e.g. beta, gamma) could simulate well the tracer transport to the stream. This is true for example for the catchments Marshall Gulch in Arizona and Gårdsjön in Sweden (van der Velde et al. 2015), Hafren in Wales (Benettin et al. 2015b; Harman 2015), and Bruntland Burn in Scotland (Benettin et al. 2017b). Although the resulting TTDs and RTD are time-varying, they rely on the assumption that a superposition of hydrological processes acting simultaneously at different temporal and spatial scales with potentially contrasting water ages can be summarized by a relatively simple and smooth SAS function of water ages or age-ranked storage. This assumption is however based on a limited number of sites where high frequency tracer data is available. This requires additional testing at more catchments with a range of climate, land use, and geology (c.f. Rinaldo et al. 2015). This is all the truer as heterogeneity in the subsurface results in complex shapes for the SAS functions (Danesh-Yazdi et al. 2018). In only a few recent studies it was shown that using only one pdf for the SAS functions could not simulate the transport of tracers well. (Queloz et al. 2015b), for example, found that the breakthrough curves of fluorobenzoate tracers in a lysimeter have multiple distinct peaks that are not fully captured by a power-law SAS function. These peaks may be associated with preferential flow on top of translatory flow in the lysimeter (Benettin et al. 2019). Kaandorp et al. (2018) found that the SAS function representing groundwater transport in one of their study catchments could not be approximated by a beta distribution, because of contributions from multiple aquifers with distinct water ages. Kirchner (2016a) showed that simplifying the composite catchment responses could in theory result in incorrect estimations of streamflow TTDs. Similarly, for a given catchment outflow (e.g. streamflow), misinterpretations could be made when a superposition of transport processes associated with contrasting travel times is approximated by a simpler SAS function consisting of only one pdf. Composite SAS functions defined as a sum of pdfs (i.e. components) may thus be necessary to advance the theory in travel time studies especially when they are used as a summary of complex streamflow generation processes.

In this work, we used a composite streamflow SAS function to calculate the time-varying streamflow TTD and catchment RTD in the Weierbach. The Weierbach is a 42 ha forested headwater catchment located in Luxembourg. Streamflow generation mechanisms in this catchment appear to be complex and need to be validated with further analyses. For example, the hydrographs are double-peaked (Martínez-Carreras et al. 2016) and associated with contrasting chemical responses through the year (Martínez-Carreras et al. 2015; Schwab et al. 2018; Wrede et al. 2015). Storage-discharge relationships are highly nonlinear and suggest the presence of a storage threshold for streamflow generation (Martínez-Carreras et al. 2016). Previous studies in this catchment did not use tracer data in transport models (Fenicia et al. 2014) or limited tracer analyses on events to investigate end-member contributions to streamflow at specific times of the year (Wrede et al. 2015). Travel times calculated from longer tracer records can help to test the emerging hypothesis of streamflow generation from a water age point of view, even though the specific transport processes to the stream cannot be identified from the TTD and SAS function alone. In this context, the SAS functions were thus used as models to test this hypothesis in a rather exploratory manner (Neuweiler and Helmig 2017; Pfister and Kirchner 2017), by inspecting the simulations of stream stable isotopes. For this we leveraged the high-frequency δ^2H record collected in the Weierbach catchment from 2015 to 2017. Travel time modeling with commonly used SAS functions did not allow satisfactory simulations of the stream δ^2H (c.f. appendix A). Building on this and on the hypothesis of streamflow generation, we increased step by step the complexity of the travel time model, by adding pdfs to the streamflow SAS function. This study reports the results of this approach and discusses the potential of composite SAS functions to describe complex streamflow generation processes in catchments.

2.2 Materials and Methods

2.2.1 Composite SAS functions

We consider a control volume encompassing a catchment. Inflows into the catchment are precipitation J and its tracer with concentration C_J. Outflows from the catchment are streamflow Q, evapotranspiration ET, and their tracer with concentrations C_Q and C_{ET}. In this study, we employ SAS functions to simulate the concentration C_Q of a tracer (here deuterium) in the stream (optionally in ET, C_{ET}) as:

$$C_Q(t) = \int_{S_T=0}^{S(t)} C_S(S_T, t)\, \omega_Q(S_T, t)\, dS_T \tag{2.1}$$

where $S(t)$ is total catchment storage (mm), S_T is age-ranked storage (mm) i.e. the amount of water in storage younger than a certain age T, $C_S(S_T, t)$ is the tracer concentration of water in storage with age T (determined from the past hydrological inputs C_J), and $\omega_Q(S_T, t)$ is the SAS function of streamflow (mm^{-1}). $\omega_Q(S_T, t)\, dS_T$ (no units) is the fraction of water of age T in streamflow at time t. We introduce composite SAS functions defined as a weighted sum of individual terms:

$$\omega_Q(S_T, t) = \sum_{m=1}^{M} \lambda_m(t)\, \omega_m(S_T, t) \tag{2.2}$$

The m=1...M components ω_m are functions that can be conceptually associated with different streamflow generation processes, each having a weight λ_m. As a consequence, equation 2.1 can be rewritten as:

$$C_Q(t) = \sum_{m=1}^{M} \lambda_m(t) \int_{S_T=0}^{S(t)} C_S(S_T,t)\,\omega_m(S_T,t)\,dS_T = \sum_{m=1}^{M} \lambda_m(t)\,C_m(t) \tag{2.3}$$

where C_m can be conceptually seen as the equivalent concentration associated with a particular streamflow generation process m. This corresponds to a generalization of the approach presented by Klaus et al. (2015a) with an arbitrary number of "mixing distributions" here defined as SAS functions. We consider that ω_m are probability distribution functions (pdf), implying $\int_{S_T=0}^{S(t)} \omega_m(S_T,t)\,dS_T = 1$. Because $\omega_Q(S_T,t)$ is also a pdf, this means that $\sum_{m=1}^{M} \lambda_m(t) = 1$ thus $\lambda_m(t) \leq 1$. Viewing equation 2.3 as a multi-component hydrograph separation (Klaus and McDonnell 2013) makes $\lambda_m(t)$ equivalent to the fraction of streamflow generated by the process m, i.e. $\lambda_m(t) = Q_m(t)/Q(t)$, where $Q_m(t)$ is the discharge (e.g. mm/h) generated by the process m. Note that a more general case would correspond to the looser requirement $\sum_{m=1}^{M} \lambda_m(t) \int_{S_T=0}^{S(t)} \omega_m(S_T,t)\,dS_T = 1$, where it is possible to have any integrable function ω_m and $\lambda_m(t) \geq 1$.

2.2.2 Transport model based on SAS functions

The transport model used in this study numerically solves the water mass balance equation (2.4) and the tracer mass balance equation (2.5) for the catchment. The water mass balance equation is:

$$\frac{dS}{dt}(t) = J(t) - Q(t) - ET(t) \tag{2.4}$$

The tracer mass balance equation is:

$$\frac{dM}{dt}(t) = C_J(t)\,J(t) - C_Q(t)\,Q(t) - C_{ET}(t)\,ET(t) \tag{2.5}$$

where $M(t)$ is tracer mass in storage. The outflow concentrations C_Q and C_{ET} are calculated from their respective SAS function according to the equivalent of equation (2.1) in the travel time domain:

$$C_*(t) = \int_{T=0}^{+\infty} C_J(t-T)\,p_*(T,t)\,dT \tag{2.6}$$

The subscript $_*$ denotes either Q or ET. $C_S(S_T,t)$ has been replaced by $C_J(S_T,t) = C_J(t-T)$, assuming no change in the tracer concentration of a water particle as it ages (ideal tracer). The variable in the integral was changed according to the equality $p_*(T,t)\,dT = \omega_*(S_T,t)\,dS_T$. The TTDs p_* are calculated from:

$$p_*(T,t) = \frac{\partial}{\partial T}(\Omega_*(S_T,t)) \tag{2.7}$$

where Ω_* is the cumulative SAS function assigned to the outflow Q or ET. When J, C_J, ET, and Q are known, equation (2.5) can be solved after assigning functions of S_T to Ω_Q and Ω_{ET} in equation (2.6), and by updating the age-ranked storage S_T at every time step. To update S_T we solve the Master Equation (ME) in the travel time domain (Botter et al. 2011) for the distribution of residence times p_S:

$$\frac{\partial}{\partial t}[S(t)\,p_S(T,t)] = -S(t)\,\frac{\partial p_S}{\partial T}(T,t) - Q(t)\,\omega_Q(T,t)\,p_S(T,t) - ET(t)\,\omega_{ET}(T,t)\,p_S(T,t) \tag{2.8}$$

and we use the relationship $S_T = S(t)\int_{u=0}^{T} p_S(u,t)\,du$. $S(t)$ is deduced from equation (2.4). Since we define the SAS functions with respect to the age-ranked storage S_T in the model, we have to use the conversion $\omega_*(T,t) = S(t)\,\omega_*(S_T,t)$ in equation (2.8). We choose to keep both variables T and t independent, unlike some recent work with SAS functions using only S_T and the method of characteristics to solve the ME (Benettin and Bertuzzo 2018; Harman 2015). Using not only S_T makes it possible to independently specify the numerical resolutions Δt and ΔT. This also allows a definition of the SAS functions ω_* directly as functions of age T if necessary. Figure 2.1 summarizes the calculations at the core of the model.

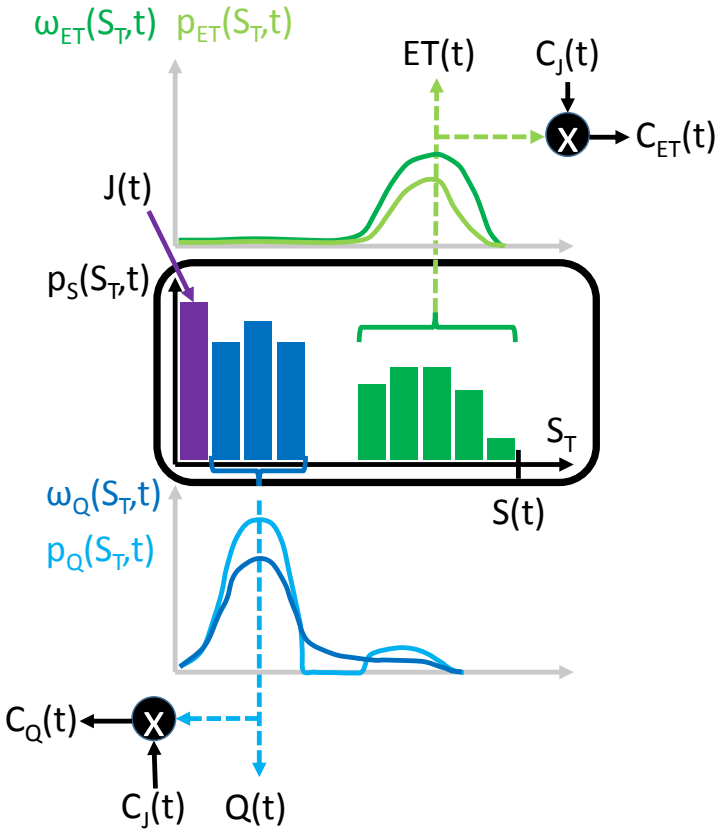

FIGURE 2.1: Conceptual representation of the model used in this study. The central black box represents catchment storage $S(t)$ with its residence time distribution p_S, as a function of time t and age-ranked storage S_T (as a proxy of age T). Precipitation J (in purple) adds new water to storage. Streamflow Q (in blue) and evapotranspiration ET (in green) remove specific ages from storage (colored bars and colored brackets) based on their chosen SAS functions ω_Q and ω_{ET}. The TTDs of Q and ET p_Q and p_{ET} (lighter colors) are thus similar to their SAS function, but modulated by the shape of p_S. The cross in the circle represents the convolution operator applied between the input concentration $C_J(t)$ and the TTDs to generate outflow tracer concentrations $C_Q(t)$ and $C_{ET}(t)$.

2.2.3 Study site description

We applied the described model to the Weierbach catchment. The Weierbach is a 42 ha forested headwater catchment located in the northwest of Luxembourg (figure

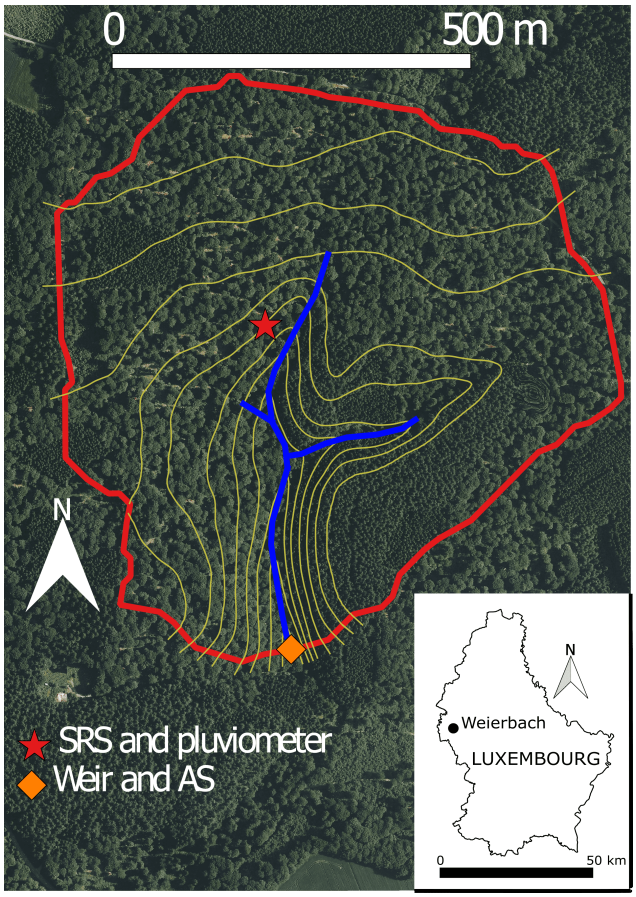

FIGURE 2.2: Map of the Weierbach catchment and its location in Luxembourg. The weir is located at coordinates (5°47′44″E, 49°49′38″N). SRS is the sequential rainfall sampler. AS is the stream autosampler. The elevation lines go by increments of 5 m from 460 m.a.s.l. close to the weir location to 510 m.a.s.l. on the plateau, close to the northern catchment divide.

2.2). It is one of the experimental catchments of the Attert basin in Luxembourg. It has recently been the focus of an increasing number of experimental studies (Antonelli et al. 2017; Glaser et al. 2018; Gourdol et al. 2018; Juilleret et al. 2016; Klaus et al. 2015b; Martínez-Carreras et al. 2015; Martínez-Carreras et al. 2016; Moragues-Quiroga et al. 2017; Pfister et al. 2017b; Scaini et al. 2017, 2018; Schwab et al. 2018) and modeling studies (Carrer et al. 2019; Fenicia et al. 2014; Glaser et al. 2019b, 2016; Kavetski et al. 2011).

The Weierbach is located in the Oesling region in Luxembourg, characterized by gently sloping plateaus cut by deep V-shaped valleys. Elevations in the catchment span from 460 to 512 m.a.s.l. Two landscape units were distinguished depending on their subsolum type and their slope (Martínez-Carreras et al. 2016). The plateau area is about 30 ha and has slopes between 0° and 5°. The hillslope area is about 12 ha and has slopes between 5° and 44°. The hillslope area includes a riparian area representing about 0.4 ha.

The vegetation in the catchment is dominated by deciduous hardwood trees (European beech and Oak), and conifers (*Picea abies* and *Pseudotsuga menziesii*). A riparian zone up to 3 m wide surrounds most of the stream and is covered by a variety of short vegetation including spinulose wood fern (*Dryopteris carthusiana*), yellow basalm (*Impatiens noli-tangere*), opposite-leaved golden saxifrage (*Chrysosplenium oppositifolium*), and wood-sorrel (*Oxalis acetosella*) (Martínez-Carreras et al. 2015).

The bedrock in the Weierbach consists mostly of Devonian slate of the Ardennes massif, phyllades, and quartzite (Juilleret et al. 2011). Pleistocene Periglacial Slope Deposits (PPSD) cover the bedrock and are oriented parallel to the slope (Juilleret et al. 2011; Juilleret et al. 2016). According the the WRB classification, the soil developed from the PPSD in the catchment is Leptic Cambisol (Humic, Ruptic, Dystric, Endosketelic, Siltic) (Juilleret et al. 2011). Many physical investigations (e.g. soil pits, core drillings, and porosity and particle size distribution measurements) were carried out in the catchment (Juilleret et al. 2016; Martínez-Carreras et al. 2016; Moragues-Quiroga et al. 2017). Two parts of the PPSD can be distinguished. From 0 to approximately 50 cm depth lies the Upper Layer (UL). The UL has a high average drainable porosity of 30% (Martínez-Carreras et al. 2016). The lower part of the PPSD is the Basal Layer (BL), between approximately 50 cm to 140 cm depth. Drainable porosity declines rapidly with depth from 30% to 10% in the BL (Gourdol et al. 2018). The coarse element content increases with depth in the BL from 30% to 75% between 50 cm and 90 cm depth, and to more than 80% between 90 cm and 140 cm depth. The orientation of the coarse material (slate stones) in the BL is mostly parallel to the bedrock surface slope (Juilleret et al. 2011). The weathered and fractured bedrock starts on average at about 140 cm depth, with fractures closing at approximately 5 m depth (Gourdol et al. 2018). The fresh bedrock starts deeper than approximately 5 m from the surface at most locations and can be considered impermeable.

The climate at the study site is governed by the interplay between seasonality in precipitation and in ET. Precipitation averaged 953 mm/yr over 2006-2014 (Pfister et al. 2017a) and is rather uniformly distributed throughout the year. The average runoff coefficient in the catchment was 50% for the period 2006-2014 (Pfister et al. 2017a). Evapotranspiration is higher in summer and creates a distinct seasonality in streamflow. Snow can accumulate for a few days in winter but it generally melts rapidly. Characteristic double peak hydrographs are observed during wetter periods (Martínez-Carreras et al. 2016; Schwab et al. 2018). The "first" peaks are flashy responses to precipitation events. They are always generated during a precipitation event. In contrast the "second" peaks are broader, delayed, longer lasting, and often occur after precipitation ended (figure 2.4). First peaks are often superimposed on second peaks during wetter periods.

2.2.4 Current understanding of streamflow generation

A perceptual model of the main flow paths through the Weierbach catchment was proposed by Wrede et al. (2015) who used a variety of field measurements and streamflow simulations. Here we update this understanding based on results from more recent experimental and modeling studies (figure 2.3). This understanding is used as hypothesis about streamflow generation that is tested with the results provided by the streamflow SAS functions.

In the current perceptual model, the first peaks of the hydrographs are generated by precipitation events causing saturation excess flow from the riparian soils into the stream (Klaus et al. 2015b). Additionally, overland flow in the riparian area and rain on the stream contribute to this first peak (Glaser et al. 2016). This is supported by hydrograph separation and end-member mixing analysis studies (Martínez-Carreras et al. 2015; Wrede et al. 2015), by observations of aerial diatoms in the stream (Antonelli et al. 2017; Klaus et al. 2015b; Pfister et al. 2009, 2017b), by high-frequency stream DOC observations (Schwab et al. 2018), and by thermal

infrared imagery (Glaser et al. 2018). Second peaks are generated by lateral subsurface flow above the PPSD/weathered bedrock interface (in the BL), where the PPSD is assumed laterally more conductive than vertically (section 2.2.3). The lateral subsurface flow is fed by infiltration and likely faster during larger precipitation events. It is initiated by a rise in groundwater levels, corresponding to a certain storage threshold in the weathered bedrock (Fenicia et al. 2014; Martínez-Carreras et al. 2016; Wrede et al. 2015) filled by accumulation of larger precipitation amounts during wetter periods. This is supported particularly well by streamflow simulations with a physically-based 3D model also constrained to surface saturation patterns (Glaser et al. 2016).

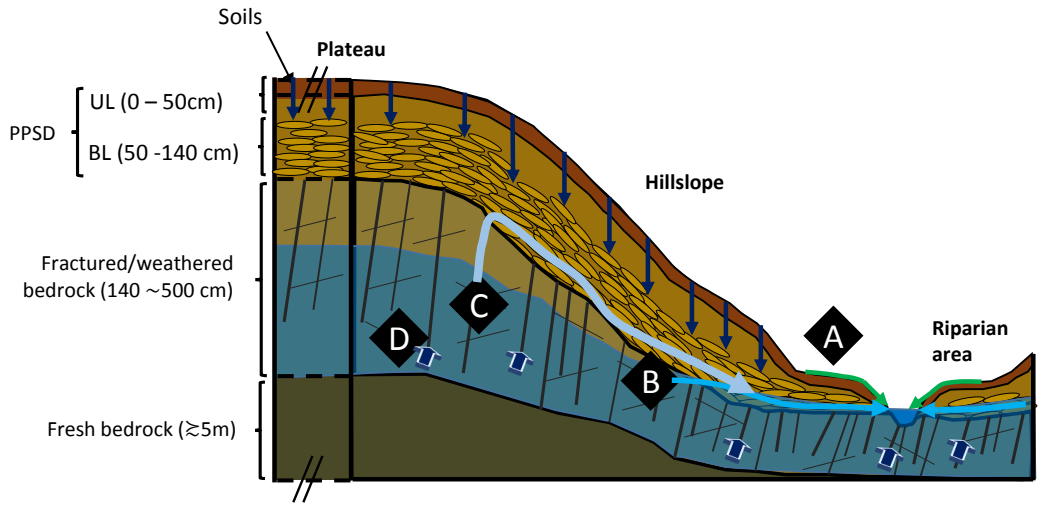

FIGURE 2.3: Perceptual model of the Weierbach catchment. Different contributions to the stream are labeled with diamonds (A to D). (A) corresponds to near stream contributions caused by saturation excess flow from the riparian soils and overland flow due to precipitation on stream plus saturated riparian areas. (B) is near-stream lateral subsurface saturated flow through the Basal Layer (BL) of the Pleistocene Periglacial Slope Deposits (PPSD) activated after a bedrock storage threshold is exceeded. (C) is the same mechanism as (B) but activated upslope. (D) is deep groundwater contributions from upstream via flows through the weathered and fractured slate bedrock with a cleavage oriented partly parallel to the stream. (D) is always active and probably sustains baseflow during extended drier periods, but it is negligible compared to lateral subsurface flow (B) and (C) during wetter periods. Flows through the PPSD are predominantly vertical (dark blue arrows) in the soils (in dark brown) and the Upper Layer (UL, in light brown). Two oblique bars in the plateau area indicate that it has a much bigger extent than the hillslope area compared to what the figure suggests.

2.2.5 Hydrometric and tracer data

We used hydrometric and tracer data from the Weierbach between October 2010 and October 2017. We focused our analysis on the period October 2015 – October 2017 with higher frequency measurements of ^{2}H. The data from October 2010 – October 2015 was used to spin-up the model (details in section 2.2.7). No major snow accumulation was observed in the catchment between October 2015 – October 2017. Therefore, snow processes are not considered in this study.

Precipitation was measured at 15 min intervals with a 200 cm^2 tipping bucket rain gauge (Young, model 52203) with 0.1 mm resolution. This device was installed in a forest clearing (red star in figure 2.2) in the Weierbach. A pressure transducer (ISCO 4120 Flow Logger) was used to record water levels at 5 min intervals at the V-notch at the outlet. Streamflow was calculated at 15 min intervals with a rating

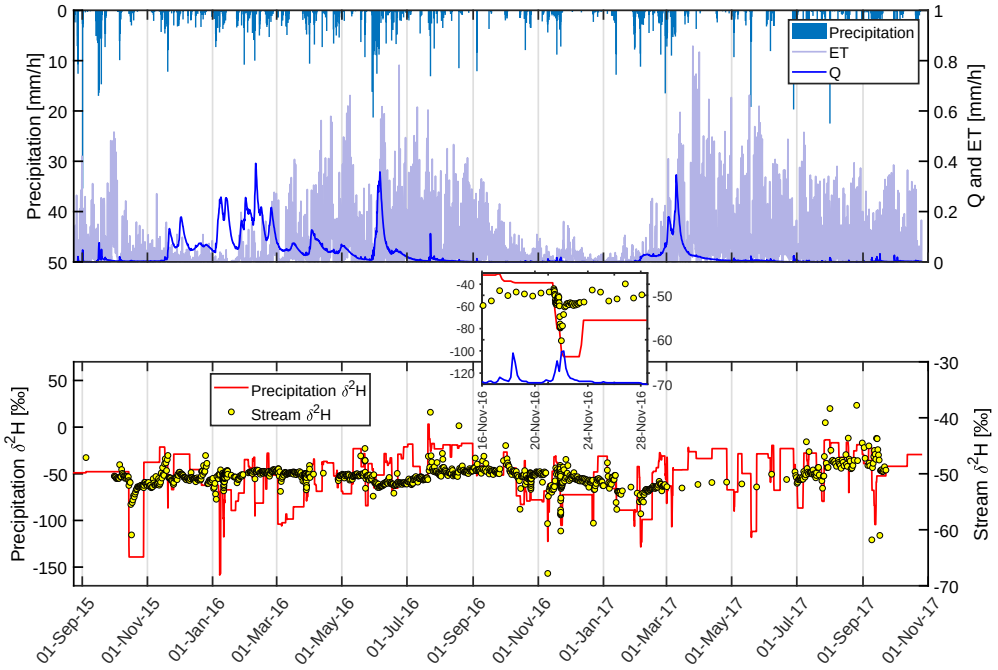

FIGURE 2.4: Hydrometric and tracer data used in this study. The upper subplot shows precipitation J, evapotranspiration ET, and streamflow Q. The lower subplot shows precipitation δ^2H and stream δ^2H. The inset shows a zoom on a typical first peak during a dry period in winter 2016 (streamflow peaking around 0.02 mm/h) that is barely visible on the upper subplot but clearly visible on the lower subplot. The axes units of the inset are the same as those of the lower subplot.

curve. Potential evapotranspiration (PET) was calculated at 15 min intervals using the reference evapotranspiration from the FAO Penman-Monteith approach (Allen et al. 1998). Details on the data used for the calculation of PET can be found in the data description of Glaser et al. (2016). Details on how actual ET was calculated from PET can be found in section 2.2.7.

Stream samples were collected at the outlet for two years (2015 – 2017) on a regular time basis, i.e. manually every second week, and automatically using ISCO autosamplers (AS, models 3700 and 6712) with a constant interval of 15 hours. Occasionally, the autosamplers were triggered by a water level rise and the sampling interval was reduced to capture flashy events. To minimize evaporation from the AS bottles, floating plastic pellets were placed in each bottle during summer. Precipitation samples were collected with a Sequential Rainfall Sampler (SRS) (c.f. Kennedy et al. 1979). The SRS collected rainfall samples approximately in 2.5 mm increments, i.e. a sampling interval of 23 hours in average. The sampling interval in precipitation can however range from a few minutes to two weeks depending on precipitation intensity. The full SRS bottles were sealed manually usually after a few days in the field. Almost all d-excess values for SRS samples were above 5, indicating no considerable fractionation by evaporation. Additionally, bulk precipitation samples were collected with a dip-in sampler approximately every two weeks in an open area 3.5 km away from the Weierbach outlet (following the IAEA/GNIP guidelines).

Aliquots of each sample were transferred to 2 mL glass vials using 0.45 µm hydrophilic polyethersulfone membrane filters (PALL Acrodisc® 13) within a few days after collection. The vials were sealed and stored in a fridge (5°C) until analysis. We analyzed the samples using a LGR Isotope Water Analyzer (model TIWA-45EP). The

deuterium ratios were converted to δ^2H values using the VSMOW reference. The analytical accuracy is taken equal to the accuracy of standards provided by LGR, i.e. 0.5‰in deuterium. The instrument precision, quantified as one standard deviation of the measured samples and standards, was maintained below 0.5‰in deuterium. This standard deviation was calculated from the last 5 injections of the samples, the 4 first injections being used only to remove the "memory effect" from the previous samples.

We aggregated the model input data to a resolution $\Delta t = 4$ hours. This resolution Δt was thus also used for the numerical discretization of the equations (section 2.2.7). $\Delta t = 4$ hours is small enough to allow the model to simulate the observed flashy δ^2H responses during events, since the sampling intervals in precipitation and in the stream are mostly above 4 hours. Smaller Δt values are possible, but they are unnecessary without higher frequency measurements, and they would increase the CPU time for the simulations considerably. We calculated the δ^2H input time series from the SRS samples and the bulk samples. We used the SRS data whenever possible, and the δ^2H values of the bulk samples were used only to fill gaps in the SRS data due to instrument failure. We interpolated the δ^2H values of precipitation between two consecutive samples (e.g. A and B) as being equal to the value of the next sample (i.e. B). For the SRS samples, the time needed to fill each bottle was determined from the measured precipitation amounts that sequentially filled the bottles (the sample volumes are known, usually ∼50 mL). For the bulk samples, this is not needed because they already represent a volume-weighted average from the time a clean empty bottle is installed until it is collected.

2.2.6 Model parsimonious development and parameterization of the SAS functions

Conceptual model structures that worked well for streamflow simulations in the Weierbach (Fenicia et al. 2014; Kavetski et al. 2011) failed to simulate deuterium responses (not shown). Essentially, lumped transport models that are not based on travel times need to use large fractions of event water at all times to simulate streamflow satisfactorily. This is in disagreement with the tracer measurements in the Weierbach suggesting larger pre-event contributions especially during wet periods (Martínez-Carreras et al. 2015; Wrede et al. 2015). This is particularly visible in the large damping of the stream δ^2H compared to precipitation δ^2H, except for short periods associated with first peaks (figure 2.4). The damped seasonal variations in stream δ^2H do not always follow the seasonal variations of δ^2H in precipitation (e.g. figure 2.4, Jan–Apr 2016, stream δ^2H decreases while precipitation δ^2H increases, and then the contrary in Jun–Jul 2016). Moreover, strong responses in δ^2H corresponding to the precipitation signatures are observed during first hydrograph peaks. These responses are added on top of the damped seasonal variations (figure 2.4). The δ^2H dynamics of the Weierbach can thus be qualified as complex, and may require models that specifically try to characterize the travel times to the stream.

Thus we used simulations relying directly on SAS functions applied to the catchment as a single control volume. Using a single control volume means that streamflow Q does not need to be simulated and can be treated as input data. We first tested commonly used SAS functions for streamflow, i.e. a power-law type (Benettin et al. 2017b), a beta distribution (van der Velde et al. 2015), and a gamma distribution (Harman 2015) (see appendix A). We first tested constant streamflow SAS functions,

and then time-variable SAS functions having parameters that change with catchment storage (as described in the papers cited above). Both these approaches could not reproduce the damped seasonal variations in δ^2H and at the same time the flashy responses in δ^2H. Thus, these SAS functions resulted in visually poor performance and Nash-Sutcliffe Efficiencies (NSE) close to 0 (see appendix A). Regardless of the low performance measures for these SAS functions, their simple functional forms and limited flexibility is not consistent with the apparent complexity of streamflow generation suggested by previous studies (see section 2.2.4). We thus interpreted these limitations as the fact that SAS functions with only one component (i.e. pdf) are unable to account for the range of occurring processes with contrasting travel times to the stream and for the flashy onset/offset of faster flow paths after precipitation events in the Weierbach. These simple models were therefore rejected mainly for their inconsistency with our "expert knowledge" in this catchment (Hrachowitz et al. 2014), used similarly to "soft data" to inform the choice of model structure beyond the limitations of "hard" performance measures (Seibert and McDonnell 2002).

Thus, we used composite SAS functions that are linear combinations of probability distribution functions (pdfs) (section 2.2.1). The complexity of the model (i.e. the number of pdfs and their parameterization) was increased step by step using process understanding until it could simulate satisfactorily stream δ^2H. We chose a gamma distribution for pdfs due to its flexibility, since its parameters can have meaningful units (here mm), and because it is defined on $[0, +\infty[$, which allows the use of age-ranked storage S_T as a variable (Harman 2015; Pangle et al. 2017). The gamma distribution is consistent with the observed $1/f^{\alpha}$ spectral density of stream chemistry (Aubert et al. 2014; Benettin et al. 2015b; Godsey et al. 2010; Harman 2015; Hrachowitz et al. 2015; Kirchner et al. 2001; Kirchner and Neal 2013; Kirchner et al. 2000). In a first step, we used a combination of two gamma distributions (e.g. Ω_2 and Ω_3, equation 2.9). This improved the simulations of the damped seasonal δ^2H dynamics, but the resulting NSE values were limited to 0.08 because the flashy δ^2H responses were still not reproduced (see appendix A). In the next step to account for onset/offset of shorter flow paths creating the flashy δ^2H responses, we added a component (Ω_1 in equation 2.9) in the streamflow SAS function for the younger ages, using a uniform distribution defined for S_T in $[0, S_u]$. S_u represents the upper limit of the storage available for Ω_1, which may be linked to the maximum length of the flow paths to the stream conceptualized by Ω_1. Using a uniform distribution avoids some numerical issues associated with the gamma distribution close to $S_T = 0$, when using shape parameters $k < 1$. As a result, the SAS function for streamflow used in this study is the following (in the cumulative form, cdf):

$$\Omega_Q(S_T, t) = \lambda_1(t)\,\Omega_1(S_T) + \lambda_2(t)\,\Omega_2(S_T) + \lambda_3(t)\,\Omega_3(S_T) \tag{2.9}$$

Ω_1 is a cumulative uniform distribution on the range $[0, S_u]$ that accounts for the flashy tracer response and is defined as:

$$\Omega_1(S_T) = \begin{cases} \frac{S_T}{S_u}, & S_T \in [0, S_u] \\ 1, & S_T > S_u \end{cases} \tag{2.10}$$

where S_u (mm) is a model parameter.

Ω_2 and Ω_3 are cumulative gamma distributions with mean parameter μ (mm) and scale parameter θ (mm), defined as:

$$\Omega_2(S_T) = \frac{\gamma(\frac{\mu_2}{\theta_2}, \frac{S_T}{\theta_2})}{\Gamma(\frac{\mu_2}{\theta_2})} \tag{2.11}$$

$$\Omega_3(S_T) = \frac{\gamma(\frac{\mu_3}{\theta_3}, \frac{S_T}{\theta_3})}{\Gamma(\frac{\mu_3}{\theta_3})} \tag{2.12}$$

γ is the incomplete gamma function, and Γ is the gamma function. The mean parameter μ is the product of the scale parameter θ and the more commonly used shape parameter k. μ and θ are treated as calibration parameters. Here we used μ because it has more meaningful units (mm) than k (no units). Furthermore, μ is visually intuitive since it is the expected value of S_T, around which the higher probabilities in Ω will generally be located. Using μ instead of k also gives more individual control on the expected mean ($k\theta$ simply becomes μ) and on the variance ($k\theta^2$ becomes $\mu\theta$) of the gamma distribution. Eventually, when randomly sampling these parameters, this makes it easier to avoid gamma components with $S_T > S(t)$, which causes numerical errors. The parameter ranges (table 2.1) correspond to shape parameters $k > 1$, giving the gamma distributions a humped shape that nevertheless allows the highest probabilities close to $S_T = 0$ for k tending to 1.

Ω_1 represents young water contributions for the considered values of S_u that are small compared to $S(t)$. The fraction of young water in the stream is thus directly affected by λ_1. Previous studies (Martínez-Carreras et al. 2015; Wrede et al. 2015) and stream δ^2H data suggested that this fraction of young water varies with catchment wetness and with the size of the precipitation events (figure 2.4). The last step of the model development thus consisted in making λ_1 time variant. We used the following equation (figure 2.5 and appendix A):

$$\lambda_1(t) = \lambda_1^* \left[f(t) + (1 - f(t))\, g(t) \right] \tag{2.13}$$

where f and g are functions with values on $[0, 1]$ defined as:

$$f(t) = f_0 \left[1 - \tanh\left(\left(\frac{S(t)}{S_{min} + S_{th}} \right)^m \right) \right] \tag{2.14}$$

$$g(t) = 1 - \exp\left(-\frac{\overline{\Delta S}(t)}{\Delta S_{th}} \right) \tag{2.15}$$

λ_1^* (no units), f_0 (no units), S_{th} (mm), and ΔS_{th} (mm) are calibration parameters (table 2.1). S_{th} is a storage threshold relative to the minimum storage, separating dry ($S(t) < S_{min} + S_{th}$) from wet periods ($S(t) > S_{min} + S_{th}$) during which second peaks occur (Martínez-Carreras et al. 2016). $m = 1000$ is a fixed coefficient found to yield a sufficiently sharp but smooth threshold behavior for $S(t)$ around the value $S_{min} + S_{th}$ (smaller values of m yield a smoother behavior). m has little effect on the simulations as long as it is sufficiently large (e.g. $m > 100$) to reproduce the intended threshold behavior. $\overline{\Delta S}(t)$ is calculated from the moving average of the storage variations $\Delta S(t)$ between $t - \Delta t^*$ and t, with Δt^* being a multiple of Δt. We chose $\Delta t^* = 2\,\Delta t = 8$ hours. This value is big enough to make $\overline{\Delta S}(t)$ more damped and smoother than $\Delta S(t)$ (larger Δt^* values make it too damped), while it is small enough to capture the strong storage variations during the catchment flashy streamflow events lasting in general a few hours to a day. Storage variations $\Delta S(t)$ are calculated from the modeled storage $S(t)$ at the resolution $\Delta t = 4$ hours fixed by the data resolution (see section 2.2.5):

$$\Delta S(t) = \Delta t \left[J(t) - Q(t) - ET(t) \right] \tag{2.16}$$

then:

$$\overline{\Delta S}(t) = \max \left(\frac{1}{3} \sum_{j=0}^{2} \Delta S(t - j\,\Delta t),\ 0 \right) \tag{2.17}$$

Essentially, $\overline{\Delta S}(t)$ is driven by large precipitation events and attenuated by high Q and/or high ET. The parameterization in equation (2.13) is a smooth version of two combined threshold functions based on storage $S(t)$ and on storage variations $\Delta S(t)$ respectively. The implemented smoothness reduces the numerical issues usually associated with thresholds in environmental modeling (Kavetski and Kuczera 2007). The following behavior is thus implemented (figure 2.5 and appendix A):

- During dry periods: $S(t) \leq S_{min} + S_{th}$ thus $f(t)$ tends to f_0.
- During wet periods: $S(t) \geq S_{min} + S_{th}$ thus $f(t)$ tends to 0.
- During large precipitation events (typically followed by first peaks): $\overline{\Delta S}(t) \geq \Delta S_{th}$ thus $g(t)$ tends to 1.
- During precipitation events with lower intensities (typically second peaks) or during recessions, storage variations are generally low: $\overline{\Delta S}(t) \leq \Delta S_{th}$ thus $g(t)$ tends to 0.

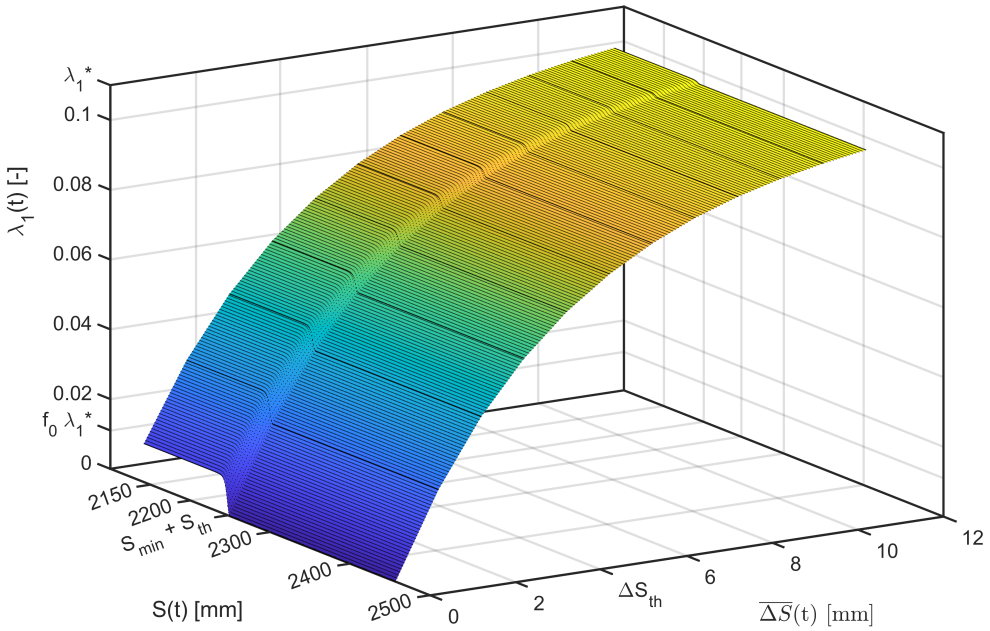

FIGURE 2.5: Relationship between $\lambda_1(t)$ and variables $S(t)$ and $\overline{\Delta S}(t)$ for all pairs of these variables between the minimum and the maximum values observed over 2015–2017 using the calibrated parameters of table 2.2.

As a result, $\lambda_1(t)$ varies during dry periods between a minimum of $f_0\lambda_1^*$ in the absence of strong precipitation events and a maximum of λ_1^* during strong precipitation events. $f_0 \in [0, 1]$ thus makes sure there is a minimum fraction of young water

in the stream during dry periods, consistent with the observed deuterium dynamics (e.g. July to September 2016 in figure 2.4) and previous work with other tracers (Martínez-Carreras et al. 2015; Wrede et al. 2015). During wet periods, $\lambda_1(t)$ varies between a minimum of 0 in the absence of strong precipitation events and a maximum of λ_1^* during strong precipitation events. The weights λ need to verify $\lambda_1(t) + \lambda_2(t) + \lambda_3(t) = 1$. For this we defined the third weight $\lambda_3(t)$ as $1 - \lambda_1(t) - \lambda_2(t)$. Finally, we considered that λ_2 is a constant calibration parameter (table 2.1). We set λ_2 constant as a first step to avoid introducing more parameters in the model than necessary. Eventually, this version of the model for Ω_Q, developed step by step to be as simple as possible, was found to be as complex as necessary to account for the superposition of streamflow generation processes and simulate δ^2H satisfactorily.

Equation (2.8) can be solved by choosing a functional form for the SAS function of ET. Stable isotope data from many catchments suggests that ET is not a perfect mixture of all soil water (Berry et al. 2018; Brooks et al. 2010). In addition, van der Velde et al. (2015) and Queloz et al. (2015b) showed the importance of having a SAS function for ET being different from random sampling for tracer simulations through hydrological systems. Therefore, we assigned a cumulative gamma distribution to the SAS function of ET:

$$\Omega_{ET}(S_T) = \frac{\gamma(\frac{\mu_{ET}}{\theta_{ET}}, \frac{S_T}{\theta_{ET}})}{\Gamma(\frac{\mu_{ET}}{\theta_{ET}})} \tag{2.18}$$

The gamma distribution was chosen due to its flexibility, for its definition on $[0, +\infty[$ allowing the use of age-ranked storage S_T as a variable, and to have parameter values and units comparable with Ω_Q.

2.2.7 Model spin-up and numerics

We initialize the simulation with an initial storage $S(t = 0) = S_{ref} > 0$. The hydrological measurements (J, Q, and ET) are balanced thus $S(t)$ fluctuates around its initial value S_{ref}. The effect of the value S_{ref} is removed by a sufficiently long spin-up period. Since the variations of $S(t)$ around S_{ref} are small compared to S_{ref}, S_{ref} is almost equal to the maximum of $S(t)$. We treated S_{ref} as a calibration parameter (c.f. Benettin and Bertuzzo 2018), and let it vary between 1500 and 2500 mm, consistent with a previous tracer-based estimate of the upper limit of catchment storage (Pfister et al. 2017a) and other field experiments. The initial deuterium mass in the system was calculated as $M(t = 0) = M_i = S_{ref} C_{ref}$ where C_{ref} was fixed at -50.6‰, the flux-weighted average stream δ^2H value (October 2015 to October 2017). Storage varies according to equation (2.4) which uses actual ET. We calculated ET from PET as:

$$ET(t) = PET(t) \tanh\left(\left(\frac{S(t)}{S_{root}}\right)^n\right) \tag{2.19}$$

Similar to m, $n = 20$ is a fixed parameter to smooth ET and avoid numerical instabilities. n has little influence on the simulations as long as small values (e.g. $n < 100$) are used. S_{root} (mm) is the model storage under which ET starts decreasing from $PET(t)$ to 0. Conceptually, S_{root} represents the storage of water available for soil evaporation and for plant transpiration until the water is held too tightly by its surrounding material. We chose $S_{root} = S_{ref} - 150$ mm. This way, S_{root} was not treated as a calibration parameter. This choice makes sure that $S(t)$ has realistic

variations around S_{ref}. Too small values of S_{root} would cause $S(t)$ to decrease continuously until it is smaller than S_{root}. As a result, ET would be overestimated for a relatively long period (e.g. several years). Too large values of S_{root} would cause $S(t)$ to increase continuously until it is bigger than S_{root} (ET would be underestimated). A similar technique to reduce ET with a similar limiting value for storage variations was used for example by (Pfister et al. 2017a) in the Weierbach and by (Fenicia et al. 2016) in the Attert basin.

Equation (2.8) requires an initial condition and a boundary condition. The boundary condition $p_S(0,t) = \frac{J(t)}{S(t)}$ was used to represent the role of precipitation in providing new water to storage (figure 2.1). The initial condition $p_S(T,0) = \frac{1}{T_0}\exp(-T/T_0)$ was used. It corresponds to an exponential initial distribution of residence times. We used $T_0 = 1.7$ years, the estimated Mean Residence Time (MRT) by (Pfister et al. 2017a). This reduces spin-up time, but the impact of the initial condition in equation (2.8) is removed by the model spin-up (100 years). The input data we used for the spin-up corresponds to the input data from October 2010 to October 2015 that we looped back over periods of 5 years. We discretized equation (2.8) in an explicit scheme using first order finite differences with respect to T and t in all derivatives of equation (2.8), keeping $\Delta t = 4$ hours and choosing $\Delta T = 8$ hours. Briefly, these choices of Δt and ΔT balance numerical precision with numerical stability, and represent a good compromise between the computational cost (i.e. CPU time) and the ability of the model to simulate high resolution dynamics of δ^2H.

2.2.8 Model calibration

Overall, 13 model parameters require calibration. These parameters are summarized in table 2.1. The parameter ranges were chosen based on their feasible values (e.g. f_0 by definition between 0 and 1), on estimates from previous studies (e.g. S_{ref} and S_{th}), on hydrological data (e.g. S_u and ΔS_{th} determined from average precipitation amounts), and on initial tests on the parameter ranges (e.g. μ and θ). These ranges guarantee a wide spectrum of shapes of SAS functions while reducing numerical errors due to for example truncation of the gamma tails.

TABLE 2.1: Parameters of the model to be calibrated

Symbol	Unit	Initial range	Description	Found in equations	Reference
S_{ref}	mm	$[1500, 2500]$	Initial storage in the model	2.4	2.2.7, Martínez-Carreras et al. (2016)
S_{th}	mm	$[80, 180]$	Storage threshold relative to the S_{min} separating dry and wet periods	2.14	2.2.6, Martínez-Carreras et al. (2016)
f_0	–	$[0, 1]$	Young water coefficient for dry periods	2.14	2.2.6
ΔS_{th}	mm	$[0.1, 5]$	Storage variation threshold for flashy events	2.15	2.2.6
λ_1^*	–	$[0, 0.5]$	Maximum value of the weight $\lambda_1(t)$	2.13	2.2.6
S_u	mm	$[1, 10]$	Range of the uniformly distributed Ω_1	2.10	2.2.6
λ_2	–	$[0, 1]$	Value of the weight $\lambda_2(t)$	2.9	2.2.6
μ_2	mm	$[300, 700]$	Mean parameter of the gamma distributed Ω_2	2.11	2.2.6
θ_2	mm	$[0, 200]$	Scale parameter of the gamma distributed Ω_2	2.11	2.2.6
μ_3	mm	$[700, 1750]$	Mean parameter of the gamma distributed Ω_3	2.12	2.2.6
θ_3	mm	$[0, 200]$	Scale parameter of the gamma distributed Ω_3	2.12	2.2.6
μ_{ET}	mm	$[300, 1100]$	Mean parameter of the gamma distributed Ω_{ET}	2.18	2.2.6
θ_{ET}	mm	$[0, 200]$	Scale parameter of the gamma distributed Ω_{ET}	2.18	2.2.6

We determined the parameter values using a standard Monte Carlo (MC) approach. This simple method was chosen because explicit numerical schemes (such as the one employed to solve equation 2.8) create numerical distortions (Kavetski and

Clark 2010) that partly prevent more advanced calibration approaches (e.g. DREAM, Vrugt 2016) from converging. N = 26,208 sets of the 13 parameters of the model were sampled randomly from uniform distributions with initial ranges defined in table 2.1. The model was run for each parameter set for a spin-up period of 100 years and for the period of interest: October 2015 to October 2017. The performance of the model was evaluated for the period October 2015–October 2017 with the Nash-Sutcliffe Efficiency (NSE) of C_{sim}, the simulated stream δ^2H:

$$NSE = 1 - \frac{\sum_{k=1}^{n}(C_{obs,k} - C_{sim,k})^2}{\sum_{k=1}^{n}(C_{obs,k} - \overline{C_{obs}})^2} \tag{2.20}$$

where k refers to time t_k, C_{obs} is the observed stream δ^2H, $\overline{C_{obs}}$ is the average C_{obs}, and n is the number of observations over October 2015–October 2017 (here n = 1,016). K = 1,000 parameter sets were retained to represent parameter uncertainties. These K behavioral sets correspond to the first 1,000 parameters ranked by decreasing NSE. We chose the NSE because it is frequently used in similar studies (allowing better comparisons) and because other objective functions (e.g. Kling Gupta Efficiency) did not yield good visual fits. Given that the high-frequency record of deuterium observations is only 2 years long, we used the whole time series as a calibration period, similar to other unsteady TTD studies with short tracer record lengths (Benettin et al. 2015b, 2017b; Hrachowitz et al. 2013; Rodriguez et al. 2018; van der Velde et al. 2015).

2.3 Results

Model performance varied between NSE = -0.58 and NSE = 0.27 for the 1,000 behavioral parameter sets. We selected the 3rd best parameter set ranked by decreasing NSE for further analysis (i.e. it has the 3rd highest NSE = 0.255). Besides having only a minor difference in NSE compared to the two better values (0.256 and 0.27), the chosen parameter set yielded a simulation that we considered visually more realistic than the two with higher NSE, because it is smoother and because it captures the damped seasonal variations in δ^2H better. Furthermore, the parameter values of this chosen set correspond better to the values with higher probabilities in the behavioral distributions (c.f. figure 2.7).

The calibrated model reproduced both the damped seasonal dynamics and the flashy event dynamics of stream δ^2H fairly well (figure 2.6). This is a noticeable improvement compared to streamflow SAS functions using only one component that were unable to do so, and resulted in visually poor performance and NSE values around 0 (see appendix A). We interpret this increased performance as the consequence of including step by step more process realism fed by experimental knowledge in the model, i.e. accounting for different ranges of travel times (via the two gamma components Ω_2 and Ω_3) that may be associated with faster and slower flows (or shorter and longer flow paths) in the hillslopes. This model also accounts for the onset/offset of young water contributions (via Ω_1). Yet the seasonality of δ^2H is not fully captured (e.g. Jun–Aug 2016). The large δ^2H amplitude of the flashy responses is also not reproduced at all instances (except for some behavioral solutions) due to the simple form of the young water translation, even though the modeled response is always concomitant with the observed one.

The ranges of values of the 1,000 behavioral parameters sets are generally similar to the initial ranges (table 2.2), suggesting noticeable parameter uncertainty. The

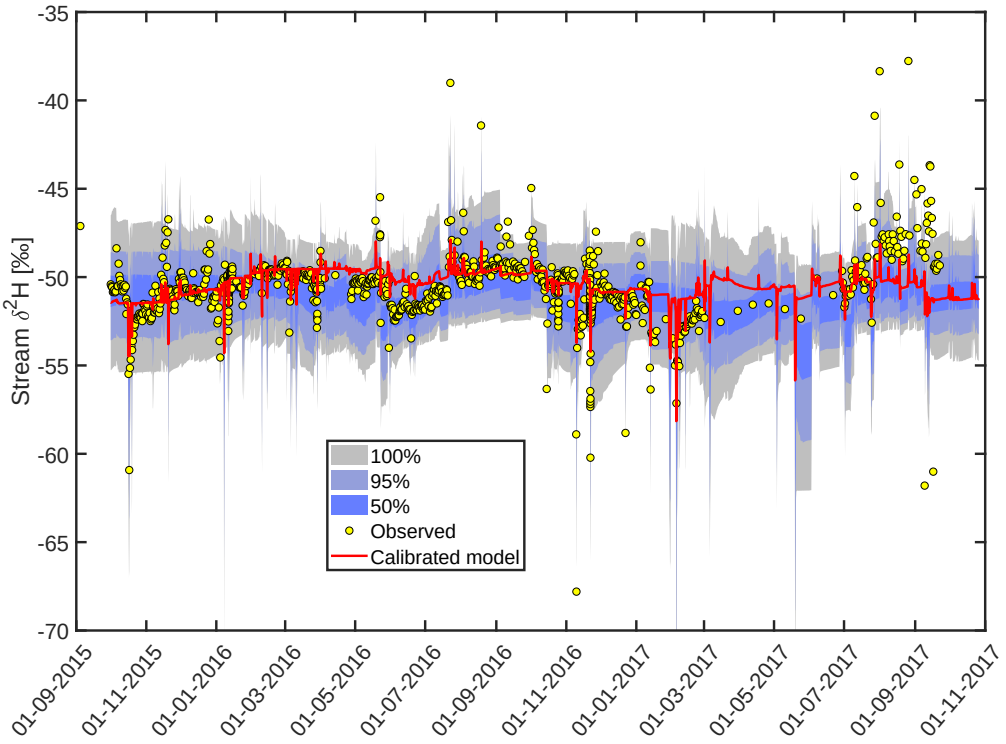

FIGURE 2.6: Simulated and observed stream δ^2H. The colored bands represent the uncertainties based on the 1,000 behavioral simulations: "100%" for all simulations, "95%" for the 2.5th to the 97.5th percentiles, and "50%" for the 25th to the 75th percentiles.

distributions of the behavioral parameter values (figure 2.7) reveal that some values were more likely to yield a better performance despite the rather wide parameter ranges. The parameter values are thus generally identifiable, except for the parameters ΔS_{th}, S_u, and θ_{ET}, whose distributions are rather flat over their entire range of values.

We calculated the time-averaged mean of the streamflow SAS function (in pdf form ω_Q for better visibility) for the October 2015 – October 2017 period (Figure 2.8). The mobilization of older water (right-hand part of the curve) is steady and dominated by the two gamma components ω_2 and ω_3. These have lower peaks than ω_1, indicating a much less selective mobilization of older water from storage than younger water. However, the peaks associated with ω_2 and ω_3 take a much larger area (almost 97%), indicating that streamflow is at all times volumetrically dominated by water older than ~2.8 months (figure 2.8). The mobilization of younger water by ω_1 is highly time-variant, and represents on average only about 1.7% of streamflow volume. It is active mostly during the flashy catchment responses (when $\lambda_1(t) > 0$), which explains why the average of ω_Q is 100 times lower than ω_1 for the youngest water. Nearly no water between 1 and 3 months old contributes to the stream. This is consistent with the damped seasonal dynamics of δ^2H showing little stream δ^2H response to the δ^2H of precipitation of the last weeks to months (figure 2.4). Eventually, the shape of the average ω_Q seems related to the multimodality of the streamflow and the stream δ^2H responses in the catchment (figure 2.4). Like streamflow, ω_Q has several peaks, and the second peak is much larger in

TABLE 2.2: Calibrated parameter values and behavioral range

Symbol	Unit	Initial range	Behavioral range[a]	Calibrated value
S_{ref}	mm	$[1500, 2500]$	$[1675, 2490]$	2426
S_{th}	mm	$[80, 180]$	$[82, 176]$	105
f_0	–	$[0, 1]$	$[0.02, 0.96]$	0.1
ΔS_{th}	mm	$[0.1, 5]$	$[0.29, 4.9]$	3.97
λ_1^*	–	$[0, 0.5]$	$[0, 0.38]$	0.11
S_u	mm	$[1, 10]$	$[1.2, 9.8]$	4.1
λ_2	–	$[0, 1]$	$[0.02, 0.86]$	0.32
μ_2	mm	$[300, 700]$	$[309, 691]$	575
θ_2	mm	$[0, 200]$	$[13, 196]$	34
μ_3	mm	$[700, 1750]$	$[717, 1543]$	1020
θ_3	mm	$[0, 200]$	$[9, 180]$	87
μ_{ET}	mm	$[300, 1100]$	$[304, 927]$	305
θ_{ET}	mm	$[0, 200]$	$[6, 195]$	77

[a] 2.5–97.5 percentile range of the distributions shown in figure 2.7

volume than the first one. However the first peak is much more visible for ω_Q than for streamflow, because it corresponds to the large amplitude of first peaks in δ^2H, made of up to 11% of the youngest water (when $\lambda_1(t)$ tends to λ_1^*).

We investigated the relationships between storage $S(t)$, streamflow $Q(t)$, and the Median Travel Time ($MdTT$) (figure 2.9). The $MdTT$ was used instead of the Mean Travel Time (MTT) because it is less sensitive to the uncertainties in the older water fractions. It was deduced from its definition satisfying the equation $\int_{T=0}^{MdTT} p_Q(T,t)\,dT = 0.5$. The $MdTT$ generally decreases for increasing $S(t)$, but the pattern is more complex than a simple linear relationship (figure 2.9). The relationship between the $MdTT$ and $Q(t)$ is however simpler and similar to an exponential one, where the $MdTT$ increases exponentially with decreasing streamflow.

2.4 Discussion

2.4.1 Does high-frequency stable isotope data imply a composite streamflow SAS function?

Until today, using only one pdf (or component) for SAS functions has been sufficient to simulate the tracer response in various catchments. High frequency tracer measurements recently supported the use of parameters that vary with catchment storage in the SAS functions (Benettin et al. 2017b; Harman 2015; van der Velde et al. 2015). Such an approach did not allow us to simulate satisfactorily the high-frequency stream δ^2H dynamics in the Weierbach (see appendix A). Streamflow SAS functions with only one component seemed unable to capture characteristic features of contrasting travel times: intermittent (on/off) flashy responses of young water superimposed to permanent volumetrically preponderant old water (figure 2.8). This old water itself consists of two ranges of travel times (figure 2.8), which are hardly captured by only one component in the streamflow SAS function. A composite streamflow SAS functions is an efficient way to model this multiplicity of ages, whereas commonly used non-composite distributions (e.g. beta or gamma) are not per se designed to represent a superposition of contrasting travel times, even with time-varying parameters. This is especially true for the on/off flashy δ^2H responses that required a part of the distribution to be active or inactive at certain

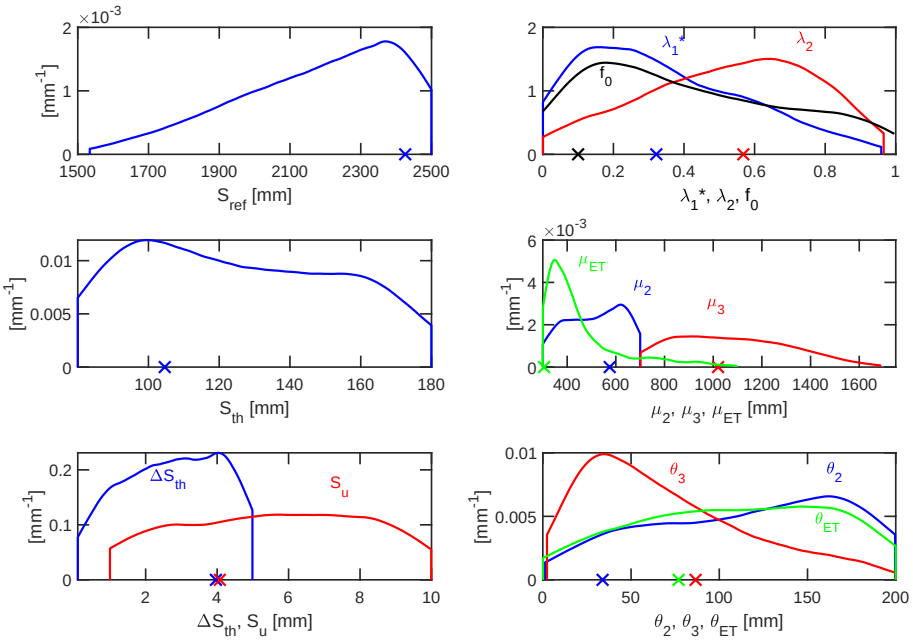

FIGURE 2.7: Probability distribution functions (pdfs) of the behavioral parameters. The pdfs were calculated by fitting an Epanechnikov kernel distribution (with default bandwidth) to the histograms of the first 1,000 parameter sets ranked by decreasing NSE (ranging from -0.58 to 0.27). The crosses indicate the corresponding chosen parameter values.

times. This could be conceptualized only with a time-varying weight (equation 2.13, figure 2.5, appendix A) representing the onset/offset of shorter flow paths. The stepwise adjustment of the streamflow SAS function, using process knowledge, resulted in a sufficiently flexible function (also complex enough) to allow a gain in NSE of roughly 0.2 for δ^2H. While the resulting performance is rather low (NSE=0.26), the improvement is relevant because it was obtained by matching the model flexibility with the complexity of streamflow generation inferred from field research (figure 2.3). In this way, the composite streamflow SAS function helped to confirm our hypothesis of streamflow generation. The increase in NSE by introducing a composite SAS function is also not the most important aspect of model development, because such an improvement in model realism cannot be judged reliably only by looking at performance metrics (Kirchner 2006). This is all the truer as better objective functions than the NSE may exist for the employed δ^2H time series here (see 2.4.4 and appendix A).

The results presented here are focused on a small headwater catchment with a particular subsurface configuration, thus generalizing the conclusions too far could be misleading. The validity of our approach will depend on the range of streamflow generation processes and their superposition. Composite SAS functions stand against the idea of parsimonious approaches and may in some cases introduce more complexity than necessary in the transport model (see 2.4.4). They may be required only when contrasting travel times to the stream are expected or observed. These may result from several contributing areas associated with different travel distances (e.g. Goderniaux et al. 2013), or from different water velocities from a same contributing area (e.g. Long and Putnam 2009). Composite steady-state TTDs have already been proposed to account for this (see references in Leray et al. 2016, p. 78).

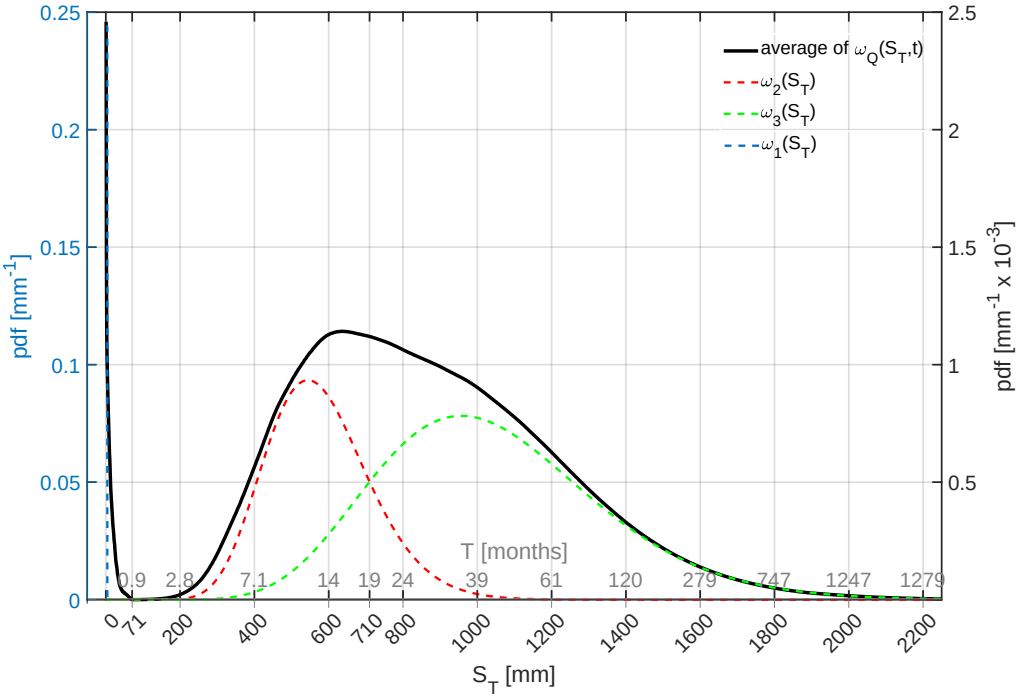

FIGURE 2.8: Time-average of the SAS function of streamflow ω_Q plotted against time-averaged age-ranked storage S_T with the corresponding average residence times T in months (averages are calculated over October 2015 – October 2017). We used a different axis (left axis, blue) for ω_1 due to its much larger magnitude than the average of ω_Q (right axis). Values on the left axis are 100 times higher than values on the right axis.

Using one pdf in the streamflow SAS function may thus be sufficient in catchments where different streamflow generation processes are associated with similar travel times, in catchments where some peaks of the streamflow TTD are associated with negligible water volumes (e.g. very narrow peaks), or where a single process dominates streamflow generation. This may happen for instance in larger catchments than the Weierbach, where small scale heterogeneities in flow paths and thus differences in travel times become negligible. In addition, coarser tracer time series (e.g. bi-weekly or monthly) may miss tracer responses associated with different flow paths and travel times, potentially making composite SAS functions unidentifiable. In any case, only inspecting the streamflow TTD and SAS function may not help to identify specific streamflow generation processes. Rather, adjusting the shape of the TTD and SAS function to the structure of the tracer data can be useful to confirm experimental knowledge about flow paths from a water age perspective.

Several studies used high frequency tracer data in this context. In the Hafren catchment in Wales (Harman 2015) a highly active fractured groundwater system hydraulically connects upslope areas to the valley (Shand et al. 2005). This is consistent with a gamma distribution centered on a travel time interval that shifts towards younger ages as the catchment gets wetter (i.e. Inverse Storage Effect) (Harman 2015). In the Bruntland Burn catchment in Scotland, an L-shaped power law distribution (Benettin et al. 2017b) is consistent with the dominant role of saturation overland flow from the peat soils in the valley that provide large fractions of young water to streamflow. In the Hafren and Bruntland Burn catchments, the oceanic climate with little seasonality in precipitation and low ET probably does not trigger

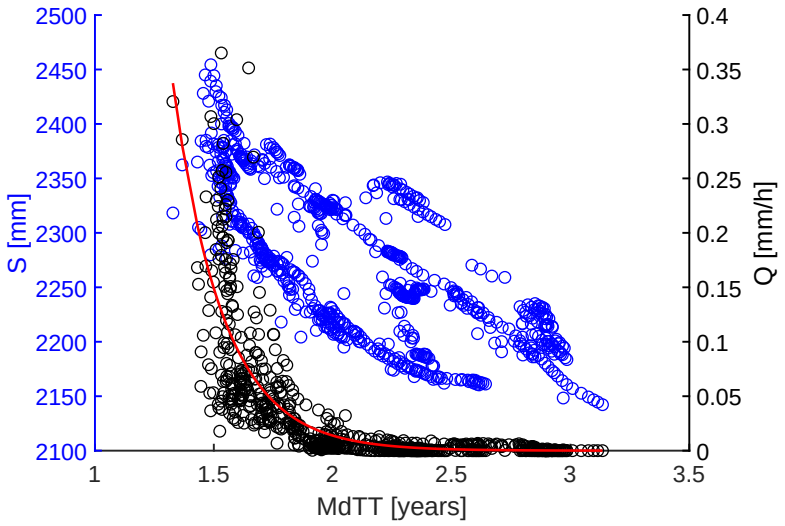

FIGURE 2.9: Relationships between the Median Travel Time ($MdTT$), catchment storage $S(t)$, and streamflow $Q(t)$. The red solid line is an exponential model of the form $f(x) = a\,e^{bx}$ fitted to the black circles.

large changes in the different flow path contributions and associated travel times across seasons, while this can be observed in a catchment with a Mediterranean climate (Rodriguez et al. 2018). In the Marshall Gulch catchment, an L-shaped beta distribution corresponds to rare but large streamflow responses dominated by water from recent and large precipitation events during the monsoon season well (van der Velde et al. 2015). Finally, in the Gårdsjön covered catchment G1, shallow but porous soils cover a mostly impervious bedrock, representing a small storage (Seibert et al. 2011) that allows a good mixing of water from many past precipitation events. The streamflow SAS function in G1 is thus similar to a uniform distribution (van der Velde et al. 2015). In conclusion, high-frequency tracer data does not always imply composite SAS functions. However, high-frequency tracer data can be used to test a hypothesis that complex streamflow generation mechanisms are superimposed in a catchment, by using composite SAS functions with increasing complexity (e.g. number of pdfs and parameterization).

2.4.2 Composite streamflow SAS function and superposition of streamflow generation processes in the Weierbach catchment

Previous studies suggested that different streamflow generation mechanisms are active in different parts of the Weierbach catchment. In particular, saturation excess flow from riparian soils and overland flow in the near-stream area (Glaser et al. 2016; Klaus et al. 2015b; Martínez-Carreras et al. 2016) were suggested to explain the high event water fraction during first peaks. Subsurface lateral flow initiated by a groundwater rise from the weathered bedrock into more conductive layers may explain the generation of delayed second peaks (Glaser et al. 2016), which consist mostly of pre-event water (Martínez-Carreras et al. 2015; Wrede et al. 2015). The complexity of the streamflow SAS function was gradually increased until high-resolution deuterium measurements were simulated satisfactorily. The resulting composite SAS function is consistent with the previous studies on the superposition of streamflow generation processes. Further identification or characterization of those processes might be possible with a distributed hydrological and transport

model (Glaser et al. 2016). The mechanisms generating streamflow (pictured as A, B, and C in figure 2.3) can be associated with the three components consisting of different travel times ω_1, ω_2, and ω_3, respectively. ω_1 conceptualizes the near-stream contributions generating the first peaks (A). ω_2 and ω_3 conceptualize the contributions by slow lateral subsurface flows at the PPSD/weathered bedrock interface. ω_2 corresponds to younger water likely to be stored closer to the stream (B) while ω_3 conceptualizes much older water likely with longer and slower flow paths from the hillslopes and the plateau area (B). Similar model performance could be obtained by using only one gamma component instead of ω_2 and ω_3, but the result would be unrealistic because most of the streamflow volume would be at least 3 years old, with a large part about being about 10 years old (appendix A). The decomposition of the streamflow SAS function into 3 components thus reflects the diversity of streamflow generation mechanisms that exist in catchments (Beven 2006; Dunne 1983). Furthermore, the shape of the calibrated ω_Q is consistent with both streamflow observations (showing a small first peak and a volumetrically dominating second peak), and with δ^2H observations (showing that first peaks have a larger proportion of young water creating the flashy responses in δ^2H).

In more detail, shortly after strong precipitation events, streamflow mobilizes the youngest 4.1 mm of water (parameter S_u in table 2.2), which is younger than 40 hours in average. First peaks are thus likely to be a mixture of current precipitation and water stored in near-stream soils (riparian area + hillslope foot) for less than 40 hours. This is further supported by the average weight of ω_1 in the composite SAS function ω_Q (1.7%) being close to the proportion of riparian area in the catchment (almost 1%). Larger pools of much older water always contribute to streamflow (figure 2.8), especially during second peaks. More than 2000 mm of water is stored in the catchment, with ages ranging from one month to several decades, and can contribute to streamflow. Two pools of water that are part of these 2000 mm can be distinguished. One pool that represents about 575 mm (parameter μ_2 in table 2.2, water between a few weeks and c.a. 2 years old) contributes more selectively to streamflow (higher values for ω_2 in figure 2.8). A second pool, twice that size, contributes less selectively to streamflow (lower values for ω_3 in figure 2.8), yet it provides larger amounts of water ($\lambda_3(t) > \lambda_2(t)$). These storage values are however uncertain because they were determined from stable isotopes, and will thus need further verification. Further work is needed to assess the added value of tritium for constraining these estimates (Stewart et al. 2010; Visser et al. 2019).

2.4.3 Parameterization of the streamflow SAS function

Previous parameterizations of SAS functions used linear functions of storage $S(t)$ or of a metric proportional to $S(t)$ (Benettin et al. 2017b; Harman 2015; van der Velde et al. 2015). In contrast, the parameterization of the streamflow SAS function in this study is nonlinear with respect to $S(t)$ (c.f. $\lambda_1(t)$ in equation 2.13). Moreover, previous studies suggested that SAS functions could vary with other state variables than $S(t)$ (Benettin et al. 2015b; Rodriguez et al. 2018). Here we found that $S(t)$ alone was not sufficient to parameterize the streamflow SAS function, since we additionally had to consider storage variations $\Delta S(t)$ (via a related metric $\overline{\Delta S}(t)$, equation 2.17) to simulate δ^2H in the stream. The parameterization of the streamflow SAS function was nonlinear with respect to $\Delta S(t)$ too. These nonlinearities suggest that water age fractions in streamflow (especially the younger fractions controlled by $\lambda_1(t)$) are not proportional to storage nor to storage variations. Especially, since $\overline{\Delta S}(t)$ is related to

the storm size, this means that young water fractions (Kirchner 2016a) may not be proportional to storm size.

The relationship between $S(t)$ and the $MdTT$ (figure 2.9) was thus not one-to-one, as seen in other catchments (Rodriguez et al. 2018; Yang et al. 2018). This means that $S(t)$ cannot be used to predict precisely the $MdTT$ and vice versa. Carrer et al. (2019) proposed to distinguish storage that is hydraulically connected to the stream $S_{CON}(t)$ from storage that is disconnected in the Weierbach and highlighted the exponential nature of the relationship between $S_{CON}(t)$ and $Q(t)$. In our work, a better relationship was found between $MdTT(t)$ and $Q(t)$ than between $MdTT(t)$ and $S(t)$. The almost exponential shape of the $MdTT(t)$–$Q(t)$ relationship (figure 2.9) suggests that the $MdTT$ may be better explained by the amount of hydraulically connected storage $S_{CON}(t)$ than by total storage $S(t)$. This is consistent with the findings of Carrer et al. (2019), who combined water balance and recession analyses to estimate $S_{CON}(t)$. We can interpret this connected storage as the amount of water stored in the weathered bedrock that becomes connected by lateral subsurface flow at the PPSD/weathered bedrock interface. Determining the location of the corresponding contributing areas and their temporal evolution would require a distributed model such as those employed by Glaser et al. (2016) in the Weierbach catchment or by Nippgen et al. (2015) at the Tenderfoot Creek Experimental Forest (central Montana). Alternatively, a dense network of local observations of subsurface storage such as water tables in groundwater wells (Jencso et al. 2009) or physical hillslope properties (Klaus and Jackson 2018) could be used to better understand the connectivity between hillslopes and the stream-riparian area at the catchment scale. Finally, we found a 105 mm storage threshold for the wet periods (table 2.2), which is in very good agreement with previous estimates based on measurements (Martínez-Carreras et al. 2016). This threshold value was thus further validated by the deuterium data.

This study therefore underlines that the "inverse" and "direct" storage effects (Harman 2015; Wilusz et al. 2017) usually defining the sensitivity of SAS functions to $S(t)$ alone need to consider other variables in the SAS functions and nonlinear parameterizations to better explain streamflow generation mechanisms (c.f. Rodriguez et al. 2018).

2.4.4 Limitations and way forward

Our work appears to use a large number of parameters (13 in total) for a solution based on a single control volume (here the entire catchment). Composite SAS functions introduced more parameters than other SAS functions (usually limited to 1–3 parameters), which goes against the need to reduce parameters. As emphasized by McGuire and McDonnell (2006), formulating a TTD that is representative of the true transport processes in a catchment is one of the most difficult tasks in travel time research. This observation is applicable to SAS functions as well. SAS functions need to express the entire range of time-varying transport processes involved in the generation of both Q and ET (Botter et al. 2010; Rinaldo et al. 2015). In a way, equation (2.8) can also be seen as a large system of coupled water balance equations that need to be solved simultaneously for as many control volumes as the considered number of age categories (usually a few hundred thousand). In that regard, a dozen parameters is a relatively small number. The number of parameters is furthermore justified by the failure of simpler transport models to simulate stream δ^2H (c.f. section 2.2.6 and appendix A). Finally, in this study over-parameterization and over-fitting are

unlikely given that the performance of the calibrated model (NSE = 0.255) is still rather limited, especially compared to the higher values (e.g. NSE > 0.5) reported in other studies simulating tracer transport using SAS functions (Benettin et al. 2017b; Harman 2015; van der Velde et al. 2015). Adding more components into the streamflow SAS function (hence more parameters) rather allowed realistic simulations of δ^2H (figure 2.6). Still, the performance of the model should be improved to better capture flashy responses (that are mostly underestimated) and for the damped seasonal variability (e.g. from May to July 2016 and from March to September 2017). For this, other types of pdfs for the components of the streamflow SAS function (e.g. beta pdfs) and/or using time-varying parameters in the components (as in Benettin et al. 2017b; Harman 2015; van der Velde et al. 2015) could be useful, especially if these parameters can depend on other variables than storage (like $\lambda_1(t)$). Furthermore, collecting precipitation samples for δ^2H at higher frequency, and decreasing the calculation time step Δt accordingly likely will improve the simulation of the flashy events. Finally, the composite SAS functions are still analytical approximations of a probably shape-free distribution (Danesh-Yazdi et al. 2018). Shape-free age distributions (e.g. Massoudieh et al. 2014) could yield better simulations and better correspond to the complexity of streamflow generation.

The increased number of parameters and their interactions resulted in the reduced identifiability of some parameters (figure 2.7). A simpler model (i.e. with fewer parameters) would have probably resulted in smaller uncertainties, but it would have larger errors since it would be unable to reproduce both the damped seasonal δ^2H dynamics and the flashiness of the δ^2H for many events (figure 2.6). The "bias-variance" trade-off refers to an optimum in model complexity that balances the increasing uncertainties ("variance") with the decreasing model errors ("bias") (Guthke 2017; James et al. 2013). A thorough evaluation of this trade-off for simulations of δ^2H dynamics would be valuable, but it is out of the scope of this paper. Nevertheless, this study shows that striving for simplicity in models may need to be balanced with flexibility (i.e. complexity) that allows for more processes to be accounted for. This is consistent with the opinion that "expert knowledge" (Gharari et al. 2014; Hrachowitz et al. 2014) or "soft data" (Seibert and McDonnell 2002) should be more often used to justify the choice of a particular model structure derived from experimental knowledge even if this means lower performance metrics, and that uninformed calibration of hydrological models does not help to gain understanding about the processes in a catchment. It is also worth pointing out that the NSE we employed here is commonly used to calibrate models to tracer data (e.g. Hrachowitz et al. 2013; Klaus et al. 2015a; McMillan et al. 2012b; van der Velde et al. 2015), whereas it is unlikely to be the most appropriate objective function. The NSE assumes that the errors between the model and the observations are normally distributed with a constant variance and no autocorrelation. Such assumptions are rarely verified a posteriori (Thyer et al. 2009) while in practice they are likely to be often incorrect (Schoups and Vrugt 2010). This is all the truer for the complex tracer time series presented in this study (appendix A). The characteristics of the δ^2H time series were challenging to simulate but highly informative about the transport processes in the catchment. Future work with similar δ^2H time series than the Weierbach may need to formulate a more appropriate objective function than the NSE. Similar tracer time series are likely to emerge in future travel time studies due to the technological progress in sampling methods (Pangle et al. 2013). New objective functions that are more adapted for the calibration of transport models to higher resolution stable isotope time series than in the past are urgently needed.

2.5 Conclusion

Simulations of the complex high-resolution δ^2H dynamics in the Weierbach were not satisfactory when relying on only one probability distribution function for the streamflow SAS function. Thus, we decomposed the SAS function into a weighted sum of pdfs, including a uniform distribution for the youngest water and two gamma distributions for the older water. We also included time variance in the SAS function and parameterized it using threshold behaviors based not only on storage, but for the first time also on a proxy of storage variations. Although this stepwise increase of complexity in the transport model goes against the trend to simplify models, it was informed by experimental knowledge and necessary for realistic δ^2H simulations in this catchment. The three components of the streamflow SAS function confirmed our hypothesis about a superposition of streamflow generation mechanisms. These mechanisms were identified in previous experimental studies in the catchment. The resulting SAS function thus offers a better correspondence between these mechanisms and a lumped description based on water ages. During flashy streamflow events (associated with flashy responses in stable isotopes), water younger than an average of 40 hours is mobilized, likely from near stream zones. Water with ages between a few weeks and a few years reaches the stream by lateral subsurface flow fed by storage in the weathered bedrock, representing likely more than 0.5 m of water. Water older than a few years likely coming from weathered bedrock storage further up in the hillslopes also contributes to streamflow by lateral subsurface flow. We found a relationship between travel times and the amount of storage connected to the stream (deduced from water balance and recession analyses) rather than total storage. These inferences were made possible by acquiring a high-resolution (sub-daily) 2-year δ^2H record in precipitation and streamflow, and by adapting the streamflow SAS function to the structure of the tracer time series showing complex dynamics. We expect that future studies will soon provide such tracer records. Composite SAS functions could thus become more often necessary to simulate more accurately the tracer response of some catchments and eventually to provide a deeper understanding of their streamflow generation mechanisms.

Acknowledgements

The authors acknowledge funding from the Luxembourg National Research Fund (FNR) in the framework of the STORE-AGE project
(FNR/CORE/C14/SR/8353440/STORE-AGE). The data used in this study is the property of the Luxembourg Institute of Science and Technology (LIST), and can be obtained from the corresponding author after approval by LIST. The authors would like to thank the members of the Catchment and Eco-Hydrology group at LIST, Gwenael Carrer, Marta Antonelli, and Jérôme Juilleret for the fruitful discussions about processes in the Weierbach catchment, Jean-François Iffly for the help in the field and for providing hydrological data used in this study, François Barnich and Gwenael Carrer for the help with the stable isotope analyses, Paolo Benettin for the very insightful discussions about SAS functions, and Uwe Ehret and Erwin Zehe for their advice. Finally, the authors would like to thank the two anonymous reviewers, the associate editor, and the editor for highly constructive feedback that helped improving the manuscript.

Chapter 3

Expanding travel time theory towards water age multimodality: Part 2

A revised version of this paper (minor revisions) was published by Wiley[1] as:

N. B. Rodriguez, P. Benettin, and J. Klaus (2020). "Multimodal water age distributions and the challenge of complex hydrological landscapes". In: *Hydrological Processes*. DOI: 10.1002/hyp.13770

Abstract

Travel time distributions (TTDs) are concise descriptions of transport processes in catchments based on water ages, and they are particularly efficient as lumped hydrological models to simulate tracers in outflows.Past studies have approximated catchment TTDs with unimodal probability distribution functions (pdf) and have successfully simulated tracers in outflows with those. However, intricate flow paths and contrasting water velocities observed in complex hydrological systems may generate multimodal age distributions. This study explores the occurrence of multimodal age distributions in hydrological systems and investigates the consequences of multimodality for tracer transport. Lumped models based on TTDs of varying complexity (unimodal and multimodal) are used to simulate tracers in the discharge of hydrological systems under well-known conditions. Specifically, we simulate tracer data from a controlled lysimeter irrigation experiment showing a multimodal response and we provide results from a virtual catchment-scale experiment testing the ability of a simplified unimodal age model to simulate a known and more complex multimodal age system. The models are based on composite StorAge Selection (SAS) functions defined as weighted sums of pdfs, that allow a straightforward implementation of uni- or multi-modal age distributions while accounting for unsteady conditions. Especially, the composite SAS functions allow a focus on multimodality that is not solely due to dynamics of inflows and outflows. These two experiments show that unimodal models provide satisfactory simulations of a given tracer. Yet, multimodal distributions are necessary to capture the detailed dynamics embedded in the observations. Furthermore, only the multimodal models are able to simulate tracers transported at time scales corresponding to the various peaks in the TTDs that are neglected by the unimodal models. Typical tracer data consisting of a single tracer measured in discharge and precipitation may not be sufficient to detect multiple peaks in the age distribution. Experimental knowledge of flow paths or the systematic use of data from multiple, independent tracers needs to be used to validate the assumption of a unimodal age distribution in the landscape of interest. Multimodal age distributions are likely to emerge in landscapes where the distributions of flow paths lengths or water velocities are themselves multimodal. Preferential flow in soils, geological heterogeneity, flow paths beyond topographical boundaries, and distributed water inputs in catchments are examples of cases where water age multimodality is likely to occur.

3.1 Introduction

The representation of hydrological systems in mathematical terms is a challenging task that deals with the complexity of the natural environment (e.g. Bras 2015; Clark et al. 2017; Dooge 1986). Spatially-implicit mathematical descriptions (e.g. lumped models) are widely used because they prove suitable to summarize hydrological processes at different scales (Hrachowitz and Clark 2017). Transfer functions that relate system inputs to system outputs by a convolution operation are particularly widespread in hydrology, notably for problems like flow and solute response of a catchment to precipitation events. These are two related but distinct problems that deal with either the celerity or the velocity at which perturbations move within a catchment (Kirchner et al. 2000; McDonnell and Beven 2014).

Transfer functions are typically modeled as probability density functions (pdf) whose shape is assumed a priori and the relevant parameters are calibrated to available data. When focusing on the hydrologic response of a catchment, the transfer function is termed instantaneous unit hydrograph (IUH) and its functional form is often assumed to be a gamma, lognormal or Weibull distribution. In some cases the shape of the IUH can be obtained starting from measurable properties of the landscape, such as its geomorphologic structure (Rinaldo et al. 1991; Rodriguez-Iturbe and Valdes 1979). This type of unit hydrographs can have complex and multimodal shapes that originate from the geomorphological complexity of a catchment.Tracer hydrology focuses on the variations in the concentrations of non-ideal tracers like solutes (e.g. chloride) or ideal tracers like constituents of the water molecule (e.g. ^{2}H) that are transported with water. The transfer functions that are relevant to tracers (ideal or not) are termed travel (or transit) time distributions (TTD). These describe the time that water molecules take to move through a hydrologic control volume (i.e. water ages in outflows) and they are used to relate tracer inputs (e.g. stable isotopes or chloride in precipitation) to tracer outputs (e.g. stable isotopes or chloride in discharge and evapotranspiration). Many studies employing TTDs as models of catchment transport have successfully simulated tracer concentrations in discharge by calibrating TTDs to tracer data under the assumption of steady-state flow conditions (see McGuire and McDonnell 2006). TTDs are often assumed to be gamma-distributed because this shape is consistent with the spectral signatures of measured solute timeseries (Kirchner and Neal 2013; Kirchner et al. 2000). Alternatively, various TTD shapes can be obtained by solving the advection-dispersion equation under different boundary conditions (Kreft and Zuber 1978). These TTDs have been employed in groundwater systems with specific aquifer configurations (e.g. Małoszewski and Zuber 1982). A comprehensive review of steady-state TTDs was provided by Leray et al. (2016), who illustrated the physical meaning and mathematical derivation of several TTD shapes.

If the assumption of steady-state flow conditions is removed (Botter et al. 2010, 2011), the resulting non-stationary (or instantaneous) TTDs are expected to have complex and irregular shapes even when the transport processes are simple (e.g. complete mixing). In transient hydrologic conditions, the shape of instantaneous TTDs cannot be assumed a priori, but StorAge Selection (SAS) functions (Botter et al. 2011; Harman 2015; Rinaldo et al. 2015; van der Velde et al. 2012) can be used to derive unsteady TTDs. SAS functions can take different mathematical forms and have usually been assumed to be beta distributions (van der Velde et al. 2012; Visser et al. 2019), power functions (Benettin et al. 2017b; Queloz et al. 2015b) or gamma distributions (Harman 2015; Wilusz et al. 2017). The resulting unsteady TTDs display various peaks corresponding to the variability of the hydrologic fluxes (i.e. precipitation or discharge). Yet, when instantaneous TTDs are averaged out to obtain an average (or marginal) TTD, simple shapes emerge (Benettin et al. 2015b; Remondi et al. 2018; van der Velde et al. 2015) and suggest that it is still possible to approximate transport processes in a catchment with a single (marginal) TTD.

Whether TTDs are derived by direct calibration to data, by averaging unsteady TTDs starting from SAS functions, or by solving simplified fluid equations, their shape is often found/assumed to have a single mode. This seems to suggest that integrating a large number of contrasting flow paths results, on average, in a seamless distribution of transit times. But should we expect TTDs to always have simple shapes in the real world? And what processes can possibly give rise to multimodality in the TTD? Hydrological landscapes transport water inputs over a wide range of

distances and in highly heterogeneous media with variable wetness conditions. It is unlikely that the existing plethora of transport processes always results in age distributions without multiple distinct and meaningful peaks. In fact, many studies (see Section 3.2) proposed to use multimodal age distributions in various hydrological systems for a number of physical reasons.

The problem of inferring the true shape of TTDs is made difficult by a major experimental issue: we cannot measure TTDs in catchments. TTDs can only be measured through tracer mass breakthrough curves, which is feasible in controlled systems like lysimeters (e.g. Pütz et al. 2016) or small experimental hillslopes (Kendall et al. 2001; Kim et al. 2016; Rodhe et al. 1996; Volkmann et al. 2018). At the catchment scale, TTDs can only be deduced indirectly through aggregated tracer information in inflows and outflows (Kirchner 2016a). In other words, hypotheses about the TTDs (e.g. functional form and parameters) cannot be tested by comparing the assumed TTDs to actual measurements of water ages, but only by comparing the tracer simulation associated with the assumed TTD to the tracer observations. The task of deducing TTDs in catchments is often transformed into a model calibration problem and this may overlook parts of the TTDs accounting for a range of transport processes not immediately visible in tracer data (figure 3.1). Indeed, there is no guarantee that an accurate tracer simulation implies an accurate TTD, unless direct measurements of water ages are available for validation. Until the day we are able to measure the true shape of an age distribution through a large-scale tracer experiment, research efforts are needed to assess how its simplification affects the simulation of tracer transport. This is particularly important as TTDs are seen as a way to allow hydrological models to simulate solute chemistry and bridge previously disconnected fields of environmental modeling (Hrachowitz et al. 2016).

This study explores the occurrence of multimodal age distributions in hydrological systems and investigates the potential of lumped transport models to account for the underlying causes and for the consequences of multimodality in tracer transport. We aim to clarify when we should expect to observe multimodal distributions and what is the effect of not accounting for this multimodality. We put a particular emphasis on the type of multimodality that is not due to time-varying hydrological fluxes by employing SAS functions and looking at marginal TTDs, which smooth the corresponding peaks out. The remainder of the paper is organized as follows: Section 3.2 reviews previous studies that reported multimodal TTD behavior. Section 3.3 reports the methodology to simulate tracer transport using TTDs and SAS functions. Section 3.4 provides modeling results for a lysimeter tracer experiment characterized by a bimodal tracer breakthrough curve, while Section 3.5 looks into the potential consequences of simplifying a multimodal TTD with a unimodal TTD by means of a catchment-scale virtual tracer experiment. A discussion of the results (Section 3.6) and a summary of the conclusions (Section 3.7) close the paper.

3.2 Acknowledging the multimodality of water age distributions

In various hydrological systems, TTDs are assumed to follow steady-state models like the piston flow, the exponential, and the dispersive models, or combinations of them (Małoszewski and Zuber 1982). These analytical distributions work well in many catchments (McGuire and McDonnell 2006; Stewart et al. 2010), where the tracers used to infer them are sampled at lower frequencies (mostly weekly to monthly).

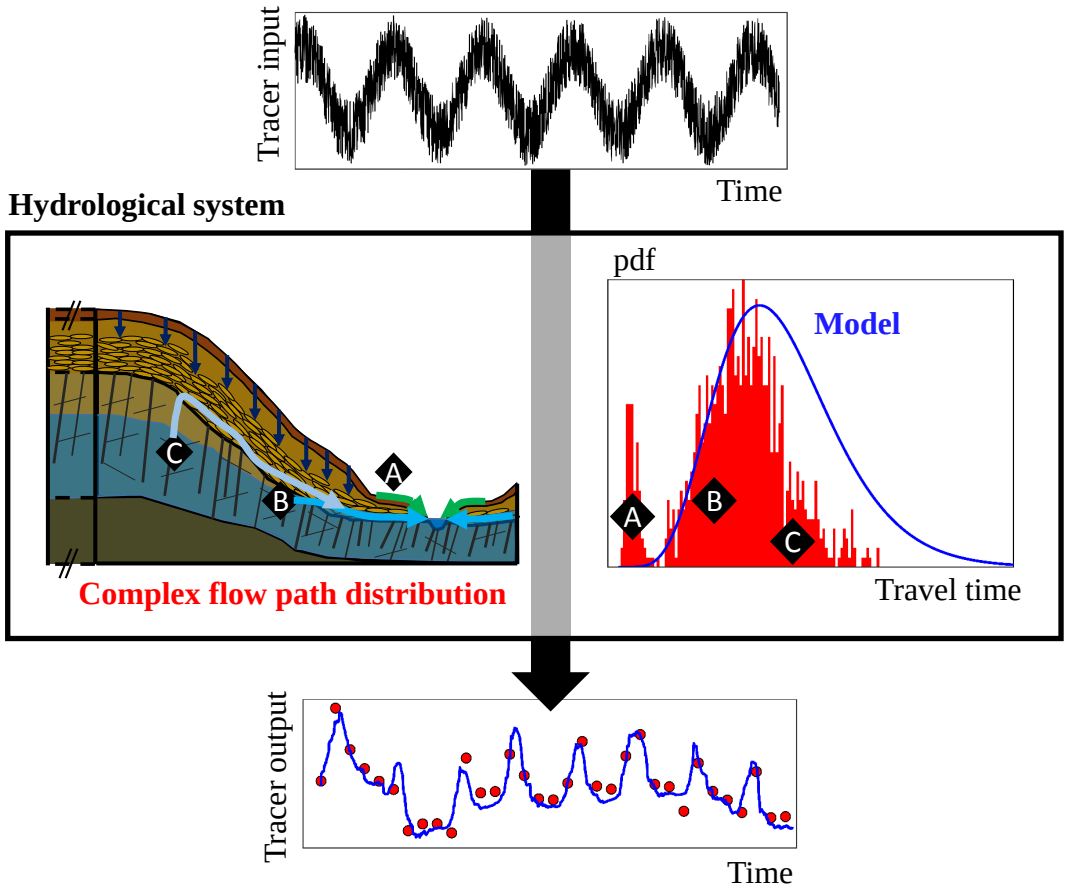

FIGURE 3.1: Visual summary of the process of modeling a hydrological system using lumped models based on TTDs and tracer input-output relationships. The complexity of discharge generation processes (A, B, and C) results in multimodal TTDs (red histogram). The TTD model (blue pdf) is a simplification of reality that works well for simulating (blue curve) a given tracer output measured in a discrete manner (red dots).

However, only strong assumptions about the system boundary conditions and its internal transport mechanisms (Leray et al. 2016; Małoszewski and Zuber 1982) can lead to such unimodal distributions. Multimodal and irregular shapes of TTDs are instead found with physically-based transport models, even in steady-state conditions (e.g. Engdahl and Maxwell 2014; Goderniaux et al. 2013).

A number of studies advocated the use of multimodal TTDs in the context of steady-state (Leray et al. 2016, p. 78, and citations below). At different scales, contrasts in water velocities or in flow path lengths may generate multimodal TTDs. Contrasting flow path lengths can for example come from: stream channel morphological features (Haggerty et al. 2002; Ward et al. 2018), competing recharge areas to a groundwater well (Cirpka et al. 2007; Massoudieh et al. 2012; Visser et al. 2013), different sources of surface water or groundwater (Duy et al. 2019; Stolp et al. 2010), or shallow and local vs. deep and regional subsurface flows (Goderniaux et al. 2013; Solomon et al. 2010; Wilusz et al. in review). Contrasting water velocities can originate from: preferential flow in soil macropores (Jackisch et al. 2017; Klaus et al. 2013; Scaini et al. 2017; Utermann et al. 1990; White et al. 1986), heterogeneous rock permeability (Desbarats 1990; Eberts et al. 2012; Ritzi Jr. et al. 2000; Weissmann et al. 2002),

or conduit flow bypassing rock matrix flow in karst geologies (Long and Putnam 2009).

In unsteady conditions, multiple peaks can be found in TTDs because of dynamic hydrological fluxes (Botter et al. 2010; van der Velde et al. 2012). However, clusters of peaks in the TTD can be related to different major transport processes to an outlet. van der Velde et al. (2010, Figure 6) found time-varying discharge TTDs with two distinct groups of peaks, the first one being intermittent and associated with tile drainage of the root zone storage (younger water) and the second being permanent and associated with slower groundwater flows in the subsurface. Recent numerical experiments on real or virtual hydrological systems (Danesh-Yazdi et al. 2018; Kaandorp et al. 2018; Remondi et al. 2018; Yang et al. 2018) showed that, even if SAS functions simplify the shapes of the age distributions compared to TTDs, they may not be described by unimodal distributions well. Rodriguez and Klaus (2019) used high frequency (i.e. sub-daily) stable isotopes in precipitation and discharge over 2 years and found that the SAS function in a slate catchment with complex flow paths may contain several peaks consistent with experimentally-identified discharge generation processes.

There is an emerging perception that TTDs in hydrological systems are complex and may be multimodal for other reasons than time-varying hydrological fluxes. In particular, different hydrological processes may generate different peaks in the age distributions, even when those are expressed as SAS functions. However, studies that supported TTD multimodality had only little or no tracer data to constrain model predictions and to validate the simulated TTD shapes. Thus, the existence of high-quality data is crucial to advance our knowledge of TTD shapes. The recent improvements in ^{2}H and ^{18}O measurement techniques (e.g. Gupta et al. 2009; Lis et al. 2008) and sampling methods (e.g. Berman et al. 2009; Herbstritt et al. 2019; Koehler and Wassenaar 2011; Munksgaard et al. 2011; Pangle et al. 2013) allowed a considerable increase in the tracer record lengths up to several decades (e.g. Hrachowitz et al. 2009b; IAEA 2019; IAEA and WMO 2019; Pfister et al. 2017a) and in the resolution of tracer data up to minutes (e.g. von Freyberg et al. 2017). This has made the detection of the entire shape of the TTDs more feasible than in the past. High-frequency tracer observations can now constrain the left-hand part of TTDs representing the short (e.g. hours to days) and intermediate (e.g. weeks) transport time scales (as in Chapter 2), while long records of tracers can constrain the right-hand tail of TTDs associated with long transport time scales (e.g. months to years, as in Chapter 4). Tracer timeseries that span several years at high resolution (e.g. Rodriguez and Klaus 2019; von Freyberg et al. 2018) will pave the way for future studies investigating water flow paths and their influence on TTDs.

3.3 Simulations of tracer concentrations through SAS functions

In this manuscript, we focus on the multimodality of the TTD of discharge, i.e. the TTD of all the water draining to a single outlet (e.g. bottom drainage of the lysimeter, virtual catchment discharge). In the following, "TTD" or "SAS function" therefore refer to the age distributions in discharge (as opposed e.g. to those of evapotranspiration ET) unless specified otherwise.

From a mathematical perspective, the TTD is the transfer function that verifies the equation:

$$C_Q(t) = \int_{T=0}^{+\infty} C_J(t-T)\, p_Q(T,t)\, dT \tag{3.1}$$

where $C_Q(t)$ is the tracer concentration in discharge at time t, $C_J(t-T)$ is the tracer concentration in precipitation at time $t-T$ and $p_Q(T,t)$ is the time-variant TTD, i.e. the distribution of water ages T (time spent since entry in the catchment) in discharge. The same equation applies to outlets other than discharge and their TTDs (e.g. the evapotranspiration flux ET the evapotranspiration TTD $p_{ET}(T,t)$). Equation 3.1 is valid only for "ideal" tracers (such as deuterium or oxygen 18 in certain conditions), for which the concentration associated with a water parcel is assumed to not change along its transport. In these circumstances, the TTD assumes the character of a lumped model representing a hypothesis about the transport processes to the outlet.

For non-ideal tracers such as solutes undergoing chemical reactions, the changes in concentration since the tracer entrance into the system need to be accounted for. Assuming that these changes are a direct function of water age T allows using equation 3.1 to simulate non-ideal tracers in outflows as well (Maher 2011; Rinaldo and Marani 1987). In such a case, $C_J(t-T)$ is replaced by $C_S(T,t)$, i.e. the tracer concentration of water in storage with age T at time t. For atmospheric tracers that are introduced through precipitation and decay with age such as tritium (Stewart et al. 2010; Visser et al. 2019), the following relationship can be implemented: $C_S(T,t) = C_J(t-T)\, e^{-T/\tau_{dec}}$, where τ_{dec} is a kinetic (decay) constant (e.g. 17.77 years for tritium). For geogenic reactive tracers like silicon and sodium (Benettin et al. 2015a; Clymans et al. 2013; Hornberger et al. 2001; Johnson et al. 1968; Maher 2010), where the input is typically tracer-free and tracer concentration can increase with age towards some limiting concentration C_{lim}, the storage concentration takes the form: $C_S(T,t) = (C_J(t-T) - C_{lim})\, e^{-T/\tau_{eq}} + C_{lim}$, where τ_{eq} is a kinetic constant (e.g. 1 year). Tracer concentrations in storage can also increase because of specific removal of tracer-free water by ET (evapoconcentration). The resulting effect on $C_S(T,t)$ is explained in detail by Queloz et al. (2015b) and can be formulated through an evapoconcentration parameter α, where $\alpha = 0$ represents pure evapoconcentration and $\alpha = 1$ represents fully conservative tracer behavior (Bertuzzo et al. 2013).

The unsteady TTDs $p_Q(T,t)$ can be obtained by solving a water balance equation for each water parcel (usually referred to as age master equation (Botter et al. 2011)). Numerical solutions to the age master equation (Benettin and Bertuzzo 2018; Harman 2015; Rodriguez and Klaus 2019; van der Velde et al. 2012; Visser et al. 2019) usually consider one control volume for the entire hydrologic system and use the "age-ranked storage" $S_T(T,t)$ as state variable. S_T represents the distribution of water volumes in storage, ranked by age. For instance, $S_T = 10$ mm for $T = 1$ day means that 10 mm of water in storage is younger than 1 day. S_T changes with time because water is aging, because new water is introduced by precipitation J, and because ages in storage are removed by outflows Q and ET. These outflows are associated with SAS functions Ω_Q and Ω_{ET} representing their preference for removing certain ages in storage. Time-varying TTDs are obtained by evaluating the SAS functions over the age-ranked storage:

$$p_Q(T,t) = \frac{\partial}{\partial T}(\Omega_Q(S_T,t)) \tag{3.2}$$

In some circumstances, it may be convenient to consider the normalized form of the ranked storage $P_S = S_T/S(t)$ and use SAS functions $\Omega_Q(P_S,t)$ that are defined

over the interval [0,1], where 0 corresponds to the youngest parcel in storage and 1 to the oldest. Tracer concentrations are obtained by coupling equation 3.2 to equation 3.1 after imposing a functional form to the SAS functions. Examples of SAS function shapes are the power function (Benettin et al. 2017b), the beta distribution (van der Velde et al. 2015), the gamma distribution (Harman 2015), or any of the functions described in section B.1. In particular, we tested composite SAS functions written as a weighted sum of probability distributions. To test our results for a larger number of TTD shapes, we used SAS functions defined both over the age-ranked storage (Section 3.5) and over its normalized form (Section 3.4).

3.4 Multimodality in a lysimeter-scale experiment

Small-scale (∼2m) experimental setups like lysimeters (e.g. Pütz et al. 2016) are currently the only hydrologic system where TTDs can be measured by labeling individual precipitation events. While tracer experiments in lysimeters often resulted in substantial single-mode tracer breakthrough curves (BTC) (e.g. Groh et al. 2018; Stumpp et al. 2009), the experiment by Queloz et al. (2015a) is one rare example showing a clear bimodal behavior with two distinct peaks in the BTC.

3.4.1 Data and Methods

Extensive information about the experiment including hydrologic and tracer data can be found in Queloz et al. (2015a). In brief, the experiment was carried out at EPFL campus (CH) between 2013 and 2014 in a large (2.5 m-depth, 1.2 m^2-diameter) vegetated lysimeter. The experiment was forced by controlled, nonstationary irrigation. Two small willow trees planted within the lysimeter triggered intense evapotranspiration fluxes that regularly exceeded 30 mm/d during warm summer days. Flow from the bottom drainage was discontinuous, with peaks following irrigation and no flow during dry days (Figure 3.2). A number of five fluorobenzoate (FBA) tracers were applied in Spring 2013 and showed different recoveries. We focus here on the tracer that had the most conservative behavior (2,6-DFBA, 60% recovery). The tracer arrived at the bottom of the lysimeter over two distinct timescales that resulted in two different major peaks with duration of weeks to months (figure 3.2). It is worth to recall here that the TTD is the normalized mass breakthrough curve (MBTC) and not the concentration breakthrough curve (CBTC). The MBTC is influenced by the variability of the hydrologic fluxes and would resemble the drainage timeseries, with multiple individual peaks corresponding to high flow events. Yet, the CBTC better reveals the two longer-lasting peaks and we consider those as representative of contrasting transport processes occurring within the lysimeter. The major of the two peaks occurred through the summer (roughly 3 months after application) when flow was very discontinuous, while the other peak occurred at the beginning of the winter (roughly six months after application), when evapotranspiration became significantly lower. Although the second peak is smaller in magnitude compared to the first peak, it drained more tracer mass because flows were higher.

Tracer transport for this experiment was previously modeled by (Queloz et al. 2015b), who used an approach based on SAS functions and conceptualized the lysimeter as two systems in series: the unsaturated soil (modeled through a power-function SAS) and the saturated gravel filter (modeled as a fully-mixed compartment). In this study, we use one single control volume for the entire lysimeter system (unsaturated and saturated) and we test a new functional shape for the SAS function $\Omega_Q(P_S)$. We

introduce the truncated normal SAS function $\Omega_T(P_S, m, s)$ (formula in B.1.1), which has 2 parameters: m, corresponding to the mode of the distribution and s, which is proportional to the standard deviation of the distribution. Depending on the value of the parameters, the normal SAS function can be bell-shaped or not. Note that even a symmetric SAS function does not imply that the resulting TTDs are symmetric. As the parameters of the normal SAS function are strongly related to the function mode, it is an appealing candidate to explore multimodality. Thus, we generated a composite SAS function as the linear combination of two normal SAS functions (see section B.1.1). The composite function allows for 2 distinct peaks and has a total of 5 calibration parameters (2 for each normal distribution and 1 for the partitioning). For comparison, we also used a single normally-distributed SAS function (2 parameters) and a power-function (1 parameter) for Ω_Q. The SAS function for the ET flux (Ω_{ET}) is assumed in all cases to be a uniform distribution over the younger storage fraction defined by the parameter u, as proposed by Harman (2015).The model further assumes that the tracer is conservative but only partially taken up by vegetation, through a parameter $\alpha \leq 1$ (see Section 3.3).The simulation was run using *tran*-SAS (Benettin and Bertuzzo 2018) and calibrated to measured tracer data through the DREAM$_{ZS}$ algorithm (ter Braak and Vrugt 2008; Vrugt et al. 2009). Calibration was based on the assumption that residuals were independent and normally distributed. Further details on the model and calibration setup can be found in Section B.2. For each of the 3 tested SAS models, the best parameter combination (Table B.1) was retained to obtain the instantaneous TTDs and compute the marginal TTD.

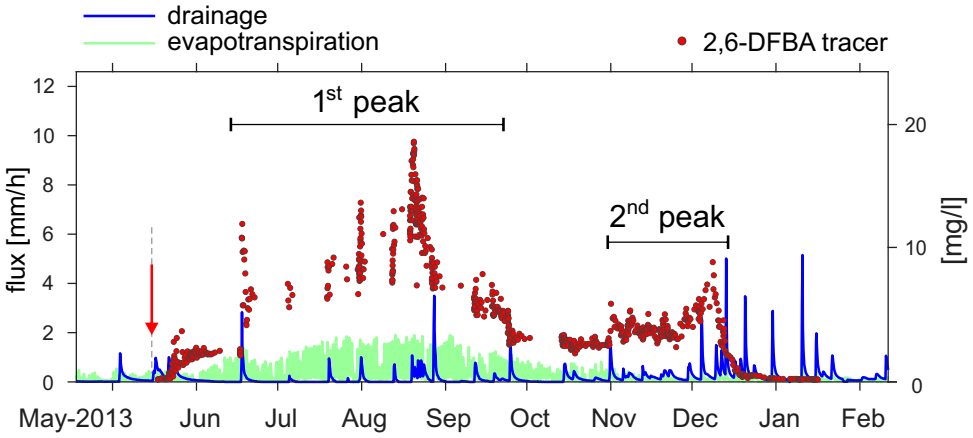

FIGURE 3.2: Experimental data from the lysimeter experiment by Queloz et al. (2015a). The vertical arrow marks the tracer application in mid May 2013. Tracer measurements highlight the occurrence of two major peaks in the tracer breakthrough: the first one occurred through the summer when flow was very discontinuous; the second occurred at the beginning of the winter when evapotranspiration became considerably lower.

3.4.2 Results

The results of the tracer breakthrough simulations are illustrated in Figure 3.3 together with their corresponding SAS function Ω_Q. Breakthrough curves are discontinuous because of the frequent absence of flow during the dry summer months. The three curves all match the beginning of the early breakthrough in mid-May 2013. This 1-month delay since tracer application is obtained through SAS functions that, in all cases (see Figure 3.3 inset), are almost null up to P_S=0.2, meaning that drainage does not comprise the 20% younger soil waters. The three simulations are also fairly

similar at reproducing the first wide peak from June to October, with the power-function simulation less-variable and slightly higher than the tracer peak, but overall broadly comparable with measurements. The main difference occurs upon the second peak, which is entirely missed by the simpler models while it is captured by the double normal SAS function. The higher accuracy obviously comes at a cost (5 SAS parameters instead of 1 or 2) but, besides a much better performance (see additional analyses in Section B.2), the double normal model is more appropriate because it allows expressing the multimodality of the data. As a result, model errors are symmetric and much less correlated than in the other cases (Figure B.2 and Table B.2). The marginal TTDs resulting from the three models are illustrated in Figure 3.4. As expected, the marginal TTD resulting from the multimodal SAS function has two distinct peaks: the first at 90-100 days and the second at 200-250 days. By contrast, the marginal TTDs resulting from other SAS functions have a single and higher peak at 90-100 days and underestimate the older water contribution in the right tail of the distribution. These simulations ultimately tell us that in all cases, water found at the bottom drainage is representative of the older water stored in the lysimeter (Figure 3.3, inset). However, two major transport mechanisms emerge from the data and the simulations: some water is relatively fast and it is able to percolate during the dry summer season (c.f. first peak of the CBTC), likely due to the occasional activation of preferential pathways; the rest of the tracer mass moves with the matrix flow and takes more than twice as much to reach the lysimeter bottom (c.f. second peak of the CBTC).

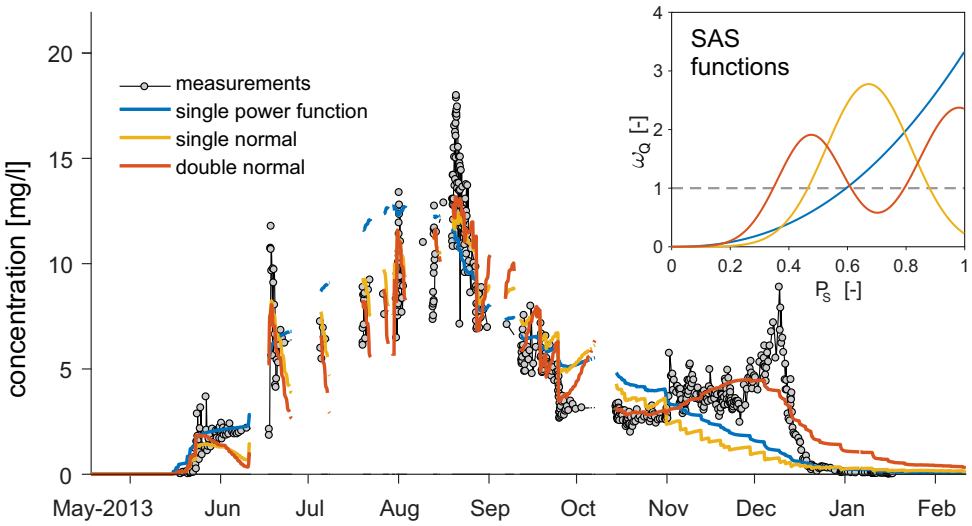

FIGURE 3.3: Tracer concentration in the lysimeter drainage as simulated by the calibrated models. Curves are discontinuous during periods with no flow. The inset reports the calibrated SAS function for each model.

3.5 Multimodality in a virtual, catchment-scale experiment

The tracer breakthrough experiment shown in section 3.4 cannot be easily applied at larger scales (e.g. hillslopes, catchments). Thus, the applicability of these results and the validity of these conclusions at larger scales require a numerical experiment.

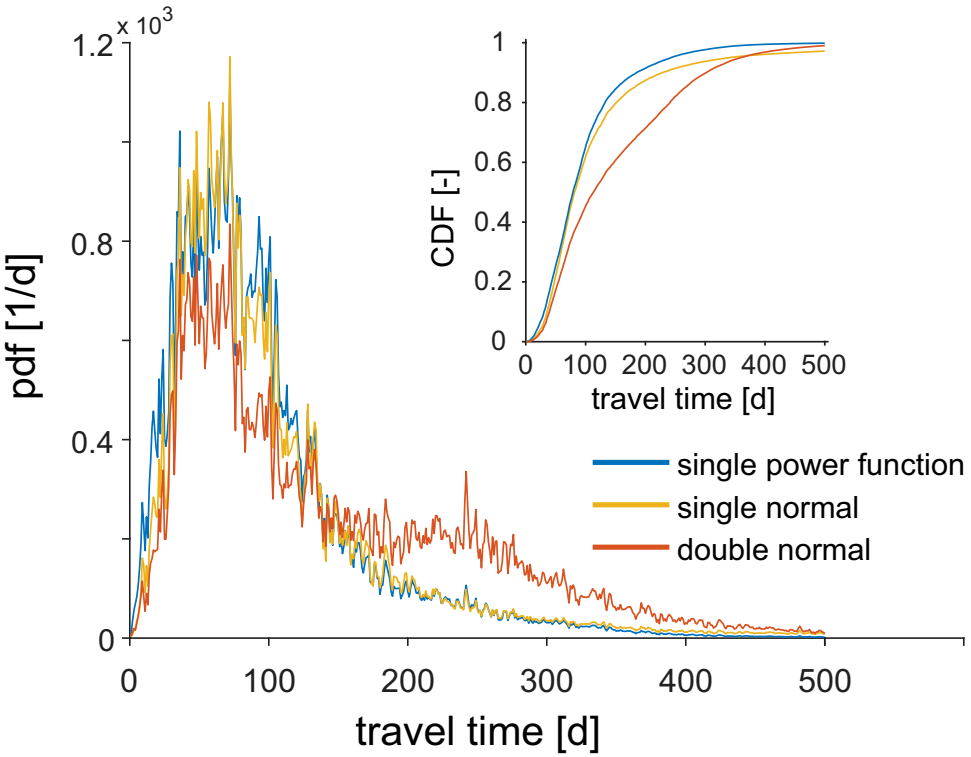

FIGURE 3.4: Marginal travel time pdf (main figure) and CDF (inset) computed from the calibrated models.

3.5.1 Methods

We consider a virtual catchment with an arbitrary scale, with known fluxes J, ET, Q, and with known deuterium content in precipitation C_J. These fluxes and tracer concentration have a yearly seasonality and have random daily variations (appendix B.3). The true transport processes to the stream are summarized by a constant, bimodal composite SAS function $\Omega^{ref}(S_T)$, made of a sum of two gamma distributions Ω_1 and Ω_2 corresponding to a young water component and an old water component (equation B.6). Here, age-ranked storage S_T is employed to allow the use of gamma distributions. To reduce the number of variables, we used a uniform distribution (random sampling of all ages) for the SAS function of ET (Ω_{ET}). The "observed" deuterium composition of discharge is generated according to equations 3.1 and 3.2 using the multimodal Ω^{ref} as SAS function. We tested the ability of a simpler, unimodal SAS function Ω_Q (one constant gamma distribution, equation B.5) to simulate the observed stream δ^2H. The two parameters of Ω_Q were calibrated against observed stream δ^2H using a Monte Carlo procedure and the Nash-Sutcliffe Efficiency to evaluate model performance (appendix B.3).

3.5.2 Results

The unimodal gamma-distributed Ω_Q simulated the observed stream δ^2H well, with a Nash-Sutcliffe Efficiency of NSE=0.89 (figure 3.5a). The model only failed to reproduce some of the rapid δ^2H responses, especially at the beginning of wetter periods. The corresponding calibrated Ω_Q is similar to the old water component Ω_2 of the SAS function Ω^{ref} (figure 3.5). As such, it clearly neglects the young water component Ω_1 ($S_T \lesssim 200$ mm), and it overestimates the old water fractions ($S_T \gtrsim 600$

mm) compared to Ω_2. This is also visible in the corresponding marginal TTDs (average TTD over 100 years) (figure 3.6), which have the largest differences over both younger ($T \lesssim 0.2$ yr) and older ($T \gtrsim 1$ yr) water components.

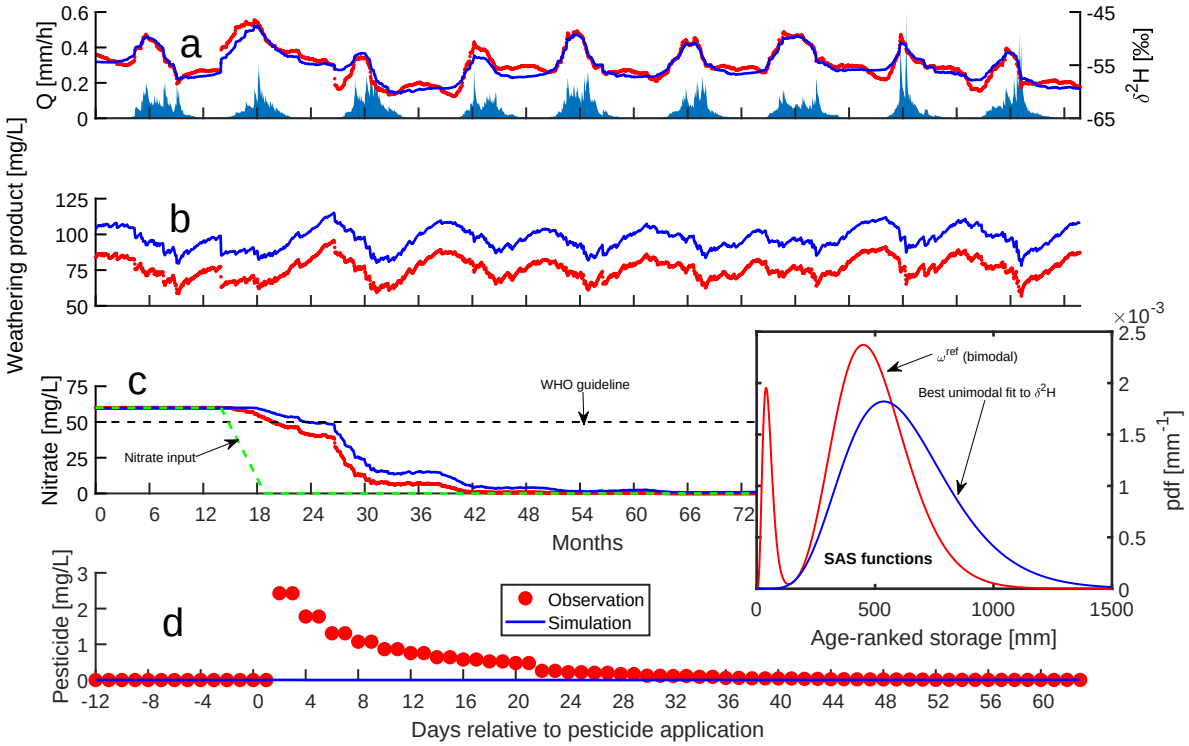

FIGURE 3.5: Simulations (blue curves) and observations (red circles) of tracers in the stream of the virtual catchment. The "observation" corresponds to the reference bimodal SAS function Ω^{ref}, while the "simulation" corresponds to the best fit to δ^2H using a gamma distributed Ω_Q (unimodal). (a) Simulation calibrated to the observed δ^2H (NSE = 0.89), (b) associated concentrations in weathering product ($C_{lim} = 200$ mg/L, $\tau_{eq} = 550$ d), (c) associated NO_3^- concentrations, and (d) associated pesticide concentrations ($\tau_{dec} = 50$ d). The nitrate input decreases linearly from 60 mg/L to 0 mg/L in 150 days. The dashed line in (c) indicates the WHO guideline of 50 mg/L as NO_3^-. The pesticide instantaneous input occurs at day 2 as 100 mm of precipitation at 50 mg/L.

The observed differences between Ω^{ref} and Ω_Q did not have a large influence on the simulations of the stream δ^2H. To assess the consequences of those differences for other tracers, we compared the simulations from those two models for 3 additional solutes in the stream:

- A virtual pesticide with a short kinetic (decay) constant $\tau_{dec} = 50$ days ($<$ 0.5 MTT, table 3.1), applied as a single pulse at 50 mg/L during a 100 mm precipitation event.
- Nitrate (NO_3^-), with a long term input of 60 mg/L, then linearly decreasing from 60 mg/L to 0 in 150 days ($\simeq$ 0.5 MTT, table 3.1).
- A weathering product (e.g. silicon) with a long kinetic constant $\tau_{eq} = 550$ days ($\simeq$ 2 MTT, table 3.1) and limit concentration $C_{lim} = 200$ mg/L.

These additional tracers have Damköhler numbers between roughly 0.5 and 6 for the considered catchment (ratio of MTT to kinetic constant, see Ameli et al.

2017; Maher and Chamberlain 2014; Oldham et al. 2013), representing from reaction-limited to transport-limited tracer dynamics, respectively. The observed differences between the unimodal Ω_Q calibrated to δ^2H and the bimodal Ω^{ref} are well reflected in the simulations of these 3 tracers. Because of the mismatch in old water fractions ($T \gtrsim 1$ year, figure 3.6), Ω_Q constantly overestimates the concentration in weathering product (figure 3.5b). The mismatch in intermediate ages ($0.25 \lesssim T \lesssim 1$ year) results in a slower middle-term response of Ω_Q to a change in the nitrate input (figure 3.5c), with a 6 months delay to return to concentrations below the WHO guideline of 50 mg/L. Finally, without a young water component ($T \lesssim 0.2$ years), the unimodal Ω_Q does not simulate the quick pesticide breakthrough (figure 3.5d) nor does it simulate a long term response because of the quick decay of the pesticide ($\tau_{dec} = 50$ days). The differences in water age between the two models are quantified through some statistics of the marginal age distributions (table 3.1). The unimodal Ω_Q results in a much larger mean travel time (MTT) (explaining the delay in middle-term response in nitrate), in a fraction of water younger than τ_{dec} 58 times lower, and in a higher fraction (13% more) of water older than τ_{eq}.

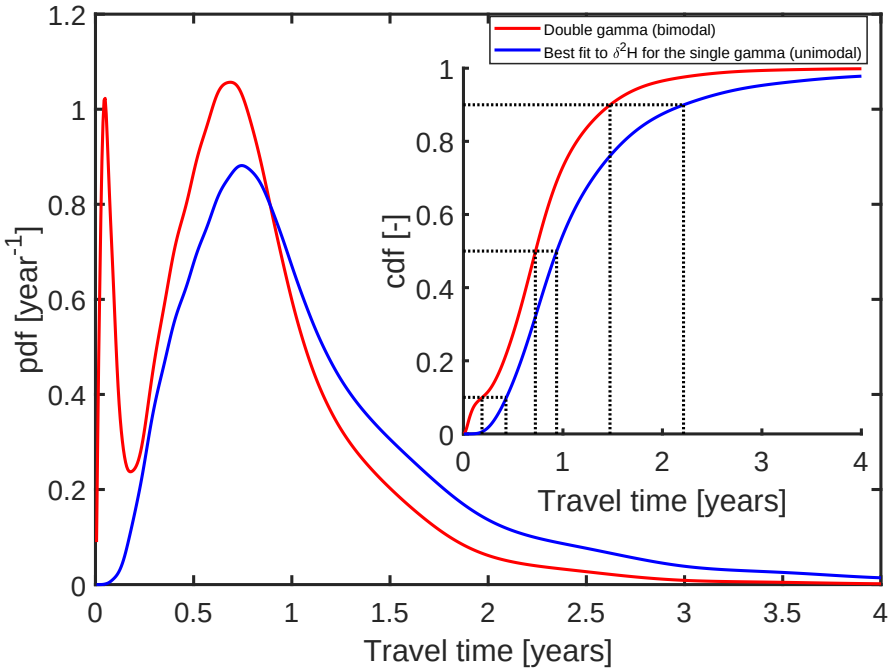

FIGURE 3.6: Marginal (averaged over 100 years) travel time pdf (main figure) and CDF (inset) for the virtual catchment computed from the "true" bimodal ω^{ref} (red) and from the unimodal gamma ω_Q calibrated to stream δ^2H (blue).

These results show that, despite the ability to simulate the observed δ^2H, the calibrated unimodal Ω_Q fails to simulate other tracers because their dynamics depend on contrasting time scales of transport that cannot be properly captured by the unimodal approximation.

3.6 Discussion

3.6.1 Relevance of water age multimodality for tracers

Our results confirm the ability of age-based lumped models to simulate tracer dynamics in outflows. Unimodal SAS functions only had 1 or 2 parameters but they

	Ω^{ref} (true multimodal gamma)	Ω_Q (calibrated unimodal gamma)
MTT [years]	0.81	1.22
$MdTT$ [years]	0.72	0.94
$F(T < 2$ days) [%]	0.049	0.00
$F(T < 7$ days) [%]	0.47	0.00
$F(T < 31$ days) [%]	6.5	0.023
$F(T < \tau_{dec} = 50$ days) [%]	8.7	0.15
$F_{yw}(T < 73$ days) [%]	10	0.82
$F(T < 2\ \tau_{dec} = 100$ days) [%]	13	2.5
$F(T < 3\ \tau_{dec} = 150$ days) [%]	20	8.7
$F(T < 365$ days) [%]	73	55
$F(T < \tau_{eq} = 550$ days) [%]	90	77
$F(T < 2\ \tau_{eq} = 1100$ days) [%]	99	95
$F(T < 3000$ days) [%]	$\simeq 100$	99.7

TABLE 3.1: Statistics of the marginal travel time distributions arising from the unimodal (Ω_Q calibrated to δ^2H) and multimodal ("true" Ω^{ref}) SAS functions. MdTT is the median travel time. MTT is the mean travel time.

provided satisfactory simulations with good performance measures. This is because they managed to capture the average tracer dynamics (e.g. as seen through a moving average) resulting from the range of travel times that characterizes the majority of the discharge. For the lysimeter, 50% of the water left via the bottom drainage in less than 100 days, which is captured by the power function and the single normal function well enough (figure 3.4). For the virtual catchment, 50% of the water left via discharge in less than 1 year, which is captured by the single gamma distribution well enough (figure 3.6). Detailed characteristics of the tracer observations such as flashy responses associated with small discharge volumes (e.g. short-term dynamics within each peak in figure 3.3, months 24 – 30, 12 – 16, and 52 – 56 in figure 3.5), or secondary peaks (Oct – Jan in figure 3.3) do not need to be reproduced to obtain satisfactory values of the standard objective functions used for calibration. Yet, these detailed characteristics represent transport processes acting at different time scales and they can be highly informative on the deficiencies of a too simple model. They could also be used as "soft data" (Seibert and McDonnell 2002) along with experimental knowledge to justify the selection of more complex age models accounting for more hydrological processes (as in Chapter 2).

Composite SAS functions increased model complexity but in return gave more accurate process representation. For instance, the double normal model applied to the lysimeter experiment (Section 3.4) was able to capture the two main timescales at which the tracer was transported through the lysimeter, while the simpler models underestimated the impact of longer transit times (Figures 3.3 and 3.4). In some cases, the increase in performance could be insufficient to justify the use additional parameters. For example, for the catchment-scale virtual experiment, the performance in δ^2H of the single gamma model does not motivate the use of a more complex model. This highlights that a single tracer measured in inflows and outflows may not allow the identification of all peaks of the TTDs. Furthermore, the virtual catchment experiment showed that the simpler model, although efficient enough in reproducing δ^2H, essentially failed to simulate the 3 other tracers. The unimodal approximation neglected the short term response of the catchment, it delayed the middle term response, and it overestimated the long term response. This shows that simulations of solute chemistry require an accurate estimation of the TTDs that goes

beyond conservative tracer input-output relationships, and that accounts for the potential multiple modes of transport associated with different age fractions (Benettin et al. 2017a). On the other hand, this also means that stream chemistry data and new available tracers (Abbott et al. 2016; Hissler et al. 2014; Pfister et al. 2017b) can be used to validate TTDs deduced from a conservative tracer input-output relationship (Guillet et al. 2019). Overall, these results highlight the need for new methods that do not calibrate TTDs using tracer input-output relationships but instead measure the TTDs directly, such as Kim et al. (2016) and Kirchner (2019).

Several characteristics shared by the two hydrological systems described in Sections 3.4–3.5 made the comparison between unimodal and multimodal models possible. Contrary to previous studies on the shape of TTDs and SAS functions, our work was not impacted by large unknowns such as total storage (Benettin et al. 2015b; Rodriguez et al. 2018) and unknown past inputs and initial conditions (Hrachowitz et al. 2011). Also, there was no uncertainty related to hydrological and tracer data regarding the spatial homogeneity of inputs, the calculations of ET, the limited tracer sampling frequency (Hrachowitz et al. 2011; Kirchner et al. 2004; Stockinger et al. 2016) and record length (McGuire and McDonnell 2006). In addition, we avoided using several common modeling assumptions that may have a noticeable impact on the inferred age distributions. These assumptions include the no-flow boundaries defining the control volume to ensure the closure of the mass balance (related to the calculation of actual ET and/or deep water losses), and the ages of water used by ET. van der Velde et al. (2015) and Ali et al. (2014) showed that the ET flux and its TTD influence the TTD of discharge. Further research is needed to improve our understanding of the role of the water ages abstracted by ET on the shapes of the discharge TTD (via the effect of Ω_{ET} on S_T or P_S hence p_Q).

3.6.2 Interpretations of the water age uni- or multi-modality

To understand the possible reasons for the multimodal character of TTDs not simply caused by time-varying inflows and outflows, it is useful to see the travel time T of an individual water parcel from a Lagrangian point of view (i.e. following the water parcel on its journey). The travel time can be written as $T = L/v$, where L is the total length of the flow path taken by that particle (from entry into the system to exit), and v is the harmonic mean of Lagrangian velocities of that parcel along the considered flow path of length L (see appendix B.4). The harmonic mean of velocities is used to obtain an average that appropriately summarizes the entire history of Lagrangian velocities along the flow path in a single measure (see appendix B.4). For instance, Kirchner et al. (2001) calculated the shape of TTDs for theoretical catchment geometries corresponding to various distributions of L (called $w(x)$, with x the distance to the stream). Seeing both L and v as random variables makes the TTD a ratio distribution. Ratio distributions generally do not have simple analytical expressions (Curtiss 1941), meaning that the shape of the TTDs can be complex, with potentially many peaks and often a heavy tail. In particular, if the distributions of L and v are multimodal for various physical reasons (see section 3.2), it can be expected that the TTD is multimodal as well (see figure 3.7). This has been recently observed in a catchment in Switzerland where the extension and the contraction of the flow network for different wetness conditions creates complex patterns of flow paths leading to multimodal TTDs (assuming a constant, homogeneous velocity field) (van Meerveld et al. 2019). In reality, the shape of TTDs will depend on the relationship between L and v. In catchments, it can be expected that the location of a given input will determine

L, but also v. For instance, in heterogeneous media, dispersion causes the distribution of v to broaden and flatten with increasing L (Le Borgne et al. 2007). This can be interpreted as an increased probability with time that the advective velocity of a water particle has changed along its transport (e.g. as in Davies et al. 2013; Scaini et al. 2019; Zehe and Jackisch 2016). Therefore, L and v are probably not independent, making the TTD, as the pdf of their ratio, difficult to estimate a priori (Curtiss 1941). Consequently, even a complete knowledge of the distributions of L and v may not allow to determine the TTD as it was done in figure 3.7 by assuming that L and v are independent.

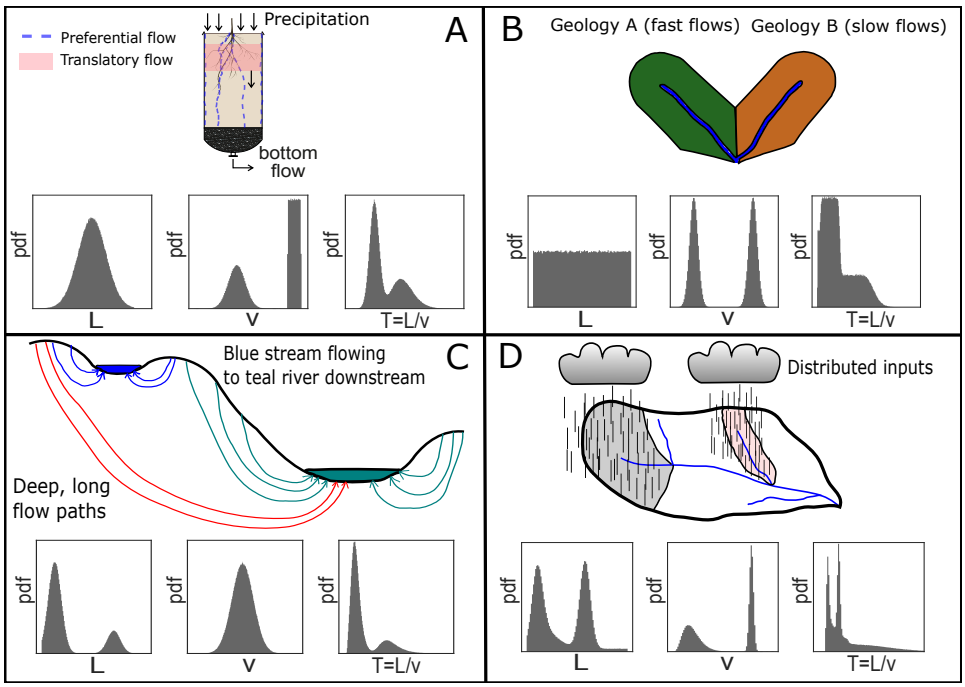

FIGURE 3.7: Four cases where multimodal TTDs are expected: preferential flow vs. translatory flow in soils (c.f. lysimeter at EPFL) (A), different geologies leading to contrasting water velocities in subcatchments with similar geometry (B), shallow, local vs. deep groundwater contributions to a river draining several subcatchments (C), and distributed precipitation over a large catchment with different geological units and different subcatchment geometries (D). Expected distributions of flow path length L and average water velocity v (harmonic mean), and the multimodal distribution of travel times T calculated as the pdf of L/v are shown.

At large scales (e.g. hillslope, catchment), L can be much larger than the scale of the heterogeneous properties of the porous media (i.e. different pore sizes, macropore network, rocks, roots, etc.). As a result, dispersion is strong and the differences in travel time T between the water particles coming from different parts of the catchment (with different L) can be attenuated. This may explain the general validity of unimodal TTDs in catchments (McGuire and McDonnell 2006). For instance, Kirchner et al. (2001) found that low Péclet numbers (i.e. advection $\ll$ dispersion) are needed to explain the gamma shape of TTDs suggested by the spectral slopes of stream chemistry. At smaller scales, L can be small compared to the scale of heterogeneities (i.e. advection $\gg$ dispersion) such that the TTDs are more likely to be multimodal. For example, preferential flow in soils can cause a strongly bimodal distribution for v. In the lysimeter at EPFL (Section 3.4), flow paths are mostly vertical and the transport distance is probably narrowly distributed around $L = 2.5$ m (the depth of the soil column). The multimodal TTDs obtained in the lysimeter are thus due to contrasting velocities coming from preferential flow vs. matrix

flow in the lysimeter (figure 3.7A), which was clearly identified from comparisons of tracers measured in soil water vs. discharge (Benettin et al. 2019). The importance of preferential flow in soils for catchment-scale TTDs may be difficult to evaluate. Glaser et al. (2019b) showed in the Weierbach catchment (Luxembourg) that rainfall-runoff simulations were not improved by modeling vertical preferential flow at the catchment scale, using the soil preferential flow parameters deduced from plot-scale tracer breakthrough experiments. However, nonuniform lateral flow (similar to lateral preferential flow) appeared to be an important mechanism for discharge generation in this catchment, and it could be implemented in the model by using a highly conductive horizontal soil layer (Glaser et al. 2019b). Different conclusions regarding preferential flow were found by van Schaik et al. (2014), who highlighted its non-negligible role for runoff simulations during extreme rainfall events in a semi-arid catchment in Spain. This suggests that preferential flow has an influence on a catchment TTD only if it provides sufficient volumes of water to the outlet.

Our results suggest that different hydrological processes generate multimodal TTDs only if these processes result in contrasting water ages and provide comparable volumes of water. For instance, if surface runoff reaches the stream in 1 hour, there may be no detectable peak in the TTD around T = 1 hour if 99% of discharge is generated by subsurface stormflow having water of at least 1 week of age. Similarly, if two recharge areas to a groundwater well provide similar water ages, the TTD of the well will be unimodal regardless of the sizes of each recharging area. Composite SAS functions provide an intuitive mathematical representation: multiple peaks are observed in the marginal TTDs if the different probability distribution functions summed together in the SAS function are far from each other in the age domain (T, P_S, or S_T), and if they are associated with weights λ (equations B.3 and B.6) that are comparable (e.g. 25% and 75%).

3.6.3 Modeling complex hydrological systems with complex age distributions

The composite TTDs (Leray et al. 2016) and the composite SAS functions proposed in this study (see also Chapter 2) have the ability to embed a larger variety of processes compared to traditional lumped models, but they are still analytical approximations of potentially more complex distributions. Shape-free distributions are more flexible and could better represent the range of transport processes in hydrological systems. Shape-free distributions can be determined from time series analyses (Turner and Macpherson 1990b), by deconvoluting the tracer output signal (Cirpka et al. 2007; Gooseff et al. 2011; Luo and Cirpka 2008; McCallum et al. 2014), from Bayesian inference using several tracers (Massoudieh et al. 2012, 2014), or from multilinear regression analysis (Kirchner 2019). However, these methods require a large number of tracers (Massoudieh et al. 2014), or tracer time series that are well conditioned for numerical reasons (Kirchner 2019), and the resulting solutions can be highly non-unique (McCallum et al. 2014). Eventually, analytical solutions can perform better than shape-free solutions in some cases (Massoudieh et al. 2014). Composite TTDs and SAS functions could thus offer a good balance between model parsimony and sufficient representation of the relevant processes at the scale of interest, which is precisely the aim of lumped models.

Spatially-explicit models allow a detailed representation of flow paths and transport media properties. These suggest that complex SAS functions and TTD shapes emerge from specific flow paths (Kaandorp et al. 2018; Pangle et al. 2017; Yang et al.

2018) or from various conductivity (or velocity) distributions in the system (Danesh-Yazdi et al. 2018; Davies et al. 2013). Distributed models also allow linking water sources to travel times, providing complementary spatial information that cannot be visualized in lumped models (Sayama and McDonnell 2009; van Huijgevoort et al. 2016). However, these models usually require more detailed initial and boundary conditions (Delavau et al. 2017; Jing et al. 2019), which get more uncertain at larger scales. In addition, these models are more computationally expensive than their lumped counterparts, and thus require efficient parallelized numerical solutions (Maxwell et al. 2018). While we often do not have the means to validate the age results from these models, they are useful to explore the possible complexities of the age distributions and relate them to the dominant hydrological processes. This is especially true as these models can much better represent the small-scale mechanisms leading to the observed hydrological processes compared to lumped models.

3.7 Conclusions

Simple TTDs with mainly one peak (or mode) have been prevalent in the last decades because more elaborate models were not needed to obtain a satisfactory tracer simulation. Nevertheless, multimodal TTDs can better describe how different groups of water ages emerge at the outlet of a catchment because of complex flow paths and water velocity distributions. Recent tracer and numerical experiments (including the two examples presented here) show that TTDs with multiple peaks can better simulate the detailed tracer dynamics observed in high-resolution tracer measurements. Moreover, using the TTDs for water chemistry simulations can expose the flaws of a too simple TTD model that underestimates/overestimates large portions of the age distribution with implications for tracer concentrations. Our work highlights that, in many hydrological systems, TTDs are likely to be multimodal because of contrasts in flowpath length or velocity (or both). We provided examples of situations where multimodal TTDs are not simply caused by the dynamics of hydrological fluxes. Multimodality can occur due to: preferential flows in soils, heterogeneity of the catchment geological substratum, deep subsurface flows beyond catchment topographical boundaries and distributed precipitation over large catchments. Different hydrological processes providing comparable volumes of waters with contrasting ages to an outlet will likely result in multimodal TTDs.

Acknowledgements

Data from the lysimeter tracer experiment can be found in Queloz et al. (2015a, appendix B). PB thanks ENAC school at EPFL for financial support. NBR (FNR/CORE/C14/SR/8353440/STORE-AGE) and JK (FNR/CORE/C17/SR/11702136/EFFECT) thank the Luxembourg National Research Fund.

Chapter 4

A dual-tracer experimental approach to determine the entire streamflow TTD

A revised version of this paper (major revisions) is under review for publication in Hydrology and Earth System Sciences (Copernicus Publications[1]) as:

N. B. Rodriguez, L. Pfister, E. Zehe, and J. Klaus (under review). "A comparison of catchment travel times and storage deduced from deuterium and tritium tracers using StorAge Selection functions". In: *Hydrology and Earth System Sciences Discussions*

Abstract

Catchment travel time distributions (TTDs) are an efficient concept to summarize the time-varying 3-dimensional transport of water and solutes to an outlet in a single function of water age and to estimate catchment storage by leveraging information contained in tracer data (e.g. ^{2}H and ^{3}H). It is argued that the increasing use of the stable isotopes of O and H compared to tritium as tracers has truncated our vision of streamflow TTDs, meaning that the long tails associated with old water are neglected. However, the reasons for the truncation of the TTD tails are still obscured by methodological and data limitations. In this study, we went beyond these limitations and tested the hypothesis that streamflow TTDs calculated using only deuterium (^{2}H) or only tritium (^{3}H) are different. We similarly tested if the mobile catchment storage (derived from the TTDs) associated with each tracer is different. For this we additionally constrained a model successfully developed to simulate high-frequency stream deuterium measurements with about 30 stream tritium measurements over the same period (2015–2017). We used data from the forested headwater Weierbach catchment (42 ha) in Luxembourg. The streamflow TTDs were estimated in unsteady conditions by using both tracers coherently within a framework based on StorAge Selection (SAS) functions. We found equal TTDs and equal mobile storage between the ^{2}H- and ^{3}H-derived estimates. The truncation hypothesis was thus rejected. The small differences we found could be explained by the calculation uncertainties and by a limited sampling frequency for tritium. Using both stable and radioactive isotopes of H as tracers reduced the age and storage uncertainties. Although tritium and stable isotopes had redundant information about younger water, using both tracers better exploited the more specific information about longer ages that ^{3}H inherently contains, and it could be even better in the next decades. The two tracers thus had overall different information contents. Tritium was however slightly more informative and cost-effective than stable isotopes for travel time analysis. We thus reiterate the call of Stewart et al. (2012) to measure tritium in the streams for travel time analysis, and emphasize the need for high-frequency tritium sampling in future studies to match the resolution in stable isotopes.

4.1 Introduction

Sustainable water resource management is based upon a sound understanding of how much water is stored in catchments, and how it is released to the streams. Isotopic tracers such as deuterium (^{2}H), oxygen 18 (^{18}O), and tritium (^{3}H) have become the cornerstone of several approaches to tackle these two critical questions (Kendall and McDonnell 1998). For instance, hydrograph separation using stable isotopes of O and H (Buttle 1994; Klaus and McDonnell 2013) has unfolded the difference between catchments hydraulic response (i.e. streamflow) and chemical response (e.g. solutes) (Kirchner 2003) related to the different concepts of water celerity and water velocity (McDonnell and Beven 2014). Isotopic tracers have also been the backbone to unravel water flow paths in soils (Sprenger et al. 2016), and to distinguish soil water going back to the atmosphere and flowing to the streams (Berry et al. 2018; Brooks et al. 2010; Dubbert et al. 2019; McCutcheon et al. 2017; McDonnell 2014).

The travel time distribution (TTD) is nevertheless the concept relying the most on isotopic tracers (McGuire and McDonnell 2006). TTDs provide a concise summary of water flow paths to an outlet by leveraging the information on storage and release contained in tracer input-output relationships. TTDs are essential to link water

quantity to water quality (Hrachowitz et al. 2016), for example by allowing calculations of stream solute dynamics from a hydrological model (Benettin et al. 2015a, 2017a; Maher 2011; Rinaldo and Marani 1987). TTDs are commonly calculated from isotopic tracers in many sub-disciplines of hydrology and thus have the potential to link the individual studies focused on the various compartments of the critical zone (e.g. groundwater and surface water) (Sprenger et al. 2019). ^{3}H has been used as an environmental tracer since the late 1950s (Begemann and Libby 1957; Dinçer et al. 1970; Eriksson 1958; Hubert 1971; Martinec 1975) and it gained particular momentum in the eighties with its use in diverse TTD models (Małoszewski and Zuber 1982; Stewart et al. 2010). It is argued that ^{3}H contains more information on the age of water than stable isotopes due to its radioactive decay (Stewart et al. 2012). For example, low tritium content generally indicates old water in which most of the ^{3}H from nuclear tests has decayed. Despite its potential, ^{3}H is used only rarely in travel time studies nowadays (Stewart et al. 2010), most likely because high precision analyses are laborious (Morgenstern and Taylor 2009) and rather expensive. In contrast, the use of stable isotopes in travel time studies has soared in the last three decades (Benettin et al. 2015a; Fenicia et al. 2010; Heidbüchel et al. 2012; Kendall and McDonnell 1998; Klaus et al. 2015a; McGuire and McDonnell 2006; Pfister et al. 2017a; Rodriguez et al. 2018). This is notably due to the fast and low-cost analyses provided by recent advances in laser spectroscopy (e.g. Gupta et al. 2009; Keim et al. 2014; Lis et al. 2008) and the associated technological progress in sampling techniques of various water sources (Berman et al. 2009; Herbstritt et al. 2019; Herbstritt et al. 2012; Koehler and Wassenaar 2011; Munksgaard et al. 2011; Pangle et al. 2013). According to Stewart et al. 2012 and Stewart and Morgenstern 2016, the limited use of ^{3}H may have cause a biased or "truncated" vision of stream TTDs, in which the the older ages associated with long TTD tails remain mostly undetected by stable isotopes. Longer mean travel times (MTT) were inferred from ^{3}H than from stable isotopes in several studies employing both tracers (Stewart et al. 2010). Longer MTTs may have profound consequences for catchment storage, usually estimated from TTDs as $S = Q \times MTT$ (with Q the flux through the catchment), assuming steady-state flow conditions (i.e. $S(t) = \overline{S(t)} = S$, $Q(t) = \overline{Q(t)} = Q$, $MTT(t) = \overline{MTT(t)} = MTT$) (Birkel et al. 2015; McGuire and McDonnell 2006; Pfister et al. 2017a; Soulsby et al. 2009). Under this assumption, a truncated TTD would result in an underestimated MTT thus an underestimated catchment storage. A different perspective on catchment storage and on its relation with travel times may however be adopted by calculating storage from unsteady TTDs.

A water molecule that reached an outlet has only one age, defined as the time it took to get there. This age is not affected by the isotopes (^{2}H, ^{3}H, and ^{18}O) carried by the molecule and used as tracers, because they do not influence its flow path or its advective velocity or its self diffusion in water (Devell 1962). The use of different methods of travel time analysis for stable isotopes of O and H and for ^{3}H (e.g. amplitudes of seasonal variations vs. radioactive decay) was hence first pointed out as a main reason for the discrepancies in MTT (Stewart et al. 2012). Further research is thus needed for developing mathematical frameworks that coherently incorporate stable isotopes of O and H and ^{3}H in travel time calculations. Moreover, several limiting assumptions were used in previous studies employing ^{3}H to derive MTTs, which are in themselves insufficient statistics to describe various aspects (e.g. shape, modes, percentiles) of the TTDs. For example, the steady-state assumption has been used in almost all ^{3}H travel time studies (Cartwright and Morgenstern 2016; Duvert et al. 2016; Gallart et al. 2016; McGuire and McDonnell 2006; Stewart et al. 2010).

Yet, time variance is a fundamental characteristic of TTDs (Botter et al. 2011; Rinaldo et al. 2015), and it has been acknowledged in simulations of stream ^{3}H only very recently (Visser et al. 2019). Recharge models are also often employed to account only indirectly for the impact of evapotranspiration fluxes (ET) on the catchment inputs in ^{3}H (Stewart et al. 2007) and for the TTD of ET. In contrast for stable isotopes, explicit considerations of ET and of the influence of its TTD on the streamflow TTD are becoming common (van der Velde et al. 2015; Visser et al. 2019). Finally, more guidance on the calibration of the TTD models against ^{3}H measurements is needed (see e.g. Gallart et al. 2016). Especially, uncertainties of ^{3}H-inferred age estimates may have been overlooked, while these could explain the differences with the stable isotope-inferred age estimates.

Besides methodological problems, the reasons for the age differences (hence apparent storage or mixing) are still not well understood, because little is known about the true age information content of ^{3}H compared to stable isotopes. First, ^{3}H sampling in catchments typically differs from stable isotope sampling in terms of frequency and flow conditions. Stable isotope records in precipitation and in the streams have lately shown increasing resolution, covering a wide range of flow conditions (Benettin et al. 2015a; Birkel et al. 2015; McGuire et al. 2005; Pfister et al. 2017a; Rodriguez and Klaus 2019; Visser et al. 2019; von Freyberg et al. 2017). Tritium records in precipitation and streams are on the other hand usually at a monthly resolution in many places around the globe (Halder et al. 2015; IAEA 2019; IAEA and WMO 2019). Only a handful of travel time studies employing ^{3}H report more than a dozen stream samples for a given site and for different conditions than baseflow (e.g. Małoszewski et al. 1983; Visser et al. 2019). This general focus on baseflow ^{3}H sampling introduces by definition a bias towards older water. Second, the natural variability of ^{3}H compared to that of stable isotopes has rarely been documented. ^{3}H in precipitation has returned to the pre-bomb levels, and like stable isotopes it shows a clear yearly seasonality (e.g. Bajjali 2012; Stamoulis et al. 2005). However, ambiguous age estimates may still be obtained with ^{3}H in the northern hemisphere because the current precipitation has similar ^{3}H concentrations than water recharged in the 1980s (Stewart et al. 2012). Higher sampling frequencies of precipitation ^{3}H are almost nonexistent. Rank and Papesch (2005) revealed a short term variability of precipitation ^{3}H likely due to different air masses, observed also during complex meteorological conditions such as hurricanes (Östlund 2013). ^{3}H in streams also show some yearly seasonality (Rank et al. 2018; Różański et al. 2001), but short term dynamics are not well understood because high frequency data sets are limited. Dinçer et al. (1970) showed that short-term stream tritium variations can be caused by the melting of the snowpack from the current and the previous winters. In addition, the seasonally higher values of precipitation ^{3}H in Spring could explain some of the ^{3}H peaks observed in the large rivers (Rank et al. 2018). More studies employing both ^{3}H and stable isotopes and comparing their age information content are therefore crucial to understand travel times in catchments from a multi-tracer perspective.

In this study, we go beyond previous work and test the hypothesis that stream TTDs and the associated catchment storage are different (considering their uncertainties) when inferred from stable isotopes or from ^{3}H measurements used in a coherent mathematical framework for both tracers. For this, we use high frequency isotopic tracer data from an experimental headwater catchment in Luxembourg. Here we focus on the stable isotope of H (deuterium ^{2}H) for which we have more precise measurements. A transport model based on TTDs was recently developed and successfully applied to simulate a two-year high frequency (sub-daily) record of ^{2}H

in the stream (Chapter 2). Here, we additionally constrain the same model within the same mathematical framework against nearly 30 stream samples of ^{3}H collected during highly varying flow conditions over the same period as for ^{2}H. We do not assume steady-state and do not rely on a recharge model by employing StorAge Selection functions to account for the type and the variability of the TTDs of Q and ET that affect the water age balance in the catchment. The tracer input-output relationships and the ^{3}H radioactive decay are accounted for in the method, which reduces ^{3}H age ambiguities. We provide guidance on how to jointly calibrate the model to both tracers and on how to derive likely ranges of storage estimates and travel time measures other than the MTT. This work addresses the following related research questions:

- Are travel times and storage inferred from a common transport model for ^{2}H and ^{3}H in disagreement?
- Are the water age information contents of ^{2}H and ^{3}H similar?

4.2 Methods

4.2.1 Study site description

This study is carried out in the Weierbach catchment, which has been the focus of an increasing number of investigations in the last few years about streamflow generation (Carrer et al. 2019; Glaser et al. 2019b, 2016; Rodriguez and Klaus 2019; Scaini et al. 2017, 2018), biogeochemistry (Moragues-Quiroga et al. 2017; Schwab et al. 2018), and pedology and geology (Gourdol et al. 2018; Juilleret et al. 2011).

The Weierbach is a forested headwater catchment of 42 ha located in northwestern Luxembourg (figure 4.1). The vegetation consists mostly of deciduous hardwood trees (European beech and Oak), and conifers (*Picea abies* and *Pseudotsuga menziesii*). Short vegetation covers a riparian area that is up to 3 m wide and that surrounds most of the stream. The catchment morphology is a deep V-shaped valley in a gently sloping plateau. The geology is essentially Devonian slate of the Ardennes massif, phyllades, and quartzite (Juilleret et al. 2011). Pleistocene Periglacial Slope Deposits (PPSD) cover the bedrock and are oriented parallel to the slope (Juilleret et al. 2011). The upper part of the PPSD ($\sim$ 0–50 cm) has higher drainable porosity than the lower part of the PPSD ($\sim$ 50–140 cm) (Gourdol et al. 2018; Martínez-Carreras et al. 2016). Fractured and weathered bedrock lies from $\sim$ 140 cm depth to $\sim$ 5 m depth on average. Below $\sim$ 5 m depth lies the fresh bedrock that can be considered impervious. The climate is temperate and semi oceanic. The flow regime is governed by the interplay of seasonality between precipitation and evapotranspiration. Precipitation is fairly uniformly distributed over the year, and averages 953 mm/yr over 2006–2014 (Pfister et al. 2017a). The runoff coefficient over the same period is 50%. Streamflow (Q) is double-peaked during wetter periods (Martínez-Carreras et al. 2016), and single peaked during drier periods occurring normally in summer when evapotranspiration (ET) is high.

Based on previous modeling (e.g. Fenicia et al. 2014; Glaser et al. 2019b) and experimental studies (e.g. Glaser et al. 2018; Juilleret et al. 2016; Martínez-Carreras et al. 2016; Scaini et al. 2017), Rodriguez and Klaus (2019) proposed a perceptual model of streamflow generation in the Weierbach. In this model, the first and flashy peaks of double-peaked hydrographs are generated by precipitation falling directly into the stream, by saturation excess flow from the near-stream soils, and by infiltration excess overland flow in the riparian area. The second peaks are generated

by delayed lateral subsurface flow. The lateral fluxes are assumed higher at the PPSD/bedrock interface due to the hydraulic conductivity contrasts (Glaser et al. 2019b, 2016). Lateral subsurface flows are thus accelerated when groundwater rises after a rapid vertical infiltration through the soils (Chapter 2). The model based on travel times presented in this study was developed in a step-wise manner based on this hypothesis of streamflow generation. The model's ability to simulate stream δ^2H dynamics helped to further confirm that these flow processes are active in the Weierbach (Chapter 2). Water flow paths and streamflow generation processes in this catchment are however not completely resolved. Other studies carried out in the Colpach catchment (containing the Weierbach) highlighted the potential role of lateral preferential flow through macropores in the highly heterogeneous soils for the generation first peaks of the hydrographs (Angermann et al. 2017; Loritz et al. 2017), contrary to the understanding from various studies in the Weierbach (Glaser et al. 2019a; Glaser et al. 2016).

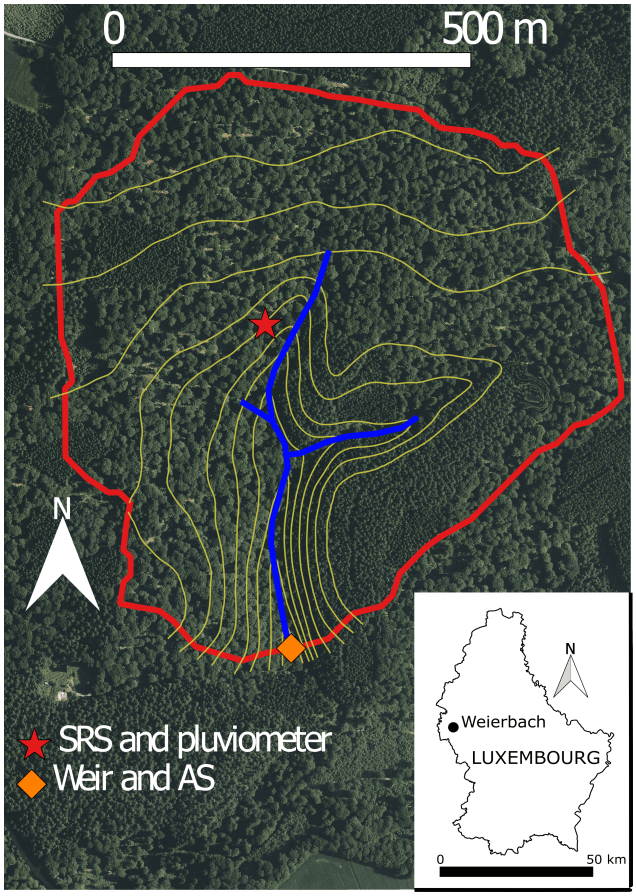

FIGURE 4.1: Map of the Weierbach catchment and its location in Luxembourg. The weir is located at coordinates (5°47′44″E, 49°49′38″N). SRS is the sequential rainfall sampler. AS is the stream autosampler. The elevation lines go by increments of 5 m from 460 m.a.s.l. downstream close to the weir location to 510 m.a.s.l. at the northern catchment divide.

4.2.2 Hydrometric and tracer data

In this study we use precipitation (J, in mm/h), ET (mm/h), Q (mm/h), and δ^2H (‰) and ^{3}H (Tritium Units, T.U.) measurements in precipitation ($C_{P,2}$ and $C_{P,3}$ respectively) and in the stream ($C_{Q,2}$ and $C_{Q,3}$ respectively). Here a subscript 2 indicates deuterium (^{2}H) and a subscript 3 indicates tritium (^{3}H). The analysis in this

study focuses on the period October 2015–October 2017 during which most samples were collected at higher frequencies than in the past (figure 4.2). Details on the hydrometric data collection (J, ET, Q), and on the ^{2}H sample collection and analysis are given in chapter 2.

The 1088 stream samples analyzed for ^{2}H were collected manually or automatically with an autosampler (AS, figure 4.1), resulting in samples every 15 hours on average over October 2015–October 2017. The 525 precipitation samples analyzed for ^{2}H were collected approximately every 2.5 mm rain increment (i.e. on average every 23 hours) with a sequential rainfall sampler (SRS) and in addition as bulk samples on a bi-weekly basis. The samples were analyzed at the Luxembourg Institute of Science and Technology (LIST) using an LGR Isotope Water Analyzer, yielding for ^{2}H an analytical accuracy of 0.5 ‰ (equal to the LGR standard accuracy), and a precision maintained <0.5 ‰ (quantified as one standard deviation of the measured samples and standards).

The 24 stream samples analyzed for ^{3}H were selected from manual bi-weekly sampling campaigns to cover various flow ranges. The samples were analyzed by the GNS Science Water Dating Laboratory (Lower Hutt, New Zealand), which provides high precision tritium measurements using electrolytic enrichment and liquid scintillation counting (Morgenstern and Taylor 2009). The precision of the stream samples varies from roughly 0.07 T.U. to roughly 0.3 T.U., but commonly around 0.1 T.U. Monthly values of ^{3}H in precipitation were obtained for the Trier station (60 km from the Weierbach) until 2016 from the WISER database of the International Atomic Energy Agency (IAEA) (IAEA and WMO 2019; Stumpp et al. 2014). The 2017 values were obtained from the Radiologie group of Bundesanstalt für Gewässerkunde (Schmidt et al. submitted). ^{3}H in precipitation before 1978 was calculated by regression with data from Vienna, Austria (Stewart et al. 2017).

For both ^{2}H and ^{3}H, the time series of tracer in precipitation was interpolated between two consecutive samples (e.g. A and B) as being equal to the value of the next sample (i.e. B). Since no measurements of J, Q, ET, and $C_{P,2}$ are available before 2010, we looped back their values of the period October 2010–October 2015 periodically before 2010 as a best estimate of their past values. We aggregated the input data (J, ET, Q, $C_{P,2}$, $C_{P,3}$) to a resolution $\Delta t = 4$ hours, which is small enough to capture the variability of flows and tracers in the input and simulate the variability of the flows and tracers in the output.

4.2.3 Mathematical framework

Mathematically, the streamflow TTD is related to the stream tracer concentrations $C_Q(t)$ according to the following equation (4.1):

$$C_Q(t) = \int_{T=0}^{+\infty} C_P^*(T,t)\, \overleftarrow{p_Q}(T,t)\, dT \tag{4.1}$$

where T is water age, t is time of observation, $C_Q(t)$ is the stream tracer concentration, $\overleftarrow{p_Q}$ (probability distribution function, p.d.f.) is the stream backward TTD (Benettin et al. 2015c), and $C_P^*(T,t)$ is the tracer concentration of the water parcel reaching the outlet at time t with age T (this parcel was in the inflow at time $t-T$, its travel time is thus T). This equation is always verified for the exact (usually unknown) TTD, because it simply expresses the fact that the stream concentration is the volume-weighted arithmetic mean of the concentrations of the water parcels with different travel times at the outlet. $C_P^*(T,t)$ depends on T and t as separate variables if the tracer concentration of a water parcel in the catchment changes between

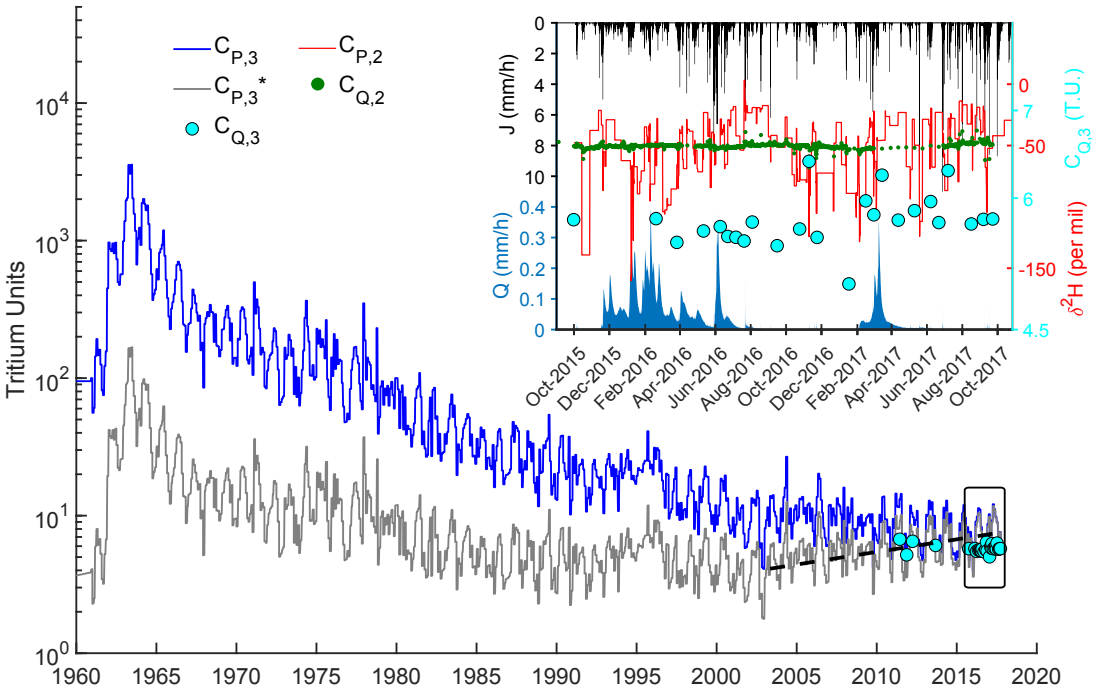

FIGURE 4.2: Data used in this study: ^{3}H in precipitation ($C_{P,3}$), the corresponding tritium activities accounting for radioactive decay until 2017 ($C^*_{P,3}$), δ^2H in precipitation $C_{P,2}$ (inset), precipitation J (inset), streamflow Q (inset), ^{3}H measurements in the stream ($C_{Q,3}$ both plots), and δ^2H in the stream ($C_{Q,2}$, inset). The period contained in the inset is represented as a rectangle in the bigger plot. The dashed line visually represents the increasing trend in $C^*_{P,3}$ that emerges as the effect of bomb peak tritium disappears (i.e. $C_{P,3}(t-T)$ stops decreasing around 2000 so $C^*_{P,3}(T,t) = C_{P,3}(t-T)\,e^{-\alpha T}$ starts decreasing with increasing T).

injection time $t - T$ and observation time t. For solutes like silicon and sodium, the concentration can increase with age (Benettin et al. 2015a). For ^{3}H, radioactive decay with a constant $\alpha = 0.0563$ yr^{-1} implies $C^*_{P,3}(T,t) = C_{P,3}(t-T)\,e^{-\alpha T}$, where $C_{P,3}(t-T)$ is the concentration in precipitation measured at $t - T$. For ^{2}H, $C^*_{P,2}(T,t) = C_{P,2}(t-T)$. Thus, the streamflow TTD simultaneously verifies equations 4.2 and 4.3:

$$C_{Q,2}(t) = \int_{T=0}^{+\infty} C^*_{P,2}(T,t)\,\overleftarrow{p_Q}(T,t)\,dT = \int_{T=0}^{+\infty} C_{P,2}(t-T)\,\overleftarrow{p_Q}(T,t)\,dT \tag{4.2}$$

$$C_{Q,3}(t) = \int_{T=0}^{+\infty} C^*_{P,3}(T,t)\,\overleftarrow{p_Q}(T,t)\,dT = \int_{T=0}^{+\infty} C_{P,3}(t-T)\,e^{-\alpha T}\,\overleftarrow{p_Q}(T,t)\,dT \tag{4.3}$$

Practically, when measurements of ^{2}H and ^{3}H are used to inversely deduce the TTD by using equations 4.2 and 4.3, different TTDs may be found. These different TTDs may be called $\overleftarrow{p_{Q,2}}$ and $\overleftarrow{p_{Q,3}}$ for instance, referring to ^{2}H and ^{3}H, respectively. To avoid introducing more variables and to avoid confusion, we do not use the names $\overleftarrow{p_{Q,2}}$ and $\overleftarrow{p_{Q,3}}$ and we instead refer to the TTDs "constrained" by a given tracer, using a common symbol $\overleftarrow{p_Q}$. We do this also to stress that the exact (true) TTD must simultaneously verify both equations 4.2 and 4.3, and that two different TTDs $\overleftarrow{p_{Q,2}}$ and $\overleftarrow{p_{Q,3}}$ cannot physically exist. This is a fundamental difference from

previous work that assumed two different TTDs, using for example equation 4.3 for ^{3}H and another method for ^{2}H (the sine-wave approach) (e.g. Małoszewski et al. 1983). The framework in this study also uses the fact that the same functional form of streamflow TTD needs to simultaneously explain both tracers to be valid, unlike previous work that used different TTD models for different tracers (Stewart and Thomas 2008).

4.2.4 Transport model based on TTDs

Most of the previous travel time studies using tritium assumed steady-state and an analytical shape for the stream TTD, and fitted the parameters of the analytical function using the framework described in 4.2.3. In this study, the TTDs are unsteady (i.e. time-varying) and cannot be analytically described. Still, they can be calculated by numerically solving the "Master Equation" (Botter et al. 2011). This method has been applied in several recent studies (e.g. Benettin et al. 2017b; Harman 2015; van der Velde et al. 2015), and is described in more details by Benettin et al. (2018). The numerical method used to solve this equation in this study is described in chapter 2.

Essentially, the Master Equation is a water balance equation where storage and fluxes are labeled with age categories. The Master Equation is thus a partial differential equation expressing the fact that the amount of water in storage having age T changes with time because of new water introduced by precipitation $J(t)$, because of water aging, and because of losses to catchment outflows $ET(t)$ and $Q(t)$. Solving the Master Equation requires knowledge (or an assumption about the shape) of the StorAge Selection (SAS) functions Ω_Q and Ω_{ET} of outflows Q and ET, which conceptually represent how likely water ages in storage are to be present in the outflows at a given time. Solving the Master Equation yields the distribution of ages in storage at every moment, that can be represented in a cumulative form with age-ranked storage S_T, defined as the amount of water in storage (e.g. 10 mm) younger than T (e.g. 1 year) at time t. $T \rightarrow S_T$ is just a mathematical change of variable, and it has no meaning respective to the location or depth of a certain water parcel with age T in the catchment. By definition $\lim_{T \to +\infty} S_T = S(t)$, where $S(t)$ is catchment storage. Ω_Q and Ω_{ET} are functions of S_T and cumulative distributions functions (c.d.f.) for numerical convenience. SAS functions are closely linked to TTDs, such that one can be found from the other using the following expression (here for Q, but valid for other outflows):

$$\overleftarrow{p_Q}(T,t) = \frac{\partial}{\partial T}(\Omega_Q(S_T,t)) \tag{4.4}$$

The partial derivative with respect to age T ensures the transition from c.d.f. to p.d.f. Assuming a parameterized form for Ω_Q and Ω_{ET} and calibrating their parameters using the framework defined in 4.2.3 yields time-varying TTDs constrained by the tracers in the outflows.

In this study, we assumed that Ω_{ET} is a function of only S_T and it is gamma distributed with a mean parameter μ_{ET} (mm) and a scale parameter θ_{ET} (mm). Rodriguez and Klaus (2019) showed that in the Weierbach, a weighted sum of three components in the streamflow SAS function is more consistent the superposition of streamflow generation processes (i.e. saturation excess flow, saturation overland flow, lateral subsurface flow, see section 4.2.1) than a single component. This means that Ω_Q is written as a weighted sum of three c.d.f.s (see appendix C.1.1 and Chapter 2):

$$\Omega_Q(S_T,t) = \lambda_1(t)\,\Omega_1(S_T) + \lambda_2(t)\,\Omega_2(S_T) + \lambda_3(t)\,\Omega_3(S_T) \tag{4.5}$$

$\lambda_1(t)$, $\lambda_2(t)$, and $\lambda_3(t)$ are time-varying weights summing to 1. Essentially, $\lambda_1(t)$ is the smallest weight and it is parameterized to increase sharply during flashy streamflow events, using parameters λ_1^*, f_0, S_{th} (mm), and ΔS_{th} (mm). $\lambda_2(t) = \lambda_2$ is calibrated, and $\lambda_3(t)$ just deduced by difference. Ω_1 is a cumulative uniform distribution over S_T in $[0, S_u]$ (with S_u a parameter in mm). Ω_1 represents the young water contributions associated with short flow paths during flashy streamflow events. Ω_1 corresponds to processes in the near stream area: saturation excess flow, saturation overland flow, and rain on the stream (Chapter 2). Ω_2 and Ω_3 are gamma-distributed with mean parameters μ_1 and μ_2 (mm), and scale parameters θ_1 and θ_2 (mm) respectively. Ω_2 and Ω_3 represent older water that is always contributing to the stream. This older water consists of groundwater stored in the weathered bedrock that flows laterally in the subsurface. Note that we used the same functional form of $\Omega_Q(S_T,t)$ for ^{2}H and ^{3}H to keep the functional form of the TTDs consistent between the tracers.

4.2.5 Model initialization and numerical details

Numerically solving the Master Equation requires an estimation of catchment mobile storage $S(t)$. In this context, $S(t)$ represents the sum of "dynamic" (or "active") storage and "inactive" (or "passive") storage (Birkel et al. 2011a; Fenicia et al. 2010; Hrachowitz et al. 2013; Soulsby et al. 2011). In this study the model is initialized in October 1915 with storage $S(t=0) = S_{ref}$. This initial value is chosen large enough to sustain Q and ET during drier periods and to store water that is sufficiently old to satisfy equation 4.1. $S(t)$ is then simply deduced from the water balance as $S(t) = S_{ref} + \int_{x=0}^{t}(J(x) - Q(x) - ET(x))\,dx$. The initial age distribution in storage is exponential with a mean of 1.7 years, the estimated Mean Residence Time (MRT) by Pfister et al. (2017a). The model is then run with time steps $\Delta t = 4$ hours and age resolution $\Delta T = 8$ hours. This way the computational cost is balanced with the resolution of the simulations in δ^2H. The period October 1915–October 2015 serves as a long spin-up period to completely remove the impact of the initial conditions. This means that S_{ref} and the initial age distribution in storage do not influence the results over October 2015 – October 2017. $ET(t)$ is taken equal to potential evapotranspiration $PET(t)$ except that it tends non-linearly towards 0 (using a constant smoothing parameter n) when storage $S(t)$ decreases below $S_{ref} - S_{root}$ (mm), where S_{root} is a parameter accounting for the water amount accessible by ET (appendix C.1.2).

4.2.6 Model calibration

The parameters of the SAS functions and the other model parameters were calibrated using a Monte Carlo technique. In total, 12 parameters were calibrated (table 4.1). The initial ranges were selected based on parameter feasible values (e.g. f_0 between 0 and 1 by definition), on previous estimations (e.g. S_{th}), on hydrological data (e.g. S_u and ΔS_{th} deduced from average precipitation depths), and on initial tests on the parameter ranges (e.g. μ and θ). These ranges allow a wide range of shapes of SAS functions while minimizing numerical errors (occurring for example for $S_T > S(t)$).

Unlike our previous modeling work in this catchment (Chapter 2), we fixed the initial storage in the model S_{ref} to 2000 mm. We did this to reduce the degrees of

TABLE 4.1: Model parameters

Symbol	Type	Unit	Initial range	Description[a]
S_{th}	Calibrated	mm	$[20, 200]$	Storage threshold relative to S_{min} separating "dry" and "wet" periods
ΔS_{th}	Calibrated	mm	$[0.1, 20]$	Threshold in short term storage changes identifying "first" peaks in hydrographs
S_u	Calibrated	mm	$[1, 50]$	Range of the uniformly distributed Ω_1
f_0	Calibrated	–	$[0, 1]$	Young water coefficient for the dry periods
λ_1^*	Calibrated	–	$[0, 1]^b$	Maximum value of the weight $\lambda_1(t)$
λ_2	Calibrated	–	$[0, 1]$	Constant[c] value of the weight $\lambda_2(t)$
μ_2	Calibrated	mm	$[0, 1600]$	Mean parameter of the gamma distributed Ω_2
θ_2	Calibrated	mm	$[0, 100]$	Scale parameter of the gamma distributed Ω_2
μ_3	Calibrated	mm	$[0, 1600]$	Mean parameter of the gamma distributed Ω_3
θ_3	Calibrated	mm	$[0, 100]$	Scale parameter of the gamma distributed Ω_3
μ_{ET}	Calibrated	mm	$[0, 1600]$	Mean parameter of the gamma distributed Ω_{ET}
θ_{ET}	Calibrated	mm	$[0, 100]$	Scale parameter of the gamma distributed Ω_{ET}
S_{root}	Constant	mm	150	Water amount accessible by ET
m	Constant	–	1000	Smoothing parameter for the calculation of $\lambda_1(t)$
n	Constant	–	20	Smoothing parameter for the calculation of $ET(t)$ from $PET(t)$
Δt^*	Constant	hours	8	Width of the moving time window used to calculate short term storage variations $\overline{\Delta S(t)}$

[a] Details about the equations involving these parameters are given in appendix C.1.1 and in Chapter 2.
[b] λ_1^* is in fact sampled between 0 and $1 - \lambda_2 \leq 1$ to ensure that $\sum_{n=1}^{3} \lambda_k(t) = 1$.
[c] $\lambda_1(t)$ varies, λ_2 is constant, and $\lambda_3(t)$ varies and it is deduced using $\lambda_3(t) = 1 - \lambda_2 - \lambda_1(t)$.

freedom when sampling the parameter space in order to limit the impact of numerical errors on the calibration. These errors are due to numerical truncation of $\Omega_Q(S_T, t)$ when a considerable part (e.g. a few percent) of its tail extends above $S(t)$. This occurs when parameters μ_2, μ_3, θ_2, and θ_3 are too large compared to S_{ref} when the latter is also randomly sampled. Choosing a constant large value for S_{ref} thus guarantees the absence of truncation errors. S_{ref} has little influence on the storage deduced from travel times since the ages sampled by streamflow are governed only by μ_2, μ_3, θ_2, and θ_3. These parameters are independent of S_{ref} as long at it allows sufficiently old water to reside in storage, which is ensured by its large value and by the long spin-up period we used (100 years).

The first step of the Monte Carlo procedure we employed consists in randomly sampling parameters from the uniform prior distributions with ranges defined in table 4.1. 12,096 sets of the 12 calibrated parameters were sampled as a Latin Hypercube (LHS, Helton and Davis 2003). The model was then run over October 1915–October 2017, and its performance was evaluated over October 2015–October 2017. We evaluated model performance in a multi-objective manner, by using separate objective functions for ^{2}H and ^{3}H. For deuterium, we used the Nash-Sutcliffe Efficiency (NSE):

$$E_2 = 1 - \frac{\sum_{k=1}^{N_2} (C_{Q,2}(t_k) - \delta^2 H(t_k))^2}{\sum_{k=1}^{N_2} (\delta^2 H(t_k) - \overline{\delta^2 H})^2} \tag{4.6}$$

where $N_2 = 1,016$ is the number of deuterium observations in the stream. For tritium, we used the Mean Absolute Error:

$$E_3 = \sum_{j=1}^{N_3} |C_{Q,3}(t_j) - {}^3H(t_j)| \tag{4.7}$$

where $N_3 = 24$ is the number of tritium observations in the stream. We used the MAE for tritium because it is common to report errors in T.U., and because of the

limited variance of stream ^{3}H (due to the low number of samples and the low variability) making the NSE less appropriate (Gallart et al. 2016). The behavioral parameter sets that are used for uncertainty calculations and further analysis were selected based on threshold values L_2 and L_3 for the performance measures E_2 and E_3 respectively. Parameter sets were considered behavioral for deuterium simulations if $E_2 > L_2 = 0$, and behavioral for tritium simulations if $E_3 < L_3 = 0.5$ T.U. We subsequently refer to these parameter sets and corresponding simulations as "constrained by deuterium", "constrained by tritium", and as "constrained by both" when both performance criteria were used. We chose these constraints to get reasonable model fits to the data, to obtain a comparable number of behavioral parameter sets for ^{2}H and ^{3}H, and to maximize the amount of information gained about the parameters when adding a constraint on the model performance for a tracer. This information gain was assessed with the Kullback-Leibler Divergence D_{KL} between the posterior parameter distributions inferred from various combinations of constraints L_2 and L_3 (section 4.2.7).

4.2.7 Information contents of ^{2}H and ^{3}H

Loritz et al. (2018) and Loritz et al. (2019) recently used information theory to detect hydrological similarity between hillslopes of the Colpach catchment and to compare topographic indexes in the Attert catchment in Luxembourg. Thiesen et al. (2019) used information theory to build an efficient predictor of rainfall-runoff events. Here we leverage information theory to evaluate our model parameter uncertainty. For this we calculated the expected information content of the posterior parameter distributions constrained by deuterium or tritium using the Shannon entropy $\mathcal{H}$:

$$\mathcal{H}(X|\,^iH) = -\sum_{k=1}^{n_I} f(I_k)\,\log_2 f(I_k) \tag{4.8}$$

In this equation, the parameter X (e.g. μ_1) takes values (e.g. 125 mm) falling in intervals I_k (e.g. $[100, 150]$ mm) that do not intersect each other and which union $\cup_{k=1}^{n_I} I_k$ equals I_X, the total interval of values on which X is defined (e.g. $[50, 500]$ mm). The definitions of the n_I intervals I_k for each parameter depend on the binning of the parameter values, given in table 4.2. The posterior probability distribution f defines the probability of the parameter X to be in a certain state (i.e. to take a value falling in an interval I_k), when constrained by the criterion $E_2 > L_2\,(i = 2)$ or $E_3 < L_3\,(i = 3)$. When using the logarithm of base 2, $\mathcal{H}$ is expressed in bits of information contained in the posterior distribution f. The uniform distribution over I_X has the maximum possible entropy. Lower values of $\mathcal{H}$ thus indicate that the posterior distribution is not flat, hence less uncertain than the uniform prior distribution. In general, lower values of $\mathcal{H}$ indicate less uncertain parameters.

We also use the Kullback-Leibler Divergence D_{KL} to evaluate the gain of information when ^{3}H is used in addition to ^{2}H to constrain model predictions or vice versa:

$$D_{KL}(X|(\,^2H \cap\,^3H),\, X|\,^iH) = \sum_{k=1}^{n_I} f(I_k)\,\log_2 \frac{f(I_k)}{g(I_k)} \tag{4.9}$$

where f is the posterior distribution constrained by $E_2 > L_2$ and $E_3 < L_3$, and g is the posterior distribution constrained only by $E_2 > L_2\,(i = 2)$ or only by $E_3 < L_3\,(i = 3)$. D_{KL} is expressed in bits of information gained when the knowledge about a parameter posterior distribution is updated by adding another tracer.

D_{KL} can also be used to evaluate the gain of information from prior to posterior parameter distributions (by using g = prior and f = posterior). Calculating D_{KL} also requires binning the parameter values to define the intervals I_k and calculate the distributions f and g. The binning for each parameter (table 4.2) was chosen such that the resulting histograms visually reveal the underlying structure of the parameter values, while avoiding uneven features and irregularities (e.g. very spiky histograms).

4.3 Results

4.3.1 Calibration results

148 parameter sets were behavioral for deuterium simulations, with E_2 ranging from $L_2 = 0$ to 0.24. 181 parameter sets were behavioral for tritium simulations, with E_3 ranging from 0.24 T.U. to $L_3 = 0.5$ T.U. Additionally, 16 parameter sets were behavioral for both tritium and deuterium simulations, with E_2 ranging from $L_2 = 0$ to 0.19 and E_3 ranging from 0.36 T.U. to $L_3 = 0.5$ T.U. These solutions show that a reasonable agreement between the model fit to ^{2}H and the model fit to ^{3}H can be found.

The behavioral posterior parameter distributions constrained by deuterium or tritium or by both generally had similar ranges than their prior distributions, except notably for μ_2, θ_2, μ_3, and θ_3 (table 4.2). To assess the reduction of parameter uncertainty, we calculated and compared the entropy of the prior and of the posterior distributions (table 4.2). A visual inspection of the posterior distributions was also made, and we show here only the parameters μ_2, θ_2, μ_3, and θ_3 (figure 4.3) that directly control the range of older water ages in streamflow, since they act mostly on the right-hand tail of the gamma components in Ω_Q. These parameters thus also have a direct influence on the catchment storage inferred via age-ranked storage S_T.

Essentially, the results (table 4.2 and figure 4.3) reveal that the parameter ranges decreased by adding information on ^{2}H or ^{3}H or both. This effect is particularly noticeable for f_0 and λ_1^*, which saw their upper boundary decrease, and for μ_2 and μ_3, which saw their lower boundary increase considerably. These results also show that the posterior distributions depart from the uniform prior distributions when considering ^{2}H alone or ^{3}H alone (i.e. $\mathcal{H}(X|\,^iH) < \mathcal{H}(X)$ in table 4.2). This effect is not very pronounced for most parameters, but clearly visible for λ_1^*, for μ_2 and μ_3 (e.g. uneven distributions of points in figure 4.3), and for μ_{ET}. The posterior distributions become considerably narrower when considering both tracers, since $\mathcal{H}(X|(\,^2H \cap \,^3H))$ is much lower than $\mathcal{H}(X)$, which is visually represented by the distribution of points tending to cluster towards a corner in figure 4.3. Generally, more was learned about the likely parameter values by adding a constraint on ^{2}H simulations after constraining ^{3}H simulations than the opposite (i.e. generally $D_{KL}(X|(\,^2H \cap \,^3H),\, X|\,^3H) \geq D_{KL}(X|(\,^2H \cap \,^3H),\, X|\,^2H)$). Noticeable exceptions to this are the parameters μ_2, θ_2, and θ_3, which are more related to the older ages in streamflow and to catchment storage than the other parameters.

Simulations of stream δ^2H captured both the slow and the fast dynamics of the observations when constrained by $E_2 > 0$ (blue bands and blue curve figure 4.4). This is not the case for δ^2H simulations constrained only by $E_3 < 0.5$ T.U. (red bands). This shows that ^{3}H contains some information that is not in common with ^{2}H about the transport processes to the stream. Yet, simulations constrained by

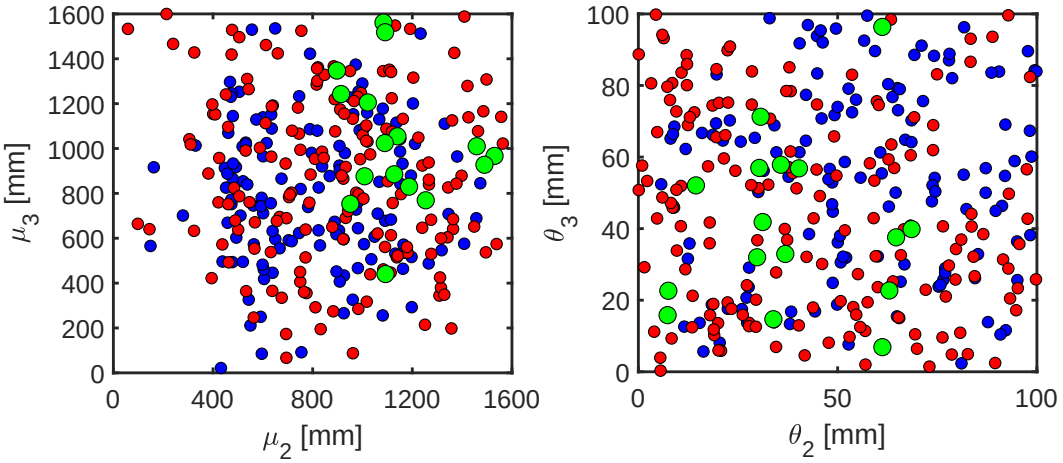

FIGURE 4.3: Distributions of SAS function mean (μ, left panel) and scale (θ, right panel) behavioral parameters directly controlling the selection of older ages by streamflow, constrained by deuterium (148 blue dots), or tritium (181 red dots), or both (16 green dots).

both criteria (green bands) have a smaller variability than those constrained only by $E_2 > 0$, suggesting that ^{3}H nevertheless contains some information that is common with ^{2}H.

Simulations of stream ^{3}H generally matched the observations better in 2017 than before 2017 (red bands and red curve in figure 4.5). Some simulations (red bands) nevertheless matched the observations before 2017 relatively well. Similar to δ^2H simulations, both the slow and the fast simulation dynamics seemed necessary to reproduce the variability in ^{3}H observations (especially in 2017), although more stream samples would be needed to confirm that the model is accurate between the current measurement points. The higher stream ^{3}H values in 2017 that are better reproduced by the model correspond to an extended dry period during which streamflow responses are mostly flashy and short-lasting hydrographs. The associated ^{3}H values are closer to precipitation ^{3}H, mostly around 10 T.U. The stream reaction to those higher values suggest a considerable influence of recent rainfall events on the stream, that steady-state TTD models relying only on tritium decay would probably struggle to simulate. This also suggests a stronger influence of old water in 2016 than in 2017 (see section 4.4.4). Simulations constrained by deuterium (blue bands) tended to overestimate stream ^{3}H. Simulations constrained by both criteria (green bands) worked well in 2017, but they overestimated stream ^{3}H before 2017. Similar to δ^2H simulations, this suggests that ^{2}H and ^{3}H have common but also distinct information contents about transport processes to the stream.

4.3.2 Storage and travel time results

For each behavioral parameter set, we calculated $\overleftarrow{P_Q}(T)$, the average stream TTD weighted by streamflow (over 2015–2017) in cumulative form (figure 4.6). Visually, there are no striking differences between $\overleftarrow{P_Q}(T)$ constrained by deuterium or by tritium, except a slightly wider spread for simulations constrained by tritium. The $\overleftarrow{P_Q}(T)$ constrained by both tracers clearly differ. The associated curves (figure 4.6c) show a much narrower spread. They are also slightly shifted towards higher ages.

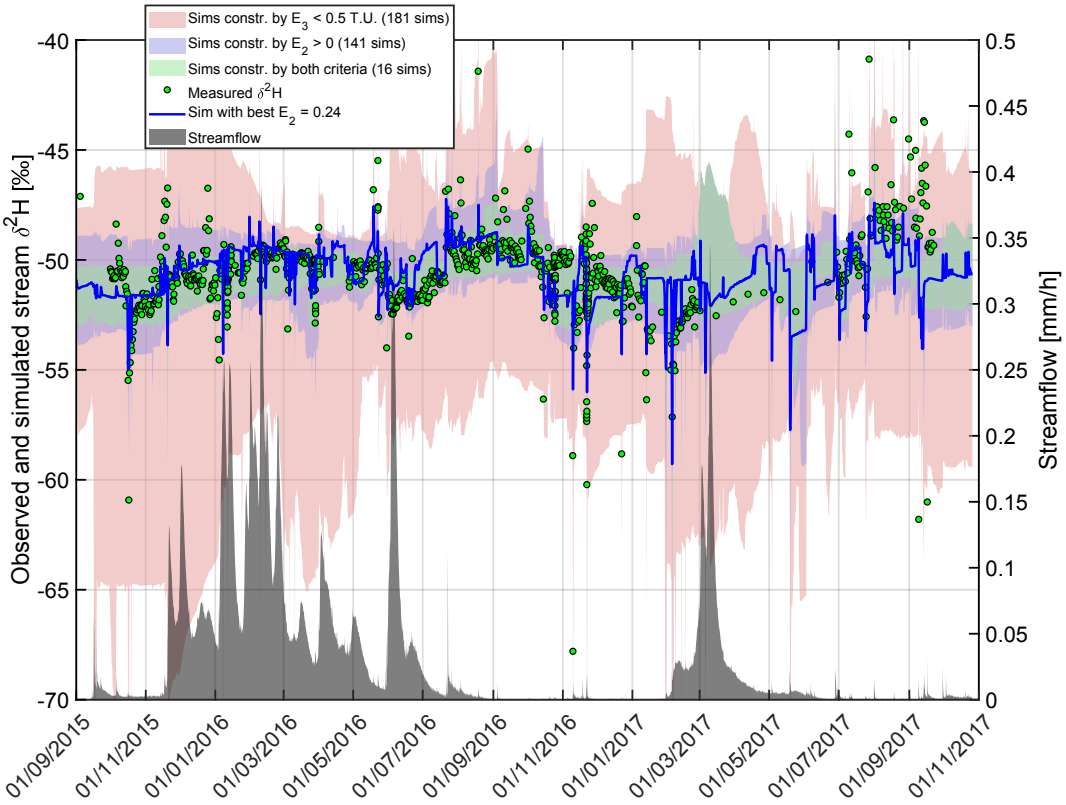

FIGURE 4.4: Simulations in deuterium. E_2 is the Nash-Sutcliffe efficiency in deuterium, and E_3 is the Mean Absolute Error in tritium units.

We calculated various statistics of the distributions $\overleftarrow{P_Q}(T)$ constrained by the different performance criteria to compare them quantitatively (table 4.3). This shows that the $\overleftarrow{P_Q}(T)$ constrained only by tritium systematically correspond to higher ages (and lower young water fractions) than those constrained only by deuterium. However, these age differences are small and could be explained by the uncertainties, which are larger for the younger age fractions, and systematically higher for tritium than for deuterium. The $\overleftarrow{P_Q}(T)$ constrained by both tracers systematically correspond to the highest ages (and the lowest young water fractions). The corresponding uncertainties are much lower than when using individual tracers.

We defined Ω_{tail} as the weighted sum of the two gamma components in Ω_Q:

$$\Omega_{tail}(S_T) = \frac{1}{\lambda_2 + \lambda_3^*} \left(\lambda_2 \, \Omega_2(S_T) + \lambda_3^* \, \Omega_3(S_T)\right) \tag{4.10}$$

where $\lambda_3^* = 1 - \lambda_2 - \lambda_1^*$. Ω_{tail} thus represents the right-hand tail of the SAS function Ω_Q, allowing us to study the asymptotic behavior of the function in detail. In particular, this asymptotic behavior is time-invariant when plotted against S_T, because Ω_2 and Ω_3 are functions of S_T only. The behavioral parameter sets were thus directly used to calculate the curves $(S_T, \Omega_{tail}(S_T))$. These curves show similar differences for ^{2}H and ^{3}H than the curves $(T, \overleftarrow{P_Q}(T))$ (figure 4.7): a slightly wider spread is observed for Ω_{tail} constrained by tritium than deuterium (figure 4.7b), and the Ω_{tail} constrained by both tracers tend to converge to a narrow envelope of curves slightly shifted towards higher storage values (figure 4.7c).

To quantitatively study the implications of different Ω_{tail} for storage estimations,

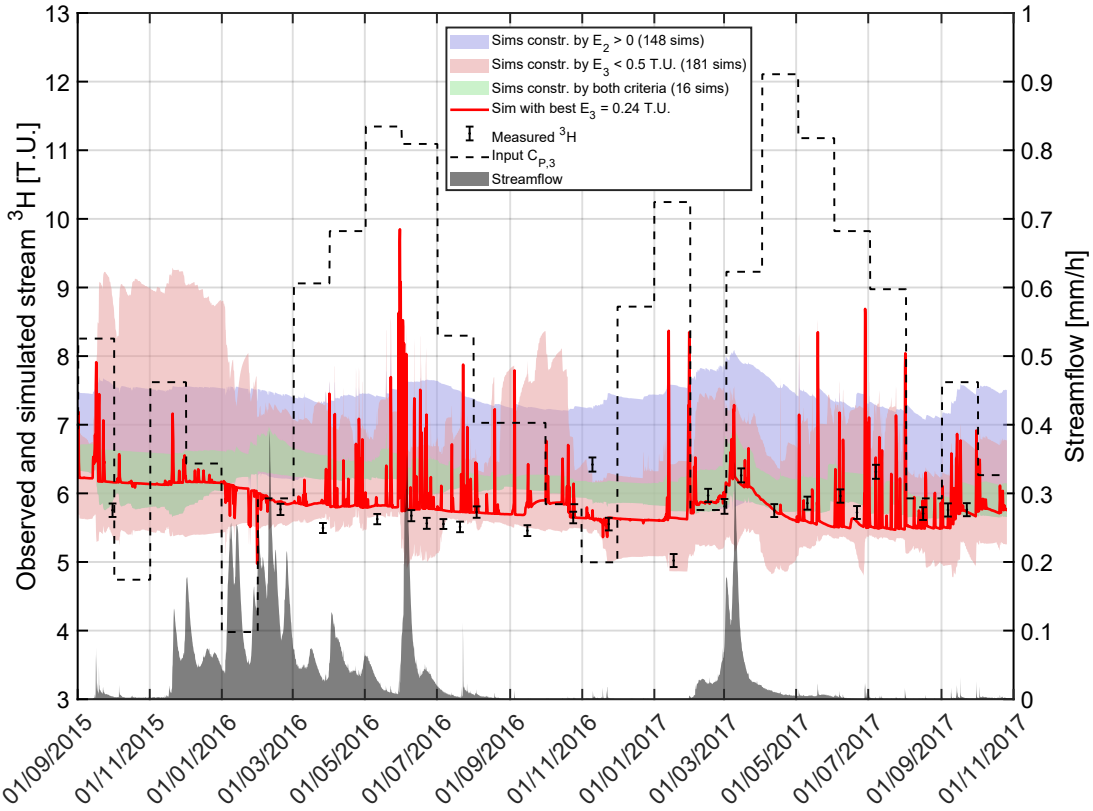

FIGURE 4.5: Simulations of stream concentrations in tritium compared to observations and to the variability in precipitation.

we computed statistics of a storage measure derived from these curves (table 4.4). The 95th percentile of Ω_{tail}, called S_{95P} (black crosses in figure 4.7) allows for estimating total mobile storage $S(t)$ from Ω_{tail}. In average, the Ω_{tail} constrained by tritium or by both tracers yielded higher mobile storage $S(t)$ and smaller spread in $S(t)$ (figure 4.7 and table 4.4). Overall, the mobile storage $S(t)$ values estimated from the tracers are mutually consistent when considering the uncertainties.

4.4 Discussion

4.4.1 Reconciliation of water ages from stable and radioactive isotopes of H

Our work shows that streamflow TTDs and the related catchment mobile storage $S(t)$ can still be estimated in unsteady conditions by using "ranked" SAS functions $\Omega(S_T, t)$ (Harman 2015). Similar to Visser et al. (2019), we propose to coherently use the measurements of stream ^{2}H and ^{3}H to calibrate the parameters of the SAS functions, here defined in the age-ranked domain $S_T \in [0, +\infty[$ instead of the cumulative residence time domain $P_S \in [0, 1]$. The calibrated tail of the streamflow SAS function Ω_Q (called here Ω_{tail}) could thus be used to approximate mobile storage $S(t)$ instead of defining the value a priori. The SAS functions also allowed us to estimate the unsteady TTDs defined in the age domain T and their statistics (mean, median, etc.). Differences between the various statistics of the TTDs were smaller than the uncertainties of the calculations when comparing the results obtained with ^{2}H alone

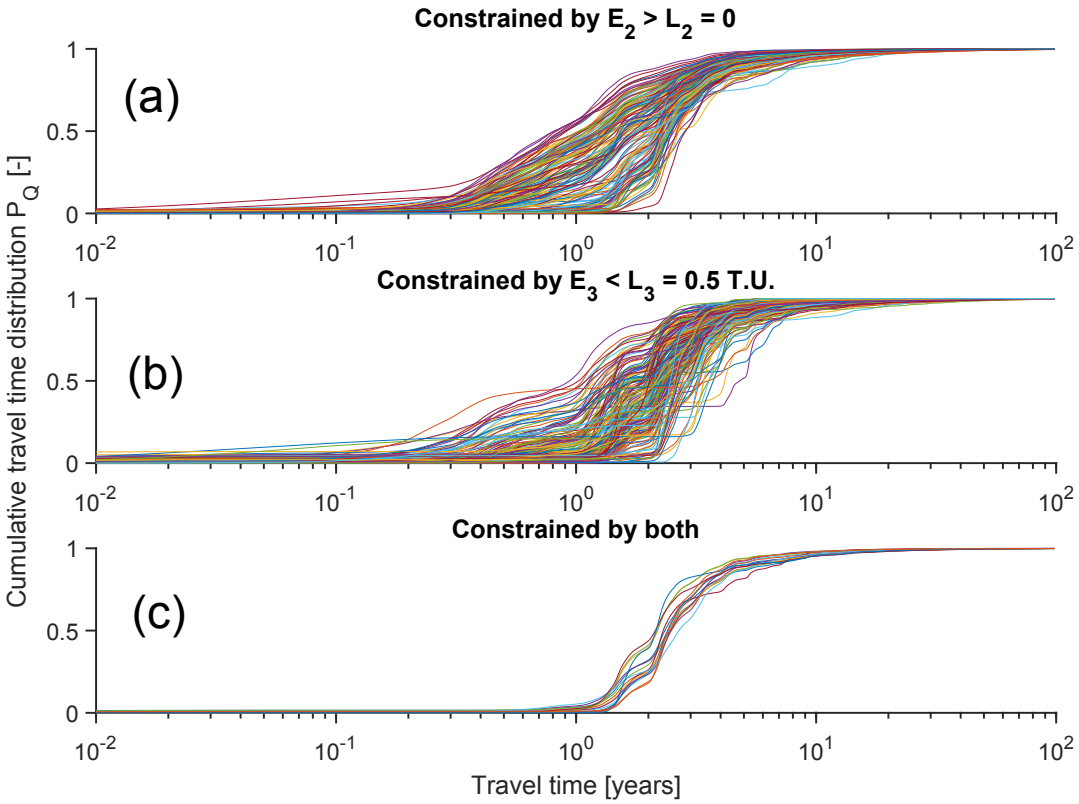

FIGURE 4.6: Flow weighted (2015–2017) cumulative stream TTDs for the behavioral parameter sets constrained by ^{2}H (a), by ^{3}H (b), and by both (c).

and with ^{3}H alone. Similarly, the derived storage estimates were consistent between ^{2}H and ^{3}H. The hypothesis of truncation of the TTD tails when using stable isotopes is therefore rejected at the present time based on the data from the Weierbach catchment. Moreover, stable isotopes do not seem to underestimate the related catchment storage compared to tritium.

These findings were made possible for a number of reasons. First, we treated ^{2}H and ^{3}H equally by calculating TTDs using a coherent mathematical framework for both tracers (i.e. same method and same functional form of TTD). Even though we clearly distinguished tritium from deuterium by accounting for the relationship between water ages and tritium activities $C_{Q,3}$ (term $\exp(-\alpha T)$ in equation 4.3), we did not use it directly to calculate T to avoid biases due to mixing of various ages at the outlet (Bethke and Johnson 2008) and to avoid the age ambiguity caused by tritium from nuclear tests (Stewart et al. 2012). Also, we did not use multiple control volumes having different TTDs determined by tracer measurements in their input and output (Małoszewski et al. 1983; Stewart and Thomas 2008; Stewart et al. 2007; Uhlenbrook et al. 2002). This way, we avoided adding large uncertainties related to difficulties in characterizing end members and gathering representative samples (Delsman et al. 2013). Second, we explicitly accounted for unsteady conditions, which has been done in only one other study using tritium (Visser et al. 2019). This allowed us to estimate realistic average TTDs corresponding to the catchment inflows, outflows, and internal flows that are highly time variant. Third, our tritium stream sampling was not focused solely on baseflow hence not biased towards old water. Fourth, we considered the entire TTDs by using various percentiles and statistics, and not only the MTT which is highly influenced by the improbable extreme values of T. This

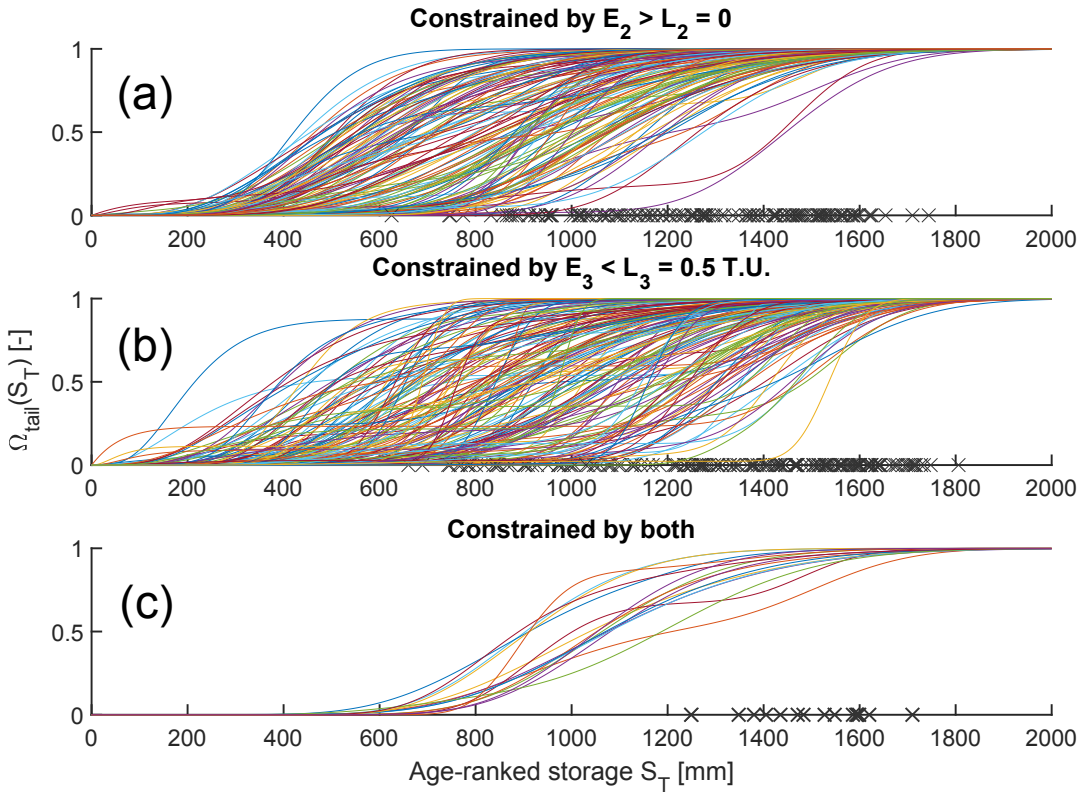

FIGURE 4.7: Cumulative right-hand tail Ω_{tail} of streamflow SAS functions for the behavioral parameter sets constrained by ^{2}H (a), by ^{3}H (b), and by both (c). Ω_{tail} is defined as the weighted sum of the two gamma components in Ω_Q. The black crosses indicate S_{95P} for each curve, i.e. the 95th percentile of Ω_{tail}

means that even if there is water older than e.g. 1,000 years in streamflow, it can be neglected if it represents less than e.g. 0.000001% of the volume. Finally, we explicitly accounted for parameter uncertainty. This is important because absolute values without an uncertainty estimate cannot be reliably interpreted.

4.4.2 Yet, tritium seems to reveal older water!

Even though the uncertainties are sufficient to account for the differences between ^{2}H- and ^{3}H-derived age and storage measures, it is worth noticing that ^{3}H systematically gave higher estimates (tables 4.3 and 4.4). The hypothesis of different transport velocities between water molecules containing deuterium and water molecules containing tritium can be rejected, because their self diffusion in water are equal (Devell 1962), and their advective velocities are the same. On the one hand, the slightly higher ages derived from tritium seem related to the apparent absence of responses of stream ^{3}H to the high precipitation ^{3}H in 2016, indicative of the dominance of old water compared to 2017. On the other hand, tritium simulations included many small peaks corresponding to flashy streamflow responses associated with young water (figure 4.5). Only some of those simulated peaks could be confirmed by the presence of stream measurements at those times, especially in 2016. More stream ^{3}H samples during these flashy events would probably support even further these simulations of young water in streamflow and shift the TTDs constrained by tritium towards younger water. We thus interpret the observed small age differences rather

as the consequence of a limited tritium sampling resolution (bi-weekly) that may still be biased towards hydrological recessions during which the youngest water fractions are absent by definition. Tritium and stable isotopes of O and H sampled synchronously at high resolution would thus pave the way for further research on stream water ages from a multi-tracer perspective.

It is also interesting that the age and storage measures estimated from a joint use of ^{2}H and ^{3}H are the highest (tables 4.3 and 4.4). In the end, tritium may have helped revealing the presence of old water in streamflow. However, it did so only when combined with deuterium. It is commonly assumed that ^{3}H is more informative about old water because of radioactive decay that relates lower tritium activities to increasing water ages (Stewart et al. 2010). However, as shown by Stewart et al. (2012) and in figure 4.2, current tritium values of the water recharged in 1980–2000 are similar to the tritium values of the water recharged today. Thus, the younger water disrupts the relationship between water age and tritium values. Adding supplementary information about the younger water in the calibration with the high frequency ^{2}H measurements may have partly helped "filtering" the currently complex relationship between water ages and tritium values, leveraging the potential of tritium for revealing the tail of the TTDs. The fact that water ages in the Weierbach are limited to about 5 years (table 4.3) could be another reason for the limited information of ^{3}H about older water. ^{3}H decays by only about 25% in 5 years, meaning that all the tritium activities of the water in the Weierbach have varied by at most $\sim$2 T.U. since water entered the catchment. This is much lower than the 10 T.U. amplitude of tritium variations in precipitation. Thus in catchments with limited residence times, radioactive decay may only give information that is redundant with the natural variability of the tracer in precipitation. In a few decades, water recharged in 1980–2000 may have completely left the catchments or may be a negligible part of storage, such that the $\log(^3H)$ of stored water may increase linearly with water age (see the recent increasing trend in $C^*_{P,3}$ in figure 4.2). Thus in a few decades, tritium could be even more informative about old water contributions because there may be no age ambiguity anymore. Furthermore, the oscillations of tritium in precipitation over long time scales (> 10 years) recently detected and related to cycles of solar magnetic activity (Palcsu et al. 2018) may give stream tritium concentrations even more age-specific meaning. Therefore it is important to re-iterate the call of Stewart et al. (2012) to start sampling tritium in streams now and for the next decades to use it in travel time analyses.

4.4.3 Age information contents of stable and radioactive isotopes

The fact that we found equal travel time and storage measures when using ^{2}H alone or ^{3}H alone does not mean that it is not worth sampling both. Our results show that more information was learned about storage and travel times (all the $D_{KL} > 0$) by using both tracers together, which resulted in lower uncertainties (lowest entropy $\mathcal{H}(X|\,(^2H \cap {}^3H))$ in table 4.2, narrower groups of curves in figures 4.6 and 4.7, lower standard deviations in tables 4.3 and 4.4). This was possible because the composite SAS functions (equation 4.5) allowed us to independently constrain different parts of the same streamflow TTD with one tracer or the other, reducing the potential trade-offs between the shapes suggested by one tracer or the other. In addition, the streamflow TTD was constrained using only stream samples. On the contrary, Stewart et al. (Table I, 2010) showed three studies where multiple TTDs corresponding to different end members (e.g. surface runoff, groundwater) are constrained by tracers sampled in the associated outlets. Although reasonable fits were shown for the

samples from the different end members, the fit of the combined TTD for the stream samples was not systematically checked (Stewart and Thomas 2008; Stewart et al. 2007; Uhlenbrook et al. 2002).

For future studies it is worth mentioning the amount of information gained per isotopic sample or per euro invested in sample analysis. This amount of information can be calculated for a given tracer by summing for all parameters the Kullback-Leibler divergences D_{KL} (see section 4.2.7) between the prior and the posterior parameter distributions. With deuterium, we learned 4.08 bits of information with 1385 samples, representing about 2.95×10^{-3} bits per sample or about 2.95×10^{-4} bits per euro. With tritium, we learned 4.47 bits of information with only 24 samples. We thus have a much higher relative information content of 0.186 bits per sample. However tritium analyses are more expensive, so the information content is only 4.33×10^{-4} bits per euro. It should be noted that for tritium the precipitation samples were not included in this cost as they were analyzed by the IAEA. Thus tritium was overall more informative than deuterium about water ages, and it was also more cost-effective. One reason for this is that tritium considerably informed us about the travel times in ET because it constrained the posterior of μ_{ET} well (table 4.2) that controls directly the ages in ET. This also highlights the importance of considering explicitly ET in streamflow travel time calculations (van der Velde et al. 2015; Visser et al. 2019) However, ^{2}H resulted in lower uncertainties for nearly all other parameters (e.g. lower Shannon entropy $\mathcal{H}(X|^2H)$, table 4.2). This is most likely due to the much higher sampling frequency for deuterium that allows for constraining the simulations better than with bi-weekly tritium measurements (see the simulation envelopes figures 4.4 and 4.5). From our experience in the Weierbach catchment, we estimate that for ^{2}H, a weekly sampling to cover the damped variations of δ^2H (i.e. about 100 samples over 2015–2017) complemented with an event-based high-frequency sampling (every 15 hours) of the flashy responses (i.e. about 300 samples over 2015–2017) could have given us as much information as the complete time series. This suggests that a more strategic sampling of ^{2}H could outperform ^{3}H in terms of cost-efficiency. The amount of information learned from the isotopic data probably scales non-linearly and probably reaches a plateau as the number of observation points grows. In the future, it would be useful to further use information theory (e.g. entropy conditional on sample size) to know how many measurements are enough and when to sample isotopes for maximum information gain on water ages. This would imply artificially re-sampling a higher-frequency isotopic time series using various strategies (e.g. Etter et al. 2018; Pool et al. 2017) and re-calibrating the model many times, which would come with an exorbitant computational price.

In the end, stable and radioactive isotopes of H have different information contents. For example, they lead to different Shannon entropy $\mathcal{H}$ for the posteriors. Also, the Kullback-Leibler divergence D_{KL} was never 0, indicating that adding one tracer after the other still allowed us to learn something about parameter values. Finally, the parameters corresponding to the best simulations in ^{2}H did not correspond to those for ^{3}H and vice versa. Our results suggest that ^{3}H is more informative about old water thanks to its radioactive decay. Yet, stable and radioactive isotopes have information in common about young water. For example, both ^{2}H and ^{3}H stream samples showed reactions to precipitation ^{2}H and ^{3}H values during flashy streamflow events, revealing the role of young water during these events. This was previously unobserved for tritium because of the sampling focused on periods outside events (Stewart et al. 2010). The theoretical span of 0–4 years pointed out in Stewart et al. (2010) should however not be taken as the only range of ages where ^{18}O, ^{2}H,

and ^{3}H have redundant information. As clearly written in Stewart et al. (2010), this limit corresponds to a steady-state exponential TTD only, while other TTD shapes (or unsteady TTDs) could yield much higher limits. More importantly, this limit can be lowered by the seasonality of the input function (see Stewart et al. 2010, p. 1647).

4.4.4 Limitations and way forward

The storage value derived from unsteady travel times constrained by tracer data (table 4.4, ∼1200–1700 mm) is noticeably larger than the maximum storage ($\simeq$ 250 mm) estimated from point measurements of porosity and water content (Martínez-Carreras et al. 2016), from water balance analyses (Pfister et al. 2017a) combined with recession techniques (Carrer et al. 2019), and from the values used in a distributed hydrological model ($\leq$ 700 mm, Glaser et al. 2016). Our storage value is more consistent with the ∼1600 mm derived from depth to bedrock and porosity data used for the Colpach catchment (containing the Weierbach) that was modeled with CATFLOW (Loritz et al. 2017). Large differences between hydrometrically-derived and tracer-derived storage estimates are not uncommon (Birkel et al. 2011a; Fenicia et al. 2010; Soulsby et al. 2009) and in fact highlight the ability of tracers to reveal the existence of stored water that is not directly involved in streamflow generation (Carrer et al. 2019; Dralle et al. 2018). This "hydraulically disconnected" storage is nevertheless important to explain the long residence times in catchments. More research is thus needed for improving the conceptualization of storage and unifying storage terminology and the various estimates obtained from tracers or other techniques. The storage value we found is not in complete contradiction with the previous estimates if we consider their uncertainties. Hydrological measurements (J, Q, and especially ET) are highly uncertain (Buttafuoco et al. 2010; Graham et al. 2010; McMahon et al. 2013; McMillan et al. 2012a; Waichler et al. 2005) and their errors are accumulated in long term water balance calculations. An explicit consideration of those uncertainties in the future could reconcile the different storage estimates. Furthermore, it is worth remembering that simplifying storage from a complex spatially-distributed quantity to a simple compact 1D water column neglects the importance of subsurface heterogeneity, surface topography, and bedrock topography for the storage and release of water. As a result, upscaling local point measurements of storage capacity that are not representative of the whole subsurface is very likely to under or overestimate the true storage capacity of the whole catchment. This is even more true if the new techniques used to scan the subsurface over larger areas such as Electrical Resistivity Tomography (ERT) are themselves associated with uncertainties, requiring adaptations (Gourdol et al. 2018) and site-specific independent knowledge (Parsekian et al. 2015).

Our conclusions rest on the assumption that the model captures the water ages in the Weierbach accurately, which was validated by the acceptable performance of the simulations. Still, the performance in δ^2H or in ^{3}H could be improved in the future by testing other models of composite SAS functions. The best NSE for deuterium simulations (called E_2) was 0.24, which is lower than the values reported in a number of studies using SAS functions (Benettin et al. 2017b; Harman 2015; van der Velde et al. 2015). It should be pointed out again the NSE may not be the most appropriate objective function to characterize performance against the δ^2H time series from the Weierbach (see Chapter 2). Future work could look for more appropriate objective functions for δ^2H, especially with respect to the information gained from model calibration. The best MAE for tritium simulations (called E_3) was 0.24. This

is slightly higher than values of RMSE (close to 0.10) reported in a number of studies using tritium (Duvert et al. 2016; Stewart and Thomas 2008; Stewart et al. 2007). However these studies had only a few stream samples, while Gusyev et al. (2013) report for instance a RMSE of 1.62 T.U. for 15 stream samples. Stream δ^2H seems to suggest larger fraction of young water than the simulations (c.f. underestimation of many flashy events in figure 4.4). Stream ^{3}H data seems to suggest larger fractions of old water than the simulations (c.f. overestimation of tritium activities over March–September 2016 in figure 4.5). A model passing through all observation points may thus show larger differences between the TTDs constrained by deuterium and the TTDs constrained by tritium. However, there are not enough ^{3}H stream samples compared to ^{2}H, so a comparison of the TTDs from this hypothetical ideal model could be misleading.

The simulations in deuterium were better for decreasing δ^2H than for increasing δ^2H (better simulations of the flashy events in δ^2H pointing downwards, figure 4.4). This is probably because the increases in δ^2H generally correspond to drier periods, during which $C_{Q,2}$ starts reacting stronger to $C_{P,2}$ indicating that young water fractions (controlled by $\lambda_1(t)$ in the model) are higher than expected. During drier periods, we expect an increase in the non-linearity of the processes delivering young water to the stream. For example, the decreasing extent of the stream network and of saturated areas observed in the Weierbach during drier conditions (Antonelli et al. in review[a],[b]) is caused by decreasing groundwater levels (Glaser et al. 2019a) and it could reduce the amounts of young water reaching the stream (c.f. van Meerveld et al. 2019). However, streamflow is lower during drier conditions, so the fractions of young water can still increase because of a less pronounced dilution of the young water in streamflow compared to wet periods. On the other hand, preferential flow observed in the soils of the Weierbach catchment and in the direct vicinity (Angermann et al. 2017; Jackisch et al. 2017; Scaini et al. 2017, 2018) may become more relevant during drier conditions and could increase the amount of young water contributing to streamflow, especially because precipitation intensities can be much higher in summer (due to thunderstorms) than in winter. The parameterization of the streamflow SAS functions via $\lambda_1(t)$ (equation C.5) includes—to some extent—the effect of wet vs. dry conditions and the role of precipitation intensity, but it seems not to fully capture how these factors influence young water fractions in the stream. Testing other parameterizations of $\lambda_1(t)$ or including other information such as soil moisture or groundwater levels in the current parameterization of $\lambda_1(t)$ may improve the simulations. Finally, the uncertainty of precipitation δ^2H could be higher during drier periods, because precipitation amounts can be too small (e.g. < 1 mm) over several weeks or because the precipitation intensities can be too high (e.g. > 5 mm/h) to be captured efficiently by the sequential rainfall sampler. This may lead to inaccuracies in the input data thus to the inability of the model to simulate the corresponding flashy events in stream δ^2H. The representation of precipitation δ^2H could thus be improved in the future by using more recent sampling techniques (e.g. Michelsen et al. 2019).

The simulations overestimated ^{3}H in the stream in 2015–2016 compared to 2017 (figure 4.5). In 2017 the simulations were better because the model used more of the young water (< 7 days old, using Ω_1) to simulate the variability and the higher values of stream ^{3}H than in 2016. The lower ^{3}H in 2015–2016 could be caused by an increased age in the older water components in 2015–2016 compared to 2017, due to changes in the importance of different subsurface flow paths in the Weierbach caused by a wetter period. The old water components Ω_2 and Ω_3 (equation

4.5) represent subsurface flows likely occurring in the lower soils and following bedrock topography (Glaser et al. 2016; Rodriguez and Klaus 2019) and in weathered bedrock fractures (Scaini et al. 2018). We used functions of S_T only for these components, meaning that the ranges of ages they select do not change considerably with time (because the distribution of S_T is rather stable). Including explicitly a dependence on time for Ω_2 and Ω_3 could help to better represent e.g. the fracture flows or deep groundwater flows in the catchment and improve ^{3}H simulations in 2015–2016. Eventually, the monthly resolution of ^{3}H in precipitation is coarser than the biweekly sampling in the stream, which can hinder accurate simulations. An increase in sampling resolution of tritium in the stream to better constrain the TTDs in the future will need to be followed by a considerable increase of sampling resolution in precipitation (Rank and Papesch 2005).

Although we found much lower deviations for the age and storage measures constrained by deuterium and tritium together (tables 4.3 and 4.4), it has to be acknowledged that this is also because there are only few accepted solutions (16), while there about 10 times more when using ^{2}H alone or ^{3}H alone. Yet, the associated curves (figures 4.6 and 4.7) fall close to each other, so the lower deviations have to be due also to lower uncertainties. A lower number of accepted solutions is in the end inevitable as it is an inherent consequence of using several performance measures independently as opposed to using a combined objective function (e.g. Hrachowitz et al. 2013; Rodriguez et al. 2018). Less strict threshold criteria for behavioral solutions could increase the number of accepted solutions but they would accept less accurate simulations, which could lead to misleading conclusions. More stream ^{3}H measurements would on the other hand allow the use of more advanced objective functions, which could lead to more accepted solutions. Eventually, the input data measured over 2010–2017 and used to spin up the model from 1960 to 2010 (J, ET, Q, and $C_{P,2}$) could be unrepresentative of the real hydrometeorological and isotopic conditions of 1960–2015 due for instance to nonstationarity or climate change. These changing conditions could affect the modeled residence times in storage and thus the estimated streamflow travel times (Wilusz et al. 2017). Different methods to spin up the model could be tested in the future (Hrachowitz et al. 2011), especially to assess the effect the effect of changing hydrometeorological and isotopic conditions on the estimation of travel times. For this, isotope tracer records that span several decades like the ones that can be reconstructed from pearl mussels shells (Pfister et al. 2019; Pfister et al. 2018) represent a crucial asset.

4.5 Conclusion

Stable isotopes of O and H and tritium are indispensable tracers to infer the streamflow TTD and derive storage estimates in catchments. Our study addressed an emerging concern about the possible deficiency of stable isotopes to infer the whole streamflow TTD. We went beyond previous data and methodological limitations and thus we did not find that stable isotopes are blind to old water fractions compared to tritium in our experimental catchment in Luxembourg. However, we found that stable isotopes and tritium do have different information contents on water ages. In fact, inferring the streamflow TTD from a joint use of both tracers better exploits their respective age information contents, which results in lower uncertainties. Even if ^{3}H appeared to be slightly more cost-effective and informative than ^{2}H, a smart sampling of the stable isotopes could outperform tritium. Future work

could compare streamflow TTD and storage from the two tracers in larger catchments where older water is expected, to give tritium more time to decay and better leverage its ability to point the presence of very old water out. We therefore recommend to: (1) keep sampling tritium in as many places as possible, as emphasized by Stewart et al. (2012); but also (2) to sample tritium at the highest frequency possible and synchronously with stable isotopes if possible. This is particularly important for the isotopic measurements in precipitation that drive all model simulations, regardless of functional forms of TTD and their parameter values. Overall this work shows that more tracer data is naturally better to gather more information about the catchments functions of storage and release.

Acknowledgements

We thank Uwe Morgenstern from GNS Science and Axel Schmidt from Bundesanstalt für Gewässerkunde (BfG) for providing access to the 2017 precipitation tritium data. We thank Laurent Gourdol for his help with the preparation of the tritium input data, and for useful discussions about estimating TTDs with tritium measurements. We thank Uwe Ehret for providing Matlab scripts to compute information theory measures. Nicolas Rodriguez and Laurent Pfister (FNR/CORE/C14/SR/8353440/STORE-AGE), and Julian Klaus (FNR/CORE/C17/SR/11702136/EFFECT) thank the Luxembourg National Research Fund. We thank Jérôme Juilleret for his help to collect tritium samples in the field. The authors acknowledge support by the state of Baden-Württemberg through bwHPC. We thank Barbara Glaser for her help with the input data (PET).

TABLE 4.2: Parameter ranges and information measures before and after calibration to isotopic data

Parameter	S_{th}	ΔS_{th}	S_u	f_0	λ_1^*	λ_2	μ_2	θ_2	μ_3	θ_3	μ_{ET}	θ_{ET}
Unit	mm	mm	mm	–	–	–	mm	mm	mm	mm	mm	mm
Range for $E_2 > L_2$	$[21, 196]$	$[0.14, 20]$	$[1.4, 50]$	$[0, 1]$	$[0, 0.76]$	$[0, 1]$	$[149, 1530]$	$[6, 100]$	$[21, 1561]$	$[2, 100]$	$[51, 926]$	$[1, 99]$
Range for $E_3 < L_3$	$[20, 200]$	$[0.1, 20]$	$[1, 50]$	$[0, 1]$	$[0, 0.92]$	$[0, 1]$	$[58, 1564]$	$[0, 100]$	$[67, 1600]$	$[0, 100]$	$[0, 959]$	$[1, 100]$
Range for $E_2 > L_2$ and $E_3 < L_3$	$[25, 177]$	$[0.17, 20]$	$[11, 49]$	$[0, 0.82]$	$[0, 0.76]$	$[0.1, 1]$	$[897, 1530]$	$[7, 69]$	$[440, 1561]$	$[7, 96]$	$[51, 120]$	$[3, 98]$
Binning[a]	[20:20:200]	[0:2:20]	[0:5:50]	[0:0.1:1]	[0:0.05:1]	[0:0.1:1]	[0:100:1600]	[0:10:100]	[0:100:1600]	[0:10:100]	[0:50:1600]	[0:10:100]
$\mathcal{H}(X)$[b]	3.17	3.32	3.32	3.32	3.71	3.32	4	3.32	4	3.32	5	3.32
$\mathcal{H}(X\|^2H)$	3.12	3.27	3.29	3.22	2.97	3.2	3.44	3.25	3.59	3.23	2.77	3.24
$\mathcal{H}(X\|^3H)$	3.1	3.3	3.31	3.3	3.23	3.24	3.73	3.25	3.83	3.25	1.53	3.22
$\mathcal{H}(X\|(^2H \cap {}^3H))$	2.52	2.95	2.53	2.91	2.65	2.91	2.48	2.18	2.91	2.75	0.7	2.48
$D_{KL}(X\|(^2H \cap {}^3H), X\|^2H)$	0.39	0.36	0.75	0.3	0.36	0.27	1.23	1.03	0.75	0.48	1.29	0.72
$D_{KL}(X\|(^2H \cap {}^3H), X\|^3H)$	0.48	0.36	0.8	0.42	0.33	0.34	1.08	0.91	0.84	0.37	2.04	0.75

[a] Binning is indicated as $[a : b : c]$, where a is the left edge of the first bin, b is the bin width, and c is the right edge of the last bin.
For instance, $[7 : 2 : 11]$ indicates data sorted with the two bins $[7, 9]$ and $[9, 11]$.
[b] $\mathcal{H}$ and D_{KL} are expressed in bits.

TABLE 4.3: Statistics of $\overleftarrow{P_Q}(T)$ constrained by deuterium or tritium

Age statistics	^{2}H ($E_2 > 0$) [mean ± std]	^{3}H ($E_3 < 0.5$ T.U.) [mean ± std]	^{2}H and ^{3}H [mean ± std]
10th percentile [years]	0.78 ± 0.49	1.10 ± 0.57	1.44 ± 0.11
25th percentile [years]	1.16 ± 0.56	1.54 ± 0.59	1.85 ± 0.22
Median age [years]	1.77 ± 0.55	2.19 ± 0.64	2.38 ± 0.15
75th percentile [years]	2.78 ± 0.61	3.07 ± 0.74	3.26 ± 0.39
90th percentile [years]	4.64 ± 1.27	4.79 ± 1.41	5.19 ± 0.86
Mean age [years]	2.90 ± 0.54	3.12 ± 0.59	3.45 ± 0.28
F_{yw}[a] [%]	1.5 ± 1.6	1.8 ± 2.3	0.61 ± 0.53
F(T < 6 months) [%]	10 ± 8.6	6.3 ± 8.2	0.75 ± 0.58
F(T < 1 year) [%]	24 ± 17	11 ± 12	2.1 ± 1.5
F(T < 3 years) [%]	77 ± 8.5	71 ± 16	70 ± 6.6

The mean and standard deviations are calculated from all retained behavioral solutions for a given criterion.

[a] Fraction of "young water" (Kirchner 2016a), younger than 0.2 years.

TABLE 4.4: Storage estimate S_{95P} constrained by deuterium or tritium

Statistics of S_{95P}	^{2}H ($E_2 > 0$)	^{3}H ($E_3 < 0.5$ T.U.)	^{2}H and ^{3}H
Mean ± st. dev. [mm]	1275 ± 245	1335 ± 279	1488 ± 135
Median ± st. dev. [mm]	1281 ± 245	1392 ± 279	1505 ± 135
Min [mm]	625	660	1249
Max [mm]	1744	1806	1710

S_{95P} is calculated as the 95th percentile of Ω_{tail} (eq. 4.10)

Conclusion

Summary

The work presented in this thesis has contributed to travel time research by addressing several connected research gaps about what controls the shape and time variability of the streamflow TTD, and how it can be determined from tracer data. A unique isotopic tracer data record was acquired in Luxembourgish catchments from extensive experimental and laboratory work. In parallel, theoretical investigations and the development of a travel time model using an existing data set (Andrews Forest, Oregon, USA) allowed first insights into the streamflow TTD. These initial findings paved the way for proposing an improved conceptualization of streamflow TTDs in complex landscapes and for testing it in an experimental catchment in Luxembourg, which in turn helped demonstrating how to leverage water age information from a dual-tracer approach. The main conclusions from the four presented studies are:

- The Inverse Storage Effect (i.e. preferential mobilization of younger water during wetter conditions) may not be always valid for streamflow. In particular, during transitions from dry to wet periods and vice versa in Mediterranean climates, the variability of streamflow travel times cannot be simply explained by storage variations. The parameterization of the streamflow TTD and SAS function may need to include more state variables than catchment storage or to include hysteresis in the storage-age relationships. [chapter 1]

- Despite their prevalent use, simple TTDs with mainly one peak or mode may not describe the transport processes in all catchments well. Composite SAS functions (leading to complex TTDs with several peaks in unsteady conditions) require more parameters and increase predictive uncertainty, but they better account for a superposition of streamflow generation processes with contrasting ages. They also allow more detailed simulations of tracer dynamics in streamflow in hydrological systems with complicated flow mechanisms [chapters 2 & 3].

- Superimposed streamflow generation processes can result in contrasting travel times at an outlet if they correspond to different flow path lengths or different water velocities, and if they provide comparable fractions of the water volumes in that outlet. Furthermore, accurate water chemistry simulations require all peaks of the TTDs to be determined, which may not be systematically achieved by using only one conservative tracer and unimodal TTD (or SAS function) models. Composite SAS functions are an efficient multimodal solution, and their different modes deduced from conservative tracers can be further validated by using complementary water chemistry data. [chapter 3].

- Stable isotopes of O and H used as tracers do not systematically truncate the long tails of the streamflow TTD compared to tritium. When the tracers are consistently used in a coherent unsteady mathematical formulation, travel times

and catchment storage deduced from ^{2}H (or ^{18}O) and ^{3}H are similar (within the uncertainty bounds). Using both tracers together is in fact advantageous since it drastically reduces the uncertainties and it may partly disambiguate the tritium-derived ages from the legacy of nuclear tests. Nevertheless, ^{3}H may still be slightly more informative about long travel times than stable isotopes thanks to its radioactive decay. [chapter 4]

In summary, this thesis contributed to the travel time research field by proposing improvements in the experimental, theoretical, and modelling areas.

Further research needs

Future research is needed to address the remaining knowledge gaps that this work pointed out. First, in order to find more meaningful state variables to parameterize time-varying TTDs, a better understanding of what controls the statistics of the average TTDs (e.g. MTT) is necessary. Geographical factors (e.g. unsaturated storage, slope aspect, topography, soil characteristics) (Broxton et al. 2009; Dunn et al. 2007; Heidbüchel et al. 2013; Hrachowitz et al. 2009a, 2010) and climatic factors (e.g. precipitation intensity and seasonality, rainfall patterns) (Hrachowitz et al. 2009a, 2010; Sayama and McDonnell 2009; von Freyberg et al. 2018; Wilusz et al. 2017) may not be the only variables controlling travel times. Internal variables such as soil moisture and groundwater levels may be obvious factors to consider. The role of storage for travel times may also be clarified by proposing a conceptualization of storage in catchments that better reflects the different pools of more mobile and less mobile water and their interactions (Birkel et al. 2011a; Carrer et al. n.d.; Carrer et al. 2019).

ET has a noticeable influence on RTDs thus on TTDs by extension (van der Velde et al. 2015; Visser et al. 2019). Estimates of ET fluxes are generally uncertain (Buttafuoco et al. 2010; Graham et al. 2010; Waichler et al. 2005), in particular due to the low resolution or absence of onsite data (McMahon et al. 2013), and due to many necessary assumptions to derive actual ET from potential ET (Allen et al. 1998). Thus, less uncertain actual ET values obtained for instance from the eddy covariance method using a flux tower could improve streamflow TTD calculations and reduce their uncertainties. A related outstanding knowledge gap is the water ages used by ET (Berry et al. 2018). Chapters 1, 2, and 4 relied on simple assumptions about the TTD of ET because no tracer data could be used to constrain it directly. Gathering tracer data representative of the ET flux (stable isotopes in plant stems and water vapor) is ongoing in the experimental catchments of Luxembourg and should tackle this unknown.

Chapters 3 and 4 highlighted that an increase in model complexity for the TTDs and SAS functions is required for more accurate tracer and solute simulations in catchment streamflow. Future work could assess more precisely the level of complexity that is justified by the tracer data, using advanced concepts such as information theory (e.g. Loritz et al. 2018). Similarly, it would be useful to investigate the trade-off between decreasing model errors and increasing parameter uncertainty from a statistical point of view ("bias-variance trade-off", Guthke 2017; James et al. 2013). More generally, a limited number of studies characterized the uncertainty of travel time calculations (Benettin et al. 2017b; Gallart et al. 2016; Jing et al. 2019; McGuire and McDonnell 2006; Stockinger et al. 2016, 2015; Timbe et al. 2014). Better dialogues between experimentalists and modellers, and the use of "soft data" (Seibert and McDonnell 2002) in the calibration of TTD models could greatly reduce travel time uncertainties, for example by helping to justify the choice of a functional

form of TTD for a particular catchment based on experimental evidence (see chapter 2). Improving our understanding of the sources and magnitude of travel time uncertainties is now required to formulate outrageous hypotheses (Burt and McDonnell 2015; Davis 1926), accept/reject them more confidently (Neuweiler and Helmig 2017; Pfister and Kirchner 2017), and thus draw stronger conclusions using TTDs.

Chapter 4 proposed to use more than one isotopic tracer to better constrain the streamflow TTD, and suggested to systematically sample tritium as often as possible (Stewart et al. 2012). A higher tracer resolution (especially in precipitation, because it is a model input) paves the way for a better characterization of the streamflow TTD. Tritium is not the only valuable tracer that can be added, as novel hydrological tracers are emerging from different subdisciplines of hydrology (e.g. Abbott et al. 2016; Hissler et al. 2014; Pfister et al. 2017b). The joint use of many different tracers for TTD calculations has an immense potential for revealing new shapes, new properties, and new controls on TTDs.

Finally, TTDs cannot be measured directly at the catchment scale, but only inferred indirectly from tracer data using assumptions on their shape. Employing other techniques that measure TTDs directly from tracer data without assumptions on their shape (Cirpka et al. 2007; Kim et al. 2016; Kirchner 2019; Massoudieh et al. 2012; Turner and Macpherson 1990a) has the potential for revealing the true complexity of TTD shapes. Artificial tracer applications at large scales (e.g. catchment scale) as shown by Rodhe et al. (1996) for the 0.63 ha Gårdsjön catchment and by Kendall et al. (2001) for the 490 m^2 Hydrohill catchment could also allow a direct measurement of the TTDs without any model. These techniques could thus better relate TTDs to the wide variety of hydrological processes.

In an ideal future, tracer data would be available at high resolution to characterize accurately and precisely the water ages in the outlets of many catchments of different sizes, shapes, land uses, and geologies around the globe. A limitless computing power would then allow to efficiently test a multitude of hypotheses about the TTDs for each site, from the most simple hypothesis to the most complex. An accurate global picture of catchment TTDs with limited uncertainties would then emerge and allow a complete understanding on what controls the shape and variability of TTDs.

Concluding remarks

TTDs are fundamental for intuitive interpretations of water storage and renewability, water flow paths, and water velocities in catchments, and they provide a simple yet comprehensive communication tool regarding water quality. TTDs are also powerful and efficient models of tracer transport through catchments that can summarize a large number of interacting processes in a single function of water age, or extract water age information about these processes from even the most intricate tracer data. Estimating TTDs in catchments has nevertheless become a catchment modelling sub-field on its own, with issues (e.g. regarding data and models) similar to other modelling fields of environmental sciences. My work contributed to tackling some of these problems by improving the conceptualization of unsteady streamflow TTDs and their estimation from isotopic tracers. The 2-year high frequency isotopic tracer data set collected in several Luxembourgish catchments, the methods, and the numerical tools developed to calculate unsteady TTDs in the framework of this thesis could be directly taken up for future investigations on Luxembourgish water resources, in collaboration with the Administration de la Gestion de l'Eau (AGE) for

instance. The research work presented in this thesis, aiming at determining more efficiently and more accurately the streamflow TTD, is therefore a relevant contribution towards improved catchment water management practices and the sustainability of natural water resources.

Appendix A

Supplementary information for Chapter 2

This appendix summarizes the modeling of stream deuterium (δ^2H) dynamics in the Weierbach catchment using StorAge Selection (SAS) functions with a single control volume encompassing the whole catchment. A brief description of the previous candidate models of SAS functions is given in section A.1. Table A.1 summarizes the best performances obtained with these models. Figures A.1 to A.9 show the corresponding simulations of stream δ^2H against the high-frequency measurements spanning 2015–2017.

A.1 SAS functions previously used for δ^2H simulations

SAS functions are usually defined using cumulative residence time P_S or age-ranked storage $S(t)\ P_S = S_T$, where $S(t)$ is catchment storage (e.g. mm). Common SAS functions include:

(A) the power-law distribution (Benettin et al. 2017b) defined with a parameter k as: $\omega_Q(P_S, t) = k\, P_S^{k-1}$

(B) the beta distribution (van der Velde et al. 2015) defined with shape parameters a and b as $\omega_Q(P_S, t) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, P_S^{a-1}\,(1-P_S)^{b-1}$

(C) and the gamma distribution (Harman 2015) defined with a shape parameter k and a scale parameter θ as: $\omega_Q(S_T, t) = \frac{1}{\Gamma(k)\,\theta^k}\, S_T^{k-1}\, e^{-S_T/\theta}$. In the manuscript, we considered the mean parameter $\mu = k\,\theta$ and calculated k as μ/θ

Additionally, time-variant versions of models A, B, and C, labeled here A_{tv}, B_{tv}, and C_{tv}, can be obtained by considering that the parameters are functions of storage $S(t)$, as detailed in the papers cited above.

Models A, B, C, and their corresponding time-variant versions A_{tv}, B_{tv}, and C_{tv} were calibrated against stream δ^2H using a similar method than described in section 2.2.8 of the manuscript.

Models A, B, and C were unable to reproduce the damped seasonal variability but also the flashy events in δ^2H. This suggests that these models are not flexible enough to account for the flow processes and the associated travel times in the Weierbach that generate the streamflow δ^2H dynamics, and/or for the time variance of these. The best simulations tended to a straight line (Figures A.1 to A.3), with associated NSE values close to 0 (Table A.1), indicating that a straight line is better in terms of NSE than any of the simulations with these models. This is not consistent with the observations showing that a straight line is not better than a simulation with a least some seasonal variability in δ^2H.

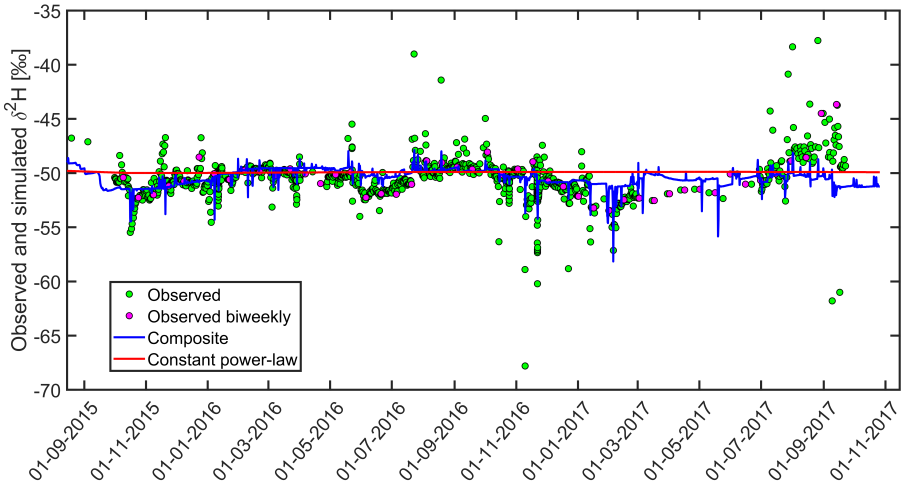

FIGURE A.1: Calibrated model A (red) and results from the composite SAS function model (blue).

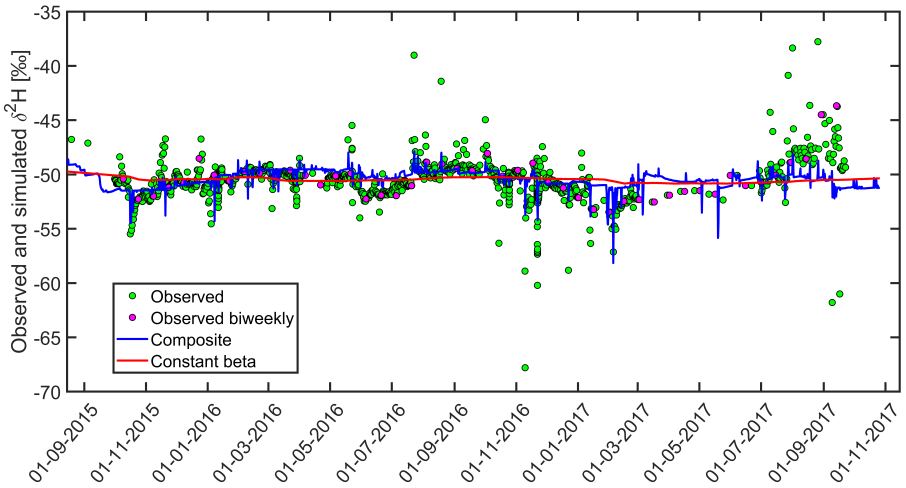

FIGURE A.2: Calibrated model B (red) and results from the composite SAS function model (blue).

Using time-varying parameters in the SAS functions increased the flexibility of the models A_{tv}, B_{tv}, and C_{tv} compared to A, B, and C. The time-variant SAS functions were therefore able to simulate some seasonal variability in δ^2H (Figures A.4 to A.6), indicating that they captured better the time variance of flow processes and the associated travel times. Despite this, they still failed to simulate accurately some periods (in particular Jan–Jun 2016) because these may be associated with other ranges of travel times than the ranges assumed by these models. These different travel times may come from the presence of shorter and longer flow paths contributing to the stream (due e.g. to storage of water closer or further from the stream in the hillslopes), or from slower and faster flows (due e.g. to different precipitation amounts or intensities). Moreover, models A_{tv}, B_{tv}, and C_{tv} still failed to simulate the flashy δ^2H events. This is because they were unable to create the rapid onset and offset of responses in δ^2H associated with the activation and deactivation of saturation excess flow in the riparian area. This resulted in NSE values that were negative or still close to 0 (Table A.1), despite visually more dynamic simulations than A, B, and C. For model C_{tv}, the increased variability of the simulation even decreased the NSE

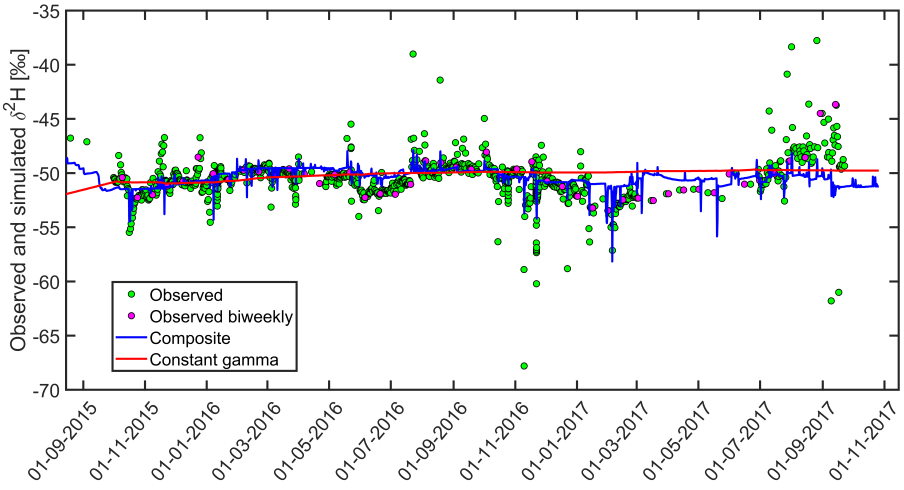

FIGURE A.3: Calibrated model C (red) and results from the composite SAS function model (blue).

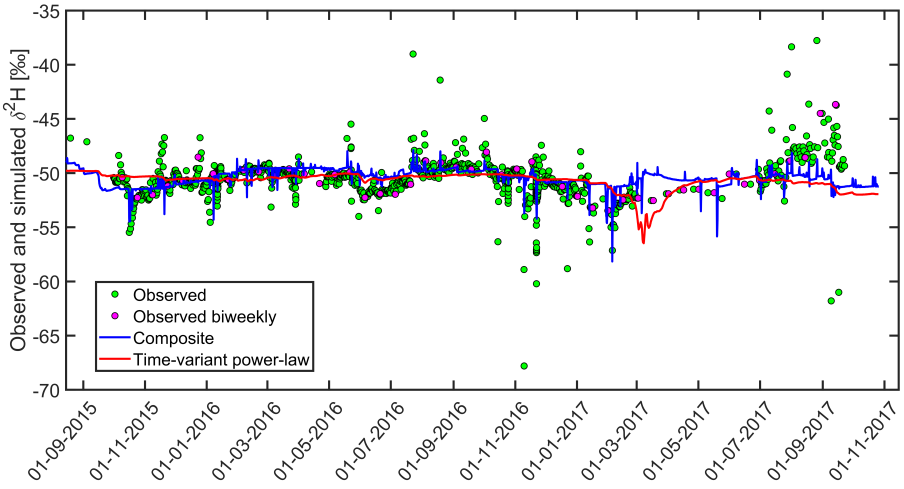

FIGURE A.4: Calibrated model A_{tv} (red) and results from the composite SAS function model (blue).

because it resulted in larger errors than model C. NSE values were also lower for B_{tv} than for B, while B_{tv} looks more realistic than B. This shows that the NSE can penalize more flexible and more realistic models if those models have larger errors in average.

To allow the model to simulate a superposition of contrasting travel times and to obtain a better performance, we introduced a second component in the SAS function. This can be seen conceptually as the addition of another streamflow generation mechanism in the SAS function. We chose to use two gamma distributed components in the SAS functions, consistent with the observed $1/f^{\alpha}$ spectral density of stream chemistry (Aubert et al. 2014; Benettin et al. 2015b; Godsey et al. 2010; Harman 2015; Hrachowitz et al. 2015; Kirchner and Neal 2013; Kirchner et al. 2000), and with the increasing use of age-ranked storage S_T (Harman 2015; Pangle et al. 2017). Thus we defined model D as:

(D) $\omega_Q(S_T,t) = \frac{\lambda}{\Gamma(k_1)\,\theta_1^{k_1}}\, S_T^{k_1-1}\, e^{-S_T/\theta_1} + \frac{1-\lambda}{\Gamma(k_2)\,\theta_2^{k_2}}\, S_T^{k_2-1}\, e^{-S_T/\theta_2}$, where k_1, k_2, θ_1, and θ_2 are the shape and scale parameters, and λ is the partition coefficient between the

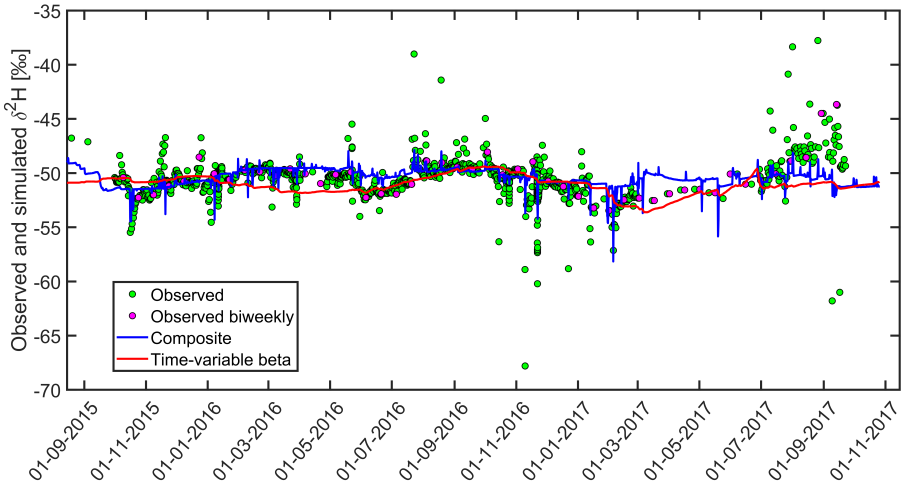

FIGURE A.5: Calibrated model B_{tv} (red) and results from the composite SAS function model (blue).

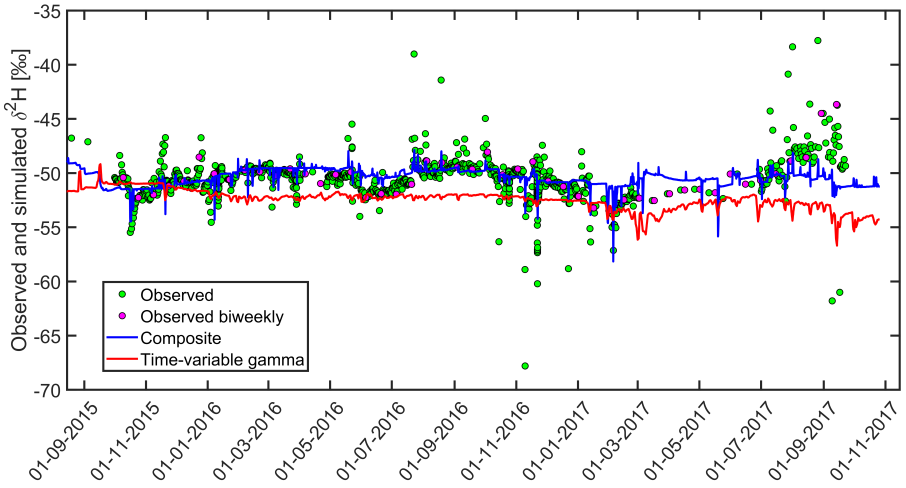

FIGURE A.6: Calibrated model C_{tv} (red) and results from the composite SAS function model (blue).

two components.

Model D simulated better the damped seasonal variability in stream δ^2H than SAS functions with one component (figure A.7), with a NSE of 0.08. Yet, it was unable to simulate the flashy δ^2H events that required a third, flashy component associated with young water. This component was later added to form the current model:

(E)

$$\begin{cases} \frac{\lambda_1(t)}{S_u} + \frac{\lambda_2(t)}{\Gamma(k_2)\,\theta_2^{k_2}}\, S_T^{k_2-1}\, e^{-S_T/\theta_2} + \frac{\lambda_3(t)}{\Gamma(k_3)\,\theta_3^{k_3}}\, S_T^{k_3-1}\, e^{-S_T/\theta_3}, & S_T \in [0, S_u] \\ \frac{\lambda_2(t)}{\Gamma(k_2)\,\theta_2^{k_2}}\, S_T^{k_2-1}\, e^{-S_T/\theta_2} + \frac{\lambda_3(t)}{\Gamma(k_3)\,\theta_3^{k_3}}\, S_T^{k_3-1}\, e^{-S_T/\theta_3}, & \text{otherwise} \end{cases}$$

Model E was eventually able to capture the damped seasonal variability in δ^2H and the flashy events as well, which explains its better score compared to all other models.

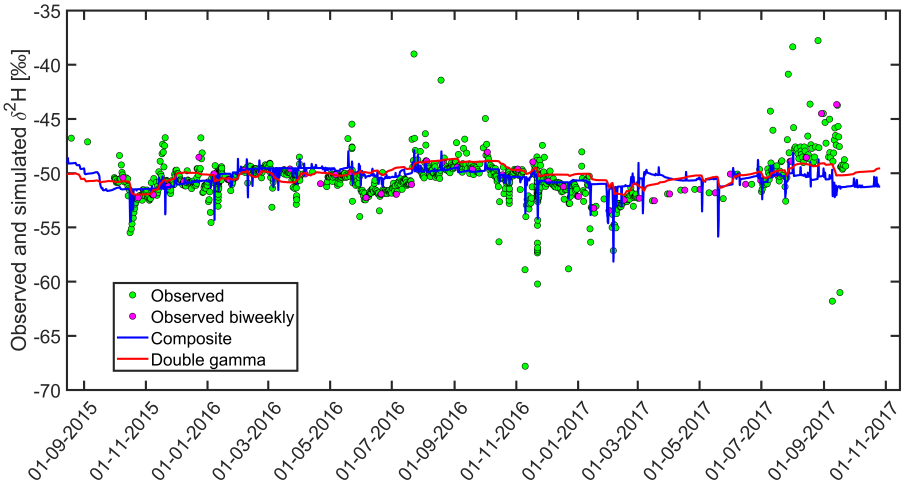

FIGURE A.7: Calibrated model D (red) and results from the composite SAS function model (blue).

We performed an additional test to see if it is possible to simplify model E by using a single gamma component (model E′, figure A.8). We then compared the gamma components from model E′ and model E (figure A.9).

(E′)

$$\begin{cases} \frac{\lambda_1(t)}{S_u} + \frac{\lambda_2(t)}{\Gamma(k_2)\,\theta_2^{k_2}}\, S_T^{k_2-1}\, e^{-S_T/\theta_2}, & S_T \in [0, S_u] \\ \frac{\lambda_2(t)}{\Gamma(k_2)\,\theta_2^{k_2}}\, S_T^{k_2-1}\, e^{-S_T/\theta_2}, & \text{otherwise} \end{cases}$$

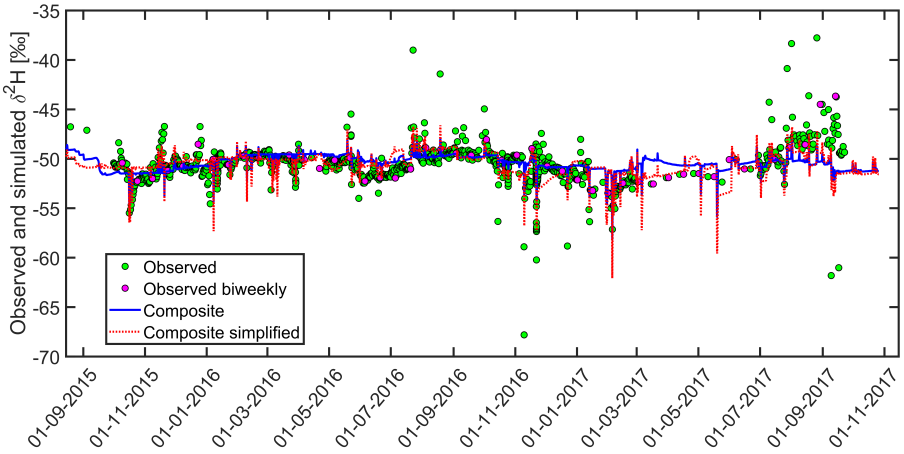

FIGURE A.8: Calibrated model E′ (red) and results from the composite SAS function model (blue).

Model E′ performs almost equally well than model E (Table A.1). However, the gamma component is shifted to larger values of S_T. Whereas ω_2 and ω_3 peaked around S_T = 600 mm and S_T = 1000 mm, which corresponds to 1-year-old water and 3-years-old water respectively (figure 2.8 in main text), the gamma pdf of model E′ peaks around S_T = 1400 mm, which corresponds to water about 10 years old. Furthermore, the gamma component in model E′ starts only around S_T = 1000 mm, which corresponds to 3-year-old water at least. It seems unrealistic that the majority

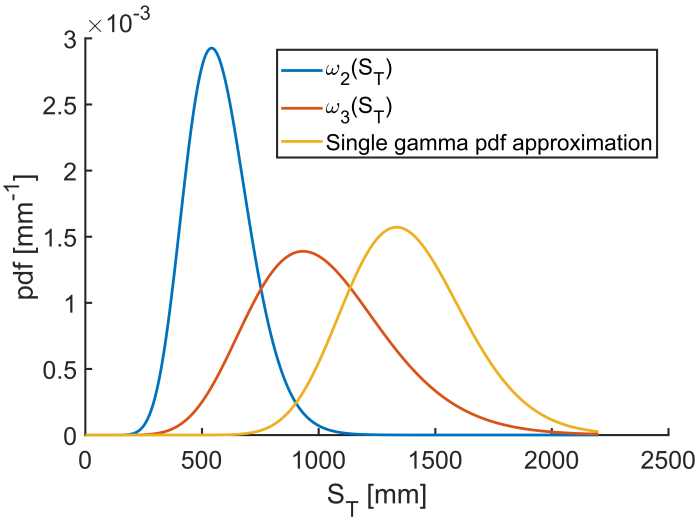

FIGURE A.9: Comparison between the two gamma components ω_2 and ω_3 of the calibrated model E and the single gamma component of the simplified model E′.

of stream water in the catchment is at least 3 years old and that a large part of it is about 10 years old, giving more credit to the results from model E.

TABLE A.1: Best NSE values for the previous candidate models

Model type	Model name	Best NSE score	Best NSE score for biweekly δ^2H
Constant power-law	A	-0.10	-0.07
Constant beta	B	0	0.05
Constant gamma	C	-0.03	-0.03
Time-variable power-law	A_{tv}	0	-0.02
Time-variable beta	B_{tv}	-0.03	0.03
Time-variable gamma	C_{tv}	-0.91	-1.71
Double gamma	D	0.08	0.25
Composite (current model)	E	0.27	0.04
Simplified composite (only 1 gamma pdf)	E′	0.24	-0.93

The NSE score for biweekly data shows that models A to C_{tv} did not perform well even without the flashy δ^2H events. This is because the slower seasonal variability in δ^2H still needs at least two components in the SAS functions to capture the different travel times to the stream (supported by the fact that model D has a NSE of 0.25 for the biweekly data). The final model E performs worse than model D (which is similar to E but without the flashy component) because the bi-weekly data does not contain flashy variations, which results in larger errors for model E than for model D. Model E′ performs worse than model E for the biweekly data because the slower seasonal variability of δ^2H is reproduced slightly better with two gamma components instead of one.

A.2 Additional model details

As mentioned in section 2.4.4, the NSE assumes normally distributed errors centered on zero and with a constant standard deviation equal to that of the observations (red histogram in figure A.10). The true model errors based on our calibrated model (E)

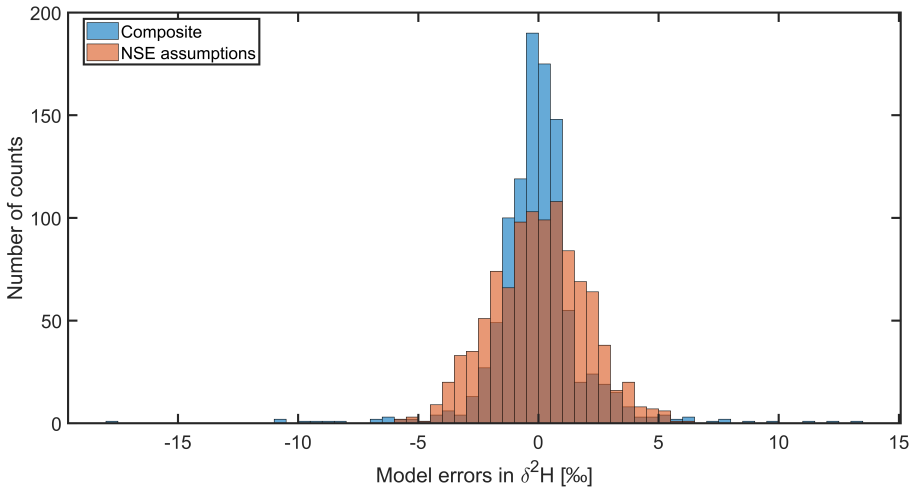

FIGURE A.10: Theoretical and observed model errors in deuterium. The observed errors are obtained from the calibrated model E. The theoretical errors are obtained from a normal distribution centered on 0 and with the standard deviation of the observations, which corresponds to the assumptions behind the NSE.

are not in agreement with these assumptions (blue histogram in figure A.10). The true errors are not symmetrical, but show two distinct groups that may correspond to varying errors magnitudes over different periods (i.e. a time-varying error variance). Model likelihood represents how likely a parameter set is, when comparing the simulations against the observations. It is calculated as the sum of the probabilities of the model errors (represented here roughly as the height of the bars in figure A.10) obtained from the assumed error model (e.g. the NSE). It can be seen that the NSE will consider that smaller errors are less likely than they really are (figure A.10, the red bars are under the blue bars for errors close to 0), which will affect the model fit to the damped seasonal variations in δ^2H. The NSE will also consider that large errors are very unlikely i.e. almost impossible (figure A.10, the red bars are negligible compared to blue bars for errors bigger than ± 5‰), which will affect the model fit to the flashy events in δ^2H. As a result, using the NSE for model calibration can limit the ability to find good model fits.

The relationships between the weight $\lambda_1(t)$ of the first uniform component (pdf) ω_1 and the storage variables $S(t)$ and $\overline{\Delta S}(t)$, described by the equations:

$$\lambda_1(t) = \lambda_1^* \left[f(t) + (1 - f(t))\, g(t) \right] \tag{A.1}$$

$$f(t) = f_0 \left[1 - \tanh\left(\left(\frac{S(t)}{S_{min} + S_{th}} \right)^m \right) \right] \tag{A.2}$$

$$g(t) = 1 - \exp\left(- \frac{\overline{\Delta S}(t)}{\Delta S_{th}} \right) \tag{A.3}$$

are represented in figures A.11 and A.12 for the values observed from 2015 to 2017 in the Weierbach. We used the calibrated parameter values $\lambda_1^* = 0.11$, $f_0 = 0.10$, $S_{th} = 105$ mm, $m = 1000$, and $\Delta S_{th} = 3.97$ mm. These figures show several things:

- $\lambda_1(t)$ is always lower than the parameter $\lambda_1^* = 0.11$

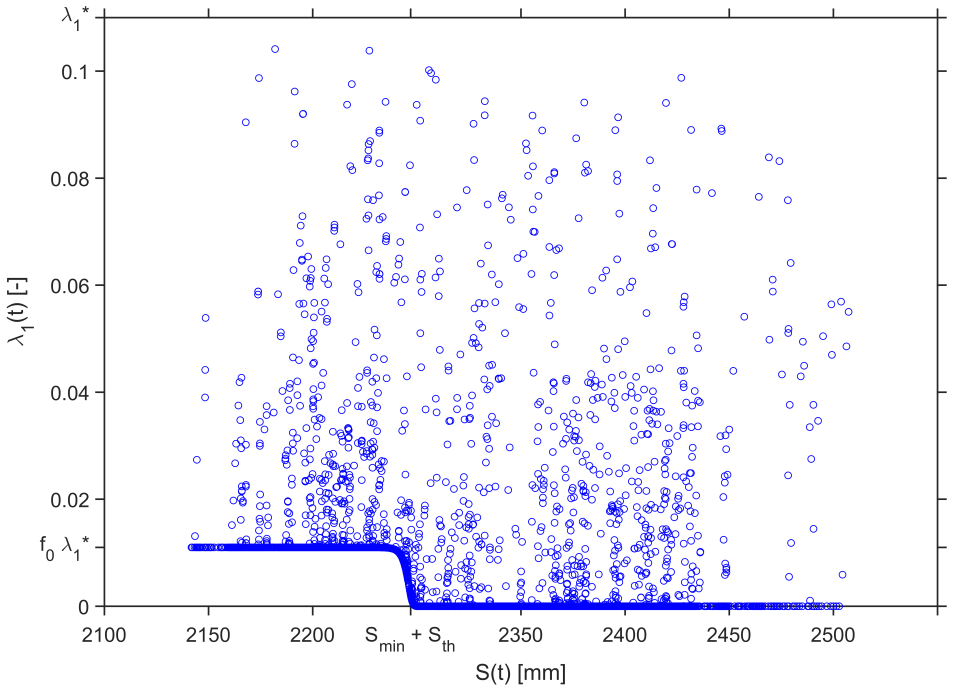

FIGURE A.11: Observed relationship between storage $S(t)$ and the weight $\lambda_1(t)$ of the first uniform component ω_1 in the composite SAS function.

- The minimum value of $\lambda_1(t)$ is slightly higher during drier periods, corresponding to $S(t) \leq S_{min} + S_{th} = 2200$ mm. This minimum value during drier periods is $f_0\,\lambda_1^*$, thus it is controlled by the parameter f_0
- The large increases of $\lambda_1(t)$ are mostly controlled by increases in $\overline{\Delta S}(t)$, which are driven by large precipitation events comparable to the parameter $\Delta S_{th} = 3.97$ mm.

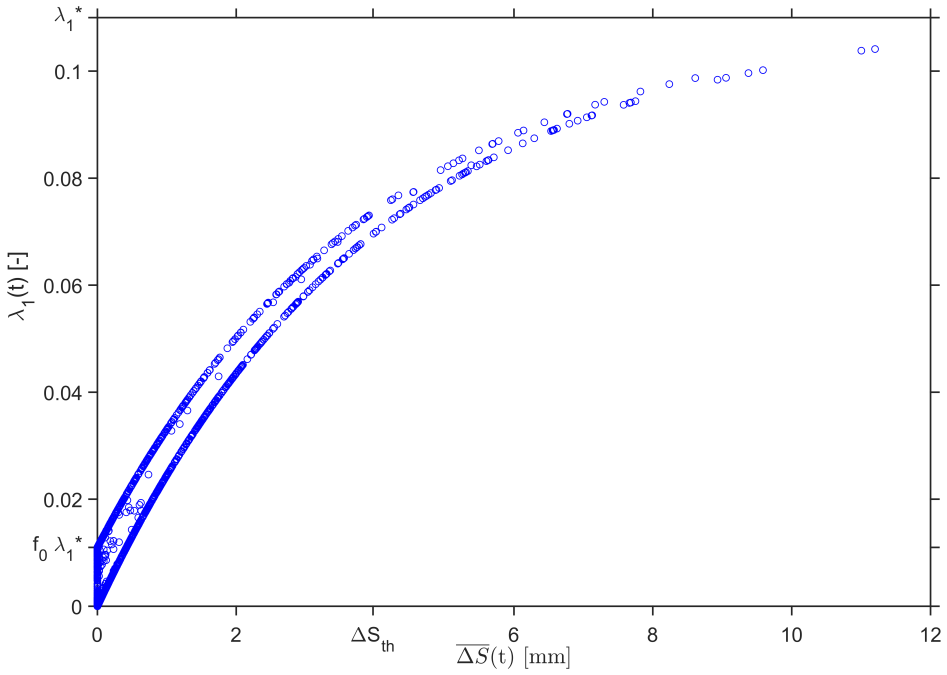

FIGURE A.12: Observed relationship between a proxy of storage variations $\overline{\Delta S}(t)$ and the weight $\lambda_1(t)$ of the first uniform component ω_1 in the composite SAS function.

Appendix B

Supplementary information for Chapter 3

B.1 Analytical formulas for the SAS functions

B.1.1 Fractional SAS functions

The SAS functions used for the lysimeter case study take the form of "fractional" SAS functions – or fSAS (Harman 2015), i.e. they are defined over the normalized ranked storage P_S, which is the cumulative distribution of ages in storage p_S. Thus, fSAS functions are probability distributions defined over the finite domain [0,1], where domain values close to 0 indicate the youngest waters in storage while values close to 1 indicate the oldest. The analytical formulas of the cumulative fSAS functions $\Omega(P_S)$ are reported below:

Single power function

$$\Omega(P_S) = (P_S)^k \tag{B.1}$$

with 1 parameter (k).

Truncated normal distribution

$$\Omega(P_S) \sim \mathcal{N}_T(P_S, m, s) = \frac{erf\left(\frac{P_s - m}{s\sqrt{2}}\right) - erf\left(\frac{m}{s\sqrt{2}}\right)}{erf\left(\frac{1-m}{s\sqrt{2}}\right) - erf\left(\frac{m}{s\sqrt{2}}\right)} \tag{B.2}$$

with 2 parameters m and s. The function erf is the Gauss error function.

Composite truncated normal distribution

$$\Omega(P_S) \sim \lambda\, \mathcal{N}_T(P_S, m_1, s_1) + (1-\lambda)\, \mathcal{N}_T(P_S, m_2, s_2) \tag{B.3}$$

with 5 parameters (m_1, m_2, s_1, s_2 and a partitioning coefficient λ).

Uniform distribution (over $[0, u]$)

$$\begin{aligned} \Omega(P_S) &= P_S/u \qquad P_S < u \\ \Omega(P_S) &= 1 \qquad\quad\ P_S \geq u \end{aligned} \tag{B.4}$$

with 1 parameter (u).

B.1.2 Ranked SAS functions

The SAS functions used in the virtual catchment case study are "ranked" SAS functions (Harman 2015). They are defined over age-ranked storage $S_T \in [0, +\infty[$. $S_T = S(t)\, p_S(T,t)$ is the amount of water in storage younger than T. $S_T \to 0$ refers to the youngest water, while $S_T \to S(t)$ refers to the oldest water. We used the following functional forms:

Single gamma distribution

$$\Omega(S_T) = \frac{1}{\Gamma(\mu/\theta)}\, \gamma(\frac{\mu}{\theta}, \frac{S_T}{\theta}) \tag{B.5}$$

with 2 parameters (mean μ and scale θ). γ is the lower incomplete gamma function. This unimodal function was used to approximate the bimodal function below:

Double gamma distribution

$$\Omega^{ref}(S_T) = \lambda\, \Omega_1(S_T, \frac{\mu_1}{\theta_1}, \theta_1) + (1-\lambda)\, \Omega_2(S_T, \frac{\mu_2}{\theta_2}, \theta_2) \tag{B.6}$$

with 5 parameters (μ_1, μ_2, θ_1, θ_2 and a partitioning coefficient λ). Ω_1 and Ω_2 are calculated using equation B.5.

B.2 Additional insights on the lysimeter model

B.2.1 Parameter calibration

Model parameters were calibrated using the Markov Chain Monte Carlo algorithm DREAM$_{ZS}$ (ter Braak and Vrugt 2008; Vrugt et al. 2009). The three models only differed in the choice of the SAS function for the bottom drainage, which was modeled either as i) a power function (1 parameter), ii) a truncated normal distribution (2 parameters) or as iii) a double truncated normal distribution (5 parameters). Analytical formulas are provided in Section B.1. In addition, all models have 3 more parameters: the total water storage S_0 at the beginning of the simulation, the parameter u for the SAS function of the evapotranspiration flux, and the evapoconcentration factor α. Calibrations were run under the assumption that model residuals were normally-distributed, which results in a log-likelihood function of the type:

$$\log(\ell) = \frac{N}{2}\log(2\pi) - N\log(\sigma_e) - \sum_{i=1}^{N} \frac{\epsilon_i^2}{2\sigma_\epsilon^2} \tag{B.7}$$

where N is the number of measurements, ϵ_i is the model residual at time step i and σ_e is the error standard deviation. To account for the correlation in model residuals (and the resulting loss of degrees of freedom), the residuals' standard deviation σ_e was multiplied by a factor $r = \sqrt{(1+\rho)/(1-\rho)}$, where ρ is the step-1 autocorrelation of the residuals' timeseries. The prior parameters' distribution was always set as uniform over a range determined from physical constraints (Table B.1). Simulations were run at time steps of 4 hours, which was a satisfactory compromise between the computational accuracy and running time. A number of 10^5 simulations was enough to guarantee the convergence of the Markov Chains and extract the parameters' posterior distributions (Figure B.1).

TABLE B.1: Parameter ranges and optimal values for the three SAS models used in the simulations. Analytical formulas for the SAS functions are reported in section B.1.1.

parameter	units	SAS model	lower bound	upper bound	optimal
k	-	single power function	0.2	10	2.9
u	-	single power function	0.01	1	0.94
S_0	mm	single power function	500	600	599
α	-	single power function	0	1	0.31
m	-	single normal	0	1	0.61
s	-	single normal	0.01	2	0.14
u	-	single normal	0.01	1	0.35
S_0	mm	single normal	500	600	504
α	-	single normal	0	1	0.31
m1	-	double normal	0	1	0.48
s1	-	double normal	0.01	2	0.11
m2	-	double normal	0	1	0.97
s2	-	double normal	0.01	2	0.15
λ	-	double normal	0	1	0.38
u	-	double normal	0.01	1	0.18
S_0	mm	double normal	500	600	503
α	-	double normal	0	1	0.23

TABLE B.2: Model performance statistics

	single power function	single normal	double normal
residual mean [mg/l]	-0.50	-0.20	0.10
residual standard dev. [mg/l]	2.14	1.97	1.37
residual autocorrelation [-]	0.87	0.82	0.58
NS efficiency [-]	0.72	0.75	0.89

B.2.2 Lysimeter model performances

For each model, the simulation with the lowest sum of squared residuals was used to compute the models' residuals and to obtain their distributions (Figure B.2a) and main statistics (Table B.2), as well as the Nash-Sutcliffe efficiency that was used as an informal indicator of model performance.

Model performance always followed model complexity. Notably, residuals were much less correlated in the double normal SAS function. The NS coefficients were generally good (NS>0.70) although these high values strongly depend on the shape of the tracer breakthrough curve which has a high variance (see Schaefli and Gupta 2007). As a more formal indicator of model performance we computed the Akaike (AIC) scores (Akaike 1974). The AIC distribution is reported in Figure B.2b and shows that, even when penalizing for the higher number of parameters, the double normal SAS function performs best. Some overlap occurs between the single power function and the single normal models, but overall the latter performs better than the former, confirming that model complexity is always justified in this case.

B.3 Additional insights on the virtual catchment model

B.3.1 Hydrological and tracer data

The daily data used for the virtual catchment was generated for a period of 100 years. This data set consists of rainfall rate J (mm/d), evapotranspiration ET (mm/d), discharge Q (mm/d), and rainfall deuterium concentration C_J (‰).

The variables J, C_J, and ET have a sinusoidal component (yearly seasonality) and a random component (daily variations):

$$J(t) = [2A_J + A_J \sin(2\,\pi\, f\,(t - \phi_J))] \times \gamma_J(t) \tag{B.8}$$

$$C_J(t) = C_0 + A_C \sin(2\,\pi\, f\,(t - \phi)) + \rho(t) \tag{B.9}$$

$$ET(t) = [2A_{ET} + A_{ET} \sin(2\,\pi\, f\,(t - \phi))] \times e_{ET}(t) \tag{B.10}$$

where f is a yearly frequency, ϕ (ϕ_J) is a phase shift for the sinusoidal component yielding minima in December (August) and maxima in June (February). A_J, A_{ET}, and A_C are sinusoidal amplitudes set to 1.1 mm/d, 0.55 mm/d, and 25 ‰, respectively. $\gamma_J(t)$ is a random number from a gamma distribution with shape parameter $k = 0.25$ and scale parameter $\theta = 4$ mm/d. $\rho(t)$ is a random number between -20 ‰ and 20 ‰ from a uniform distribution. $e_{ET}(t)$ is a random number from an exponential distribution with mean parameter $\mu = 2$ mm/d. Values below 1 mm/d for J were randomly set to 0 with an 80% probability to obtain more days without precipitation.

Discharge was generated as:

$$Q(t) = a \left(\frac{S(t)}{S_0} \right)^b \tag{B.11}$$

where $a = 1.64$ mm/d and $b = 30$ are storage-discharge parameters, and storage $S(t)$ was deduced from the water balance:

$$S(t) = S(t_0) + \int_{u=t_0}^{t} (J(u) - Q(u) - ET(u))\, du \tag{B.12}$$

and $S(t_0) = S_0 = 1000$ mm.

Observations of stream deuterium content δ^2H for the virtual catchment were generated by running the model forward using the 100-year data and by assigning the reference SAS function Ω^{ref} to discharge (section B.1.2). We used the parameters $\lambda = 0.1$, $\mu_1 = 50$ mm, $\mu_2 = 500$ mm, $\theta_1 = 10$ mm, and $\theta_2 = 50$ mm in the double gamma model described in section B.1.2. This ensures that the two modes of the SAS function are clearly separated from each other and that their pdfs have similar amplitudes.

B.3.2 Parameter calibration and model performance

The theoretical framework described in the main text and in section B.1.2 was applied to simulate tracer concentrations in the discharge of the virtual catchment with the single gamma SAS function and to calibrate its parameters. For the virtual experiment, the age Master Equation was numerically solved following the same procedure as *tran*-SAS (used for the lysimeter case study), differing only in the numerical scheme. Briefly, instead of using the method of characteristics that considers only one variable S_T, the Master Equation was discretized with respect to time t and age T separately. This has no implications for the results of this study, since both numerical schemes are stable and accurate for the virtual catchment, and the underlying equation that is solved is exactly the same.

The parameters of the single gamma SAS function used for the discharge of the virtual catchment (section B.1.2) were calibrated against the stream δ^2H observations with a Monte Carlo algorithm. 15,000 pairs of the mean parameter μ and the shape parameter θ were sampled from uniform prior distributions with ranges $[0, 700]$ mm and $[0, 500]$ mm respectively. Model performance was evaluated against the 31,970 δ^2H daily values (100-year period minus the first ~12.5 years as spin-up) with the Nash-Sutcliffe Efficiency, which assumes normally distributed residuals (similar to equation B.7) with a constant variance equal to that of the measurements and no residual autocorrelation. We selected the best performing (NSE = 0.89) parameter set to analyze the results from the single gamma model and compare them to the "true" Ω^{ref} model. The parameters were well identifiable (figure B.3). Parameter uncertainty was calculated using the 150 best simulations with respect to their NSE (range: 0.75–0.89) and it was found to be small (figure B.3). The best model corresponds to $\mu = 627$ mm and $\theta = 90$ mm.

B.4 Travel times, flow paths, and velocities in a Lagrangian framework

Let's consider a water parcel travelling from its entry point into the system (e.g. as precipitation touching the soil surface) along a flow path until the exit point (e.g. the stream channel at the catchment outlet). The flow path has a total length L, and the position of the water parcel along the flow path can be summarized with the parametric coordinate ℓ, defined as the length travelled by the parcel along the flow path since its entry point. ℓ therefore goes from 0 to L. $d\ell$ symbolizes the infinitesimal increment in ℓ made by the water parcel in the infinitesimal amount of time dT. Note that $d\ell$ is the same at every time, while dT varies according to the Lagrangian velocity $v(\ell)$ of the water parcel. The travel time to the exit point is obtained from:

$$T = \int_{\text{flowpath}} dT \tag{B.13}$$

dT is calculated as:

$$dT = \frac{d\ell}{v(\ell)} \tag{B.14}$$

such that:

$$T = \int_{\ell=0}^{L} \frac{d\ell}{v(\ell)} \tag{B.15}$$

From this expression it is clear that only particles that do not become immobile along the flow path can be considered in this framework (to avoid division by 0). We therefore only refer to the travel times of mobile water parcels.

We can now use the harmonic mean to simplify this expression. Let's recall the expression of the harmonic mean $\mathcal{H}$ for a continuous, invertible function f over the domain $[a, b]$:

$$\mathcal{H} = \frac{b-a}{\int_{x=a}^{b} \frac{dx}{f(x)}} \tag{B.16}$$

which implies:

$$\int_{x=a}^{b} \frac{dx}{f(x)} = \frac{b-a}{\mathcal{H}} \tag{B.17}$$

Let's call v the harmonic mean of all Lagrangian velocities $v(\ell)$ along the flow path of length L. Equation B.15 is analogous to equation B.17, we can thus rewrite it as:

$$T = \int_{\ell=0}^{L} \frac{d\ell}{v(\ell)} = \frac{L}{v} \tag{B.18}$$

This proves that the travel time can be written in a Lagrangian framework as the ratio of the total flow path length and the average velocity along the flow path (using the harmonic mean).

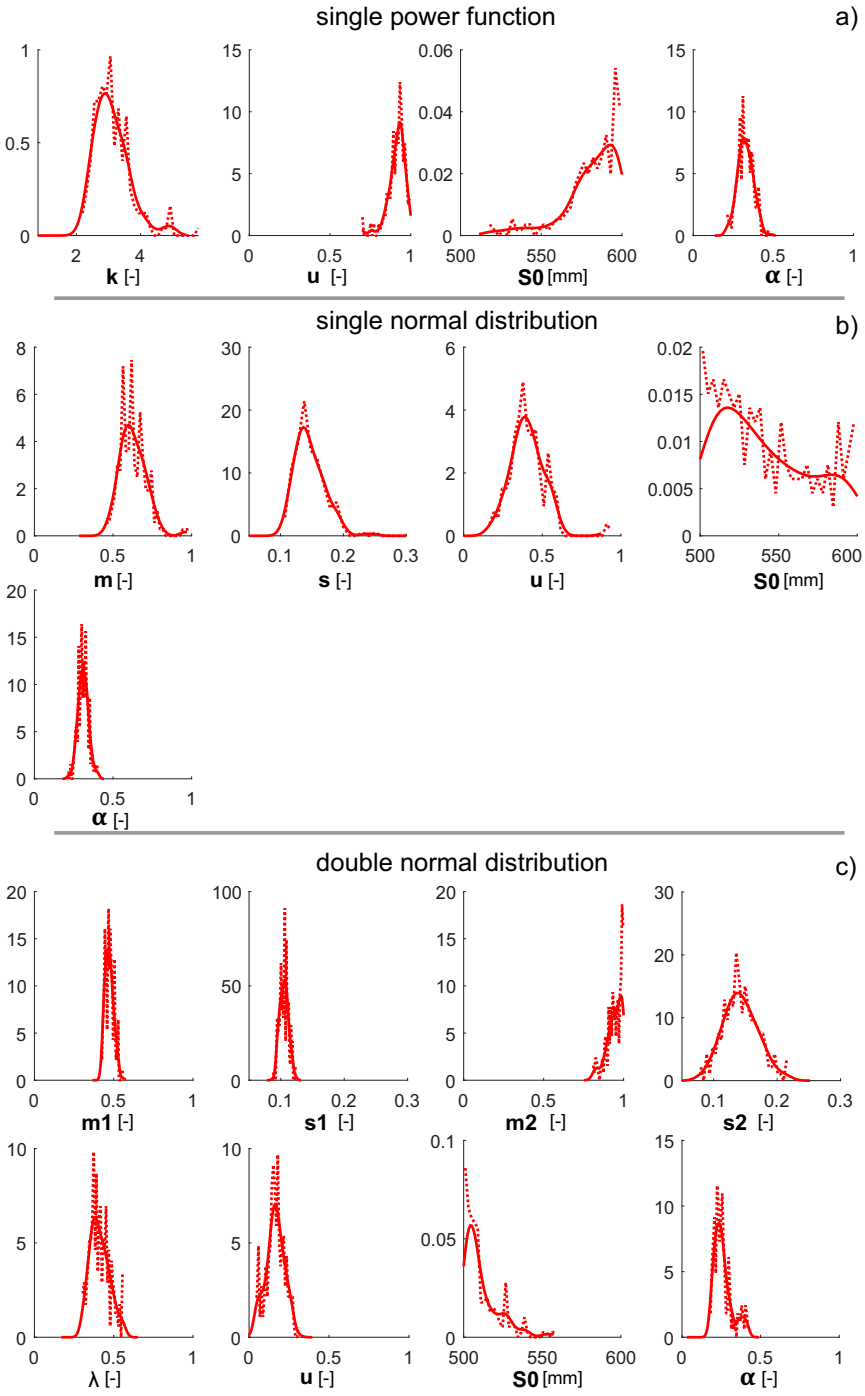

FIGURE B.1: Posterior parameter distributions for the three lysimeter models. The pdf's were extracted from the converged Markov Chains (dashed lines) and then smoothed (solid lines).

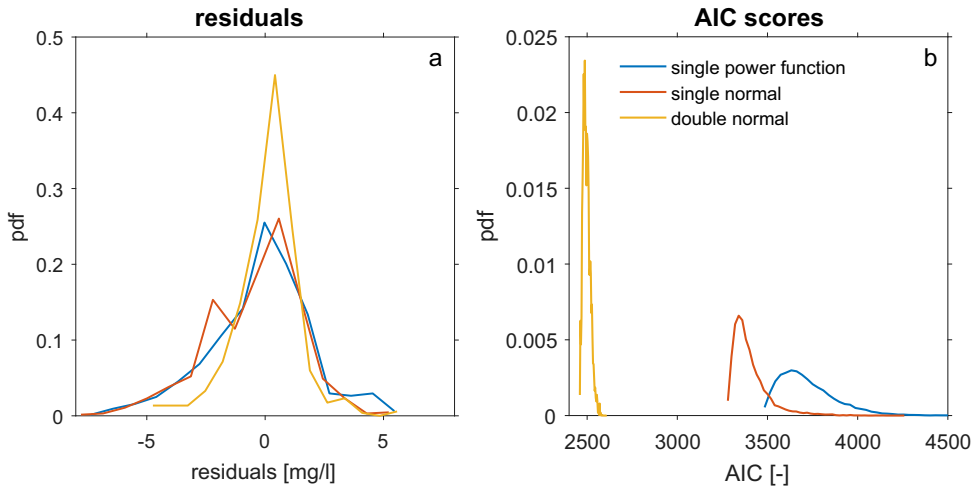

FIGURE B.2: Performance indicators for the three models: residuals distributions (a) and Akaike (AIC) scores (b). The more complex models perform better even when they are penalized for the extra number of parameters.

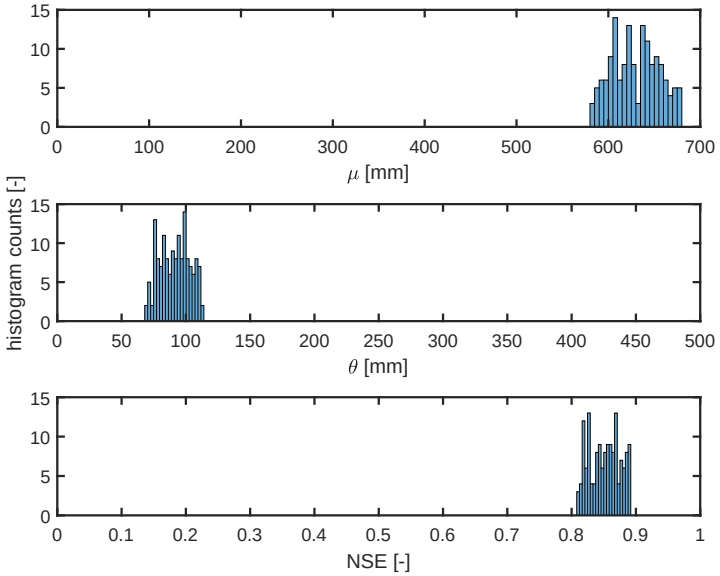

FIGURE B.3: Posterior parameter histograms for the single gamma SAS function in the virtual catchment. The histograms were extracted from the parameter sets with the 150 highest NSE values.

Appendix C

Supplementary information for Chapter 4

C.1 Model equations

C.1.1 Parameterization of the SAS functions

In this section we provide further details on the equations used in the model. The composite streamflow SAS function Ω_Q used in this study is:

$$\Omega_Q(S_T, t) = \lambda_1(t)\,\Omega_1(S_T) + \lambda_2(t)\,\Omega_3(S_T) + \lambda_3(t)\,\Omega_1(S_T) \tag{C.1}$$

$\Omega_1(S_T)$ is a cumulative uniform distribution for S_T in $[0, S_u]$, where S_u (mm) is a calibrated parameter representing the amount of stored young water potentially contributing to flashy streamflow responses. Thus:

$$\Omega_1(S_T) = \begin{cases} \frac{S_T}{S(t)}, & S_T \in [0, S_u] \\ 1, & S_T > S_u \end{cases} \tag{C.2}$$

$\Omega_2(S_T)$ and $\Omega_3(S_T)$ are direct functions of S_T and are gamma-distributed:

$$\Omega_2(S_T) = \frac{1}{\Gamma(\frac{\mu_2}{\theta_2})}\,\gamma(\frac{\mu_2}{\theta_2}, \frac{S_T}{\theta_2}) \tag{C.3}$$

$$\Omega_3(S_T) = \frac{1}{\Gamma(\frac{\mu_3}{\theta_3})}\,\gamma(\frac{\mu_3}{\theta_3}, \frac{S_T}{\theta_3}) \tag{C.4}$$

where Γ is the gamma function, γ is the lower incomplete gamma function, μ_2 and μ_3 (mm) are mean parameters (calibrated), and θ_2 and θ_3 (mm) are scale parameters (calibrated).

$\lambda_1(t)$, $\lambda_2(t)$, and $\lambda_3(t)$ sum to 1. These are simply time-varying weights giving each component (i.e. c.d.f. Ω) a dynamic contribution to streamflow generation. In particular, $\lambda_1(t)$ is made highly time-variant to represent the flashy hydrographs that have an on-off type of response to precipitation. $\lambda_2(t)$ is considered constant and calibrated to keep the parameterization parsimonious. $\lambda_3(t) = 1 - \lambda_2 - \lambda_1(t)$ is deduced by difference for parsimony as well. Since $\Omega_1(S_T)$ represents young water contributions and previous studies in the Weierbach showed that event water contributions depend on the catchment wetness and on precipitation intensity (Martínez-Carreras et al. 2015; Wrede et al. 2015), $\lambda_1(t)$ was parameterized using storage $S(t)$ and a proxy storage variations $\overline{\Delta S(t)}$ (see Rodriguez and Klaus (2019) for more details):

$$\lambda_1(t) = \lambda_1^* \left[f(t) + (1 - f(t))\, g(t) \right] \tag{C.5}$$

where $\lambda_1^* \in [0,1]$ (no units) is a calibrated parameter representing the maximum value of $\lambda_1(t)$, and $f(t) \in [0,1]$ and $g(t) \in [0,1]$ are given by:

$$f(t) = f_0 \left(1 - \tanh \left[\left(\frac{S(t)}{S_{min} + S_{th}} \right)^m \right] \right) \tag{C.6}$$

$$g(t) = 1 - \exp \left(- \frac{\overline{\Delta S(t)}}{\Delta S_{th}} \right) \tag{C.7}$$

$f_0 \in [0,1]$ (no units) is a calibrated parameter guaranteeing a minimum for $\lambda_1(t)$ during dry periods, $S_{min} = \min(S(t))$, and S_{th} (mm, calibrated parameter) is a storage threshold relative the to minimum storage S_{min} separating wet ($S(t) > S_{min} + S_{th}$) from dry periods ($S(t) < S_{min} + S_{th}$). $m = 1000$ is a fixed parameter used to smooth the function f with respect to $S(t)$. $\overline{\Delta S(t)}$ is a proxy of storage variations calculated as a moving average of storage variations over a time window $\Delta t^* = 2\,\Delta t$:

$$\overline{\Delta S(t)} = \max \left(\frac{1}{3} \sum_{j=0}^{2} \Delta S(t - j\Delta t), 0 \right) \tag{C.8}$$

with $\Delta S(t) = \Delta t\, (J(t) - Q(t) - ET(t))$. $\overline{\Delta S(t)}$ essentially increases during precipitation events and decreases when $Q(t)$ or $ET(t)$ are high. ΔS_{th} is a threshold in $\overline{\Delta S(t)}$ above which $g(t)$ tends to 1, allowing $\lambda_1(t)$ to increase and decrease sharply during flashy streamflow events.

C.1.2 Actual evapotranspiration

Actual evapotranspiration $ET(t)$ is calculated from potential evapotranspiration $PET(t)$ using the formula:

$$ET(t) = PET(t)\, \tanh \left[\left(\frac{S(t)}{S_{root}} \right)^n \right] \tag{C.9}$$

where $S_{root} = S_{ref} - 150$ is a fixed parameter (mm) representing the storage threshold $S(t) = S_{root}$ below which $ET(t)$ starts decreasing from $PET(t)$ towards 0. This decrease is smoothed by the fixed coefficient $n = 20$. S_{root} accounts for the water available for evaporation and plant transpiration until the capillary forces offer too much resistance. This formula thus represents the decrease in water losses to the atmosphere under water limited conditions.

References

Abbott, B. W., V. Baranov, C. Mendoza-Lera, M. Nikolakopoulou, A. Harjung, T. Kolbe, M. N. Balasubramanian, T. N. Vaessen, F. Ciocca, A. Campeau, M. B. Wallin, P. Romeijn, M. Antonelli, J. Gonçalves, T. Datry, A. M. Laverman, J.-R. de Dreuzy, D. M. Hannah, S. Krause, C. Oldham, and G. Pinay (2016). "Using multi-tracer inference to move beyond single-catchment ecohydrology". In: *Earth-Science Reviews* 160, pp. 19–42. DOI: 10.1016/j.earscirev.2016.06.014.

Aggarwal, P. K. (2002). "Isotope hydrology at the International Atomic Energy Agency". In: *Hydrological Processes* 16.11, pp. 2257–2259. DOI: 10.1002/hyp.5043.

Akaike, H. (1974). "A new look at the statistical model identification". In: *IEEE Transactions on Automatic Control* 19.6, pp. 716–723. DOI: 10.1109/TAC.1974.1100705.

Ali, M., A. Fiori, and D. Russo (2014). "A comparison of travel-time based catchment transport models, with application to numerical experiments". In: *Journal of Hydrology* 511.0, pp. 605–618. DOI: 10.1016/j.jhydrol.2014.02.010.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith (1998). *Crop evapotranspiration: guidelines for computing crop water requirements*. FAO Irrigation and drainage paper 56. Food and Agriculture Organization of the United Nations.

Ameli, A. A., K. Beven, M. Erlandsson, I. F. Creed, J. J. McDonnell, and K. Bishop (2017). "Primary weathering rates, water transit times, and concentration-discharge relations: A theoretical analysis for the critical zone". In: *Water Resources Research* 53.1, pp. 942–960. DOI: 10.1002/2016WR019448.

Angermann, L., C. Jackisch, N. Allroggen, M. Sprenger, E. Zehe, J. Tronicke, M. Weiler, and T. Blume (2017). "Form and function in hillslope hydrology: characterization of subsurface flow based on response observations". In: *Hydrology and Earth System Sciences* 21.7, pp. 3727–3748. DOI: 10.5194/hess-21-3727-2017.

Antonelli, M., B. Glaser, T. R., J. Klaus, and L. Pfister (in review[a]). "Saturated areas through the lens: 1. Spatio-temporal variability of surface saturation documented through Thermal Infrared imagery". In: *Hydrological Processes*.

— (in review[b]). "Saturated areas through the lens: 2. Spatio-temporal variability of streamflow generation and its relationship with surface saturation". In: *Hydrological Processes*.

Antonelli, M., C. E. Wetzel, L. Ector, A. J. Teuling, and L. Pfister (2017). "On the potential for terrestrial diatom communities and diatom indices to identify anthropic disturbance in soils". In: *Ecological Indicators* 75, pp. 73–81. DOI: 10.1016/j.ecolind.2016.12.003.

Aubert, A., C. Gascuel-Odoux, G. Gruau, N. Akkal, M. Faucheux, Y. Fauvel, C. Grimaldi, Y. Hamon, A. Jaffrezic, M. Le Coz-Bouhnik, J. Molénat, P. Petitjean, L. Ruiz, and P. Mérot (2013). "Solute transport dynamics in small, shallow groundwater-dominated agricultural catchments: insights from a high-frequency, multisolute 10 yr-long monitoring study". In: *Hydrology and Earth System* 17.4, pp. 1379–1391. DOI: 10.5194/hess-17-1379-2013.

Aubert, A. H., J. W. Kirchner, C. Gascuel-Odoux, M. Faucheux, G. Gruau, and P. Mérot (2014). "Fractal Water Quality Fluctuations Spanning the Periodic Table in

an Intensively Farmed Watershed". In: *Environmental Science and Technology* 48.2. PMID: 24328425, pp. 930–937. DOI: 10.1021/es403723r.

Bajjali, W. (2012). "Spatial variability of environmental isotope and chemical content of precipitation in Jordan and evidence of slight change in climate". In: *Applied Water Science* 2.4, pp. 271–283. DOI: 10.1007/s13201-012-0046-1.

Barnard, H. R., C. B. Graham, W. J. Van Verseveld, J. R. Brooks, B. J. Bond, and J. J. McDonnell (2010). "Mechanistic assessment of hillslope transpiration controls of diel subsurface flow: a steady-state irrigation approach". In: *Ecohydrology* 3.2, pp. 133–142. DOI: 10.1002/eco.114.

Begemann, F. and W. Libby (1957). "Continental water balance, ground water inventory and storage times, surface ocean mixing rates and world-wide water circulation patterns from cosmic-ray and bomb tritium". In: *Geochimica et Cosmochimica Acta* 12.4, pp. 277–296. DOI: 10.1016/0016-7037(57)90040-6.

Benettin, P. and E. Bertuzzo (2018). "*tran*-SAS v1.0: a numerical model to compute catchment-scale hydrologic transport using StorAge Selection functions". In: *Geoscientific Model Development* 11.4, pp. 1627–1639. DOI: 10.5194/gmd-11-1627-2018.

Benettin, P., S. W. Bailey, J. L. Campbell, M. B. Green, A. Rinaldo, G. E. Likens, K. J. McGuire, and G. Botter (2015a). "Linking water age and solute dynamics in streamflow at the Hubbard Brook Experimental Forest, NH, USA". In: *Water Resources Research* 51.11, pp. 9256–9272. DOI: 10.1002/2015WR017552.

Benettin, P., S. W. Bailey, A. Rinaldo, G. E. Likens, K. J. McGuire, and G. Botter (2017a). "Young runoff fractions control streamwater age and solute concentration dynamics". In: *Hydrological Processes* 31.16, pp. 2982–2986. DOI: 10.1002/hyp.11243.

Benettin, P., J. W. Kirchner, A. Rinaldo, and G. Botter (2015b). "Modeling chloride transport using travel time distributions at Plynlimon, Wales". In: *Water Resources Research* 51.5, pp. 3259–3276. DOI: 10.1002/2014WR016600.

Benettin, P., P. Queloz, M. Bensimon, J. J. McDonnell, and A. Rinaldo (2018). "Velocities, residence times, tracer breakthroughs in a vegetated lysimeter: a multitracer experiment". In: *Water Resources Research*. in review.

— (2019). "Velocities, residence times, tracer breakthroughs in a vegetated lysimeter: a multitracer experiment". In: *Water Resources Research* 55.1, pp. 21–33. DOI: 10.1029/2018wr023894.

Benettin, P., A. Rinaldo, and G. Botter (2013a). "Kinematics of age mixing in advection-dispersion models". In: *Water Resources Research* 49.12, pp. 8539–8551. DOI: 10.1002/2013WR014708.

— (2015c). "Tracking residence times in hydrological systems: forward and backward formulations". In: *Hydrological Processes*. DOI: 10.1002/hyp.15034.

Benettin, P., C. Soulsby, C. Birkel, D. Tetzlaff, G. Botter, and A. Rinaldo (2017b). "Using SAS functions and high-resolution isotope data to unravel travel time distributions in headwater catchments". In: *Water Resources Research* 53.3, pp. 1864–1878. DOI: 10.1002/2016WR020117.

Benettin, P., Y. van der Velde, S. E. A. T. M. van der Zee, A. Rinaldo, and G. Botter (2013b). "Chloride circulation in a lowland catchment and the formulation of transport by travel time distributions". In: *Water Resources Research* 49.8, pp. 4619–4632. DOI: 10.1002/wrcr.20309.

Berman, E. S. F., M. Gupta, C. Gabrielli, T. Garland, and J. J. McDonnell (2009). "High-frequency field-deployable isotope analyzer for hydrological applications". In: *Water Resources Research* 45.10. DOI: 10.1029/2009WR008265.

Berry, Z. C., J. Evaristo, G. Moore, M. Poca, K. Steppe, L. Verrot, H. Asbjornsen, L. S. Borma, M. Bretfeld, P. Hervé-Fernández, M. Seyfried, L. Schwendenmann, K. Sinacore, L. De Wispelaere, and J. McDonnell (2018). "The two water worlds hypothesis: Addressing multiple working hypotheses and proposing a way forward". In: *Ecohydrology* 11.3. e1843 ECO-16-0180.R2, e1843. DOI: 10.1002/eco.1843.

Bertuzzo, E., M. Thomet, G. Botter, and A. Rinaldo (2013). "Catchment-scale herbicides transport: Theory and application". In: *Advances in Water Resources* 52.0, pp. 232–242. DOI: 10.1016/j.advwatres.2012.11.007.

Bethke, C. M. and T. M. Johnson (2008). "Groundwater Age and Groundwater Age Dating". In: *Annual Review of Earth and Planetary Sciences* 36.1, pp. 121–152. DOI: 10.1146/annurev.earth.36.031207.124210.

Beven, K. J. (2006). *Streamflow generation processes. Benchmark papers in hydrology, 1*. English. Working Paper. International Association of Hydrological Sciences.

Birkel, C., S. M. Dunn, D. Tetzlaff, and C. Soulsby (2010a). "Assessing the value of high-resolution isotope tracer data in the stepwise development of a lumped conceptual rainfall–runoff model". In: *Hydrological Processes* 24.16, pp. 2335–2348. DOI: 10.1002/hyp.7763.

Birkel, C., D. Tetzlaff, S. M. Dunn, and C. Soulsby (2010b). "Towards a simple dynamic process conceptualization in rainfall–runoff models using multi-criteria calibration and tracers in temperate, upland catchments". In: *Hydrological Processes* 24.3, pp. 260–275. DOI: 10.1002/hyp.7478.

Birkel, C. and C. Soulsby (2015). "Advancing tracer-aided rainfall–runoff modelling: a review of progress, problems and unrealised potential". In: *Hydrological Processes* 29.25, pp. 5227–5240. DOI: 10.1002/hyp.10594.

— (2016). "Linking tracers, water age and conceptual models to identify dominant runoff processes in a sparsely monitored humid tropical catchment". In: *Hydrological Processes* 30.24, pp. 4477–4493. DOI: 10.1002/hyp.10941.

Birkel, C., C. Soulsby, and D. Tetzlaff (2011a). "Modelling catchment-scale water storage dynamics: reconciling dynamic storage with tracer-inferred passive storage". In: *Hydrological Processes* 25.25, pp. 3924–3936. DOI: 10.1002/hyp.8201.

— (2015). "Conceptual modelling to assess how the interplay of hydrological connectivity, catchment storage and tracer dynamics controls nonstationary water age estimates". In: *Hydrological Processes*. DOI: 10.1002/hyp.10414.

Birkel, C., D. Tetzlaff, S. M. Dunn, and C. Soulsby (2011b). "Using lumped conceptual rainfall–runoff models to simulate daily isotope variability with fractionation in a nested mesoscale catchment". In: *Advances in Water Resources* 34.3, pp. 383–394. DOI: 10.1016/j.advwatres.2010.12.006.

Bond, B. J., J. A. Jones, G. Moore, N. Phillips, D. Post, and J. J. McDonnell (2002). "The zone of vegetation influence on baseflow revealed by diel patterns of streamflow and vegetation water use in a headwater basin". In: *Hydrological Processes* 16.8, pp. 1671–1677. DOI: 10.1002/hyp.5022.

Botter, G. (2012). "Catchment mixing processes and travel time distributions". In: *Water Resources Research* 48. DOI: 10.1029/2011WR011160.

Botter, G., E. Bertuzzo, and A. Rinaldo (2010). "Transport in the hydrologic response: Travel time distributions, soil moisture dynamics, and the old water paradox". In: *Water Resources Research* 46. DOI: 10.1029/2009WR008371.

— (2011). "Catchment residence and travel time distributions: The master equation". In: *Geophysical Research Letters* 38. DOI: 10.1029/2011GL047666.

Bras, R. L. (2015). "Complexity and organization in hydrology: A personal view". In: *Water Resources Research* 51.8, pp. 6532–6548. DOI: 10.1002/2015WR016958.

Broad, W. J. (2005). "With a Push From the U.N., Water Reveals Its Secrets". In: *The New York Times*.

Brooks, J. R., H. R. Barnard, R. Coulombe, and J. J. McDonnell (2010). "Ecohydrologic separation of water between trees and streams in a Mediterranean climate". In: *Nature Geoscience* 3.2, pp. 100–104. DOI: 10.1038/ngeo722.

Broxton, P. D., P. A. Troch, and S. W. Lyon (2009). "On the role of aspect to quantify water transit times in small mountainous catchments". In: *Water Resources Research* 45.8. DOI: 10.1029/2008WR007438.

Burt, T. P. and J. J. McDonnell (2015). "Whither field hydrology? The need for discovery science and outrageous hydrological hypotheses". In: *Water Resources Research* 51.8, pp. 5919–5928. DOI: 10.1002/2014WR016839.

Buttafuoco, G., T. Caloiero, and R. Coscarelli (2010). "Spatial uncertainty assessment in modelling reference evapotranspiration at regional scale". In: *Hydrology and Earth System Sciences* 14.11, pp. 2319–2327. DOI: 10.5194/hess-14-2319-2010.

Buttle, J. (1994). "Isotope hydrograph separations and rapid delivery of pre-event water from drainage basins". In: *Progress in Physical Geography: Earth and Environment* 18.1, pp. 16–41. DOI: 10.1177/030913339401800102.

Cain, M. R., A. S. Ward, and M. Hrachowitz (2019). "Ecohydrologic separation alters interpreted hydrologic stores and fluxes in a headwater mountain catchment". In: *Hydrological Processes* 0.0. DOI: 10.1002/hyp.13518.

Capell, R., D. Tetzlaff, A. J. Hartley, and C. Soulsby (2012). "Linking metrics of hydrological function and transit times to landscape controls in a heterogeneous mesoscale catchment". In: *Hydrological Processes* 26.3, pp. 405–420. DOI: 10.1002/hyp.8139.

Carrer, G. E., J. Klaus, and L. Pfister (n.d.). "Review and assessment of catchment storage estimation approaches". In: *in prep.* ().

Carrer, G. E., J. Klaus, and L. Pfister (2019). "Assessing the Catchment Storage Function Through a Dual-Storage Concept". In: *Water Resources Research* 55.1, pp. 476–494. DOI: 10.1029/2018WR022856.

Cartwright, I. and U. Morgenstern (2016). "Contrasting transit times of water from peatlands and eucalypt forests in the Australian Alps determined by tritium: implications for vulnerability and the source of water in upland catchments". In: *Hydrology and Earth System Sciences* 20.12, pp. 4757–4773. DOI: 10.5194/hess-20-4757-2016.

Cirpka, O. A., M. N. Fienen, M. Hofer, E. Hoehn, A. Tessarini, R. Kipfer, and P. K. Kitanidis (2007). "Analyzing Bank Filtration by Deconvoluting Time Series of Electric Conductivity". In: *Groundwater* 45.3, pp. 318–328. DOI: 10.1111/j.1745-6584.2006.00293.x.

Clark, M. P., M. F. P. Bierkens, L. Samaniego, R. A. Woods, R. Uijlenhoet, K. E. Bennett, V. R. N. Pauwels, X. Cai, A. W. Wood, and C. D. Peters-Lidard (2017). "The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism". In: *Hydrology and Earth System Sciences* 21.7, pp. 3427–3440. DOI: 10.5194/hess-21-3427-2017.

Clymans, W., G. Govers, E. Frot, B. Ronchi, B. Van Wesemael, and E. Struyf (2013). "Temporal dynamics of bio-available Si fluxes in a temperate forested catchment (Meerdaal forest, Belgium)". In: *Biogeochemistry* 116.1-3, pp. 275–291. DOI: 10.1007/s10533-013-9858-9.

Criss, R. E. and W. E. Winston (2008). "Do Nash values have value? Discussion and alternate proposals". In: *Hydrological Processes* 22.14, pp. 2723–2725. DOI: 10.1002/hyp.7072.

Curtiss, J. H. (Dec. 1941). "On the Distribution of the Quotient of Two Chance Variables". In: *The Annals of Mathematical Statistics* 12.4, pp. 409–421. DOI: 10.1214/aoms/1177731679.

Danesh-Yazdi, M., J. Klaus, L. E. Condon, and R. M. Maxwell (2018). "Bridging the gap between numerical solutions of travel time distributions and analytical storage selection functions". In: *Hydrological Processes* 32.8, pp. 1063–1076. DOI: 10.1002/hyp.11481.

Darracq, A., G. Lindgren, and G. Destouni (2008). "Long-term development of phosphorus and nitrogen loads through the subsurface and surface water systems of drainage basins". In: *Global Biogeochemical Cycles* 22.3. DOI: 10.1029/2007GB003022.

Davies, J., K. Beven, A. Rodhe, L. Nyberg, and K. Bishop (2013). "Integrated modeling of flow and residence times at the catchment scale with multiple interacting pathways". In: *Water Resources Research* 49.August, pp. 4738–4750. DOI: 10.1002/wrcr.20377.

Davis, W. M. (1926). "The value of outrageous geological hypotheses". In: *Science* 63.1636, pp. 463–468. DOI: 10.1126/science.63.1636.463.

Delavau, C. J., T. Stadnyk, and T. Holmes (2017). "Examining the impacts of precipitation isotope input ($\delta^{18}O_{ppt}$) on distributed, tracer-aided hydrological modelling". In: *Hydrology and Earth System Sciences* 21.5, pp. 2595–2614. DOI: 10.5194/hess-21-2595-2017.

Delsman, J. R., G. H. P. O. Essink, K. J. Beven, and P. J. Stuyfzand (2013). "Uncertainty estimation of end-member mixing using generalized likelihood uncertainty estimation (GLUE), applied in a lowland catchment". In: *Water Resources Research* 49.8, pp. 4792–4806. DOI: 10.1002/wrcr.20341.

Desbarats, A. J. (1990). "Macrodispersion in sand-shale sequences". In: *Water Resources Research* 26.1, pp. 153–163. DOI: 10.1029/WR026i001p00153.

Devell, L. (1962). "Measurements of the Self-diffusion of Water in Pure Water, H2O-D2O Mixtures and Solutions of Electrolytes". In: *Acta Chemica Scandinavica* 16, pp. 2177–2188. DOI: 10.3891/acta.chem.scand.16-2177.

Dinçer, T., B. R. Payne, T. Florkowski, J. Martinec, and E. Tongiorgi (1970). "Snowmelt runoff from measurements of tritium and oxygen-18". In: *Water Resources Research* 6.1, pp. 110–124. DOI: 10.1029/WR006i001p00110.

Dooge, J. C. I. (1986). "Looking for hydrologic laws". In: *Water Resources Research* 22.9S, 46S–58S. DOI: 10.1029/WR022i09Sp0046S.

Dralle, D. N., W. J. Hahm, D. M. Rempe, N. J. Karst, S. E. Thompson, and W. E. Dietrich (2018). "Quantification of the seasonal hillslope water storage that does not drive streamflow". In: *Hydrological Processes* 32.13, pp. 1978–1992. DOI: 10.1002/hyp.11627.

Dubbert, M., M. C. Caldeira, D. Dubbert, and C. Werner (2019). "A pool-weighted perspective on the two-water-worlds hypothesis". In: *New Phytologist* 0.0. DOI: 10.1111/nph.15670.

Duffy, C. J. (2010). "Dynamical modelling of concentration–age–discharge in watersheds". In: *Hydrological Processes* 24.12, pp. 1711–1718. DOI: 10.1002/hyp.7691.

Dunn, S. M., J. R. Bacon, C. Soulsby, D. Tetzlaff, M. I. Stutter, S. Waldron, and I. A. Malcolm (2008). "Interpretation of homogeneity in δ18O signatures of stream water in a nested sub-catchment system in north-east Scotland". In: *Hydrological Processes* 22.24, pp. 4767–4782. DOI: 10.1002/hyp.7088.

Dunn, S. M., J. J. McDonnell, and K. B. Vaché (2007). "Factors influencing the residence time of catchment waters: A virtual experiment approach". In: *Water Resources Research* 43.6. DOI: 10.1029/2006WR005393.

Dunne, T. (1983). "Relation of field studies and modeling in the prediction of storm runoff". In: *Journal of Hydrology* 65.1. Scale Problems in Hydrology, pp. 25–48. DOI: 10.1016/0022-1694(83)90209-3.

Duvert, C., M. K. Stewart, D. I. Cendón, and M. Raiber (2016). "Time series of tritium, stable isotopes and chloride reveal short-term variations in groundwater contribution to a stream". In: *Hydrology and Earth System Sciences* 20.1, pp. 257–277. DOI: 10.5194/hess-20-257-2016.

Duy, N. L., N. V. Dung, I. Heidbüchel, H. Meyer, M. Weiler, B. Merz, and H. Apel (2019). "Identification of groundwater mean transit times of precipitation and riverbank infiltration by two-component lumped parameter models". In: *Hydrological Processes* 0.ja. DOI: 10.1002/hyp.13549.

Dyrness, C. T. (1969). *Hydrologic properties of soils on three small watersheds in the western Cascades of Oregon*. Res. Note PNW-111. Pacific Northwest Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture.

Eberts, S. M., J. K. Böhlke, L. J. Kauffman, and B. C. Jurgens (2012). "Comparison of particle-tracking and lumped-parameter age-distribution models for evaluating vulnerability of production wells to contamination". In: *Hydrogeology Journal* 20.2, pp. 263–282. DOI: 10.1007/s10040-011-0810-6.

Ehrhardt, S., R. Kumar, J. H. Fleckenstein, S. Attinger, and A. Musolff (2018). "Decadal trajectories of nitrate input and output in three nested catchments along a land use gradient". In: *Hydrology and Earth System Sciences Discussions* 2018, pp. 1–36. DOI: 10.5194/hess-2018-475.

Engdahl, N. B. and R. M. Maxwell (2014). "Approximating groundwater age distributions using simple streamtube models and multiple tracers". In: *Advances in Water Resources* 66, pp. 19–31. DOI: 10.1016/j.advwatres.2014.02.001.

Eriksson, E. (1958). "The Possible Use of Tritium' for Estimating Groundwater Storage". In: *Tellus* 10.4, pp. 472–478. DOI: 10.1111/j.2153-3490.1958.tb02035.x.

Erostate, M., F. Huneau, E. Garel, M. Lehmann, T. Kuhn, L. Aquilina, V. Vergnaud-Ayraud, T. Labasque, S. Santoni, S. Robert, D. Provitolo, and V. Pasqualini (2018). "Delayed nitrate dispersion within a coastal aquifer provides constraints on land-use evolution and nitrate contamination in the past". In: *Science of The Total Environment* 644, pp. 928–940. DOI: 10.1016/j.scitotenv.2018.06.375.

Etter, S., B. Strobl, J. Seibert, and H. J. I. van Meerveld (2018). "Value of uncertain streamflow observations for hydrological modelling". In: *Hydrology and Earth System Sciences* 22.10, pp. 5243–5257. DOI: 10.5194/hess-22-5243-2018.

Evaristo, J., M. Kim, J. van Haren, L. A. Pangle, C. J. Harman, P. A. Troch, and J. J. McDonnell (2019). "Characterizing the Fluxes and Age Distribution of Soil Water, Plant Water, and Deep Percolation in a Model Tropical Ecosystem". In: *Water Resources Research* 55.4, pp. 3307–3327. DOI: 10.1029/2018WR023265.

Farlin, J., T. Gallé, M. Bayerle, D. Pittois, C. Braun, H. El Khabbaz, M. Elsner, and P. Maloszewski (2013). "Predicting Pesticide Attenuation in a Fractured Aquifer Using Lumped-Parameter Models". In: *Groundwater* 51.2, pp. 276–285. DOI: 10.1111/j.1745-6584.2012.00964.x.

Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2006). "Is the groundwater reservoir linear? Learning from data in hydrological modelling". In: *Hydrology and Earth System Sciences* 10.1, pp. 139–150. DOI: 10.5194/hess-10-139-2006.

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?" In: *Hydrological Processes* 28.4, pp. 2451–2467. DOI: 10.1002/hyp.9726.

Fenicia, F., D. Kavetski, H. H. G. Savenije, and L. Pfister (2016). "From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions". In: *Water Resources Research* 52.2, pp. 954–989. DOI: 10.1002/2015WR017398.

Fenicia, F., J. J. McDonnell, and H. H. G. Savenije (2008). "Learning from model improvement: On the contribution of complementary data to process understanding". In: *Water Resources Research* 44.6. DOI: 10.1029/2007WR006386.

Fenicia, F., S. Wrede, D. Kavetski, L. Pfister, L. Hoffmann, H. H. G. Savenije, and J. J. McDonnell (2010). "Assessing the impact of mixing assumptions on the estimation of streamwater mean residence time". In: *Hydrological Processes* 24.12, pp. 1730–1741. DOI: 10.1002/hyp.7595.

Gabrielli, C., J. McDonnell, and W. Jarvis (2012). "The role of bedrock groundwater in rainfall–runoff response at hillslope and catchment scales". In: *Journal of Hydrology* 450-451, pp. 117–133. DOI: 10.1016/j.jhydrol.2012.05.023.

Gallart, F., M. Roig-Planasdemunt, M. K. Stewart, P. Llorens, U. Morgenstern, W. Stichler, L. Pfister, and J. Latron (2016). "A GLUE-based uncertainty assessment framework for tritium-inferred transit time estimations under baseflow conditions". In: *Hydrological Processes* 30.25, pp. 4741–4760. DOI: 10.1002/hyp.10991.

Gharari, S., M. Hrachowitz, F. Fenicia, H. Gao, and H. H. G. Savenije (2014). "Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration". In: *Hydrology and Earth System Sciences* 18.12, pp. 4839–4859. DOI: 10.5194/hess-18-4839-2014.

Glaser, B., M. Antonelli, M. Chini, L. Pfister, and J. Klaus (2018). "Technical note: Mapping surface-saturation dynamics with thermal infrared imagery". In: *Hydrology and Earth System Sciences* 22.11, pp. 5987–6003. DOI: 10.5194/hess-22-5987-2018.

Glaser, B., M. Antonelli, L. Hopp, and J. Klaus (2019a). "Intra-catchment variability of surface saturation – insights from longterm observations and simulations". In: *Hydrology and Earth System Sciences Discussions* 2019, pp. 1–22. DOI: 10.5194/hess-2019-203.

Glaser, B., C. Jackisch, L. Hopp, and J. Klaus (2019b). "How Meaningful are Plot-Scale Observations and Simulations of Preferential Flow for Catchment Models?" In: *Vadose Zone Journal* 18.1. DOI: 10.2136/vzj2018.08.0146.

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". In: *Water Resources Research* 52.10, pp. 8317–8342. DOI: 10.1002/2015WR018414.

Goderniaux, P., P. Davy, E. Bresciani, J.-R. de Dreuzy, and T. Le Borgne (2013). "Partitioning a regional groundwater flow system into shallow local and deep regional flow compartments". In: *Water Resources Research* 49.4, pp. 2274–2286. DOI: 10.1002/wrcr.20186.

Godsey, S. E., W. Aas, T. a. Clair, H. a. de Wit, I. J. Fernandez, J. S. Kahl, I. a. Malcolm, C. Neal, M. Neal, S. J. Nelson, S. a. Norton, M. C. Palucis, B. L. Skjelkvåle, C. Soulsby, D. Tetzlaff, and J. W. Kirchner (2010). "Generality of fractal 1/f scaling in catchment tracer time series, and its implications for catchment travel time distributions". In: *Hydrological Processes* 24.12, pp. 1660–1671. DOI: 10.1002/hyp.7677.

Gooseff, M. N., D. A. Benson, M. A. Briggs, M. Weaver, W. Wollheim, B. Peterson, and C. S. Hopkinson (2011). "Residence time distributions in surface transient storage zones in streams: Estimation via signal deconvolution". In: *Water Resources Research* 47.5. DOI: 10.1029/2010WR009959.

Gourdol, L., R. Clément, J. Juilleret, L. Pfister, and C. Hissler (2018). "Large-scale ERT surveys for investigating shallow regolith properties and architecture". In: *Hydrology and Earth System Sciences Discussions* 2018, pp. 1–39. DOI: 10.5194/hess-2018-519.

Graham, C. B., W. van Verseveld, H. R. Barnard, and J. J. McDonnell (2010). "Estimating the deep seepage component of the hillslope and catchment water balance within a measurement uncertainty framework". In: *Hydrological Processes* 24.25, pp. 3631–3647. DOI: 10.1002/hyp.7788.

Greenland, D. (1994). "The Pacific Northwest regional context of the climate of the H. J. Andrews experimental forest long-term ecological research site". In: *Northwest Science*.

Groh, J., S. C., A. Lücke, T. Pütz, J. Vanderborght, and H. Vereecken (2018). "Inverse Estimation of Soil Hydraulic and Transport Parameters of Layered Soils from Water Stable Isotope and Lysimeter Data". In: *Vadose Zone Journal* 17.1. DOI: 10.2136/vzj2017.09.0168.

Guillet, G., J. L. Knapp, S. Merel, O. A. Cirpka, P. Grathwohl, C. Zwiener, and M. Schwientek (2019). "Fate of wastewater contaminants in rivers: Using conservative-tracer based transfer functions to assess reactive transport". In: *Science of The Total Environment* 656, pp. 1250–1260. DOI: 10.1016/j.scitotenv.2018.11.379.

Gupta, P., D. Noone, J. Galewsky, C. Sweeney, and B. H. Vaughn (2009). "Demonstration of high-precision continuous measurements of water vapor isotopologues in laboratory and remote field deployments using wavelength-scanned cavity ring-down spectroscopy (WS-CRDS) technology". In: *Rapid Communications in Mass Spectrometry* 23.16, pp. 2534–2542. DOI: 10.1002/rcm.4100.

Gusyev, M. A., M. Toews, U. Morgenstern, M. Stewart, P. White, C. Daughney, and J. Hadfield (2013). "Calibration of a transient transport model to tritium data in streams and simulation of groundwater ages in the western Lake Taupo catchment, New Zealand". In: *Hydrology and Earth System Sciences* 17.3, pp. 1217–1227. DOI: 10.5194/hess-17-1217-2013.

Guthke, A. (2017). "Defensible Model Complexity: A Call for Data-Based and Goal-Oriented Model Choice". In: *Groundwater* 55.5, pp. 646–650. DOI: 10.1111/gwat.12554.

Haggerty, R., S. M. Wondzell, and M. A. Johnson (2002). "Power-law residence time distribution in the hyporheic zone of a 2nd-order mountain stream". In: *Geophysical Research Letters* 29.13, pp. 18-1-18–4. DOI: 10.1029/2002GL014743.

Halder, J., S. Terzer, L. I. Wassenaar, L. J. Araguás-Araguás, and P. K. Aggarwal (2015). "The Global Network of Isotopes in Rivers (GNIR): integration of water isotopes in watershed observation and riverine research". In: *Hydrology and Earth System Sciences* 19.8, pp. 3419–3431. DOI: 10.5194/hess-19-3419-2015.

Hale, V. C. and J. J. McDonnell (2016). "Effect of bedrock permeability on stream base flow mean transit time scaling relations: 1. A multiscale catchment intercomparison". In: *Water Resources Research* 52.2, pp. 1358–1374. DOI: 10.1002/2014WR016124.

Haraldsson, H. V. and H. Sverdrup (2013). "Finding Simplicity in Complexity in Biogeochemical Modelling". In: *Environmental Modelling*. John Wiley and Sons, Ltd. Chap. 17, pp. 277–289. DOI: 10.1002/9781118351475.ch17.

Harman, C. J., A. S. Ward, and A. Ball (2016). "How does reach-scale stream-hyporheic transport vary with discharge? Insights from rSAS analysis of sequential tracer injections in a headwater mountain stream". In: *Water Resources Research* 52.9, pp. 7130–7150. DOI: 10.1002/2016WR018832.

Harman, C. J. (2015). "Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed". In: *Water Resources Research* 51.1, pp. 1–30. DOI: 10.1002/2014WR015707.

Harr, R. D. (1977). "Water flux in soil and subsoil on a steep forested slope". In: *Journal of Hydrology* 33.1, pp. 37–58. DOI: 10.1016/0022-1694(77)90097-X.

Harr, R. D. and F. M. McCorison (1979). "Initial effects of clearcut logging on size and timing of peak flows in a small watershed in western Oregon". In: *Water Resources Research* 15.1, pp. 90–94. DOI: 10.1029/WR015i001p00090.

Heidbüchel, I., P. A. Troch, and S. W. Lyon (2013). "Separating physical and meteorological controls of variable transit times in zero-order catchments". In: *Water Resources Research* 49.11, pp. 7644–7657. DOI: 10.1002/2012WR013149.

Heidbüchel, I., P. A. Troch, S. W. Lyon, and M. Weiler (2012). "The master transit time distribution of variable flow systems". In: *Water Resources Research* 48. DOI: 10.1029/2011WR011293.

Helton, J. and F. Davis (2003). "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems". In: *Reliability Engineering and System Safety* 81.1, pp. 23–69. DOI: 10.1016/S0951-8320(03)00058-9.

Herbstritt, B., B. Gralher, and M. Weiler (2019). "Continuous, near-real-time observations of water stable isotope ratios during rainfall and throughfall events". In: *Hydrology and Earth System Sciences* 23.7, pp. 3007–3019. DOI: 10.5194/hess-23-3007-2019.

Herbstritt, B., B. Gralher, and M. Weiler (2012). "Continuous in situ measurements of stable isotopes in liquid water". In: *Water Resources Research* 48.3. DOI: 10.1029/2011WR011369.

Hissler, C., P. Stille, C. Guignard, J. F. Iffly, and L. Pfister (2014). "Rare Earth Elements as Hydrological Tracers of Anthropogenic and Critical Zone Contributions: A Case Study at the Alzette River Basin Scale." In: *Procedia Earth and Planetary Science* 10. Geochemistry of the Earth's surface GES-10 Paris France, 18-23 August, 2014., pp. 349–352. DOI: 10.1016/j.proeps.2014.08.036.

Hornberger, G., T. Scanlon, and J. Raffensperger (2001). "Modelling transport of dissolved silica in a forested headwater catchment: the effect of hydrological and chemical time scales on hysteresis in the concentration-discharge relationship". In: *Hydrological Processes* 15.10, 2029–2038. DOI: 10.1002/hyp.254.

Hrachowitz, M. and M. P. Clark (2017). "HESS Opinions: The complementary merits of competing modelling philosophies in hydrology". In: *Hydrology and Earth System Sciences* 21.8, pp. 3953–3973. DOI: 10.5194/hess-21-3953-2017.

Hrachowitz, M., O. Fovet, L. Ruiz, T. Euser, S. Gharari, R. Nijzink, J. Freer, H. H. G. Savenije, and C. Gascuel-Odoux (2014). "Process consistency in models: The importance of system signatures, expert knowledge, and process complexity". In: *Water Resources Research* 50.9, pp. 7445–7469. DOI: 10.1002/2014WR015484.

Hrachowitz, M., H. Savenije, T. A. Bogaard, D. Tetzlaff, and C. Soulsby (2013). "What can flux tracking teach us about water age distribution patterns and their temporal dynamics?" In: *Hydrology and Earth System Sciences* 17.2, 533–564. DOI: 10.5194/hess-17-533-2013.

Hrachowitz, M., C. Soulsby, D. Tetzlaff, J. J. C. Dawson, and I. A. Malcolm (2009a). "Regionalization of transit time estimates in montane catchments by integrating landscape controls". In: *Water Resources Research* 45.5. DOI: 10.1029/2008WR007496.

Hrachowitz, M., C. Soulsby, D. Tetzlaff, J. Dawson, S. Dunn, and I. Malcolm (2009b). "Using long-term data sets to understand transit times in contrasting headwater catchments". In: *Journal of Hydrology* 367.3, pp. 237–248. DOI: 10.1016/j.jhydrol.2009.01.001.

Hrachowitz, M., C. Soulsby, D. Tetzlaff, and I. A. Malcolm (2011). "Sensitivity of mean transit time estimates to model conditioning and data availability". In: *Hydrological Processes* 25.6, pp. 980–990. DOI: 10.1002/hyp.7922.

Hrachowitz, M., C. Soulsby, D. Tetzlaff, I. A. Malcolm, and G. Schoups (2010). "Gamma distribution models for transit time estimation in catchments: Physical interpretation of parameters and implications for time-variant transit time assessment". In: *Water Resources Research* 46.10. DOI: 10.1029/2010WR009148.

Hrachowitz, M., P. Benettin, B. M. van Breukelen, O. Fovet, N. J. Howden, L. Ruiz, Y. van der Velde, and A. J. Wade (2016). "Transit times—the link between hydrology and water quality at the catchment scale". In: *Wiley Interdisciplinary Reviews: Water* 3.5, pp. 629–657. DOI: 10.1002/wat2.1155.

Hrachowitz, M., O. Fovet, L. Ruiz, and H. H. G. Savenije (2015). "Transit time distributions, legacy contamination and variability in biogeochemical $1/f^{\alpha}$ scaling: how are hydrological response dynamics linked to water quality at the catchment scale?" In: *Hydrological Processes* 29.25, pp. 5241–5256. DOI: 10.1002/hyp.10546.

Hubert, P. (1971). "Etude par le tritium de la dynamique des eaux du Lac Léman. Apport du tritium à la limnologie physique". PhD thesis. Université de Paris.

IAEA (2019). *Global Network of Isotopes in Rivers. The GNIR Database*.

IAEA and WMO (2019). *Global Network of Isotopes in Precipitation. The GNIP Database*.

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". In: *Hydrology and Earth System Sciences* 21.7, pp. 3749–3775. DOI: 10.5194/hess-21-3749-2017.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An Introduction to Statistical Learning*.

Jasechko, S., J. W. Kirchner, J. M. Welker, and J. J. McDonnell (2016). "Substantial proportion of global streamflow less than three months old". In: *Nature Geoscience* 9, pp. 126–129. DOI: 10.1038/ngeo2636.

Jencso, K. G., B. L. McGlynn, M. N. Gooseff, S. M. Wondzell, K. E. Bencala, and L. A. Marshall (2009). "Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale". In: *Water Resources Research* 45.4. DOI: 10.1029/2008WR007225.

Jing, M., F. Heße, R. Kumar, O. Kolditz, T. Kalbacher, and S. Attinger (2019). "Influence of input and parameter uncertainty on the prediction of catchment-scale groundwater travel time distributions". In: *Hydrology and Earth System Sciences* 23.1, pp. 171–190. DOI: 10.5194/hess-23-171-2019.

Johnson, N. M., G. E. Likens, F. Bormann, and R. S. Pierce (1968). "Rate of chemical weathering of silicate minerals in New Hampshire". In: *Geochimica et Cosmochimica Acta* 32.5, pp. 531–545. DOI: 10.1016/0016-7037(68)90044-6.

Juilleret, J., J. Iffly, L. Pfister, and C. Hissler (2011). "Remarkable Pleistocene periglacial slope deposits in Luxembourg (Oesling): pedological implication and geosite potential". In: *Bulletin de la Société des naturalistes luxembourgeois* 112, pp. 125–130.

Juilleret, J., S. Dondeyne, K. Vancampenhout, J. Deckers, and C. Hissler (2016). "Mind the gap: A classification system for integrating the subsolum into soil surveys". In: *Geoderma* 264. Soil mapping, classification, and modelling: history and future directions, pp. 332–339. DOI: 10.1016/j.geoderma.2015.08.031.

Kaandorp, V. P., P. G. B. de Louw, Y. van der Velde, and H. P. Broers (2018). "Transient Groundwater Travel Time Distributions and Age-Ranked Storage-Discharge Relationships of Three Lowland Catchments". In: *Water Resources Research* 54.7, pp. 4519–4536. DOI: 10.1029/2017wr022461.

Kavetski, D. and M. P. Clark (2010). "Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction". In: *Water Resources Research* 46.10. DOI: 10.1029/2009WR008896.

Kavetski, D., F. Fenicia, and M. P. Clark (2011). "Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment". In: *Water Resources Research* 47.5. DOI: 10.1029/2010WR009525.

Kavetski, D. and G. Kuczera (2007). "Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration". In: *Water Resources Research* 43.3. DOI: 10.1029/2006WR005195.

Kazemi, G. A., J. H. Lehr, and P. Perrochet (2005). "The Applications of Groundwater Age Data". In: *Groundwater Age*. John Wiley and Sons, Ltd. Chap. 3, pp. 51–80. DOI: 10.1002/0471929514.ch3.

Keim, R. F., C. Kendall, and A. Jefferson (2014). "The Expanding Utility of Laser Spectroscopy". In: *Eos, Transactions American Geophysical Union* 95.17, pp. 144–144. DOI: 10.1002/2014EO170007.

Kendall, C. and J. J. McDonnell (1998). *Isotope tracers in catchment hydrology*. Elsevier, Amsterdam. DOI: 10.1016/B978-0-444-81546-0.50001-X.

Kendall, C., J. J. McDonnell, and W. Gu (2001). "A look inside 'black box' hydrograph separation models: a study at the Hydrohill catchment". In: *Hydrological Processes* 15.10, pp. 1877–1902. DOI: 10.1002/hyp.245.

Kennedy, V. C., G. W. Zellweger, and R. J. Avanzino (1979). "Variation of rain chemistry during storms at two sites in northern California". In: *Water Resources Research* 15.3, pp. 687–702. DOI: 10.1029/WR015i003p00687.

Kim, M., L. A. Pangle, C. Cardoso, M. Lora, T. H. M. Volkmann, Y. Wang, C. J. Harman, and P. A. Troch (2016). "Transit time distributions and StorAge Selection functions in a sloping soil lysimeter with time-varying flow paths: Direct observation of internal and external transport variability". In: *Water Resources Research*. DOI: 10.1002/2016WR018620.

Kirchner, J. W. (2019). "Quantifying new water fractions and transit time distributions using ensemble hydrograph separation: theory and benchmark tests". In: *Hydrology and Earth System Sciences* 23.1, pp. 303–349. DOI: 10.5194/hess-23-303-2019.

Kirchner, J. W., X. Feng, and C. Neal (2001). "Catchment-scale advection and dispersion as a mechanism for fractal scaling in stream tracer concentrations". In: *Journal of Hydrology* 254.1-4, 82–101. DOI: 10.1016/S0022-1694(01)00487-5.

Kirchner, J. W. and C. Neal (2013). "Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection." In: *Proceedings of the National Academy of Sciences of the United States of America* 110.30, pp. 12213–8. DOI: 10.1073/pnas.1304328110.

Kirchner, J. W. (2003). "A double paradox in catchment hydrology and geochemistry". In: *Hydrological Processes* 17.4, pp. 871–874. DOI: 10.1002/hyp.5108.

— (2006). "Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology". In: *Water Resources Research* 42.3. DOI: 10.1029/2005WR004362.

— (2016a). "Aggregation in environmental systems - Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments". In: *Hydrology and Earth System Sciences* 20.1, pp. 279–297. DOI: 10.5194/hess-20-279-2016.

Kirchner, J. W. (2016b). "Aggregation in environmental systems - Part 2: Catchment mean transit times and young water fractions under hydrologic nonstationarity". In: *Hydrology and Earth System Sciences* 20.1, pp. 299–328. DOI: 10.5194/hess-20-299-2016.

Kirchner, J. W., X. Feng, and C. Neal (2000). "Fractal stream chemistry and its implications for contaminant transport in catchments". In: *NATURE* 403.6769, 524–527. DOI: 10.1038/35000537.

Kirchner, J. W., X. Feng, C. Neal, and A. J. Robson (2004). "The fine structure of water-quality dynamics: the (high-frequency) wave of the future". In: *Hydrological Processes* 18.7, pp. 1353–1359. DOI: 10.1002/hyp.5537.

Klaus, J. and J. McDonnell (2013). "Hydrograph separation using stable isotopes: Review and evaluation". In: *Journal of Hydrology* 505, pp. 47–64. DOI: 10.1016/j.jhydrol.2013.09.006.

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". In: *Hydrology and Earth System Sciences* 17.1, pp. 103–118. DOI: 10.5194/hess-17-103-2013.

Klaus, J., K. P. Chun, K. J. McGuire, and J. J. McDonnell (2015a). "Temporal dynamics of catchment transit times from stable isotope data". In: *Water Resources Research* 51.6, pp. 4208–4223. DOI: 10.1002/2014WR016247.

Klaus, J. and C. R. Jackson (2018). "Interflow Is Not Binary: A Continuous Shallow Perched Layer Does Not Imply Continuous Connectivity". In: *Water Resources Research* 54.9, pp. 5921–5932. DOI: 10.1029/2018WR022920.

Klaus, J., C. E. Wetzel, N. Martínez-Carreras, L. Ector, and L. Pfister (2015b). "A tracer to bridge the scales: on the value of diatoms for tracing fast flow path connectivity from headwaters to meso-scale catchments". In: *Hydrological Processes* 29.25, pp. 5275–5289. DOI: 10.1002/hyp.10628.

Knighton, J., V. Souter-Kline, T. Volkman, P. A. Troch, M. Kim, C. Harman, C. Morris, B. Buchanan, and M. T. Walter (2019). "Seasonal and Topographic Variations in Ecohydrological Separation Within a Small, Temperate, Snow-Influenced Catchment". In: *Water Resources Research* 0.0. DOI: 10.1029/2019WR025174.

Koehler, G. and L. I. Wassenaar (2011). "Realtime Stable Isotope Monitoring of Natural Waters by Parallel-Flow Laser Spectroscopy". In: *Analytical Chemistry* 83.3. PMID: 21214188, pp. 913–919. DOI: 10.1021/ac102584q.

Kreft, A. and A. Zuber (1978). "On the physical meaning of the dispersion equation and its solutions for different initial and boundary conditions". In: *Chemical Engineering Science* 33.11, pp. 1471–1480. DOI: 10.1016/0009-2509(78)85196-3.

Lamb, R. and K. Beven (1997). "Using interactive recession curve analysis to specify a general catchment storage model". In: *Hydrology and Earth System Sciences* 1.1, pp. 101–113. DOI: 10.5194/hess-1-101-1997.

Le Borgne, T., J.-R. de Dreuzy, P. Davy, and O. Bour (2007). "Characterization of the velocity field organization in heterogeneous media by conditional correlation". In: *Water Resources Research* 43.2. DOI: 10.1029/2006WR004875.

Legout, C., J. Molenat, L. Aquilina, C. Gascuel-Odoux, M. Faucheux, Y. Fauvel, and T. Bariac (2007). "Solute transfer in the unsaturated zone-groundwater continuum of a headwater catchment". In: *Journal of Hydrology* 332.3, pp. 427–441. DOI: 10.1016/j.jhydrol.2006.07.017.

Leray, S., N. B. Engdahl, A. Massoudieh, E. Bresciani, and J. McCallum (2016). "Residence time distributions for hydrologic systems: Mechanistic foundations and steady-state analytical solutions". In: *Journal of Hydrology* 543. Residence times

in subsurface hydrological systems: Signature of hydrological processes and impact on environmental applications, pp. 67–87. DOI: 10.1016/j.jhydrol.2016.01.068.

Lindgren, G. A., G. Destouni, and A. V. Miller (2004). "Solute transport through the integrated groundwater-stream system of a catchment". In: *Water Resources Research* 40.3. DOI: 10.1029/2003WR002765.

Lis, G., L. I. Wassenaar, and M. J. Hendry (2008). "High-Precision Laser Spectroscopy D/H and 18O/16O Measurements of Microliter Natural Water Samples". In: *Analytical Chemistry* 80.1. PMID: 18031060, pp. 287–293. DOI: 10.1021/ac701716q.

Long, A. J. and L. D. Putnam (2009). "Age-distribution estimation for karst groundwater: Issues of parameterization and complexity in inverse modeling by convolution". In: *Journal of Hydrology* 376.3, pp. 579–588. DOI: 10.1016/j.jhydrol.2009.07.064.

Loritz, R., H. Gupta, C. Jackisch, M. Westhoff, A. Kleidon, U. Ehret, and E. Zehe (2018). "On the dynamic nature of hydrological similarity". In: *Hydrology and Earth System Sciences* 22.7, pp. 3663–3684. DOI: 10.5194/hess-22-3663-2018.

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". In: *Hydrology and Earth System Sciences* 21.2, pp. 1225–1249. DOI: 10.5194/hess-21-1225-2017.

Loritz, R., A. Kleidon, C. Jackisch, M. Westhoff, U. Ehret, H. Gupta, and E. Zehe (2019). "A topographic index explaining hydrological similarity by accounting for the joint controls of runoff formation". In: *Hydrology and Earth System Sciences* 23.9, pp. 3807–3821. DOI: 10.5194/hess-23-3807-2019.

Luo, J. and O. A. Cirpka (2008). "Traveltime-based descriptions of transport and mixing in heterogeneous domains". In: *Water Resources Research* 44.9. DOI: 10.1029/2007WR006035.

Maher, K. (2010). "The dependence of chemical weathering rates on fluid residence time". In: *Earth and Planetary Science Letters* 294.1–2, pp. 101–110. DOI: 10.1016/j.epsl.2010.03.010.

— (2011). "The role of fluid residence time and topographic scales in determining chemical fluxes from landscapes". In: *Earth and Planetary Science Letters* 312, pp. 48–58. DOI: 10.1016/j.epsl.2011.09.040.

Maher, K. and C. P. Chamberlain (2014). "Hydrologic Regulation of Chemical Weathering and the Geologic Carbon Cycle". In: *Science* 343.6178, pp. 1502–1504. DOI: 10.1126/science.1250770.

Małoszewski, P. and A. Zuber (1982). "Determining the turnover time of groundwater systems with the aid of environmental tracers: 1. Models and their applicability". In: *Journal of Hydrology* 57.3, pp. 207–231. DOI: 10.1016/0022-1694(82)90147-0.

Małoszewski, P., W. Rauert, W. Stichler, and A. Herrmann (1983). "Application of flow models in an alpine catchment area using tritium and deuterium data". In: *Journal of Hydrology* 66.1, pp. 319–330. DOI: 10.1016/0022-1694(83)90193-2.

Martinec, J. (1975). "Subsurface flow from snowmelt traced by tritium". In: *Water Resources Research* 11.3, pp. 496–498. DOI: 10.1029/WR011i003p00496.

Martínez-Carreras, N., C. E. Wetzel, J. Frentress, L. Ector, J. J. McDonnell, L. Hoffmann, and L. Pfister (2015). "Hydrological connectivity inferred from diatom transport through the riparian-stream system". In: *Hydrology and Earth System Sciences* 19.7, pp. 3133–3151. DOI: 10.5194/hess-19-3133-2015.

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". In: *Journal of Hydrology* 543, pp. 255–269. DOI: 10.1016/j.jhydrol.2016.10.004.

Massoudieh, A., S. Sharifi, and D. K. Solomon (2012). "Bayesian evaluation of groundwater age distribution using radioactive tracers and anthropogenic chemicals". In: *Water Resources Research* 48.9. DOI: 10.1029/2012WR011815.

Massoudieh, A., A. Visser, S. Sharifi, and H. P. Broers (2014). "A Bayesian modeling approach for estimation of a shape-free groundwater age distribution using multiple tracers". In: *Applied Geochemistry* 50, pp. 252–264. DOI: 10.1016/j.apgeochem.2013.10.004.

Maxwell, R. M., L. E. Condon, M. Danesh-Yazdi, and L. A. Bearup (2018). "Exploring source water mixing and transient residence time distributions of outflow and evapotranspiration with an integrated hydrologic model and Lagrangian particle tracking approach". In: *Ecohydrology*, e2042. DOI: 10.1002/eco.2042.

McCallum, J. L., N. B. Engdahl, T. R. Ginn, and P. G. Cook (2014). "Nonparametric estimation of groundwater residence time distributions: What can environmental tracer data tell us about groundwater residence time?" In: *Water Resources Research* 50.3, pp. 2022–2038. DOI: 10.1002/2013WR014974.

McCutcheon, R. J., J. P. McNamara, M. J. Kohn, and S. L. Evans (2017). "An evaluation of the ecohydrological separation hypothesis in a semiarid catchment". In: *Hydrological Processes* 31.4, pp. 783–799. DOI: 10.1002/hyp.11052.

McDonnell, J. J., K. McGuire, P. Aggarwal, K. J. Beven, D. Biondi, G. Destouni, S. Dunn, A. James, J. Kirchner, P. Kraft, S. Lyon, P. Maloszewski, B. Newman, L. Pfister, A. Rinaldo, A. Rodhe, T. Sayama, J. Seibert, K. Solomon, C. Soulsby, M. Stewart, D. Tetzlaff, C. Tobin, P. Troch, M. Weiler, A. Western, A. Wörman, and S. Wrede (2010). "How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis". In: *Hydrological Processes* 24.12, pp. 1745–1754. DOI: 10.1002/hyp.7796.

McDonnell, J. J. (2014). "The two water worlds hypothesis: ecohydrological separation of water between streams and trees?" In: *Wiley Interdisciplinary Reviews: Water* 1.4, pp. 323–329. DOI: 10.1002/wat2.1027.

McDonnell, J. and K. Beven (2014). "Debates on Water Resources: The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph". In: *Water Resources Research* 50.6, pp. 5342–5350. DOI: 10.1002/2013WR015141.

McGuire, K. J., J. J. McDonnell, M. Weiler, C. Kendall, B. L. McGlynn, J. M. Welker, and J. Seibert (2005). "The role of topography on catchment-scale water residence time". In: *Water Resources Research* 41.5. DOI: 10.1029/2004WR003657.

McGuire, K. J. and J. J. McDonnell (2006). "A review and evaluation of catchment transit time modeling". In: *Journal of Hydrology* 330.3-4, pp. 543–563. DOI: 10.1016/j.jhydrol.2006.04.020.

— (2010). "Hydrological connectivity of hillslopes and streams: Characteristic time scales and nonlinearities". In: *Water Resources Research* 46.10. DOI: 10.1029/2010WR009341.

McGuire, K. J., M. Weiler, and J. McDonnell (2007). "Integrating tracer experiments with modeling to assess runoff processes and water transit times". In: *Advances in Water Resources* 30.4, pp. 824–837. DOI: 10.1016/j.advwatres.2006.07.004.

McMahon, T. A., M. C. Peel, L. Lowe, R. Srikanthan, and T. R. McVicar (2013). "Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis". In: *Hydrology and Earth System Sciences* 17.4, pp. 1331–1363. DOI: 10.5194/hess-17-1331-2013.

McMillan, H., T. Krueger, and J. Freer (2012a). "Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality". In: *Hydrological Processes* 26.26, pp. 4078–4111. DOI: 10.1002/hyp.9384.

McMillan, H., D. Tetzlaff, M. Clark, and C. Soulsby (2012b). "Do time-variable tracers aid the evaluation of hydrological model structure? A multimodel approach". In: *Water Resources Research* 48.5. DOI: 10.1029/2011WR011688.

Michelsen, N., G. Laube, J. Friesen, S. M. Weise, A. B. A. Bait Said, and T. Müller (2019). "Technical note: A microcontroller-based automatic rain sampler for stable isotope studies". In: *Hydrology and Earth System Sciences* 23.6, pp. 2637–2645. DOI: 10.5194/hess-23-2637-2019.

Monteith, J. L. (1981). "Evaporation and surface temperature". In: *Quarterly Journal of the Royal Meteorological Society* 107.451, pp. 1–27. DOI: 10.1002/qj.49710745102.

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

Morgenstern, U. and C. B. Taylor (2009). "Ultra low-level tritium measurement using electrolytic enrichment and LSC". In: *Isotopes in Environmental and Health Studies* 45.2. PMID: 20183224, pp. 96–117. DOI: 10.1080/10256010902931194.

Munksgaard, N. C., C. M. Wurster, and M. I. Bird (2011). "Continuous analysis of δ^{18}O and δD values of water by diffusion sampling cavity ring-down spectrometry: a novel sampling device for unattended field monitoring of precipitation, ground and surface waters". In: *Rapid Communications in Mass Spectrometry* 25.24, pp. 3706–3712. DOI: 10.1002/rcm.5282.

Nash, J. and J. Sutcliffe (1970). "River flow forecasting through conceptual models part I — A discussion of principles". In: *Journal of Hydrology* 10.3, pp. 282–290. DOI: 10.1016/0022-1694(70)90255-6.

Neuweiler, I. and R. Helmig (2017). "Debates—Hypothesis testing in hydrology: A subsurface perspective". In: *Water Resources Research* 53.3, pp. 1784–1791. DOI: 10.1002/2016WR020047.

Nippgen, F., B. L. McGlynn, and R. E. Emanuel (2015). "The spatial and temporal evolution of contributing areas". In: *Water Resources Research* 51.6, pp. 4550–4573. DOI: 10.1002/2014WR016719.

Oldham, C. E., D. E. Farrow, and S. Peiffer (2013). "A generalized Damkohler number for classifying material processing in hydrological systems". In: *Hydrology and Earth System Sciences* 17.3, pp. 1133–1148. DOI: 10.5194/hess-17-1133-2013.

Östlund, G. H. (2013). "Hurricane Tritium I: Preliminary Results on Hilda 1964 and Betsy 1965". In: *Isotope Techniques in the Hydrologic Cycle*. American Geophysical Union (AGU), pp. 58–60. DOI: 10.1029/GM011p0058.

Oudijk, G. and L. M. Schmitt (2000). "Age dating of a chlorinated solvent plume in groundwater". In: *Tracers and Modelling in Hydrogeology, IAHS publication no. 262*. International Atomic Energy Agency (IAEA).

Page, T., K. Beven, J. Freer, and C. Neal (2007). "Modelling the chloride signal at Plynlimon, Wales, using a modified dynamic TOPMODEL incorporating conservative chemical mixing (with uncertainty)". In: *Hydrological Processes* 21.3, pp. 292–307. DOI: 10.1002/hyp.

Palcsu, L., U. Morgenstern, J. Sültenfuss, G. Koltai, E. László, M. Temovski, Z. Major, J. T. Nagy, L. Papp, C. Varlam, I. Faurescu, M. Túri, L. Rinyu, G. Czuppon, E. Bottyán, and A. J. T. Jull (2018). "Modulation of Cosmogenic Tritium in Meteoric Precipitation by the 11-year Cycle of Solar Magnetic Field Activity". In: *Scientific Reports* 8.1, p. 12813. DOI: 10.1038/s41598-018-31208-9.

Pangle, L. A., M. Kim, C. Cardoso, M. Lora, A. A. Meira Neto, T. H. M. Volkmann, Y. Wang, P. A. Troch, and C. J. Harman (2017). "The mechanistic basis for storage-dependent age distributions of water discharged from an experimental hillslope". In: *Water Resources Research* 53.4, pp. 2733–2754. DOI: 10.1002/2016WR019901.

Pangle, L. A., J. Klaus, E. S. F. Berman, M. Gupta, and J. J. McDonnell (2013). "A new multisource and high-frequency approach to measuring δ^2H in hydrological field studies". In: *Water Resources Research* 49.11, pp. 7797–7803. DOI: 10.1002/2013WR013743.

Parsekian, A. D., K. Singha, B. J. Minsley, W. S. Holbrook, and L. Slater (2015). "Multiscale geophysical imaging of the critical zone". In: *Reviews of Geophysics* 53.1, pp. 1–26. DOI: 10.1002/2014RG000465.

Pfister, L., C. Grave, J.-N. Beisel, and J. J. McDonnell (2019). "A global assessment of freshwater mollusk shell oxygen isotope signatures and their relation to precipitation and stream water". In: *Scientific Reports* 9.1, p. 4312. DOI: 10.1038/s41598-019-40369-0.

Pfister, L. and J. W. Kirchner (2017). "Debates—Hypothesis testing in hydrology: Theory and practice". In: *Water Resources Research* 53.3, pp. 1792–1798. DOI: 10.1002/2016WR020116.

Pfister, L., N. Martínez-Carreras, C. Hissler, J. Klaus, G. E. Carrer, M. K. Stewart, and J. J. McDonnell (2017a). "Bedrock geology controls on catchment storage, mixing, and release: A comparative analysis of 16 nested catchments". In: *Hydrological Processes* 31.10, pp. 1828–1845. DOI: 10.1002/hyp.11134.

Pfister, L., J. J. McDonnell, S. Wrede, D. Hlúbiková, P. Matgen, F. Fenicia, L. Ector, and L. Hoffmann (2009). "The rivers are alive: on the potential for diatoms as a tracer of water source and hydrological connectivity". In: *Hydrological Processes* 23.19, pp. 2841–2845. DOI: 10.1002/hyp.7426.

Pfister, L., F. Thielen, E. Deloule, N. Valle, E. Lentzen, C. Grave, J.-N. Beisel, and J. J. McDonnell (2018). "Freshwater pearl mussels as a stream water stable isotope recorder". In: *Ecohydrology* 11.7, e2007. DOI: 10.1002/eco.2007.

Pfister, L., C. E. Wetzel, J. Klaus, N. Martínez-Carreras, M. Antonelli, A. J. Teuling, and J. J. McDonnell (2017b). "Terrestrial diatoms as tracers in catchment hydrology: a review". In: *Wiley Interdisciplinary Reviews: Water* 4.6, e1241. DOI: 10.1002/wat2.1241.

Pool, S., D. Viviroli, and J. Seibert (2017). "Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration?" In: *Journal of Hydrology* 554, pp. 613–622. DOI: 10.1016/j.jhydrol.2017.09.037.

Pütz, T., R. Kiese, U. Wollschläger, J. Groh, H. Rupp, S. Zacharias, E. Priesack, H. H. Gerke, R. Gasche, O. Bens, E. Borg, C. Baessler, K. Kaiser, M. Herbrich, J.-C. Munch, M. Sommer, H.-J. Vogel, J. Vanderborght, and H. Vereecken (2016). "TERENO-SOILCan: a lysimeter-network in Germany observing soil processes and plant diversity influenced by climate change". In: *Environmental Earth Sciences* 75.18, p. 1242. DOI: 10.1007/s12665-016-6031-5.

Queloz, P., E. Bertuzzo, L. Carraro, G. Botter, F. Miglietta, P. Rao, and A. Rinaldo (2015a). "Transport of fluorobenzoate tracers in a vegetated hydrologic control volume: 1. Experimental results". In: *Water Resources Research* 51.4, pp. 2773–2792. DOI: 10.1002/2014WR016433.

Queloz, P., L. Carraro, P. Benettin, G. Botter, A. Rinaldo, and E. Bertuzzo (2015b). "Transport of fluorobenzoate tracers in a vegetated hydrologic control volume: 2. Theoretical inferences and modeling". In: *Water Resources Research* 51.4, pp. 2793–2806. DOI: 10.1002/2014WR016508.

Rank, D. and W. Papesch (2005). "Isotopic composition of precipitation in Austria in relation to air circulation patterns and climate". In: *Isotopic composition of precipitation in the Mediterranean basin in relation to air circulation patterns and climate*. International Atomic Energy Agency (IAEA). Chap. 2, pp. 19–35.

Rank, D., S. Wyhlidal, K. Schott, S. Weigand, and A. Oblin (2018). "Temporal and spatial distribution of isotopes in river water in Central Europe: 50 years experience with the Austrian network of isotopes in rivers". In: *Isotopes in Environmental and Health Studies* 54.2. PMID: 29082751, pp. 115–136. DOI: 10.1080/10256016.2017.1383906.

Remondi, F., M. Botter, P. Burlando, and S. Fatichi (2019). "Variability of transit time distributions with climate and topography: A modelling approach". In: *Journal of Hydrology* 569, pp. 37–50. DOI: 10.1016/j.jhydrol.2018.11.011.

Remondi, F., J. W. Kirchner, P. Burlando, and S. Fatichi (2018). "Water Flux Tracking With a Distributed Hydrological Model to Quantify Controls on the Spatiotemporal Variability of Transit Time Distributions". In: *Water Resources Research* 54.4, pp. 3081–3099. DOI: 10.1002/2017wr021689.

Rigon, R., M. Bancheri, and T. R. Green (2016). "Age-ranked hydrological budgets and a travel time description of catchment hydrology". In: *Hydrology and Earth System Sciences* 20.12, pp. 4929–4947. DOI: 10.5194/hess-20-4929-2016.

Rinaldo, A. and A. Marani (1987). "Basin scale-model of solute transport". In: *Water Resources Research* 23.11, 2107–2118. DOI: 10.1029/WR023i011p02107.

Rinaldo, A., K. J. Beven, E. Bertuzzo, L. Nicotina, J. Davies, A. Fiori, D. Russo, and G. Botter (2011). "Catchment travel time distributions and water flow in soils". In: *Water Resources Research* 47. DOI: 10.1029/2011WR010478.

Rinaldo, A., P. Benettin, C. J. Harman, M. Hrachowitz, K. J. McGuire, Y. van der Velde, E. Bertuzzo, and G. Botter (2015). "Storage selection functions: A coherent framework for quantifying how catchments store and release water and solutes". In: *Water Resources Research* 51.6, pp. 4840–4847. DOI: 10.1002/2015WR017273.

Rinaldo, A., A. Marani, and R. Rigon (1991). "Geomorphological Dispersion". In: *Water Resources Research* 27.4, pp. 513–525.

Ritzi Jr., R. W., D. F. Dominic, A. J. Slesers, C. B. Greer, E. C. Reboulet, J. A. Telford, R. W. Masters, C. A. Klohe, J. L. Bogle, and B. P. Means (2000). "Comparing statistical models of physical heterogeneity in buried-valley aquifers". In: *Water Resources Research* 36.11, pp. 3179–3192. DOI: 10.1029/2000WR900143.

Rodhe, A., L. Nyberg, and K. Bishop (1996). "Transit Times for Water in a Small Till Catchment from a Step Shift in the Oxygen 18 Content of the Water Input". In: *Water Resources Research* 32.12, pp. 3497–3511. DOI: 10.1029/95WR01806.

Rodriguez, N. B., L. Pfister, E. Zehe, and J. Klaus (under review). "A comparison of catchment travel times and storage deduced from deuterium and tritium tracers using StorAge Selection functions". In: *Hydrology and Earth System Sciences Discussions*.

Rodriguez, N. B., P. Benettin, and J. Klaus (2020). "Multimodal water age distributions and the challenge of complex hydrological landscapes". In: *Hydrological Processes*. DOI: 10.1002/hyp.13770.

Rodriguez, N. B. and J. Klaus (2019). "Catchment Travel Times From Composite StorAge Selection Functions Representing the Superposition of Streamflow Generation Processes". In: *Water Resources Research* 55.11, pp. 9292–9314. DOI: 10.1029/2019WR024973.

Rodriguez, N. B., K. J. McGuire, and J. Klaus (2018). "Time-Varying Storage-Water Age Relationships in a Catchment With a Mediterranean Climate". In: *Water Resources Research* 54.6, pp. 3988–4008. DOI: 10.1029/2017wr021964.

Rodriguez-Iturbe, I. and J. B. Valdes (1979). "The geomorphologic structure of hydrologic response". In: *Water Resources Research* 15.6, pp. 1409–1420. DOI: 10.1029/WR015i006p01409.

Rothacher, J., R. L. Fredricksen, and C. T. Dyrness (1967). *Hydrologic and related characteristics of three small watersheds in the Oregon Cascades*. Publication number 344. Pacific Northwest Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture.

Rothacher, J. S. (2016). *Stream discharge in gaged watersheds at the Andrews Experimental Forest, 1949 to present. Environmental Data Initiative*. URL: https://doi.org/10.6073/pasta/ce70a42e1f637cd430d7b0fb64faefb4.

Różański, K., K. Froehlich, and W. G. Mook (2001). *Environmental Isotopes in the Hydrological Cycle, Principles and Applications. Volume III: Surface water*. IAEA and UNESCO.

Sayama, T. and J. J. McDonnell (2009). "A new time-space accounting scheme to predict stream water residence time and hydrograph source components at the watershed scale". In: *Water Resources Research* 45.7. DOI: 10.1029/2008WR007549.

Scaini, A., N. Amvrosiadi, C. Hissler, L. Pfister, and K. Beven (2019). "Following tracer through the unsaturated zone using a multiple interacting pathways model: Implications from laboratory experiments". In: *Hydrological Processes* 0.0. DOI: 10.1002/hyp.13466.

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". In: *Journal of Hydrology* 546, pp. 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. Beven (2018). "Hillslope response to sprinkling and natural rainfall using velocity and celerity estimates in a slate-bedrock catchment". In: *Journal of Hydrology* 558, pp. 366–379. DOI: 10.1016/j.jhydrol.2017.12.011.

Schaefli, B. and H. V. Gupta (2007). "Do Nash values have value?" In: *Hydrological Processes* 21.15, pp. 2075–2080. DOI: 10.1002/hyp.6825.

Schmidt, A., G. Frank, W. Stichler, L. Duester, T. Steinkopff, and C. Stumpp (submitted). "Tritium in precipitation and surface waters in Germany". In: *Hydrological Processes*.

Schoups, G. and J. A. Vrugt (2010). "A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors". In: *Water Resources Research* 46.10. DOI: 10.1029/2009WR008933.

Schwab, M. P., J. Klaus, L. Pfister, and M. Weiler (2018). "Diel fluctuations of viscosity-driven riparian inflow affect streamflow DOC concentration". In: *Biogeosciences* 15.7, pp. 2177–2188. DOI: 10.5194/bg-15-2177-2018.

Segura, C., A. L. James, D. Lazzati, and N. T. Roulet (2012). "Scaling relationships for event water contributions and transit times in small-forested catchments in Eastern Quebec". In: *Water Resources Research* 48.7. DOI: 10.1029/2012WR011890.

Seibert, J., K. Bishop, L. Nyberg, and A. Rodhe (2011). "Water storage in a till catchment. I: Distributed modelling and relationship to runoff". In: *Hydrological Processes* 25.25, pp. 3937–3949. DOI: 10.1002/hyp.8309.

Seibert, J. and J. J. McDonnell (2002). "On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration". In: *Water Resources Research* 38.11, pp. 23-1-23–14. DOI: 10.1029/2001WR000978.

Shand, P., a. H. Haria, C. Neal, K. J. Griffiths, D. C. Gooddy, a. J. Dixon, T. Hill, D. K. Buckley, and J. E. Cunningham (2005). "Hydrochemical heterogeneity in

an upland catchment: further characterisation of the spatial, temporal and depth variations in soils, streams and groundwaters of the Plynlimon forested catchment, Wales". In: *Hydrology and Earth System Sciences* 9.6, pp. 621–644. DOI: 10.5194/hess-9-621-2005.

Sollins, P. and F. M. McCorison (1981). "Nitrogen and carbon solution chemistry of an old growth coniferous forest watershed before and after cutting". In: *Water Resources Research* 17.5, pp. 1409–1418. DOI: 10.1029/WR017i005p01409.

Solomon, D. K., R. J. Poreda, P. G. Cook, and A. Hunt (1995). "Site Characterization Using 3H/3 He Ground-Water Ages, Cape Cod, MA". In: *Groundwater* 33.6, pp. 988–996. DOI: 10.1111/j.1745-6584.1995.tb00044.x.

Solomon, D. K., D. P. Genereux, L. N. Plummer, and E. Busenberg (2010). "Testing mixing models of old and young groundwater in a tropical lowland rain forest with environmental tracers". In: *Water Resources Research* 46.4. DOI: 10.1029/2009WR008341.

Son, K. and M. Sivapalan (2007). "Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data". In: *Water Resources Research* 43.1. DOI: 10.1029/2006WR005032.

Soulsby, C., C. Birkel, J. Geris, J. Dick, C. Tunaley, and D. Tetzlaff (2015a). "Stream water age distributions controlled by storage dynamics and nonlinear hydrologic connectivity: Modeling with high-resolution isotope data". In: *Water Resources Research* 51.9, pp. 7759–7776. DOI: 10.1002/2015WR017888.

Soulsby, C., C. Birkel, J. Geris, and D. Tetzlaff (2015b). "Spatial aggregation of time-variant stream water ages in urbanizing catchments". In: *Hydrological Processes* 29.13, pp. 3038–3050. DOI: 10.1002/hyp.10500.

Soulsby, C., K. Piegat, J. Seibert, and D. Tetzlaff (2011). "Catchment-scale estimates of flow path partitioning and water storage based on transit time and runoff modelling". In: *Hydrological Processes* 25.25, pp. 3960–3976. DOI: 10.1002/hyp.8324.

Soulsby, C., D. Tetzlaff, and M. Hrachowitz (2009). "Tracers and transit times: windows for viewing catchment scale storage?" In: *Hydrological Processes* 23.24, SI, 3503–3507. DOI: 10.1002/hyp.7501.

Sprenger, M., H. Leistert, K. Gimbel, and M. Weiler (2016). "Illuminating hydrological processes at the soil-vegetation-atmosphere interface with water stable isotopes". In: *Reviews of Geophysics* 54.3, pp. 674–704. DOI: 10.1002/2015RG000515.

Sprenger, M., C. Stumpp, M. Weiler, W. Aeschbach, S. T. Allen, P. Benettin, M. Dubbert, A. Hartmann, M. Hrachowitz, J. W. Kirchner, J. J. McDonnell, N. Orlowski, D. Penna, S. Pfahl, M. Rinderer, N. Rodriguez, M. Schmidt, and C. Werner (2019). "The Demographics of Water: A Review of Water Ages in the Critical Zone". In: *Reviews of Geophysics* 0.0. DOI: 10.1029/2018RG000633.

Sprenger, M., D. Tetzlaff, J. Buttle, H. Laudon, and C. Soulsby (2018). "Water ages in the critical zone of long-term experimental sites in northern latitudes". In: *Hydrology and Earth System Sciences* 22.7, pp. 3965–3981. DOI: 10.5194/hess-22-3965-2018.

Stamoulis, K., K. Ioannides, P. Kassomenos, and A. Vlachogianni (2005). "Tritium Concentration in Rainwater Samples in Northwestern Greece". In: *Fusion Science and Technology* 48.1, pp. 512–515. DOI: 10.13182/FST05-A978.

Stewart, M. K., U. Morgenstern, M. A. Gusyev, and P. Małoszewski (2017). "Aggregation effects on tritium-based mean transit times and young water fractions in spatially heterogeneous catchments and groundwater systems". In: *Hydrology and Earth System Sciences* 21.9, pp. 4615–4627. DOI: 10.5194/hess-21-4615-2017.

Stewart, M. K., U. Morgenstern, J. J. McDonnell, and L. Pfister (2012). "The 'hidden streamflow' challenge in catchment hydrology: a call to action for stream water transit time analysis". In: *Hydrological Processes* 26.13, pp. 2061–2066. DOI: 10.1002/hyp.9262.

Stewart, M. K. and J. T. Thomas (2008). "A conceptual model of flow to the Waikoropupu Springs, NW Nelson, New Zealand, based on hydrometric and tracer (^{18}O, Cl,^{3}H and CFC) evidence". In: *Hydrology and Earth System Sciences* 12.1, pp. 1–19. DOI: 10.5194/hess-12-1-2008.

Stewart, M. K., J. Mehlhorn, and S. Elliott (2007). "Hydrometric and natural tracer (oxygen-18, silica, tritium and sulphur hexafluoride) evidence for a dominant groundwater contribution to Pukemanga Stream, New Zealand". In: *Hydrological Processes* 21.24, pp. 3340–3356. DOI: 10.1002/hyp.6557.

Stewart, M. K. and U. Morgenstern (2016). "Importance of tritium-based transit times in hydrological systems". In: *Wiley Interdisciplinary Reviews: Water* 3.2, pp. 145–154. DOI: 10.1002/wat2.1134.

Stewart, M. K., U. Morgenstern, and J. J. McDonnell (2010). "Truncation of stream residence time: how the use of stable isotopes has skewed our concept of streamwater age and origin". In: *Hydrological Processes* 24.12, pp. 1646–1659. DOI: 10.1002/hyp.7576.

Stockinger, M. P., H. R. Bogena, A. Lücke, B. Diekkrüger, T. Cornelissen, and H. Vereecken (2016). "Tracer sampling frequency influences estimates of young water fraction and streamwater transit time distribution". In: *Journal of Hydrology* 541, pp. 952–964. DOI: 10.1016/j.jhydrol.2016.08.007.

Stockinger, M. P., A. Lücke, J. J. McDonnell, B. Diekkrüger, H. Vereecken, and H. R. Bogena (2015). "Interception effects on stable isotope driven streamwater transit time estimates". In: *Geophysical Research Letters* 42.13, pp. 5299–5308. DOI: 10.1002/2015GL064622.

Stolp, B. J., D. K. Solomon, A. Suckow, T. Vitvar, D. Rank, P. K. Aggarwal, and L. F. Han (2010). "Age dating base flow at springs and gaining streams using helium-3 and tritium: Fischa-Dagnitz system, southern Vienna Basin, Austria". In: *Water Resources Research* 46.7. DOI: 10.1029/2009WR008006.

Stumpp, C., J. Klaus, and W. Stichler (2014). "Analysis of long-term stable isotopic composition in German precipitation". In: *Journal of Hydrology* 517, pp. 351–361. DOI: 10.1016/j.jhydrol.2014.05.034.

Stumpp, C., G. Nützmann, S. Maciejewski, and P. Maloszewski (2009). "A comparative modeling study of a dual tracer experiment in a large lysimeter under atmospheric conditions". In: *Journal of Hydrology* 375.3-4, pp. 566–577. DOI: 10.1016/j.jhydrol.2009.07.010.

Swanson, F. J. and M. E. James (1967). *Geology and geomorphology of the H. J. Andrews Experimental Forest, western Cascades, Oregon*. Res. Pap. PNW-188. Pacific Northwest Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture.

Ter Braak, C. J. F. and J. A. Vrugt (2008). "Differential Evolution Markov Chain with snooker updater and fewer chains". In: *Statistics and Computing* 18.4, pp. 435–446. DOI: 10.1007/s11222-008-9104-9.

Tetzlaff, D., C. Birkel, J. Dick, J. Geris, and C. Soulsby (2014). "Storage dynamics in hydropedological units control hillslope connectivity, runoff generation, and the evolution of catchment transit time distributions". In: *Water Resources Research* 50.2, pp. 969–985. DOI: 10.1002/2013WR014147.

Tetzlaff, D., J. Seibert, K. J. McGuire, H. Laudon, D. A. Burns, S. M. Dunn, and C. Soulsby (2009). "How does landscape structure influence catchment transit time

across different geomorphic provinces?" In: *Hydrological Processes* 23.6, pp. 945–953. DOI: 10.1002/hyp.7240.

Thiesen, S., P. Darscheid, and U. Ehret (2019). "Identifying rainfall-runoff events in discharge time series: a data-driven method based on information theory". In: *Hydrology and Earth System Sciences* 23.2, pp. 1015–1034. DOI: 10.5194/hess-23-1015-2019.

Thornthwaite, C. W. (1948). "An Approach toward a Rational Classification of Climate". In: *Geographical Review* 38.1, pp. 55–94. DOI: 10.2307/210739.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009). "Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis". In: *Water Resources Research* 45.12. DOI: 10.1029/2008WR006825.

Timbe, E., D. Windhorst, P. Crespo, H.-G. Frede, J. Feyen, and L. Breuer (2014). "Understanding uncertainties when inferring mean transit times of water trough tracer-based lumped-parameter models in Andean tropical montane cloud forest catchments". In: *Hydrology and Earth System Sciences* 18.4, pp. 1503–1523. DOI: 10.5194/hess-18-1503-2014.

Turner, J. V., D. K. Macpherson, and S. R. A. (1987). "The mechanisms of catchment flow processes using natural variations in deuterium and oxygen-18". In: *Journal of Hydrology* 94.1. Hydrology and Salinity in the Collie River Basin, Western Australia, pp. 143–162. DOI: 10.1016/0022-1694(87)90037-0.

Turner, J. V. and C. J. Barnes (1998). "Chapter 21 - Modeling of Isotope and Hydrogeochemical Responses in Catchment Hydrology". In: *Isotope Tracers in Catchment Hydrology*. Ed. by C. Kenfall and J. J. McDonnell. Amsterdam: Elsevier, pp. 723–760. DOI: 10.1016/B978-0-444-81546-0.50028-8.

Turner, J. V. and D. K. Macpherson (1990a). "Mechanisms Affecting Streamflow and Stream Water Quality: An Approach via Stable Isotope, Hydrogeochemical,and Time Series Analysis". In: *Water Resources Research* 26.12, pp. 3005–3019. DOI: 10.1029/WR026i012p03005.

— (1990b). "Mechanisms Affecting Streamflow and Stream Water Quality: An Approach via Stable Isotope, Hydrogeochemical,and Time Series Analysis". In: *Water Resources Research* 26.12, pp. 3005–3019. DOI: 10.1029/WR026i012p03005.

Uhlenbrook, S., M. Frey, C. Leibundgut, and P. Maloszewski (2002). "Hydrograph separations in a mesoscale mountainous basin at event and seasonal timescales". In: *Water Resources Research* 38.6, pp. 31-1-31–14. DOI: 10.1029/2001WR000938.

Utermann, J., E. J. Kladivko, and W. A. Jury (1990). "Evaluating Pesticide Migration in Tile-Drained Soils with a Transfer Function Model". In: *Journal of Environmental Quality* 19. 4, pp. 707–714. DOI: 10.2134/jeq1990.00472425001900040013x.

Vaché, K. and J. McDonnell (2006). "A process-based rejectionist framework for evaluating catchment runoff model structure". In: *Water Resources Research* 42.2. DOI: 10.1029/2005WR004247.

Van der Velde, Y., G. H. de Rooij, J. C. Rozemeijer, F. C. van Geer, and H. P. Broers (2010). "Nitrate response of a lowland catchment: On the relation between stream concentration and travel time distribution dynamics". In: *Water Resources Research* 46. DOI: 10.1029/2010WR009105.

Van der Velde, Y., P. J. J. F. Torfs, S. E. A. T. M. van der Zee, and R. Uijlenhoet (2012). "Quantifying catchment-scale mixing and its effect on time-varying travel time distributions". In: *Water Resources Research* 48. W06536. DOI: 10.1029/2011WR011310.

Van der Velde, Y., I. Heidbüchel, S. W. Lyon, L. Nyberg, A. Rodhe, K. Bishop, and P. A. Troch (2015). "Consequences of mixing assumptions for time-variable travel

time distributions". In: *Hydrological Processes* 29.16, pp. 3460–3474. DOI: 10.1002/hyp.10372.

Van Huijgevoort, M. H. J., D. Tetzlaff, E. H. Sutanudjaja, and C. Soulsby (2016). "Using high resolution tracer data to constrain water storage, flux and age estimates in a spatially distributed rainfall-runoff model". In: *Hydrological Processes*, n/a–n/a. DOI: 10.1002/hyp.10902.

Van Meerveld, H. J. I., J. W. Kirchner, M. J. P. Vis, R. S. Assendelft, and J. Seibert (2019). "Expansion and contraction of the flowing stream network changes hillslope flowpath lengths and the shape of the travel time distribution". In: *Hydrology and Earth System Sciences Discussions* 2019, pp. 1–18. DOI: 10.5194/hess-2019-218.

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". In: *Hydrological Processes* 28.24, pp. 5805–5816. DOI: 10.1002/hyp.10086.

Van Verseveld, W. J., H. R. Barnard, C. B. Graham, J. J. McDonnell, J. R. Brooks, and M. Weiler (2017). "A sprinkling experiment to quantify celerity–velocity differences at the hillslope scale". In: *Hydrology and Earth System Sciences* 21.11, pp. 5891–5910. DOI: 10.5194/hess-21-5891-2017.

Visser, A., M. Thaw, A. Deinhart, R. Bibby, M. Safeeq, M. Conklin, B. Esser, and Y. Van der Velde (2019). "Cosmogenic Isotopes Unravel the Hydrochronology and Water Storage Dynamics of the Southern Sierra Critical Zone". In: *Water Resources Research* 55.2, pp. 1429–1450. DOI: 10.1029/2018WR023665.

Visser, A., H. P. Broers, R. Purtschert, J. Sültenfuß, and M. de Jonge (2013). "Groundwater age distributions at a public drinking water supply well field derived from multiple age tracers (85Kr, 3H/3He, and 39Ar)". In: *Water Resources Research* 49.11, pp. 7778–7796. DOI: 10.1002/2013WR014012.

Volkmann, T. H. M., A. Sengupta, L. A. Pangle, K. Dontsova, G. A. Barron-Gafford, C. J. Harman, G.-Y. Niu, L. K. Meredith, N. Abramson, A. A. M. Neto, Y. Wang, J. R. Adams, D. D. Breshears, A. Bugaj, J. Chorover, A. Cueva, S. B. DeLong, M. Durcik, T. P. A. Ferre, E. A. Hunt, T. E. Huxman, M. Kim, R. M. Maier, R. K. Monson, J. D. Pelletier, M. Pohlmann, C. Rasmussen, J. Ruiz, S. R. Saleska, M. G. Schaap, M. Sibayan, M. Tuller, J. L. M. van Haren, X. Zeng, and P. A. Troch (2018). "Controlled Experiments of Hillslope Coevolution at the Biosphere 2 Landscape Evolution Observatory: Toward Prediction of Coupled Hydrological, Biogeochemical, and Ecological Change". In: *Hydrology of Artificial and Controlled Experiments*. Ed. by J.-F. Liu and W.-Z. Gu. Rijeka: IntechOpen. Chap. 2. DOI: 10.5772/intechopen.72325.

Von Freyberg, J., S. T. Allen, S. Seeger, M. Weiler, and J. W. Kirchner (2018). "Sensitivity of young water fractions to hydro-climatic forcing and landscape properties across 22 Swiss catchments". In: *Hydrology and Earth System Sciences* 22.7, pp. 3841–3861. DOI: 10.5194/hess-22-3841-2018.

Von Freyberg, J., B. Studer, and J. W. Kirchner (2017). "A lab in the field: high-frequency analysis of water quality and stable isotopes in stream water and precipitation". In: *Hydrology and Earth System Sciences* 21.3, pp. 1721–1739. DOI: 10.5194/hess-21-1721-2017.

Vrugt, J., C. T. Braak, C. Diks, B. Robinson, J. Hyman, and D. Higdon (2009). "Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling". In: *International Journal of Nonlinear Sciences and Numerical Simulation* 10.3, pp. 271–288. DOI: 10.1515/IJNSNS.2009.10.3.273.

Vrugt, J. A. (2016). "Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation". In: *Environmental Modelling and Software* 75, pp. 273–316. DOI: 10.1016/j.envsoft.2015.08.013.

Wagener, T., M. Sivapalan, P. Troch, and R. Woods (2007). "Catchment Classification and Hydrologic Similarity". In: *Geography Compass* 1.4, pp. 901–931. DOI: 10.1111/j.1749-8198.2007.00039.x.

Waichler, S. R., B. C. Wemple, and M. S. Wigmosta (2005). "Simulation of water balance and forest treatment effects at the H.J. Andrews Experimental Forest". In: *Hydrological Processes* 19.16, pp. 3177–3199. DOI: 10.1002/hyp.5841.

Ward, A. S., N. M. Schmadel, and S. M. Wondzell (2018). "Time-Variable Transit Time Distributions in the Hyporheic Zone of a Headwater Mountain Stream". In: *Water Resources Research* 54.3, pp. 2017–2036. DOI: 10.1002/2017WR021502.

Weissmann, G. S., Y. Zhang, E. M. LaBolle, and G. E. Fogg (2002). "Dispersion of groundwater age in an alluvial aquifer system". In: *Water Resources Research* 38.10, pp. 16-1-16–13. DOI: 10.1029/2001WR000907.

White, R. E., J. S. Dyson, R. A. Haigh, W. A. Jury, and G. Sposito (1986). "A Transfer Function Model of Solute Transport Through Soil: 2. Illustrative Applications". In: *Water Resources Research* 22.2, pp. 248–254. DOI: 10.1029/WR022i002p00248.

Wilusz, D. C., C. J. Harman, and W. P. Ball (2017). "Sensitivity of Catchment Transit Times to Rainfall Variability Under Present and Future Climates". In: *Water Resources Research*. in press. DOI: 10.1002/2017WR020894.

Wilusz, D., C. J. Harman, W. Ball, R. Maxwell, and A. Buda (in review). "Using particle tracking to understand catchment flow paths, age distributions, and the paradoxical origins of the inverse storage effect". In:

Wrede, S., F. Fenicia, N. Martínez-Carreras, J. Juilleret, C. Hissler, A. Krein, H. H. G. Savenije, S. Uhlenbrook, D. Kavetski, and L. Pfister (2015). "Towards more systematic perceptual model development: a case study using 3 Luxembourgish catchments". In: *Hydrological Processes* 29.12, pp. 2731–2750. DOI: 10.1002/hyp.10393.

Yang, J., I. Heidbüchel, A. Musolff, F. Reinstorf, and J. H. Fleckenstein (2018). "Exploring the Dynamics of Transit Times and Subsurface Mixing in a Small Agricultural Catchment". In: *Water Resources Research* 54.3, pp. 2317–2335. DOI: 10.1002/2017wr021896.

Zal, N., C. Whalley, T. Christiansen, P. Kristensen, and F. Néry (2018). *European waters - assessment of status and pressures 2018*. European Environmental Agency (EEA), pp. 1–92. DOI: 10.2800/303664.

Zehe, E., H. Elsenbeer, F. Lindenmaier, K. Schulz, and G. Blöschl (2007). "Patterns of predictability in hydrological threshold systems". In: *Water Resources Research* 43.7. DOI: 10.1029/2006WR005589.

Zehe, E. and C. Jackisch (2016). "A Lagrangian model for soil water dynamics during rainfall-driven conditions". In: *Hydrology and Earth System Sciences* 20.9, pp. 3511–3526. DOI: 10.5194/hess-20-3511-2016.