Water Resources Research

Catchment Functioning Under Prolonged Drought Stress: Tracer-Aided Ecohydrological Modeling in an Intensively Managed Agricultural Catchment

Xiaoqiang Yang[1,2] ![ORCID], Doerthe Tetzlaff[1,3,4] ![ORCID], Chris Soulsby[1,4], Aaron Smith[1], and Dietrich Borchardt[2] ![ORCID]

[1]Department of Ecohydrology, Leibniz Institute of Freshwater Ecosystem and Inland Fisheries – IGB, Berlin, Germany, [2]Department of Aquatic Ecosystem Analysis and Management, Helmholtz Centre for Environmental Research – UFZ, Magdeburg, Germany, [3]Department of Geography, Humboldt University of Berlin, Berlin, Germany, [4]Northern Rivers Institute, School of Geosciences, University of Aberdeen, Aberdeen, UK

Abstract High spatial heterogeneity of catchment properties enhances the variability of ecohydrological responses to changing natural and anthropogenic conditions, like the European-wide droughts in 2018–2019. Based on new adaptations of a tracer-aided, process-based ecohydrological model (EcH$_2$O-iso), we investigated drought-induced nonstationary ecohydrological behavior in a small agricultural headwater catchment (1.44 km^2) in central Germany. Multiple environmental time-series helped inform various aspects of catchment functioning that have been impacted by agricultural activity and changing climate conditions and helped to further constrain model calibration. Multi-criteria calibration showed that data collected during drought years were highly informative in reproducing the changes in stream water dynamics. Further, inclusion of δ^2H and δ^{18}O data was valuable for reducing model uncertainty and increasing confidence in simulations of green- and blue-water flux partitioning and storage-flux-age interactions. Using the best-performing calibrations, we further analyzed the high spatiotemporal variability of internal ecohydrological processes and the varying responses of fluxes and associated water ages to prolonged drought stress. Under drought conditions, modeled stream runoff contributed from deeper, older storages increased significantly after a particularly wet season, resulting in a sharp increase in stream water age. Unlike relatively minor changes in soil evaporation, seasonally typical transpiration fluxes were initially maintained in April–June but dramatically decreased as the drought further developed in July–September. Importantly, the tracer-based transpired water age was much older after April, providing a potential indicator of drought impacts and the need for precautionary management responses. Our findings are important for similar agricultural headwater ecosystems in other drought-sensitive regions.

1. Introduction

Recent extreme droughts in Europe have provided strong research foci in the hydrological community to better understand and forecast periods of water stress (Albergel et al., 2020; Bakke et al., 2020; Kleine et al., 2020; Smith et al., 2020a). In line with long-term climate change projections, the extreme drought in 2018–2019 provided a unique opportunity to investigate expected regional climate change impacts both in terms of monitoring (e.g., the TERENO observatories in Germany [Heinrich et al., 2019; Wollschläger et al., 2016]) and modeling (e.g., Samaniego et al., 2018; Smith et al., 2020a) flux-storage dynamics. As drought impacts propagate through a catchment's ecohydrological cycle (Wilhite & Glantz, 1985), increasingly dry conditions (i.e., a meteorological drought of reduced precipitation and high temperatures) typically deplete soil water storage, gradually supressing "green" (i.e., soil evaporation and vegetation transpiration) and "blue" (i.e., groundwater recharge and runoff generation) water fluxes. However, resilience to droughts and subsequent recovery patterns differ for different compartments of the ecohydrological cycle. Orth and Destouni (2018) showed that blue water fluxes generally decrease more strongly and rapidly than green water fluxes as soil moisture deficits increase. In prolonged droughts, impacts propagate to aquifers, and it generally takes a longer time for recharge fluxes to recover during rewetting periods (Orth & Destouni, 2018). Of course, the general patterns vary under specific physiographic and climate conditions. In regions where surface waters are substantially fed by deeper aquifer storage, blue water fluxes likely

show greater resilience but longer recovery than green water fluxes under severe droughts (e.g., Sawada et al., 2014). The resilience and recovery can also be affected by different combinations of vegetation and soil properties (Smith et al., 2020a).

Therefore, a comprehensive understanding of catchment ecohydrological functioning and its spatio-temporal dynamics is of particular importance to maintaining ecosystem services provisioned by hydrological processes and to assess potential impacts of extreme climate conditions like droughts. The interaction of human and ecohydrological systems is especially strong in intensively managed agricultural catchments (Porporato et al., 2015). Humans have managed water, energy, and nutrient cycling (through irrigation, fertilizer application, tile drainage systems, etc.) to improve agro-ecosystem productivity, while potentially reducing biodiversity and resilience to extreme external disturbances (Lin, 2011; Schmidhuber & Tubiello, 2007; Weise et al., 2020). Ecohydrology-based modeling approaches create opportunities to inform guidelines for sustainable use of water and land resources in agricultural systems (Porporato et al., 2015) and, more generally, to advance our understanding of plant-water-nutrient interactions (Newman et al., 2006).

Spatially distributed, process-based ecohydrological models are valuable tools to address these knowledge gaps. Such model conceptualizations seek to balance model complexity and process representations (Clark et al., 2017; Tetzlaff et al., 2008). For a reliable representation of internal function, process-based models should be carefully parameterized and calibrated against a wide range of observations (i.e., using both "hard" and "soft" information [Seibert & McDonnell, 2002; Winsemius et al., 2009]). Large variations in information content, as well as associated uncertainty, reside in different data types and different periods of data collection (Beven & Smith, 2015; Westerberg & McMillan, 2015). Therefore, multi-criteria calibration based on contrasting data sources at different scales is important, especially for spatially distributed models (Fatichi et al., 2016; Holmes et al., 2020; Kuppel et al., 2018a). In particular, information on stable isotopes

of water (deuterium and oxygen-18 ratios; δ^2H and δ^{18}O, respectively) has been increasingly integrated in ecohydrological modeling, from which storage-flux interactions, precipitation partitioning and surface water dynamics can be tracked and verified (Birkel & Soulsby, 2015; McGuire & McDonnell, 2015; Tetzlaff et al., 2015). Water ages of different storages and fluxes are additionally of interest in describing the quantity and quality of waters traveling through ecohydrological systems (Porporato & Calabrese, 2015; Sprenger et al., 2019). Therefore, tracer-aided ecohydrological modeling can better resolve the quantification of the partitioning of blue and green water fluxes and help evaluate impacts of disturbances from changing climatic and anthropogenic conditions (Birkel et al., 2014).

The EcH$_2$O-iso model (Kuppel et al., 2018b; Maneta & Silverman, 2013) explicitly integrates energy balances, water transfers, and vegetation dynamics, following the basic philosophy of providing a physically based, yet computationally efficient process representation. The EcH$_2$O-iso model acts as a "middle path" (as termed in Kuppel et al., 2018b) for linking spatio-temporal water dynamics with conservative tracer transport, which can be applied from the plot-scale soil-vegetation-atmosphere continuum to regional-scale

earth system modeling. The model tracks stable isotopes of water (δ^2H and δ^{18}O) and water ages in each ecohydrological compartment, and therefore, it provides a more complete picture capturing both water particle transport (velocity) and the speed of catchment response (celerity). The model has been applied and verified at both plot scale (Douinot et al., 2019; Knighton et al., 2020; Smith et al., 2020a) and catchment scale (Kuppel et al., 2018a, 2018b; Lozano-Parra et al., 2014; Maneta & Silverman, 2013; Smith et al., 2020b).

In this study, we make a fundamental adaptation to the model for the Schäfertal catchment, an intensively cultivated headwater located in the Harz mountains of the German Central Uplands, which is highly sensitive to climate variability (Anis & Rode, 2015). As a long-term experimental site (Wollschläger et al., 2016), the Schäfertal catchment has been part of multiple observatory projects collecting a range of ecohydrological variables (e.g., the International Hydrological Decade of UNESCO since 1960s, the Terrestrial Environmental Observatories [TERENO] and the follow-up Modular Observation Solutions for Earth Systems [MOSES]). Intensive human activities have heavily affected soil water (e.g., due to tile drainage) and groundwater level dynamics (due to nearby mining activities). In addition, hydroclimatic conditions have changed dramatically, leading to severe droughts in recent years (The German Weather Service, https://www.dwd.de/EN/press/press_node.html). We investigate the following questions:
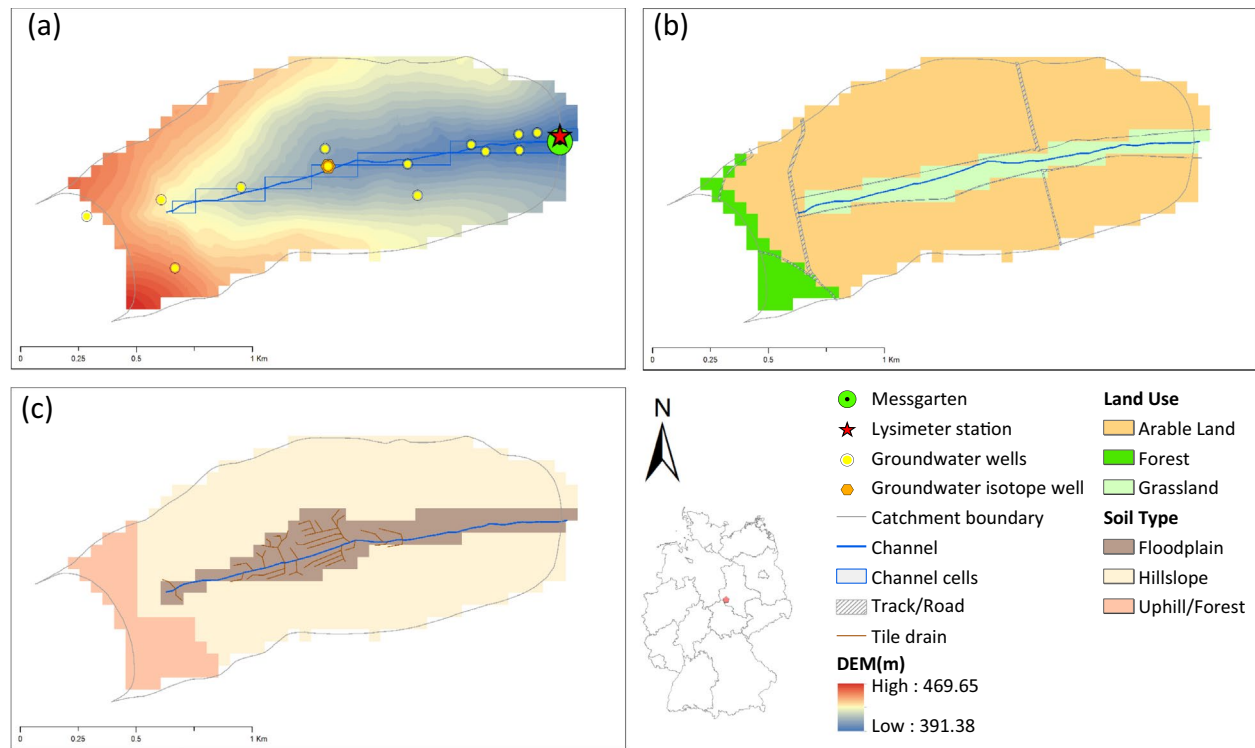
Figure 1. The Schäfertal catchment and monitoring locations ((a) DEM and monitoring sites, (b) Land use types, and (c) soil types and the tile drain network). A V-notch weir is also installed in the Messgarten climate station for discharge measuring and stream water sampling. The lysimeter station for soil moisture data is located ca. 100 m north-east of the Messgarten.

(1) How can model structure be adapted to conceptualize anthropogenic activities that impact catchment ecohydrological functioning?

(2) How valuable is the information brought by different calibration data, particularly stable isotopes of water as conservative tracers?

(3) Can a multi-criteria calibration of a process-based (ecohydrological) model reproduce catchment responses of flow and stable isotopes of water and capture internal catchment dynamics of water fluxes, storages, and ages?

(4) Particularly, what is the value of data collected under drought stress in terms of modeling ecohydrological responses under changing conditions?

The unique characteristics and data sets of the Schäfertal catchment provide the opportunity to investigate responses of catchment functions to the changing anthropogenic and climatic conditions. Through tracer-aided ecohydrological modeling, we aim to assess how process-based modeling can be adapted to capture the changing catchment functioning and quantify the spatiotemporal variability of each ecohydrological compartment under the drought stress.

2. Study Site and Data

2.1. The Schäfertal Catchment

The Schäfertal catchment is an agricultural headwater catchment (1.44 km^2, Figure 1), located in the lower Harz Mountains in central Germany. Elevation ranges from 391 to 470 m. The north- and south-facing hillslopes, with an average slope of 11°, are intensively cultivated (crops rotated among winter wheat, triticale, rapeseed, etc.); the middle valley bottom is dominated by grassland with channels; the upstream hilltop is occupied by sparse forest (Figure 1b). The underlying Paleozoic greywacke and Devonian shale are covered by periglacial layers with different fractions of loess and rock fragments, resulting in complex geomorphological structures through the soil profile (Kistner et al., 2013).

Field soil measurements in the hillslopes (Altermann & Mautschke, 1970; Graeff et al., 2009) indicate relatively high porosity and hydraulic conductivity (e.g., $9.95 \times 10^{-6} \text{ m s}^{-1}$) in the top soil layer, and a base layer at depth deeper than circa 0.4 m with lower hydraulic conductivity (e.g., $2.31 \times 10^{-7} \text{ m s}^{-1}$; Graeff et al., 2009). Soil properties are generally homogenous based on the German soil database (BGR, 2018) and field soil core sampling (Schrön, 2017), while a certain degree of spatial variability is due to the detailed topographic characteristics (e.g., reported in Ollesch et al., 2005; Anis & Rode, 2015). Moreover, artificial tile drains are extensive throughout the central valley bottom (Figure 1c). Here, we refined the soil map from previous studies into three types, that is, "Floodplain" soils for Fluvisols and Gleysols with artificial drainage in the valley bottom, "Hillslope" soils for Gleyic Cambisols and Luvisols in the agricultural hillslopes, and "uphill" soils for Rankers and Cambisols in the uphill and forest areas (Figure 1c).

Being located in the sheltered leeward regions of the Harz Mountains (peaked at 1,141 m a.s.l), as well as the transition area of the German Northern Plain and the Central Uplands, the catchment exhibits a prevailing continental climate, and therefore, is vulnerable to climatic variability, including summer droughts (Anis & Rode, 2015; Wollschläger et al., 2016). Long-term annual precipitation is 626 mm (averaged in 1969–2019), and average air temperature is 6.7°C (ranging from −1.4°C in January to 15.3°C in June) with significant snowmelt in spring (Ollesch et al., 2005).

In the study years 2010–2019, annual precipitation decreased to an average value of 538 mm (significantly lower than that of 1968–2009; ANOVA test, $p < 0.05$); monthly mean air temperature significantly increased in August–December (seasonal Mann-Kendall test using the *R "trend"* package (Pohlert, 2020), $p < 0.05$), although the annual average was slightly increased (7.3°C). Moreover, the catchment experienced recent prolonged climatic drought stress starting from 2017: the average annual precipitation was only 67.5% of the long-term average, with a historically low value occurring in 2018 (335 mm); the total number of warm days (i.e., the maximum temperature exceeding 30°C) was 28 days in 2018–2019, compared to four days per year in 2010–2017.

The catchment has been heavily impacted by a legacy of anthropogenic activities. Tile drainage networks were constructed at ca. 1 m depth in the soils of central valley bottoms in 1972 (Figure 1c). This has largely impacted soil properties and soil water dynamics, which also results in land-use change from previous pasture to arable land since 1979 (Ollesch et al., 2010). Mining activities were also historically undertaken in areas surrounding the catchment (started from 1970s and stopped since 1991), and de-watering had affected groundwater levels that were artificially lowered up to 4.7 m below land surface in the valley bottom. Long-term discharge and groundwater table records (1968–2019) showed that the groundwater level has been mostly recovered since 2009 (Ollesch et al., 2010), and surface discharge has been continuously observed, even in recent drought years.

2.2. Multi-Source Data Collection and Analysis

The Schäfertal catchment has been intensively monitored since 1960s by the University of Applied Science Magdeburg-Stendal (HS-MD), including a climate station ("Messgarten," Figure 1a), discharge gauging station and a dense groundwater observation network (Reinstorf et al., 2010). Since 2010, it has further been part of the TERENO/MOSES observatories, as one of the headwater research sites for advancing understandings of ecosystem processes and catchment functions (Wollschläger et al., 2016), including the implementation of a point-scale lysimeter station (Pütz et al., 2016).

We used data from multiple monitoring projects at different scales (Table 1). Daily time series were derived from measurements with varying temporal resolutions. Daily meteorological forcing data (i.e., precipitation, air temperature [mean, minimum, and maximum], mean relative humidity, mean wind speed, and global radiation) were aggregated from 1-min interval measurements at the Messgarten climate and hydrological station (Table 1). Mean daily longwave radiation data were obtained from German Weather Service (DWD) climate stations Dresden-Klotzsche (51.12°N, 13.75°E) and Leipzig-Holzhausen (51.31°N, 12.44°E). Streamflows were gauged at a V-notch weir installed in the Messgarten, where water levels were originally recorded (the "SEBAPlus 15" radar sensor, SEBA Hydrometrie GmbH) and converted into discharge (according to the device-specific rating relation). Note that some implausible discharge observations in specific

Table 1
Summary of the Collected Multiple Data Sources

Data category	Data type	Description and processing	Source
Meteorology	Precipitation (P) Air temperature (Tair) Relative Humidity (RH) Wind Speed (WS) Shortwave Radiation (SR)	Original data were measured at 1-min interval in 2010–2019, from which the daily time series were calculated (summation for P; means for RH, WS, SR; daily mean, max and min for Tair)	Messgarten station, TERENO (HS-MD, UFZ)
	Longwave Radiation	Mean daily values from nearby DWD stations Dresden-Klotzsche (51.12°N, 13.75°E) and Leipzig-Holzhausen (51.31°N, 12.44°E)	German Weather Service (DWD)
Streamflow	Discharge	1-h interval in 2010–2012; 10-min interval in 2013–2019	TERENO (HS-MD, UFZ)
Stable isotopes of water (^{2}H and ^{18}O)	Precipitation	Monthly in 2013–2017/10; Daily weekly in 2017/11–2019	Messgarten station, TERENO (HS-MD, UFZ)
	Stream water	Daily weekly in 2017/11–2019	
	Groundwater	2019/07–2019/11; Groundwater well 61 (51.65°N, 11.03°E)	
Plot-scale	Soil moisture content	10-min interval in 2012–2019 TDR sensor at 10, 30, and 50 cm depth	Lysimeter station TERENO-SoilCan, UFZ
	Soil temperature	5-min interval in 2010–2019, Measured at ground surface, 10, 20, and 30 cm depth	Messgarten station, TERENO (HS-MD, UFZ)
Long-term	Precipitation, air temperature and discharge	Daily data (Data Set S1)	HS-MD, Anis and Rode (2015)

Abbreviations: P, Precipitation; Tair, Air temperature; RH, Relative Humidity; WS, Wind Speed; SR, Shortwave Radiation; TERENO, Terrestrial Environmental Observatories.

short periods (e.g., the extremely low-value period after a device failure in September 2015), were retained since no clear correction notes were documented. Soil temperature and soil moisture at different depths were collected from the measuring points in the Messgarten station and the lysimeter station, respectively (Table 1). Additionally, groundwater water level data were collected from several groundwater wells distributed within the catchment (Figure S1). Long-term daily time series (1968–2019) of precipitation, air temperature and discharge were also used to assess long-term changes of climate condition and anthropogenic activities and their impacts on the catchment functioning (supporting information Data Set S1).

Rainfall and stream water samples for stable isotopes of water have been sampled at Messagarten station (Figure 1a) and analyzed using a laser-based analyzer (analytical precision of 1.0‰ and 0.3‰ for δ^2H and δ^{18}O, respectively [Lutz et al., 2018]) by the Department of Catchment Hydrology, Helmholtz Center for Environmental Research - UFZ. We used two available data sets: the monthly precipitation data in 2013–2017/10 (collected from an evaporation-free collector, PALMEX; see Lutz et al., 2018 for sampling details), and the daily data in 2017/11–2019/08 (sampled from both precipitation and stream water). The regressed local meteoric water line (LMWL) was $\delta^2 H = 7.2 \cdot \delta^{18} O + 2.1$ ($R^2 = 0.96$, Figure S2). Continuous daily δ^2H and δ^{18}O in precipitation was generated differently according to the two isotopic data sets. For 2010–2017/10, daily values were randomly generated based on the normal distributions (i.e., mean and standard deviation were calculated for each month from the monthly data set considering seasonal variations). For 2017/11–2019, missing values were filled with the values of the next measured time points if the gap period is shorter than 7 days, otherwise daily values were randomly generated based on the "daily" data set (Figure S3). Additional groundwater isotopes were also sampled from a groundwater well ID 61 (Figure 1a) during 2019/07–2019/11.

3. Methods

3.1. The EcH₂O-iso Model

The fully distributed ecohydrological model EcH$_2$O (Maneta & Silverman, 2013) integrates vertical energy balances, vertical and lateral water transfers and vegetation dynamics. The model formed the platform for further module development for tracing water isotopes and water ages (i.e., the tracer-aided EcH$_2$O-iso model; Kuppel et al., 2018b). Here, we provide a brief description of the model conceptualizations and tracer/age tracking in different ecohydrological compartments (for comprehensive model description, please refer to the original publications Maneta & Silverman, 2013 and Kuppel et al., 2018b).

The model resolves each grid cell in a vertical sequence through the canopy, surface, and soil levels. The energy balance is solved first for each vegetation type at the canopy and land surface levels, from which the evapotranspiration-related latent heat fluxes are calculated. Canopy and soil evaporation are determined by the latent heat at the canopy level and the first soil layer, respectively, while constrained by available water storages (i.e., the maximum canopy storage and the difference between first layer soil porosity and residual moisture, respectively). Transpiration associated latent heat is controlled by the canopy temperature and the canopy conductance. A Jarvis-type stomatal conductance model is used, considering vegetation-dependent maximum stomatal conductance and the sensitivity of vapor pressure deficit at the leaf surface, light availability, and root-zone averaged soil moisture availability. Transpiration is contributed from all three layers and partitioned according to the water availability and the root proportion of each soil layer. The land surface energy balances (including surface net radiation, ground heat flux, and evaporation related latent heat flux; see Equations in Maneta & Silverman, 2013) are formed as nonlinear functions of surface skin temperature, which are solved using the Newton-Raphson method. The ground heat flux calculation is assumed to have two thermal layers along the soil profile, with no diurnal temperature variations below a damping depth (Liebethal & Foken, 2007). The soil temperature is therefore represented by the temperature at the damping depth.

Water balances at all levels are conceptualized by a bucket-type approach (Figure S4). Precipitation is split into intercepted water and throughfall based on the maximum canopy interception storage. All intercepted water can only be evaporated, without further interactions with surface and subsurface storages. Upon exceeding the canopy storage capacity, throughfall enters the surface water storage (snowpack and ponding water) where snowmelt and infiltration processes are considered. Surface water infiltration is calculated using the Green-Ampt method (Mein & Larson, 1973). Gravitational soil water (i.e., soil water content above field capacity) further infiltrates into the underlying layers, with a leakage parameter controlling the vertical seepage that leaves the system. In the third layer, subsurface lateral flow exchanges are estimated using a kinematic-wave based scheme (Equation S1, Text S1). Because of this, up-ward return flows can occur if deeper soil layers become saturated. The remaining surface-ponded water is considered as surface flow for lateral transport to the down-slope cell and further re-infiltration. Both lateral flow components are transported to the down-slope cell, following the topographic drainage network. In grid cells where stream channels are present, all surface ponding water becomes surface runoff, and subsurface runoff generation is controlled by a seepage parameter (soil-type dependent), saturated hydraulic conductivity, and gravitational soil water storage at the third layer (Equation S2, Text S1). Stream surface water routing along the stream network is calculated by solving the kinematic-wave equation.

Ecohydrological fluxes and storage-fluxes exchanges are tracked by isotopes and water ages (Kuppel et al., 2018b). For each grid cell, the tracer (i.e., δ^2H, δ^{18}O and ages) module is fully integrated into the calculations of each water flux and each water storage, following the full-mixing assumption. Specific implementations are listed as follows:

- The throughfall water is assumed to have the same isotopic signature as the same time step precipitation and the age of zero.
- Isotopic compositions and ages of canopy and soil evaporations are equivalent to those in the canopy and first-layer soil water storages, respectively.
- Isotopic fractionation due to soil evaporation is considered using the Craig-Gordon model (Craig & Gordon, 1965), adjusted for saturation in a porous medium (see details in Kuppel et al., 2018b).

- Transpiration is considered as a nonfractionating process, meaning full conservative mixing of transpired water contribution from each soil layer (according to the root fractions).
- Water age is calculated primarily following the same principle as that of conservative isotopic tracers (without fractionation process). Ages in all water storages increase by one time-step after each modeling time step, and reset to zero if the storage is empty.

3.2. Model Adaptations for Tile Drainage and Deeper Groundwater Dynamics

As previously mentioned, the lower boundary of EcH$_2$O-iso has been set as that of the third soil layer, meaning that deeper groundwater dynamics have largely been excluded in previous applications. Such implementations are sufficient for regions with humid climate conditions and strong topographic gradients where shallow groundwater plays a crucial role. This was found to be unsuitable for catchments where stream surface waters are considerably impacted by deeper groundwater dynamics, like the Schäfertal catchment (Graeff et al., 2009; Ollesch et al., 2010). Moreover, the shallow depth of tile drains in the Schäfertal catchment (ca. 1 m depth) constrains subsurface lateral fluxes to the depth of tile drains, which precludes the model's consideration of deeper water dynamics. Therefore, we fundamentally changed the model configuration/parameterization to represent tile drainage impacts and added an extra groundwater storage to conceptualize the deeper groundwater dynamics.

First, the total depth of soil was set-up as 1 m to compensate the approximate depth of tile drainage and the EcH$_2$O design that subsurface lateral fluxes occur only at the bottom of the third soil layer. This set-up has also been supported by the groundwater level data from near stream wells (e.g., well ID 61, Figure S1), showing that the near stream groundwater levels mostly fluctuated within 1 m below ground surface. Second, instead of specific governing equations for tile drainage, we conventionally considered the impacts based on parameter adaptation. The overall impacts on catchment drainage were reflected by a higher horizontal hydraulic conductivity, compared to field soil measurements (e.g., in Graeff et al., 2009). Therefore, the corresponding parameter of the impacted floodplain soils was artificially elevated, that is, the initial range of hydraulic conductivity of top-soil layer (parameter Kstop_1, supporting information Data Set S2) was given one order of magnitude higher than the other unaffected soil types.

The extra groundwater storage was added under the soil layers for each grid cell (Figure S4). The storage was further divided into the hydrologically active storage for hydrological dynamics ("dynamic" storage) and the "passive" storage for mixing and retention of tracers and chemical solutes (Benettin et al., 2015; Birkel et al., 2015; X. Yang et al. 2018). A proportionality of the dynamic storage was introduced as a global model parameter (*HydroFrac_ExtraGW*, Data Set S2). The groundwater storage received vertical leakage fluxes from the soil water storage and lateral baseflow fluxes from up-slope grid cell(s). The lateral baseflow flux to a down-slope grid cell was calculated using similar kinematic-wave approach as the subsurface flux, considering only the dynamic storage (Equation S3, Text S1). The calculation of baseflow export to the stream (i.e., baseflow discharge) was adopted from that of subsurface runoff export, but we introduced a new seepage parameter (Equation S4, Text S1). The extra groundwater storage had no upward feedbacks to the overlying soil water system.

Tracers (both isotopes and ages) were assumed to fully mix in the total extra groundwater storage and be conservatively transported in association with baseflow dynamics. The tracer signatures were updated as follows:

$$\frac{dC_{new}}{dt} = \frac{1}{ExtraGW}\left(F_L \cdot C_L + \sum_i F_{IN,i} \cdot C_{IN,i} - F_{OUT} \cdot C_{new} \right),$$ (1)

where C and F denote generic tracer signature (i.e., δ^2H, δ^{18}O, and water age), and water flux, respectively, for soil water leakage (L), up-slope inflow flux(es) (IN) and outflow flux (OUT); *ExtraGW* denotes the total extra groundwater storage. The equation is solved using the fourth-order Runge-Kutta method. The updated tracer signature of baseflow flux (C_{new}) was also assigned to the baseflow runoff component if a stream channel is present.

3.3. Model Parameterization and Multi-Criteria Calibration

Model parameters in EcH$_2$O-iso represent soil properties and ecohydrological dynamics (see Maneta & Silverman, 2013). Balancing the spatial variation of processes and the complexity of model parameterization (Clark et al., 2017), parameters were considered as global, soil, or vegetation dependent (the number being 4, 16, and 12, respectively). In total, 88 parameters were defined in this study (see Supporting Information Data Set S2), with initial ranges estimated based on literature (e.g., Kuppel et al., 2018a), field data (Graeff et al., 2009) and specific considerations for anthropogenic impacts (see Section 5.1). To reduce the number of calibrated parameters, a sensitivity analysis was conducted using the Morris method (Morris, 1991) for parameter ranking and screening. Sensitivity indices were generated based on 100 trajectories (i.e., a total of 8,900 model evaluations which is sufficient according to general criteria suggested in Pianosi et al., 2016) and the radial-based Latin-Hypercube sampling strategy, which has a better sampling efficiency and coverage within the parameter space (Campolongo et al., 2011; van Griensven et al., 2006). We analyzed the parameter rankings separately for multiple data sets (i.e., discharge at the outlet, soil moistures, soil temperature, and stream water isotopes) and summarized a common set of 61 most sensitive parameters for further model calibration (see Figure S5).

Model calibration was conducted using a multi-criteria method modified from Ala-aho et al. (2017). A large number of parameter sets was first sampled using Latin-Hypercube sampling. We generated 500,000 parameter sets for model calibrations. Then, model performance metric for each of the multiple constraining data was calculated. We used the KGE metric (Gupta et al., 2009) for soil temperature and log-transformed discharge data, and the MAE (Mean Absolute Error) for soil moisture and stream water isotope data. We designed a two-stage approach: (1) the "best" 10,000 simulations based only on discharge performance were first retained from the 500,000 runs; (2) among the 10,000 simulations, performance metrics for multiple constraining data were ordered respectively (i.e., obtaining the cumulative distribution functions of data-specific metrics), and the common threshold quantile for all cumulative distribution functions was iteratively determined above which the common set of 100 simulations could be found (Ala-aho et al., 2017). The final common 100 runs were therefore the best 100 runs simultaneously fulfilling all criteria. The 90% confidence interval ($CI = 95\%quantile - 5\%quantile$) among the best 100 runs was referred to as the simulation uncertainty, and the overall predictive uncertainty (PU) was quantified as $PU = \sum_{t=1}^{n} CI_t / \overline{SIM}_t$ (where CI_t and $\overline{SIM}_t$ are the simulated confidence interval and median value, respectively, at time step t ($t = 1,2,\ldots,n$); n is the total number of time steps) (Kuppel et al., 2018a). All computations were conducted in the iDiv high-performance computing (HPC) cluster at Helmholtz Center for Environmental Research - UFZ.

3.4. Model Set-Up and Evaluation

The spatial resolution was set as 50×50 m, and all geographical information was resampled and unified accordingly (Figure 1). Grid cells that contain channels were generated from the drainage network based on the resampled 50 m DEM and corresponded with the observed stream channels from the water authority of Saxony-Anhalt State (LHW; Figure 1a). Three soil layers were defined as 0–20, 20–50, and 50–100 cm, considering common agricultural tillage depth (ca. 20 cm), soil base-layer depth (Section 2.1), and the tile drain depth, respectively. The initial storages of the extra groundwater were roughly set as 2 and 5 m for hillslope and valley bottom grid cells, respectively.

Daily time-step modeling was set up in 2010–2019. In addressing the potential impacts from unrepresentative initial conditions (e.g., initial soil moisture, groundwater storage, tracer signatures in different storage, etc.), the first two years 2010–2012 were used as the spin-up period (repeated for 10 times) and excluded from further model evaluation. The vegetation dynamic module was deactivated in this study.

The Schäfertal catchment experienced a prolonged drought during 2018–2019. To investigate the value of the information content of different monitoring data sets in different periods, we designed three model calibration and validation schemes:

- *Scheme I.* The model was calibrated in 2013–2016 and validated in 2017–2019, against measured data of discharge, soil moisture at layer 1 and layer 2, and soil temperature.

Table 2
Statistics of Model Performances According to the Multi-Criteria Calibration Schemes Against Different Data sources

		Scheme I		Scheme II		Scheme III	
		Calibration	Validation	Calibration	Validation	Calibration	Validation
Data		2013–2016	2017–2019	2016–2019	2013–2015	2016–2019	2013–2015
Discharge	KGE	0.65 ± 0.02	0.43 ± 0.30	0.78 ± 0.02	0.50 ± 0.06	0.77 ± 0.02	0.49 ± 0.05
(log-transformed)	PU	0.21	0.32	0.21	0.23	0.22	0.22
[a]Isotope ^{2}H	MAE	23.2 ± 34.4		24.7 ± 67.0		7.13 ± 1.06	
	PU	1.45		1.40		0.26	
[a]Isotope ^{18}O	MAE	3.38 ± 4.45		3.51 ± 9.16		1.20 ± 0.29	
	PU	1.55		1.50		0.30	
Soil moisture at Layer 1	MAE	0.051 ± 0.006	0.055 ± 0.005	0.057 ± 0.005	0.059 ± 0.010	0.059 ± 0.005	0.061 ± 0.011
	PU	0.43	0.45	0.50	0.52	0.53	0.53
Soil moisture at Layer 2	MAE	0.043 ± 0.005	0.070 ± 0.008	0.062 ± 0.006	0.050 ± 0.009	0.064 ± 0.008	0.051 ± 0.012
	PU	0.34	0.41	0.48	0.42	0.50	0.44
Soil	KGE	0.69 ± 0.03	0.69 ± 0.09	0.73 ± 0.03	0.67 ± 0.05	0.70 ± 0.05	0.64 ± 0.06
temperature	PU	0.43	0.38	0.35	3.21	0.48	1.35
[b]Surface temperature	KGE	0.52 ± 0.18		0.67 ± 0.13		0.70 ± 0.17	
	PU	1.09		2.36		1.37	

The mean ± standard deviation of the evaluation metrics and Predictive Uncertainty were calculated based on the best 100 runs.

Abbreviations: KGE, Kling-Gupta Efficiency; MAE, Mean Absolute Error; PU, Predictive Uncertainty.

[a]Stream-water stable isotopes were only available in 2017–2019. [b]Surface temperature in 2013–2019 was exclusively used for validation.

- *Scheme II*. The model was calibrated in 2016–2019 and validated in 2013–2015, against the same data sources as Scheme I.
- *Scheme III*. The same as Scheme II, but adding on the available stream water isotopic data for calibration.

Therefore, *Scheme I* used only the information during the "normal" years (2013–2016) for model calibration, while *Scheme II* used the same data types but covering both normal and drought (2018–2019) periods. Finally, the value of incorporating isotope data in the calibration could be assessed from the comparisons between *Scheme II* and *Scheme III*.

4. Results

4.1. Multi-Criteria Calibration and Validation Based on Different Monitoring Data

Model performances were evaluated and compared between the three multi-criteria calibration schemes (Table 2). Although stream discharge showed large seasonal variations (e.g., high flow peaks >0.1 m^3 s^{-1}, whereas low flows were as low as 0.001 m^3 s^{-1}), discharge simulations generally captured the seasonal patterns well. Mean KGEs in the respective calibration periods were 0.65, 0.78, and 0.77 for the three schemes. In the validation period of *Scheme I* (2017–2019), discharge performance decreased largely in terms of the lowered overall goodness-of-fit (mean KGE = 0.43) and the larger simulation uncertainty (PU = 0.32, compared to 0.21 in 2013–2016). Especially during low-flow periods, simulations were often underestimated, and uncertainty bounds were widened (Figure S6). In contrast, as being part of the calibrations, discharge simulations of *Scheme II* and *Scheme III* in 2017–2019 performed among the best, with a well-constrained uncertainty during low-flow periods (Figure 2b). The performance degradations in the validation period of *Scheme II* and *Scheme III* (KGEs = ca. 0.50 in 2013–2015) were partly due to the more complex seasonal rainfall-runoff patterns (KGEs were relatively low in 2013–2016 even being calibrated in *Scheme I*) and some short-term suspicious discharge measurements (see Section 2.2).
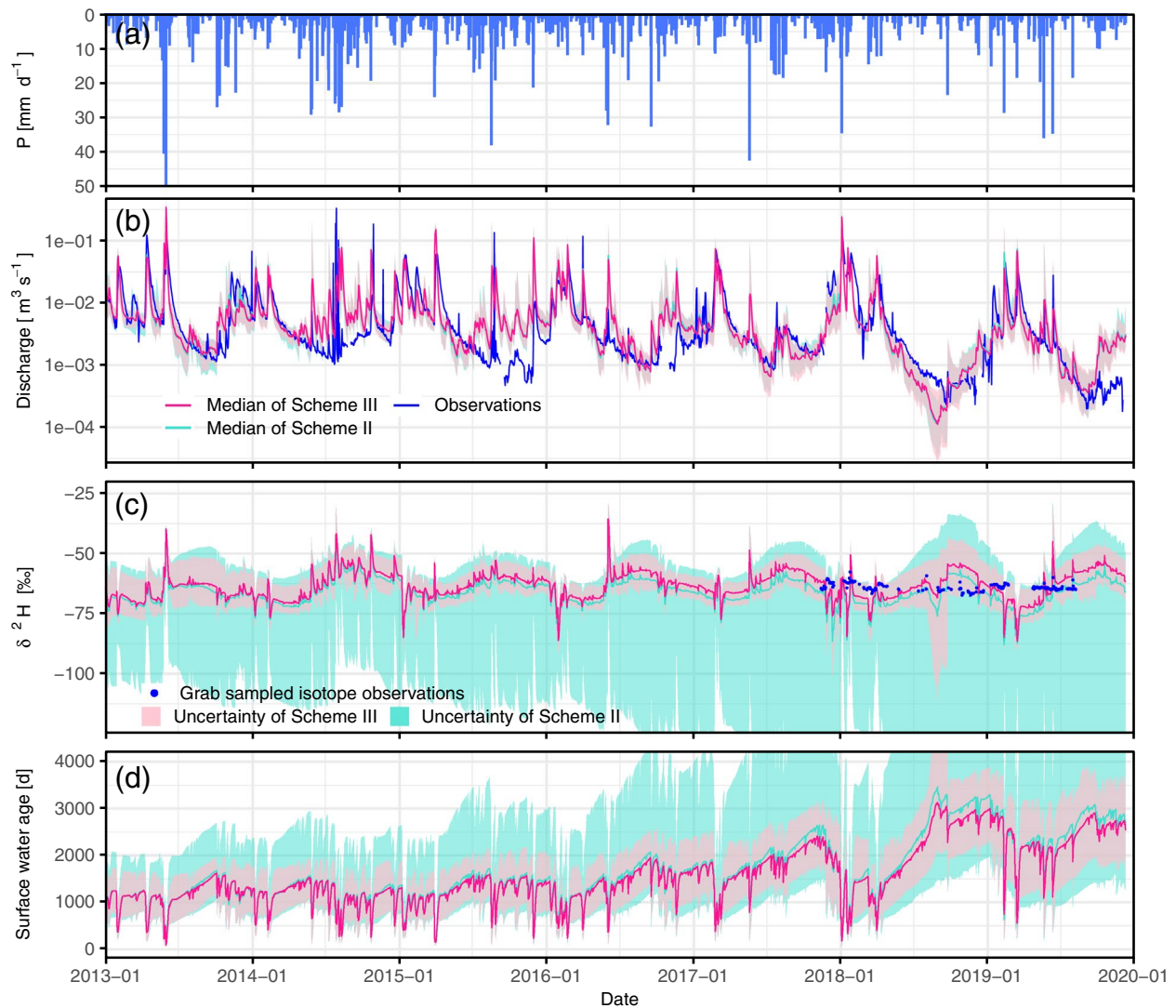
Figure 2. Model performances of stream water at the catchment outlet based on multi-criteria calibration Schemes II and III ((a) precipitation (P), (b) discharge, (c) isotopes (δ^2H), and (d) surface water ages). The median values and uncertainty bounds are the medians and 90% confidence intervals, respectively, of the best 100 model runs.

Compared to *Scheme II*, *Scheme III* additionally incorporated stream water isotopes in the calibration, and model simulations of isotopes (δ^2H and δ^{18}O) improved significantly (Table 2, and Figures 2c and S7). Calibrated against the measured isotopes, the mean absolute error (the goodness-of-fit metrics) decreased from 24.7 ± 67.0 and 3.51 ± 9.16 (‰), for δ^2H and δ^{18}O, respectively, to 7.13 ± 1.06 and 1.20 ± 0.29 (‰); the PU decreased by ca. 5 times. Based on the well-constrained simulation uncertainty of *Scheme III*, measured isotope variations were mostly captured within the uncertainty bounds (75% and 64% for δ^2H and δ^{18}O, respectively, of the 170 data points). The median values of the best simulations from *Scheme III* were generally in line with the measurements, with more marked dynamics as responses of rainfall events during winter-spring and more steady states during summer-autumn (Figure 2c). Similarly, the isotope information helped to reduce simulation uncertainty of stream water age (Figure 2d). Based on *Scheme III*, stream water age simulations revealed a general dominance of waters aged around 4 years (the mean of the simulated median ages), with significantly younger waters during rainfall events (Figures 2a and 2d). Moreover, stable stream water age of 3.3 years was estimated before 2017; while due to the extreme drought summer of 2018, a dramatic increase could be observed, and a higher level (the mean age of 6.0 years) was maintained until the end of 2019.
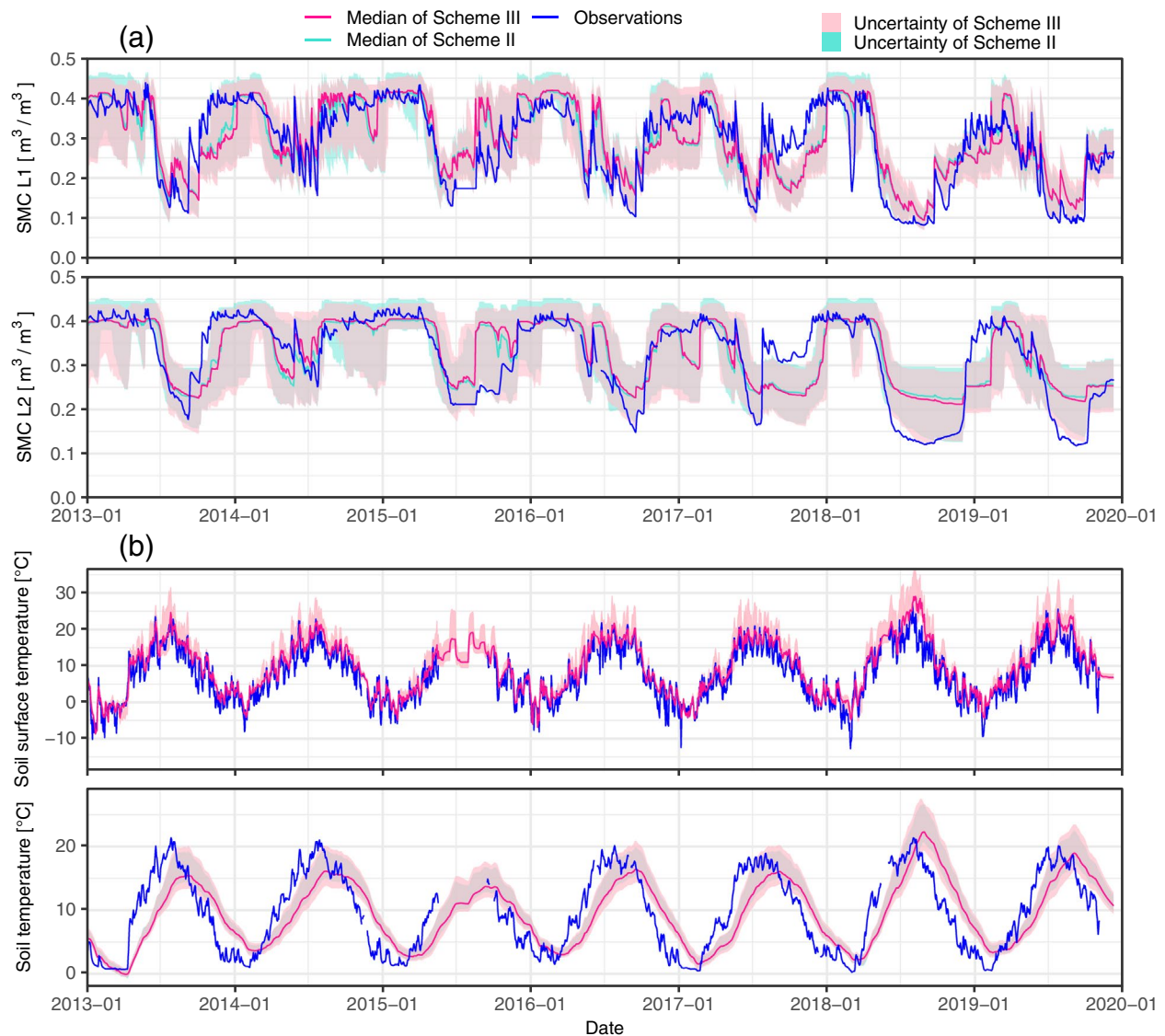
Figure 3. Model performances at the plot scale. (a) Soil moisture content (SMC) at Layer 1 (1–20 cm; upper) and Layer 2 (20–50 cm; lower); observations were collected at soil depth of 10 and 30 cm for Layer 1 and Layer 2, respectively. (b) Soil surface temperature (upper) and soil temperature (lower); Observations were ground surface temperature and the mean soil temperature (measured at depths of 10, 20, and 30 cm), respectively. SMC, Soil moisture content.

At the plot scale, model performances were generally good and the differences of the three calibration schemes were not as large as those at catchment scale (Table 2). Compared to *Scheme II* and *Scheme III*, soil moisture performances of *Scheme I* were slightly better in Layer 1 (0–20 cm), while in Layer 2 (20–50 cm), the mean MAE performances decreased from 0.043 in the calibration to 0.070 in the validation period. Despite the inclusion in the calibrations of *Scheme II* and *Scheme III*, Layer 2 simulations under drought conditions still performed worse than those in previous normal years (mean MAEs = ca. 0.063 and 0.050, respectively). Nevertheless, all calibration schemes reproduced the prolonged low soil moisture periods in 2018–2019, and the uncertainty bound of the best performed *Scheme III* covered most of the lysimeter observations (Figure 3a). Due to the impacts of higher air temperature in 2018–2019, surface temperature simulations of *Scheme II* and *Scheme III* performed much better (mean KGEs = 0.67 and 0.70, respectively), compared to those of *Scheme I* (KGEs = 0.52 ± 0.18). However, soil temperature simulations were similarly good among the three schemes (the mean KGEs >0.65), possibly due to the limited drought impacts on soil temperature as revealed in the observations (Figure 3b). As referenced by available isotope measurements,
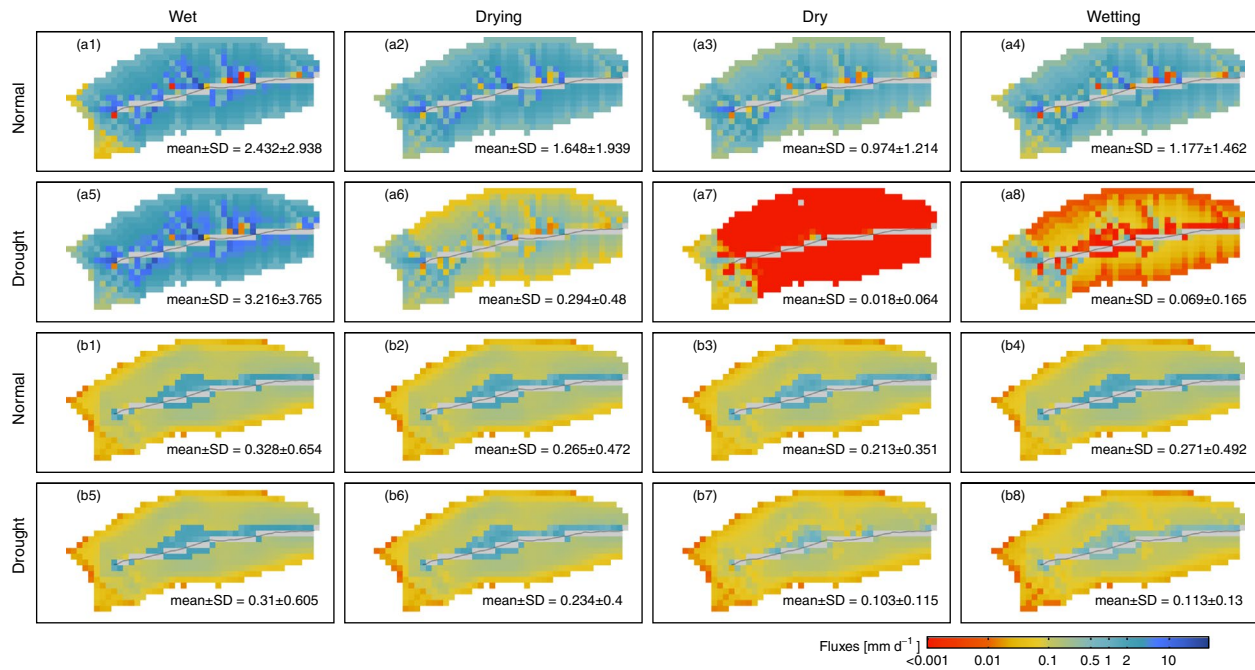
Figure 4. Spatial distributions of terrestrial surface (a1–a8) and subsurface (b1–b8) water fluxes in different seasonal periods (Wet: January–March, Drying: April–June, Dry: July–September, and Wetting: October–December) in the normal (2013–2016) and drought (2018–2019) years. Note the log-transformed, color-coded mean fluxes, with the spatial means and standard deviations (SDs) shown in each subplot. SD, standard deviation.

plot-scale tracer simulations of *Scheme III* were reasonable in terms of ranges and temporal variations in different soil water storages (e.g., the simulations at the lysimeter site shown in Figure S8).

Overall, simulations of *Scheme III* performed the best both from catchment and plot scale perspectives. Therefore, all our modeling analyses in the following sections were based on the results of the best 100 simulations of *Scheme III*.

4.2. Spatiotemporal Variability of Water Fluxes and Ages at the Catchment Scale

The well-constrained model performances, especially reducing uncertainty in the tracer simulations, added confidence in the analysis of the spatiotemporal variability of the ecohydrological partitioning at the catchment scale. Given the strong seasonality of measured soil moisture and simulated stream water ages, we defined four hydrological periods, that is, Wet (January–March), Drying (April–June), Dry (July–September), and Wetting (October–December) periods (Figure S9). We further compared the spatiotemporal patterns between the normal years (2013–2016) and drought years (2018–2019).

In the terrestrial phase, surface flow fluxes occurred mostly during heavy rainfall and snowmelt events in winter and spring. Their seasonal patterns were in accordance with those of the regional climate, with higher fluxes during the wet period and lower during the dry period (e.g., mean value of 2.5 vs. 1.0 mm d^{-1} in the normal years, Figure 4a). EcH$_2$O-iso assumed that all ponded water flows to the down-slope grid cell, therefore, the spatial pattern of surface flow fluxes generally followed that of the topographically driven flow accumulation throughout the drainage network (e.g., in the normal years, Figures 4a1–4a4). The spatial variations were generally large in all seasons (as indicated by the high standard deviations(SDs) in each subplot). Notably, such spatial and seasonal variations were further increased during the drought years (Figures 4a5–4a8). After the wet period, surface flow fluxes decreased sharply to 0.29 mm d^{-1} in the drying period, and further to 0.018 mm d^{-1} in the dry period. Interestingly, in the wetting period, surface fluxes were still very low and did not achieve a similar recovery as for the normal years (Figures 4a8 vs. 4a4); the flux declines were more severe in the tile-drain managed floodplain areas than in the unmanaged hillslopes. Subsurface fluxes showed high spatial variations in tile-drain managed floodplain areas (mostly ranged in 0.50–3.16 mm d^{-1}), hillslopes (0.03–0.25 mm d^{-1}), and upper hilltops (0.01–0.03 mm d^{-1}, Figures 4b1–4b8),
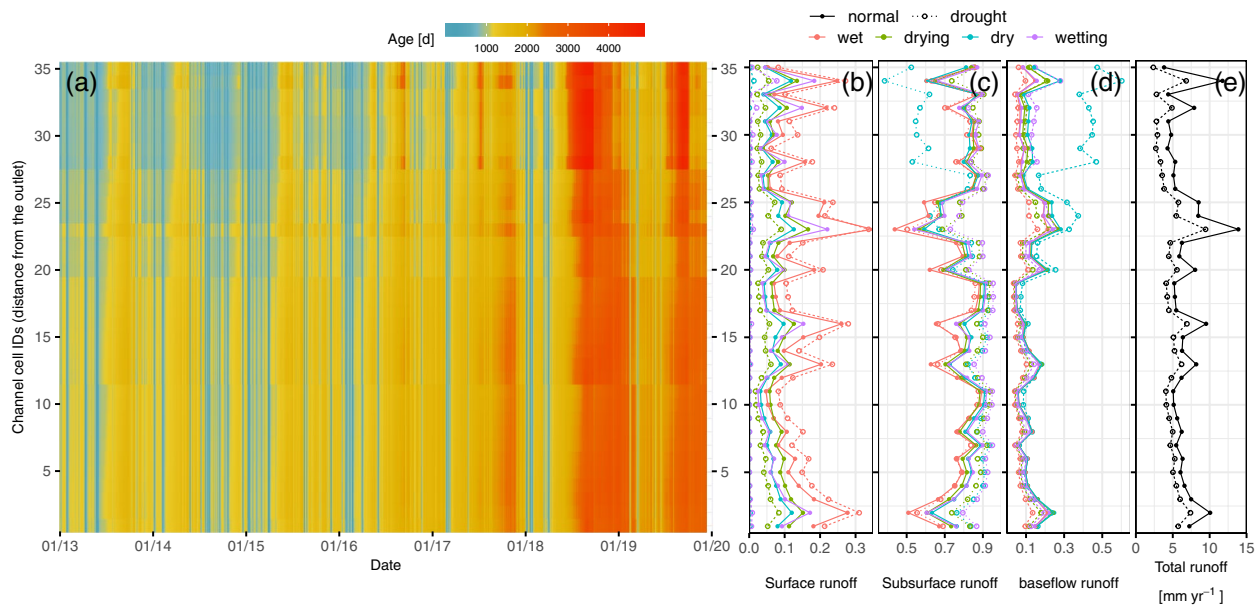
Figure 5. Daily variations of simulated stream water ages (a), the mean runoff partitioning in the four hydrological periods of the normal and drought years (proportions of surface runoff (b), subsurface runoff (c) and baseflow discharge (d), and the annual total runoff (e)) along the stream channels. Channel cell (50 × 50 m) IDs are ordered from the outlet to the up-stream segments, representing the increasing distances from the outlet.

mainly following the spatial pattern of soil types. The seasonal variations were generally not high during the whole simulation period, except for the tile-drain managed floodplain areas where subsurface fluxes decreased to the range of 0.25–1.12 mm d^{-1} during the dry and wetting periods of the drought years (Figures 4b7 and 4b8). Since the newly added extra groundwater storage in EcH$_2$O-iso does not allow upward fluxes back into the soil layer, the baseflow fluxes therefore did not affect soil water dynamics.

For stream water dynamics, runoff contributions along the stream channel and the partitioning of the three runoff components (surface runoff, subsurface runoff, and baseflow discharge) varied both seasonally and inter-annually, and thereby, the overall stream water ages showed high longitudinal and temporal variations (Figure 5). During the normal years (2013–2016), subsurface runoff predominated in most stream segments for all seasonal periods (the mean percentages were more than 75% of total runoff, Figure 5c). Surface runoff proportions were significantly higher in the wet period than in other periods (the mean percentages were 13% and 6%–9%, respectively, $p < 0.01$; Figure 5b). Such seasonal variations of surface runoff changed stream water ages significantly (ANOVA test, $p < 0.01$), with generally the youngest in the wet period (946 ± 268 days), the oldest in the dry period (1,276 ± 374 days) and intermediate in the drying and wetting periods (1,136 ± 303 and 1,181 ± 290 days, respectively; Figure 5a). Moreover, heavy rainfall events throughout the year caused short-term predominance of surface runoff, which dramatically reduced water ages to the minima of 13–50 days. During the drought years (2018–2019), surface runoff proportions decreased to 4% in the drying period and <0.5% in the dry and wetting periods (Figure 5b). Subsurface runoff still predominated in most stream segments (with slightly higher proportions compared to the values during the normal years), while the percentages in the upper one-fourth stream segments decreased to 40%–60% in the dry period (with the baseflow runoff contributions increased to 60%–40% accordingly, Figures 5c and 5d). Combining the decreased younger surface runoff contribution and the reduced stream discharge, stream water ages increased largely during the drought years (Figures 2d and 5a). Moreover, several stream segments (e.g., for the stream cell IDs 13, 16, and 23) showed dramatic decreases of subsurface runoff proportions and increases of surface and baseflow runoffs accordingly (Figures 5b–5d). Such changes in runoff partitioning also caused longitudinal changes in stream water ages (Figure 5a).

Compared to the blue water fluxes, the Schäfertal catchment exhibited a higher proportion of green water outputs (i.e., transpiration and soil evaporation). Driven by seasonal dynamics of the energy balance, evaporation showed clear seasonal patterns with much higher fluxes in summer (evaporation was
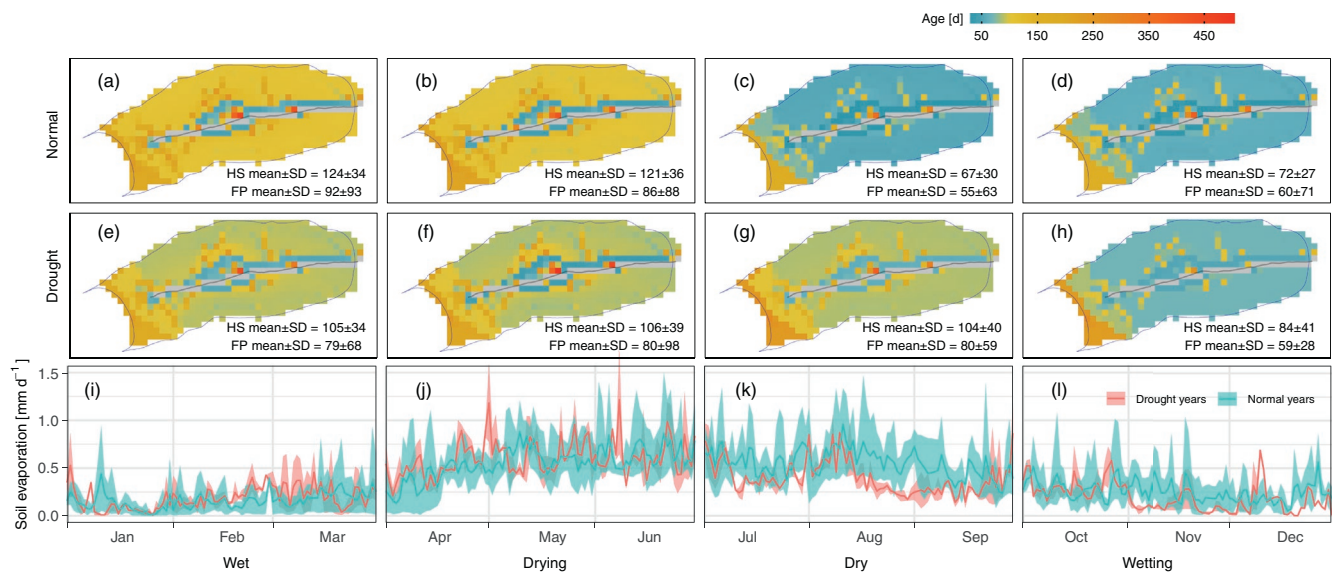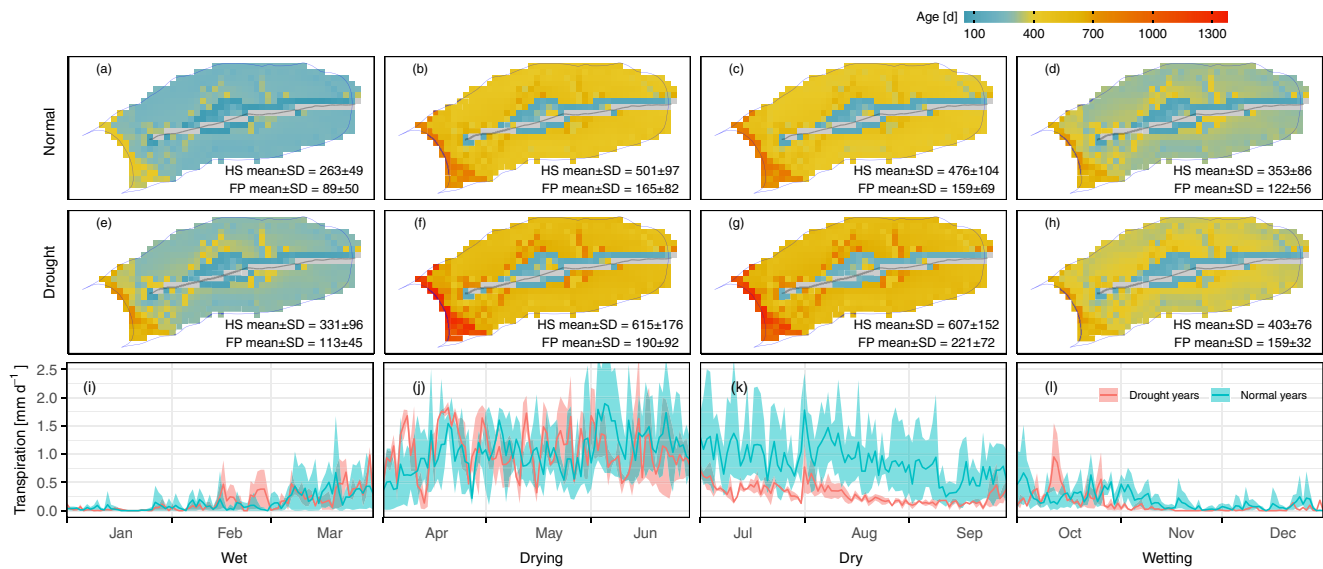
Figure 6. Spatial distributions of the mean evaporated water ages in the four hydrological periods during the normal (subplots (a)–(d)) and drought years (subplots (e)–(h)), and the total evaporated fluxes of the whole catchment (subplots (i)–(l)). The spatial means and SDs were shown in each map (values of HS and FP were calculated for the hillslopes/up-hill and tile-drained floodplain grid cells, respectively). FP, floodplain; HS, hillslope; SD, standard deviation.

0.53 ± 0.18 and 0.20 ± 0.12 mm d^{-1} in April–September and October–March, respectively, Figure 6i–6l). Compared to the normal years, soil evaporation fluxes were only significantly decreased in the dry and wetting periods (ANOVA test, $p < 0.001$ and 0.01, respectively), meanwhile with significantly older evaporated water ages during the drought years (Figures 6c and 6d vs. 6g and 6h, $p < 0.001$). In contrast, in the wet and drying periods, evaporated water ages were significantly older during the normal years than the drought years (Figures 6a and 6b vs. 6e and 6f, $p < 0.001$), although their fluxes were similar (Figures 6i–5j). The ages showed clear seasonal changes in the up-hill forest and hillslope grid cells (HS values in Figures 6a–6h), while in the tile-drain impacted floodplain grid cells, they were generally younger and more uniform across seasons and years (only the differences in the dry period between the normal and drought years were slightly significant; see floodplain cell values [FP] in Figures 6c vs. 6g, $p = 0.02$). Despite variations of absolute values, the spatial pattern of water ages generally followed the distributions of soil types and topographic drainage network.

Compared to soil evaporation (derived from the first soil layer), transpiration was influenced by soil water dynamics across the whole root zone, as well as impacts of aboveground vegetation properties. During the growing seasons of the normal years (April–September, Figures 7j–7k), transpired water fluxes reached up to 0.96 ± 0.32 mm d^{-1} (accounting for 54 ± 18% of total precipitation); while in the drought years, the fluxes dramatically decreased from a similar level as the normal years in April–June to 0.29 ± 0.16 mm d^{-1} in July–September. The seasonal patterns of the transpired water ages were similar in both normal and drought years (i.e., the oldest in the drying period, getting younger in the dry and wetting periods until the youngest in the wet periods). The absolute ages were significantly younger in the normal years than in the drought years (Figures 7a–7d vs. 7e–7h, ANOVA test, $p < 0.001$). Transpired water ages in the tile-drain impacted floodplain sub-areas were generally much younger (0.5–1.0 years) and responded differently to the drought stress. The ages were not significantly increase in the drying period (FP values in Figure 7b vs. 7f, $p > 0.1$), while they further increased in the dry period, which was different from the general seasonal pattern. The general spatial patterns of ages were jointly determined by soil properties, subsurface drainage networks and vegetation types, showing oldest water in the forest areas and youngest in the floodplain areas.

4.3. Ecohydrological Functioning under Drought Stress

As noted, the Schäfertal catchment experienced prolonged drought stress in 2018–2019, which heavily impacted the catchment functioning in terms of both blue and green water dynamics. Although the absolute

Figure 7. Spatial distributions of the mean transpired water ages in the four hydrological periods during the normal ((a)–(d)) and drought years ((e)–(h)), and the total transpired fluxes of the whole catchment ((i)–(l)). The spatial means and SDs were shown in each subplot (values of HS and FP were calculated for the hillslopes/up-hill and tile-drained floodplain grid cells, respectively). FP, floodplain; HS, hillslope; SD, standard deviation.

values of fluxes/storages and their ages varied largely, the relative spatial patterns were nearly consistent during the whole study period, driven by soil and vegetation characteristics. Therefore, we summarized and compared the temporal dynamics of each ecohydrological compartment in the normal and drought years (Figure 8). Precipitation was significantly reduced in the drought years (mean of 399 mm, accounting for 68% of long-term average). The drought situation was worsened by the unevenly distributed temporal patterns of inputs (e.g., more frequent heavy rainfall events and extremely low precipitation, in total of 62 mm during the dry period), and the significantly increased air temperature (e.g., the annual mean increased from 7.3 to 8.5°C and the daily maximum temperatures were mostly >20°C during the whole drying and dry periods, Figure 8a). Such dramatic changes affected the water fluxes, soil water storages, as well as their ages (Figures 8b–8d).

The green water fluxes were the dominant outputs, accounting for ca. 59% and 68% of annual precipitation in the normal and drought years, respectively (Figure 8b). Soil evaporation and transpiration were heavily impacted by the drought stress with much higher seasonal variations and stronger short-term fluctuation. Their seasonal patterns were likely driven by the pattern of air temperature in the normal years, whereas by that of precipitation in the drought years (Figures 8a and 8b). During the beginning of the drought period (April–June), the total green water fluxes were very similar. However, soil water decreased dramatically from 196 mm (for the upper 50 cm soil water storage) and 63 mm (for the lower 50–100 cm gravitational water storage) to 134 and 25 mm, respectively (Figure 8c). Moreover, the transpired water ages were significantly older during this period (545 ± 35 vs. 453 ± 35 days, respectively, $p < 0.001$; Figure 8b). Followed by the driest period (July–September), the two soil water storages further reduced to 123 ± 5.8 and 13.9 ± 4.7 mm, which accounted for 77% and 38% of amounts in the normal years, respectively (Figure 8c). Such high soil water deficits in the drought years strongly limited green water dynamics in the dry period, in terms of reduced fluxes (reduced by 28% and 67% for evaporation and transpiration, respectively) and older water ages (the means of 96 vs. 64 days for evaporation and 554 vs. 428 days for transpiration).

In the drought years, the limited soil water availability continued to the wetting period, which further caused much lower overall stream discharge and altered the seasonal patterns (Figure 8d). The annual runoff was only 119 mm (56% of that in the normal years). Among the four hydrological periods, total streamflow during the dry and wetting periods was 12.6 mm, accounting for only 15% of that in the normal years. Moreover, the runoff partitioning was also changed with higher contributions from the deeper groundwater system, which largely increased stream water ages (see Figure 5).
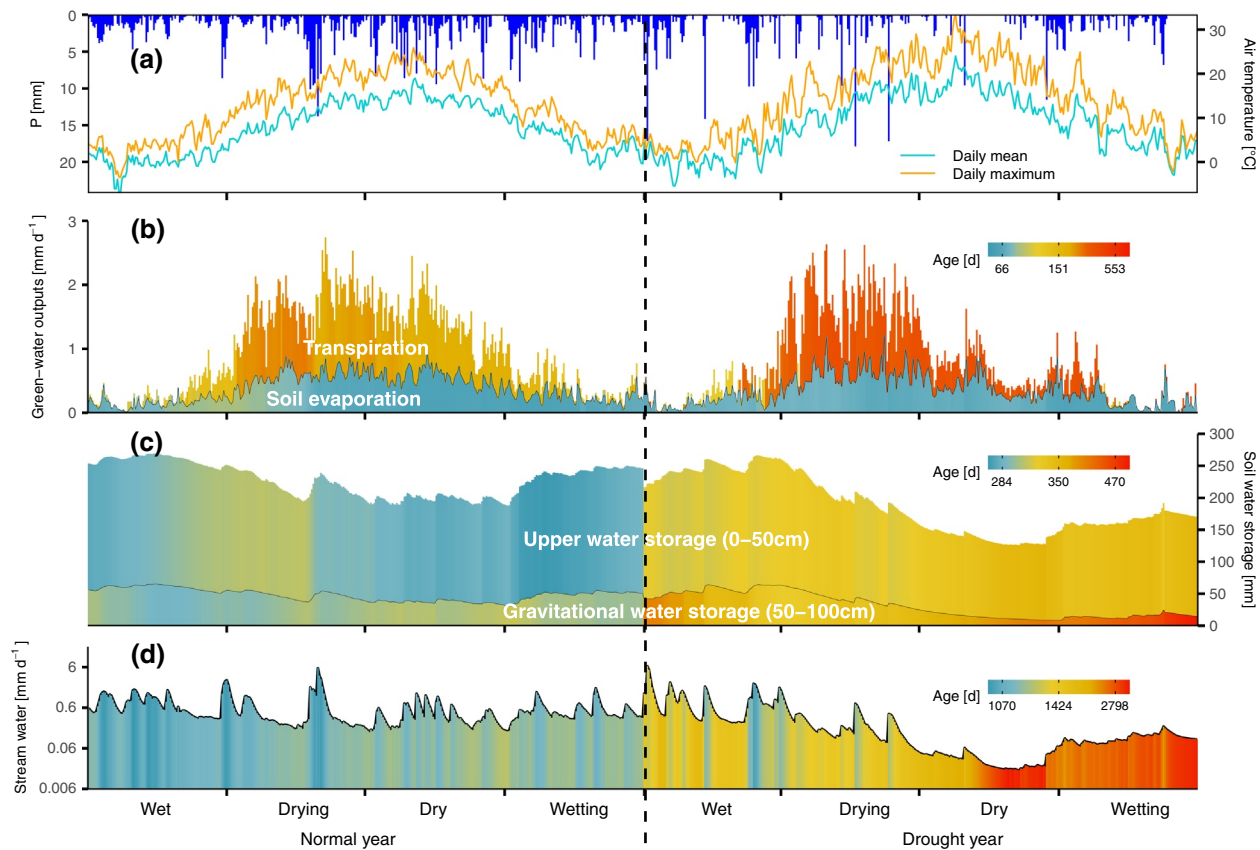
Figure 8. Summary of catchment ecohydrological responses to the drought stress ((a) precipitation and air temperature, (b) catchment-wide transpiration and soil evaporation, (c) soil water storage and (d) discharge at the catchment outlet during the normal (2013–2016) and drought years (2018–2019)). For each ecohydrological compartment, water storage/fluxes were shown as mean daily values, and the color code was normalized from the water-weighted mean daily ages (the age values of 10%, 50%, and 90% quantiles were shown in the legend, respectively). The upper water storage (total soil water amount in the upper 50 cm) and gravitational water storage (freely draining water above the field capacity in the third-layer 50–100 cm) were calculated from the most hydrologically active floodplain areas.

5. Discussion

5.1. Adaptations of the EcH₂O-iso Model Conceptualization and Parameterization

The soil properties and geological formations of the Schäfertal catchment have been well examined (Altermann & Mautschke, 1970; Graeff et al., 2009). The functioning of the agriculturally dominated catchment has been impacted by artificial tile drainage. Instead of developing a separate submodule for the tile drain flow, we adjusted the generic EcH₂O-iso parameterizations to capture effects of the artificially elevated subsurface lateral drainage. We assigned a higher horizontal hydraulic conductivity to the floodplain soils than in previous studies (Graeff et al., 2009). Based on the best-performed calibration *Scheme III*, the calibrated top-soil value (mean ± SD) was $4.83 \times 10^{-4} \pm 2.44 \times 10^{-4} \text{ m}^3 \text{ s}^{-1}$ (Figure S5), which is one to two orders of magnitudes higher than the calibrated values of the hillslope and uphill soils. Meanwhile, the calibrated ratios of vertical to horizontal hydraulic conductivities were significantly lower for the floodplain soils (mean ± SD = 0.039 ± 0.095, compared to 0.093 ± 0.125 and 0.071 ± 0.116 for hillslope and uphill soils, respectively). This ensured more realistic vertical soil water exchanges between soil layers in the tile drained soils.

For agricultural catchments, the construction of tile drainage systems is a common measure to improve crop productivity, which can result in marked changes in water and nutrient pathways (Li et al., 2010). Some models have therefore developed specific routines for that (Guo et al., 2018), while the absence of detailed tile drain information (e.g., locations, physical characteristics, and the functional efficiency of pipe

networks) commonly obscures precise simulations of tile drainage flows (De Schepper et al., 2017). Moreover, it is still not well known how the cumulative effects of such local tile drains affect the overall drainage response when upscaling to the catchment scale (De Schepper et al., 2015; Li et al., 2010). Therefore, with the lack of tile drain information, common practices in simulating catchment drainage networks are either to introduce more general empirical equations (e.g., Li et al., 2010) or to adjust model configuration/parameterization based on conventional structures (e.g., De Schepper et al., 2017). In our study, the tile drain networks were extensively constructed within the floodplain soils (Figure 1c), but no records were kept of their details or maintenance. Still, the adaptations of the EcH$_2$O-iso model parameterization and configuration in this study provide an illustration of how such agricultural anthropogenic changes can be conventionally integrated in process-based ecohydrological models. As for general tile drain process representation, appropriate modeling considerations are always subjective to information availability, research scales and objectives (e.g., see discussions in De Schepper et al., 2015).

In addition, long-term hydroclimatic and groundwater level data showed that the Schäfertal catchment was heavily impacted by nearby mining activities before the 1990s. Annual runoff decreased by ca. 70% in the 1980s and the stream completely dried out in dry seasons due largely to the lower groundwater levels (see Data Set S1 and Ollesch et al., 2010). At the event scale, bimodal hydrographs with strongly damped second peaks were commonly observed before 1975 but much less frequent afterward (Graeff et al., 2009; Ollesch et al., 2010). It has been argued that the second peaks are not a result from interflow but due to the fast displacement of the groundwater levels (Graeff et al., 2009). After the recovery of groundwater levels (since 2009), flowing water is now perennially observed in the stream, even under the recent extreme droughts. Moreover, the strongly damped stream water isotopic data revealed that a relatively large catchment storage is likely involved in catchment mixing dynamics (Soulsby et al., 2011; Tetzlaff et al., 2014). These pieces of evidence confirmed that deeper groundwater dynamics are influential in the Schäfertal catchment. Given that tile drain impacts and corresponding EcH$_2$O-iso soil parameter adaptations, it was necessary to provide a fundamental change to model structure to capture deeper groundwater dynamics. With the new structure, the model simulations confirmed that the baseflow constantly accounts for ca. 10% of annual runoff, with much higher proportions during the driest periods of the drought years (Figure 5d).

This new structure allows deeper groundwater storage exchanges and lateral baseflow fluxes to be considered in EcH$_2$O-iso, and therefore, greatly extends the model transferability to catchments that are impacted by both rapid, shallow subsurface stormflow and deeper, slower responding groundwater systems. The introduced passive component of the deeper storage allows higher flexibility for isotope and water age dynamics, which plausibly invoke orders of magnitude higher catchment storage compared to the seasonal flow dynamics, especially under extreme hydroclimatic conditions (Birkel et al., 2015). Implementations of the deeper groundwater storages were largely simplified based on reasonable assumptions in the study catchment. Further developments might be necessary for better representations of ecohydrological impacts of deep groundwater dynamics (e.g., the downward extension of plant/crop roots, the upward fluxes to soil layers due to groundwater rise, etc.).

5.2. The Information Content Brought by Different Data Time Series in Tracer-Aided Ecohydrological Modeling

Long-term data from experimental catchments are valuable in advancing ecohydrological process understanding and revealing catchment responses to environmental changes (Tetzlaff et al., 2017). As discussed in Section 5.1, long-term groundwater level data (their disturbance and recovery due to mining activities) have revealed that stream water in the Schäfertal catchment was significantly influenced by the deeper groundwater system. In addition, long-term hydroclimatic data (1968–2019, Data Set S1) showed that large rainfall events (e.g., daily precipitation >10 mm) occurred irregularly across different hydrological periods. The outlet discharge in wet-drying periods was higher than 0.1 m^3 s^{-1}, whereas in dry-wetting periods it was as low as 0.001 m^3 s^{-1}. This revealed strong intra- and inter-annual variabilities of catchment water storage and green-blue water partitioning. Furthermore, other "soft" information of agricultural management is also informative, though details are normally missing for quantitative investigation. The extensively constructed tile drainage in the middle valley bottom (Figure 1c) changed land use from pastures to cultivated arable fields, indicating that this artificial drainage has largely altered soil hydrological effects from more

clay and silty floodplain soils to a new state with much higher hydrological connectivity. This likely further caused the valley bottoms to become "hotspots" for subsurface lateral water exchanges (Figure 4). Such long-term historical data and soft information essentially informed the identification of dominant processes (as argued e.g., in Seibert & McDonnell, 2002 and McMillan et al., 2011) and further guided our model adaptations and set-up in this study.

Besides the general structure reflecting process-based catchment understanding, a model has to be comprehensively calibrated and validated against multiple data sources to allow a detailed evaluation of model behavior (Wagener et al., 2001), especially for multiple output fluxes (Gupta et al., 1998). In this study, we designed a two-stage calibration strategy, that is, the multicriteria calibrations (simultaneously against soil moisture, soil temperature, stream discharge, and/or stream water isotopes) were conducted after the first-round, single-objective calibration (against discharge). Streamflow integrates information from multiple convolutional processes (Beven & Binley, 1992; Kirchner, 2009). In turn, the stream discharge data informs the general water balance, which is of particular importance for the Schäfertal catchment given its strong temporal variability of blue- and green-water dynamics. Therefore, discharge observations were first used to constrain overall water balance modeling (i.e., obtaining the best 10,000 out of 500,000 runs). Of course, streamflow information is known to be too ambiguous to disentangle internal processes and quantify their catchment-wide configurations (Finger et al., 2015; Güntner et al., 1999; Kuppel et al., 2018a). Hence, multiple criteria based on plot- and catchment-scale data were simultaneously considered at the second stage of calibration, which provided reasonable model simulations and uncertainty estimates at both plot and catchment scales. The calibrated parameters (Figure S5) were well constrained within initial ranges, although some of them were poorly identified presumably due to "equifinality" effects given the large number of parameters (Beven & Freer, 2001), or trade-off effects of the multi-objective calibration (e.g., see Holmes et al., 2020). Nevertheless, final simulation uncertainties of all calibrated outputs were mostly well constrained (Table 2 and corresponding figures).

Amongst multiple monitoring data sets, the inclusion of isotope data in calibration (i.e., comparing *Scheme II* and *Scheme III*) increased the isotope simulation performances (better goodness-of-fit and lower uncertainty; Table 2 and Figure 2). The high isotope simulation uncertainty of *Scheme II* was likely mainly derived from the extra groundwater storage because the deeper groundwater contributed significantly to stream water dynamics (Section 5.1). The calibrated proportions of the dynamic storage (parameter $HydroFrac\_ExtraGW$) were 0.047 ± 0.028 across all calibration schemes (Figure S5c), which is consistent with Soulsby et al. (2011) who found that total storage for tracer mixing can be two orders of magnitude higher than dynamic storage. The actual total storage cannot be sufficiently inferred by hydrological data alone (Birkel et al., 2011). This has probably caused much higher uncertainties in the tracer simulations, especially for the floodplain soils with active subsurface exchanges and considerable baseflow seepage (Figures 4b and 5d). However, isotope data provide "orthogonal" information on catchment behavior (Fenicia et al., 2008; Kirchner, 2016) and integrating them into spatially distributed modeling opens the opportunity of satisfying both velocity (the transport of water/solute particles) and celerity (the pressure wave propagation) perspectives (Holmes et al., 2020; Kuppel et al., 2018b). Therefore, the added information content of stream water isotope data is particularly valuable in constraining our modeling uncertainty, as for *Scheme III*. The isotope simulations at different water storages at plot scale were likely in a reasonable range as referred by available isotope data (e.g., the lysimeter-located grid cell in Figure S8), although the uncertainty in the extra groundwater storage was relatively high compared to that of upper soil storages. Moreover, other model performances at catchment (e.g., discharge) and plot (e.g., soil moisture) scales were maintained (and even slightly better for surface temperature, Table 2). This supports the argument that the added tracer information facilitated a more robust capturing of internal processes (e.g., storage-flux interactions, Figure 4; flow pathways, Figure 5), which is consistent with other studies that evaluated the value of isotope data for model calibration (He et al., 2019; Holmes et al., 2020; Stadnyk & Holmes, 2020).

The plausible tracer simulations based on *Scheme III* enabled more reliable estimates of water ages in stream water and ecohydrological fluxes. The simulated median stream water ages converged at around 3.3 years (Figure 2d), due to the predominance of subsurface runoff in stream runoff generation, with baseflow from the extra groundwater storage contributing consistently ca. 10% (Figures 5b–5d). Significantly younger waters dominated during the wet and drying periods, as well as in response to heavy rainfall events due to the

largely increased proportions of surface runoff (Figures 2d and 5b). Such seasonal patterns are in line with other studies (Birkel et al., 2015; Kuppel et al., 2018b; Sprenger et al., 2019). We noted that our simulated ages were significantly older than those by J. Yang, et al. (2018), who studied the Schäfertal's water transit times during 1997–2007 based on the theory of StorAge Selection (SAS) functions. During their investigation period, fast recessions after the main high-flow seasons were observed, and the stream temporally dried out during driest days of the year (Data Set S1). Despite the potential impacts of varying methodological representations (see discussions in Kuppel et al., 2018b), this presumably mainly reflects that groundwater level had not yet fully recovered from mine de-watering until 2009 (Section 5.1). Therefore, deeper groundwater dynamics had limited impacts on stream water dynamics, compared to that in our simulation period.

5.3. The Value of Observations and Modeling under Drought Stress in Assessing Ecohydrological Responses under Changing Conditions

Compared to *Scheme II* and *Scheme III*, simulations of *Scheme I* failed to reproduce the streamflow dynamics in the drought years (Table 2), demonstrating that non-stationary behavior of catchment ecohydrological functioning has been induced by prolonged droughts. In the drier years 2018–2019, soil moisture was severely depleted as informed from the lysimeter data (Figure 3a), whereas groundwater levels in the valley bottoms were only slightly decreased (Figure S1). This implied that deeper groundwater storage, with longer residence times, maintained stream flow. Besides model adaptations of the new extra groundwater storage, inclusion of data collected under drought stress was highly important in constraining our ecohydrological modeling as highlighted by the good streamflow performances of *Scheme II* and *Scheme III*. Our findings agree with the strategy proposed by Westra et al. (2014) that improving model ability under changing hydroclimatic conditions should be conducted from data, parameterization and model structure perspectives.

In addition to the changes of water quantity, our tracer-aided modeling further showed a high degree of non-stationary in water ages under the drought stress (Figure 8), which can have major implications for water quality. Runoff contributions from deeper groundwater increased up to 60% of total runoff, which further resulted in a sharp increase of stream water ages by ca. 200%, to ca. 6.0 years (Figure 5). Despite the good simulations from calibration *Scheme III*, we recognized that tracer simulations in the drought years need to be critically assessed. The simulation uncertainty of stable isotopes of water increased during the drought periods (Figures 2c and S7), which is associated with the large uncertainty of estimating the deeper groundwater storage and its water ages (Birkel et al., 2015; Stewart et al., 2010). This has been further complicated by the spatially differentiated responses of groundwater levels to the drought stress (Figure S1). Moreover, the heavy damping of observed isotope signals indicated a strong mixing that might limit interpretations of the input-output signal/noises relationship (Stewart et al., 2010), and our simulated stream water ages were around the upper limit of travel times that can be inferred from stable isotopes of water (Sprenger et al., 2019; Stewart et al., 2010). Nonetheless, the mean residence time of such drought-invoked deeper storage (i.e., the groundwater differences between normal and drought years) was roughly estimated from the decreases of measured groundwater levels during July–December of 2018–2019 (Text S2). The estimated ages of 3,300–5,200 days were in line with our tracer-based simulations (Figure 5).

Green water dynamics in the Schäfertal catchment accounted for more than half of the annual rainfall, and that proportion further increased to ca. 70% in the drought years (Figure 8). Moreover, transpiration from deeper groundwater was very limited due to the relatively shallow effective root zone for the main crops (e.g., the depth of 50–100 cm for wheat and barley (Fan et al., 2016)). Therefore, the ecohydrological outputs (i.e., soil evaporation and transpiration) were more sensitive to the drought stress compared to the blue water dynamics, as reported for other regions (e.g., Sawada et al., 2014; Orth & Destouni, 2018). Furthermore, the detailed spatiotemporal comparisons of the modeled fluxes and water ages under the normal and drought years (Figures 6–8) provided in-depth insights into anticipated ecohydrological responses to the changing climate conditions, especially for drought-sensitive agricultural catchments. Soil evaporation showed minor quantitative differences, indicating that the uppermost soils can generally be replenished by precipitation (Smith et al., 2020a); while the spatial and seasonal variability of the evaporated water ages was presumably determined by the proportions of storage that have been renewed by recent precipitation and upslope lateral surface inputs, and the variations of soil properties (Figure 6). In addition to the consistently older water ages in the drought years (Figure 7), transpiration responded differently in different stages

of the drought development (Figure 8): at the beginning of the crop growing season (the drying period), although precipitation declined by 36%, transpiration fluxes were maintained at a similar (or even slightly higher) level to the normal years, due to the replenished soil water storages after rewetting seasons and presumably the higher radiation availability (Orth & Destouni, 2018). Consequently, all soil water storages were dramatically depleted and this severely limited transpiration as the drought further developed until the end of the growing season. In contrast, the drought-induced impacts affected the transpired water ages at the beginning of the growing season, significantly increasing the age compared to the normal years. Therefore, the water age information is of particular use as an indicator warning for drought-induced impacts on agronomy systems and further guiding precautions of mitigation-adaptation measures.

As a common agricultural practice, artificial tile drains largely accelerated catchment drainage (e.g., the elevated subsurface fluxes in the floodplain areas in Figure 4b), while altering soil water dynamics of the entire soil profile further affected the ecohydrological green water usage in terms of water ages of transpiration and soil evaporation (from the entire soil profile and the top soil layer, respectively). The ages in the tile-drain managed floodplain areas were generally much younger and less seasonally varying than those in the unmanaged hillslope/uphill areas (Figures 6 and 7), presumably due to the accelerated soil water exchanges and thereby lower soil water residence times in the managed floodplain areas. Moreover, the tile-drain management altered responses of the green-water ages under drought stress, especially for transpiration. In contrast to the above discussed catchment-wide early responses of transpired water ages (Figure 8b), the increase in the managed floodplain sub-areas was not statistically significant in the drying period (Figures 7f vs. 7b). The soil water was likely fully replenished after a rewetting season (i.e., transpiration fluxes generally aged within 0.5–1.0 years, compared to ages of >1.0 years in the unmanaged hillslopes) and sustained comparable ecohydrological outputs at the beginning of the drought. However, this resulted in the severely decreased soil water storage (Figure 8c) and limited transpiration as the drought further developed in the dry period (reflected by the abnormally older transpired water ages than those in the drying, Figures 7g vs. 7f). This implies that the whole soil root zone of tile-drain managed agricultural areas becomes more sensitive to hydroclimatic variability, which differed from the normal inter-annual "memory effect" of transpiration (Kuppel et al., 2020). Of course, this study is limited by the catchment's spatial organization and relatively small areal proportion of tile-drain managed areas. Further research is needed for comprehensive investigations on ecohydrological impacts of the tile-drain management.

6. Conclusions

The tracer-aided, process-based ecohydrological model, EcH$_2$O-iso, was applied to the intensively managed, long-term experimental Schäfertal catchment. The functioning of this agricultural catchment has been impacted by artificial tile drainage and groundwater system dynamics, as reflected in long-term historical information (i.e., hydroclimatic data and anthropogenic activities). Accordingly, model parameterization of soil properties was adapted, and a revised model structure for deeper groundwater dynamics was developed. These model adaptations were of particular importance for process representation under the prolonged drought stress in 2018–2019, which has induced an extreme catchment response in terms of ecohydrological functioning. A two-stage strategy and alternative schemes were designed for comprehensive model calibration and validation, which make the most use of multi-source monitoring time-series, including catchment-scale hydroclimatic data, plot-scale soil moisture and soil temperature measurements and tracer (stable isotope of water δ^2H and δ^{18}O) data. The added value of the isotope data was highlighted in achieving more robust green- and blue-water partitioning and disentangling of internal storage-flux-age interactions. In particular, detailed spatially and temporally varying responses of green and blue water dynamics under the drought stress could be investigated.

The tracer-aided ecohydrological modeling in this study demonstrated that (1) understandings of catchment functioning can be informed from uniquely valuable long-term hydrological data and "soft" information (see also e.g., Tetzlaff et al., 2017 and Seibert & McDonnell, 2002), (2) anthropogenic activities can fundamentally alter natural ecohydrological processes and should be appropriately considered in process-based modeling, especially for agricultural catchments, (3) model calibration can be strengthened by the use of a variety of data under different climatic and anthropogenic conditions, and (4) reliable ecohydrological mod-

eling needs to be based on critical assessment of the information content contained in multi-source data. Based on robust model evaluations (e.g., calibration *Scheme III*), the EcH$_2$O-iso model provided more reliable estimates of water fluxes and ages of individual ecohydrological stores. Under drought stress, green and blue water dynamics showed spatiotemporally varying responses, partly due to the tile drain management; moreover, the tracer-based water ages mostly responded earlier than the fluxes (e.g., transpired water ages were much older than in normal years at the beginning of the droughts, while the transpired fluxes were comparable). Therefore, our tracer-based ecohydrological modeling also has insightful value for identifying drought-induced impacts on ecosystem function, especially for drought sensitive agricultural regions. In particular, water flux ages provided sensitive early indicators of drought induced eco-hydrological alterations across a wide range of temporal and spatial scales.

Data Availability Statement

The model code of EcH$_2$O-iso is publically available at http://bitbucket.igb-berlin.de:7990/users/ech2o.

References

Ala-aho, P., Tetzlaff, D., McNamara, J. P., Laudon, H., & Soulsby, C. (2017). Using isotopes to constrain water flux and age estimates in snow-influenced catchments using the STARR (Spatially distributed Tracer-Aided Rainfall–Runoff) model. *Hydrology and Earth System Sciences*, *21*(10), 5089–5110. https://doi.org/10.5194/hess-21-5089-2017

Albergel, C., Zheng, Y., Bonan, B., Dutra, E., Rodríguez-Fernández, N., Munier, S., et al. (2020). Data assimilation for continuous global assessment of severe conditions over terrestrial surfaces. *Hydrology and Earth System Sciences*, *24*(9), 4291–4316. https://doi.org/10.5194/hess-24-4291-2020

Altermann, M., & Mautschke, J. (1970). *Ergebnisbericht über quartärgeologische- bodenhydrologische Untersuchungen im Gebiet des Schäfertals und Waldbachs bei Siptenfelde*. Halle Germany: VEG GFE.

Anis, M. R., & Rode, M. (2015). Effect of climate change on overland flow generation: A case study in central Germany. *Hydrological Processes*, *29*(11), 2478–2490. https://doi.org/10.1002/hyp.10381

Bakke, S. J., Ionita, M., & Tallaksen, L. M. (2020). The 2018 northern European hydrological drought and its drivers in a historical perspective. *Hydrology and Earth System Sciences Discussions*, 1–44. https://doi.org/10.5194/hess-2020-239

Benettin, P., Kirchner, J. W., Rinaldo, A., & Botter, G. (2015). Modeling chloride transport using travel time distributions at Plynlimon, Wales. *Water Resources Research*, *51*(5), 3259–3276. https://doi.org/10.1002/2014wr016600

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, *249*(1), 11–29. https://doi.org/10.1016/S0022-1694(01)00421-8

Beven, K., & Smith, P. (2015). Concepts of information content and likelihood in parameter calibration for hydrological simulation models. *Journal of Hydrologic Engineering*, *20*(1), A4014010. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000991

BGR. (2018). *Soil map of Germany 1:200 000 (BÜK200)*. In Bundesanstalt für Geowissenschaften und Rohstoffe (BGR). Germany.

Birkel, C., & Soulsby, C. (2015). Advancing tracer-aided rainfall–runoff modeling: A review of progress, problems and unrealized potential. *Hydrological Processes*, *29*(25), 5227–5240. https://doi.org/10.1002/hyp.10594

Birkel, C., Soulsby, C., & Tetzlaff, D. (2011). Modeling catchment-scale water storage dynamics: Reconciling dynamic storage with tracer-inferred passive storage. *Hydrological Processes*, *25*(25), 3924–3936. https://doi.org/10.1002/hyp.8201

Birkel, C., Soulsby, C., & Tetzlaff, D. (2014). Developing a consistent process-based conceptualization of catchment functioning using measurements of internal state variables. *Water Resources Research*, *50*(4), 3481–3501. https://doi.org/10.1002/2013wr014925

Birkel, C., Soulsby, C., & Tetzlaff, D. (2015). Conceptual modeling to assess how the interplay of hydrological connectivity, catchment storage and tracer dynamics controls nonstationary water age estimates. *Hydrological Processes*, *29*(13), 2956–2969. https://doi.org/10.1002/hyp.10414

Campolongo, F., Saltelli, A., & Cariboni, J. (2011). From screening to quantitative sensitivity analysis. A unified approach. *Computer Physics Communications*, *182*(4), 978–988. https://doi.org/10.1016/j.cpc.2010.12.039

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., et al. (2017). The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. *Hydrology and Earth System Sciences*, *21*(7), 3427–3440. https://doi.org/10.5194/hess-21-3427-2017

Craig, H., & Gordon, L. I. (1965). Deuterium and Oxygen 18 variations in the ocean and the marine atmosphere. In *Stable isotopes in Oceanographic studies and paleotemperatures*. Pisa: Laboratorio di geologia numcleare.

De Schepper, G., Therrien, R., Refsgaard, J. C., & Hansen, A. L. (2015). Simulating coupled surface and subsurface water flow in a tile-drained agricultural catchment. *Journal of Hydrology*, *521*, 374–388. https://doi.org/10.1016/j.jhydrol.2014.12.035

De Schepper, G., Therrien, R., Refsgaard, J. C., He, X., Kjaergaard, C., & Iversen, B. V. (2017). Simulating seasonal variations of tile drainage discharge in an agricultural catchment. *Water Resources Research*, *53*(5), 3896–3920. https://doi.org/10.1002/2016wr020209

Douinot, A., Tetzlaff, D., Maneta, M., Kuppel, S., Schulte-Bisping, H., & Soulsby, C. (2019). Ecohydrological modeling with EcH2O-iso to quantify forest and grassland effects on water partitioning and flux ages. *Hydrological Processes*, *33*(16), 2174–2191. https://doi.org/10.1002/hyp.13480

Fan, J., McConkey, B., Wang, H., & Janzen, H. (2016). Root distribution by depth for temperate agricultural crops. *Field Crops Research*, *189*, 68–74. https://doi.org/10.1016/j.fcr.2016.02.013

Fatichi, S., Pappas, C., & Ivanov, V. Y. (2016). Modeling plant–water interactions: An ecohydrological overview from the cell to the global scale. *WIREs Water*, *3*(3), 327–368. https://doi.org/10.1002/wat2.1125

Fenicia, F., McDonnell, J. J., & Savenije, H. H. G. (2008). Learning from model improvement: On the contribution of complementary data to process understanding. *Water Resources Research*, *44*(6). https://doi.org/10.1029/2007wr006386

Finger, D., Vis, M., Huss, M., & Seibert, J. (2015). The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments. *Water Resources Research*, *51*(4), 1939–1958. https://doi.org/10.1002/2014wr015712

Graeff, T., Zehe, E., Reusser, D., Lück, E., Schröder, B., Wenk, G., et al. (2009). Process identification through rejection of model structures in a mid-mountainous rural catchment: Observations of rainfall–runoff response, geophysical conditions and model inter-comparison. *Hydrological Processes*, *23*(5), 702–718. https://doi.org/10.1002/hyp.7171

Güntner, A., Uhlenbrook, S., Seibert, J., & Leibundgut, C. (1999). Multi-criterial validation of TOPMODEL in a mountainous catchment. *Hydrological Processes*, *13*(11), 1603–1620. https://doi.org/10.1002/(sici)1099-1085(19990815)13:11<1603::Aid-hyp830>3.0.Co;2-k

Guo, T., Gitau, M., Merwade, V., Arnold, J., Srinivasan, R., Hirschi, M., & Engel, B. (2018). Comparison of performance of tile drainage routines in SWAT 2009 and 2012 in an extensively tile-drained watershed in the Midwest. *Hydrology and Earth System Sciences*, *22*(1), 89–110. https://doi.org/10.5194/hess-22-89-2018

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling. *Journal of Hydrology*, *377*(1), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751–763. https://doi.org/10.1029/97wr03495

He, Z., Unger-Shayesteh, K., Vorogushyn, S., Weise, S. M., Kalashnikova, O., Gafurov, A., et al. (2019). Constraining hydrological model parameters using water isotopic compositions in a glacierized basin, Central Asia. *Journal of Hydrology*, *571*, 332–348. https://doi.org/10.1016/j.jhydrol.2019.01.048

Heinrich, I., Balanzategui, D., Bens, O., Blume, T., Brauer, A., Dietze, E., et al. (2019). Regionale auswirkungen des globalen wandels: Der Extremsommer 2018 in nordostdeutschland. *System Erde*, *9*(1), 38–47. https://doi.org/10.2312/GFZ.syserde.09.01.6

Holmes, T., Stadnyk, T. A., Kim, S. J., & Asadzadeh, M. (2020). Regional calibration with isotope tracers using a spatially distributed model: A comparison of methods. *Water Resources Research*, *56*(9), e2020WR027447. https://doi.org/10.1029/2020wr027447

Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research*, *45*(2). https://doi.org/10.1029/2008wr006912

Kirchner, J. W. (2016). Aggregation in environmental systems – Part 2: Catchment mean transit times and young water fractions under hydrologic nonstationarity. *Hydrology and Earth System Sciences*, *20*(1), 299–328. https://doi.org/10.5194/hess-20-299-2016

Kistner, I., Ollesch, G., Meissner, R., & Rode, M. (2013). Spatial-temporal dynamics of water soluble phosphorus in the topsoil of a low mountain range catchment. *Agriculture, Ecosystems & Environment*, *176*, 24–38. https://doi.org/10.1016/j.agee.2013.05.016

Kleine, L., Tetzlaff, D., Smith, A., Wang, H., & Soulsby, C. (2020). Using water stable isotopes to understand evaporation, moisture stress, and re-wetting in catchment forest and grassland soils of the summer drought of 2018. *Hydrology and Earth System Sciences*, *24*(7), 3737–3752. https://doi.org/10.5194/hess-24-3737-2020

Knighton, J., Kuppel, S., Smith, A., Soulsby, C., Sprenger, M., & Tetzlaff, D. (2020). Using isotopes to incorporate tree water storage and mixing dynamics into a distributed ecohydrologic modeling framework. *Ecohydrology*, *13*(3), e2201. https://doi.org/10.1002/eco.2201

Kuppel, S., Tetzlaff, D., Maneta, M. P., & Soulsby, C. (2018). What can we learn from multi-data calibration of a process-based ecohydrological model? *Environmental Modelling & Software*, *101*, 301–316. https://doi.org/10.1016/j.envsoft.2018.01.001

Kuppel, S., Tetzlaff, D., Maneta, M. P., & Soulsby, C. (2018). EcH2O-iso 1.0: Water isotopes and age tracking in a process-based, distributed ecohydrological model. *Geoscientific Model Development*, *11*(7), 3045–3069. https://doi.org/10.5194/gmd-11-3045-2018

Kuppel, S., Tetzlaff, D., Maneta, M. P., & Soulsby, C. (2020). Critical zone storage controls on the water ages of ecohydrological outputs. *Geophysical Research Letters*, *47*(16), e2020GL088897. https://doi.org/10.1029/2020GL088897

Li, H., Sivapalan, M., Tian, F., & Liu, D. (2010). Water and nutrient balances in a large tile-drained agricultural catchment: A distributed modeling study. *Hydrology and Earth System Sciences*, *14*(11), 2259–2275. https://doi.org/10.5194/hess-14-2259-2010

Liebethal, C., & Foken, T. (2007). Evaluation of six parameterization approaches for the ground heat flux. *Theoretical and Applied Climatology*, *88*(1), 43–56. https://doi.org/10.1007/s00704-005-0234-0

Lin, B. B. (2011). Resilience in agriculture through crop diversification: Adaptive management for environmental change. *BioScience*, *61*(3), 183–193. https://doi.org/10.1525/bio.2011.61.3.4

Lozano-Parra, J., Maneta, M. P., & Schnabel, S. (2014). Climate and topographic controls on simulated pasture production in a semiarid Mediterranean watershed with scattered tree cover. *Hydrology and Earth System Sciences*, *18*(4), 1439–1456. https://doi.org/10.5194/hess-18-1439-2014

Lutz, S. R., Krieg, R., Müller, C., Zink, M., Knöller, K., Samaniego, L., & Merz, R. (2018). Spatial patterns of water age: Using young water fractions to improve the characterization of transit times in contrasting catchments. *Water Resources Research*, *54*(7), 4767–4784. https://doi.org/10.1029/2017WR022216

Maneta, M. P., & Silverman, N. L. (2013). A Spatially distributed model to simulate water, energy, and vegetation dynamics using information from regional climate models. *Earth Interactions*, *17*(11), 1–44. https://doi.org/10.1175/2012ei000472.1

McGuire, K. J., & McDonnell, J. J. (2015). Tracer advances in catchment hydrology. *Hydrological Processes*, *29*(25), 5135–5138. https://doi.org/10.1002/hyp.10740

McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M., & Woods, R. A. (2011). Hydrological field data from a modeler's perspective: Part 1. Diagnostic tests for model structure. *Hydrological Processes*, *25*(4), 511–522. https://doi.org/10.1002/hyp.7841

Mein, R. G., & Larson, C. L. (1973). Modeling infiltration during a steady rain. *Water Resources Research*, *9*(2), 384–394. https://doi.org/10.1029/WR009i002p00384

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*(2), 161–174. https://doi.org/10.2307/1269043

Newman, B. D., Wilcox, B. P., Archer, S. R., Breshears, D. D., Dahm, C. N., Duffy, C. J., et al. (2006). Ecohydrology of water-limited environments: A scientific vision. *Water Resources Research*, *42*(6). https://doi.org/10.1029/2005wr004141

Ollesch, G., Hermann, J., Meissner, R., & Reinstrof, F. (2010). Anthropogenic changes of water balance and runoff processes in a small low-mountain catchment. In A. Herrmann, & S. Schumann (Eds.), *Status and perspectives of hydrology in small basins. Proceedings of the international workshop held in Goslar-Hahnenklee* (pp. 65–70). Germany: IAHS.

Ollesch, G., Sukhanovski, Y., Kistner, I., Rode, M., & Meissner, R. (2005). Characterization and modeling of the spatial heterogeneity of snowmelt erosion. *Earth Surface Processes and Landforms*, *30*(2), 197–211. https://doi.org/10.1002/esp.1175

Orth, R., & Destouni, G. (2018). Drought reduces blue-water fluxes more strongly than green-water fluxes in Europe. *Nature Communications*, *9*(1), 3602. https://doi.org/10.1038/s41467-018-06013-7

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Pohlert, T. (2020). *Trend: Non-parameteric trend test and change-point detection, edited, R Package version 1.1.2*.

Porporato, A., & Calabrese, S. (2015). On the probabilistic structure of water age. *Water Resources Research*, *51*(5), 3588–3600. https://doi.org/10.1002/2015wr017027

Porporato, A., Feng, X., Manzoni, S., Mau, Y., Parolari, A. J., & Vico, G. (2015). Ecohydrological modeling in agroecosystems: Examples and challenges. *Water Resources Research*, *51*(7), 5081–5099. https://doi.org/10.1002/2015wr017289

Pütz, T., Kiese, R., Wollschläger, U., Groh, J., Rupp, H., Zacharias, S., et al. (2016). TERENO-SOILCan: A lysimeter-network in Germany observing soil processes and plant diversity influenced by climate change. *Environmental Earth Sciences*, *75*(18), 1242. https://doi.org/10.1007/s12665-016-6031-5

Reinstorf, F., Tiedge, J., Bauspiess, J., John, H., & Ollesch, G. (2010). Time series modeling in the Schaefertal catchment in the Lower Harz Mountains, central Germany. In A. Herrmann, & S. Schumann (Eds.), *Status and perspectives of hydrology in small basins. Proceedings of the international workshop held in Goslar-Hahnenklee* (pp. 249–253). Germany: IAHS.

Samaniego, L., Thober, S., Kumar, R., Wanders, N., Rakovec, O., Pan, M., et al. (2018). Anthropogenic warming exacerbates European soil moisture droughts. *Nature Climate Change*, *8*(5), 421–426. https://doi.org/10.1038/s41558-018-0138-5

Sawada, Y., Koike, T., & Jaranilla-Sanchez, P. A. (2014). Modeling hydrologic and ecologic responses using a new eco-hydrological model for identification of droughts. *Water Resources Research*, *50*(7), 6214–6235. https://doi.org/10.1002/2013wr014847

Schmidhuber, J., & Tubiello, F. N. (2007). Global food security under climate change. *Proceedings of the National Academy of Sciences*, *104*(50), 19703. https://doi.org/10.1073/pnas.0701976104

Schrön, M. (2017). *Cosmic-ray neutron sensing and its applications to soil and land surface hydrology*, PhD dissertation Thesis. University of Potsdam.

Seibert, J., & McDonnell, J. J. (2002). On the dialogue between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research*, *38*(11), 23-1–23-14. https://doi.org/10.1029/2001wr000978

Smith, A., Tetzlaff, D., Kleine, L., Maneta, M., & Soulsby, C. (2020). Isotope-aided modeling of ecohydrologic fluxes and water ages under mixed land use in Central Europe: The 2018 drought and its recovery. *Hydrological Processes*, *34*(16), 3406–3425. https://doi.org/10.1002/hyp.13838

Smith, A., Tetzlaff, D., Kleine, L., Maneta, M., & Soulsby, C. (2020). Upscaling land-use effects on water partitioning and water ages using tracer-aided ecohydrological models. *Hydrology and Earth System Sciences Discussions*, 1–36. *2020* https://doi.org/10.5194/hess-2020-539

Soulsby, C., Piegat, K., Seibert, J., & Tetzlaff, D. (2011). Catchment-scale estimates of flow path partitioning and water storage based on transit time and runoff modeling. *Hydrological Processes*, *25*(25), 3960–3976. https://doi.org/10.1002/hyp.8324

Sprenger, M., Stumpp, C., Weiler, M., Aeschbach, W., Allen, S. T., Benettin, P., et al. (2019). The demographics of water: A review of water ages in the critical zone. *Reviews of Geophysics*, *57*(3), 800–834. https://doi.org/10.1029/2018rg000633

Stadnyk, T. A., & Holmes, T. L. (2020). On the value of isotope-enabled hydrological model calibration. *Hydrological Sciences Journal*, *65*(9), 1525–1538. https://doi.org/10.1080/02626667.2020.1751847

Stewart, M. K., Morgenstern, U., & McDonnell, J. J. (2010). Truncation of stream residence time: How the use of stable isotopes has skewed our concept of streamwater age and origin. *Hydrological Processes*, *24*(12), 1646–1659. https://doi.org/10.1002/hyp.7576

Tetzlaff, D., Birkel, C., Dick, J., Geris, J., & Soulsby, C. (2014). Storage dynamics in hydropedological units control hillslope connectivity, runoff generation, and the evolution of catchment transit time distributions. *Water Resources Research*, *50*(2), 969–985. https://doi.org/10.1002/2013wr014147

Tetzlaff, D., Buttle, J., Carey, S. K., van Huijgevoort, M. H. J., Laudon, H., McNamara, J. P., et al. (2015). A preliminary assessment of water partitioning and ecohydrological coupling in northern headwaters using stable isotopes and conceptual runoff models. *Hydrological Processes*, *29*(25), 5153–5173. https://doi.org/10.1002/hyp.10515

Tetzlaff, D., Carey, S. K., McNamara, J. P., Laudon, H., & Soulsby, C. (2017). The essential value of long-term experimental data for hydrology and water management. *Water Resources Research*, *53*(4), 2598–2604. https://doi.org/10.1002/2017wr020838

Tetzlaff, D., McDonnell, J. J., Uhlenbrook, S., McGuire, K. J., Bogaart, P. W., Naef, F., et al. (2008). Conceptualizing catchment processes: Simply too complex? *Hydrological Processes*, *22*(11), 1727–1730. https://doi.org/10.1002/hyp.7069

van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., & Srinivasan, R. (2006). A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of Hydrology*, *324*(1), 10–23. https://doi.org/10.1016/j.jhydrol.2005.09.008

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., & Sorooshian, S. (2001). A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, *5*(1), 13–26. https://doi.org/10.5194/hess-5-13-2001

Weise, H., Auge, H., Baessler, C., Bärlund, I., Bennett, E. M., Berger, U., et al. (2020). Resilience trinity: Safeguarding ecosystem functioning and services across three different time horizons and decision contexts. *Oikos*, *129*(4), 445–456. https://doi.org/10.1111/oik.07213

Westerberg, I. K., & McMillan, H. K. (2015). Uncertainty in hydrological signatures. *Hydrology and Earth System Sciences*, *19*(9), 3951–3968. https://doi.org/10.5194/hess-19-3951-2015

Westra, S., Thyer, M., Leonard, M., Kavetski, D., & Lambert, M. (2014). A strategy for diagnosing and interpreting hydrological model nonstationarity. *Water Resources Research*, *50*(6), 5090–5113. https://doi.org/10.1002/2013wr014719

Wilhite, D. A., & Glantz, M. H. (1985). Understanding: the Drought Phenomenon: The Role of Definitions. *Water International*, *10*(3), 111–120. https://doi.org/10.1080/02508068508686328

Winsemius, H. C., Schaefli, B., Montanari, A., & Savenije, H. H. G. (2009). On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information. *Water Resources Research*, *45*(12). https://doi.org/10.1029/2009wr007706

Wollschläger, U., Attinger, S., Borchardt, D., Brauns, M., Cuntz, M., Dietrich, P., et al. (2016). The Bode hydrological observatory: A platform for integrated, interdisciplinary hydro-ecological research within the TERENO Harz/Central German Lowland Observatory. *Environmental Earth Sciences*, *76*(1), 29. https://doi.org/10.1007/s12665-016-6327-5

Yang, J., Heidbüchel, I., Musolff, A., Reinstorf, F., & Fleckenstein, J. H. (2018). Exploring the dynamics of transit times and subsurface mixing in a small agricultural catchment. *Water Resources Research*, *54*(3), 2317–2335. https://doi.org/10.1002/2017wr021896

Yang, X., Jomaa, S., Zink, M., Fleckenstein, J. H., Borchardt, D., & Rode, M. (2018). A new fully distributed model of nitrate transport and removal at catchment scale. *Water Resources Research*, *54*(8), 5856–5877. https://doi.org/10.1029/2017wr022380