Influence of input and parameter uncertainty on the prediction of catchment-scale groundwater travel time distributions

Miao Jing[1,2], Falk Heße[1], Rohini Kumar[1], Olaf Kolditz[3,4], Thomas Kalbacher[3], and Sabine Attinger[1,5]

[1]Department of Computational Hydrosystems, UFZ – Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany
[2]Institute of Geosciences, Friedrich Schiller University Jena, Burgweg 11, 07749 Jena, Germany
[3]Department of Environmental Informatics, UFZ – Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany
[4]Applied Environmental Systems Analysis, Technische Universität Dresden, Dresden, Germany
[5]Institute of Earth and Environmental Sciences, University of Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany

Correspondence: Miao Jing (miao.jing@ufz.de); Falk Heße (falk.hesse@ufz.de)

Abstract. Groundwater travel time distributions (TTDs) provide a robust description of the subsurface mixing behavior and hydrological response of a subsurface system. Lagrangian particle tracking is often used to derive the groundwater TTDs. The reliability of this approach is subjected to the uncertainty of external forcings, internal hydraulic properties, and the interplay between them. Here, we evaluate the uncertainty of catchment groundwater TTDs in an agricultural catchment using a 3-D groundwater model with an overall focus on revealing the relationship between external forcing, internal hydraulic properties, and TTD predictions. Eight recharge realizations are sampled from a high-resolution dataset of land surface fluxes and states. Calibration-constrained hydraulic conductivity fields (K_s fields) are stochastically generated using the null-space Monte Carlo (NSMC) method for each recharge realization. The random walk particle tracking (RWPT) method is used to track the pathways of particles and compute travel times. Moreover, an analytical model under the random sampling (RS) assumption is fitted against the numerical solutions, serving as a reference for the mixing behavior of the model domain. The StorAge Selection (SAS) function is used to interpret the results in terms of quantifying the systematic preference for discharging young/old water. The simulation results reveal the primary effect of recharge on the predicted mean travel time (MTT). The different realizations of calibration-constrained K_s fields moderately magnify or attenuate the predicted MTTs. The analytical model does not properly replicate the numerical solution, and it underestimates the mean travel time. Simulated SAS functions indicate an overall preference for young water for all realizations. The spatial pattern of recharge controls the shape and breadth of simulated TTDs and SAS functions by changing the spatial distribution of particles' pathways. In conclusion, overlooking the spatial nonuniformity and uncertainty of input (forcing) will result in biased travel time predictions. We also highlight the worth of reliable observations in reducing predictive uncertainty and the good interpretability of SAS functions in terms of understanding catchment transport processes.

1 Introduction

Travel/transit time distributions (TTDs) of groundwater provide a description of how aquifers store and release water and pollutants under external forcing conditions, which has significant implications for interdisciplinary environmental studies. For example, remarkable time-lags of the reaction of streamflow with outer forcings and considerable amounts of "old water" (i.e., water with an age of decades or longer) in streamflow have been observed in many studies (Howden et al., 2010; Stewart et al., 2012). Moreover, the legacy nitrogen in groundwater storage may dominate the annual nitrogen loads in agricultural basins (Wang et al., 2016; Van Meter et al., 2016; Van Meter et al., 2017). Groundwater TTDs offer important insights into the vulnerability of aquifers to pollution spreading, and they are critically important for the environmental assessment of non-point-source agricultural contamination (Böhlke and Denver, 1995; Böhlke, 2002; Molnat and Gascuel-Odoux, 2002; Eberts et al., 2012). TTDs shed light on the quantification of the long-term influence of agricultural contamination, which is crucial for water quality and sustainability.

The accurate quantification of groundwater travel time at a regional scale is extremely challenging. A primary difficulty is that the complex geometric, topographic, meteorologic, and hydraulic properties of hydrologic systems control the flow and mixing processes and therefore define the unique shape of the travel time distribution (TTD) (Leray et al., 2016; Hale and McDonnell, 2016; Engdahl et al., 2016). The other difficulty is that the groundwater system is intricately and tightly coupled to land surface hydrologic processes. The fundamental characteristics and the coupled nature determine the response of a catchment to outer forcings such as anthropogenic climate change, artificial abstraction, and agricultural and chemical contamination (Tetzlaff et al., 2014; van der Velde et al., 2015; Heße et al., 2017).

The techniques for determining groundwater TTDs can be categorized into two groups: geochemical approaches and numerical modeling approaches (McCallum et al., 2014). In geochemical approaches, the lumped parameter models are often used to interpret the catchment-scale observation of an environmental tracer concentration. Environmental tracer datasets can be divided into those representing the concentration distribution of young water (e.g., ^{3}H, SF_6, ^{85}K, and CFCs) and those representing the concentration distribution of old water (e.g., ^{36}Cl, ^{4}He, ^{39}Ar, and ^{14}C). Additionally, the analytical StorAge Selection (SAS) function is a cutting-edge tool for characterizing transport processes in lumped, time-varying hydrologic systems at the hillslope/catchment scale (Botter et al., 2011; Rinaldo et al., 2011; Van Der Velde et al., 2012; Harman, 2015; Danesh-Yazdi et al., 2018). This framework provides a clear distinction between the travel time (the time spent by a water parcel or a solute from its entrance to the control volume until its exit) and the residence time (the age of the water parcel or the solute existing in the control volume at a particular time). The SAS function has been successfully applied to interpret environmental tracer data through some assumptions of the mixing mechanism (Benettin et al., 2015, 2017). However, analytical approaches fall short in representing the dispersion of transport processes caused by catchment heterogeneity. Strong heterogeneity leads to significant aggregation error of mean travel times (MTTs) when using analytical models to interpret the tracer data (Kirchner, 2016; Stewart et al., 2016).

In contrast to such an analytical approach, physically based numerical models can explicitly describe the geometry, topography, and geological structures, and they can represent the flow paths of individual water particles. Physically based numerical

models are structurally complex and computationally expensive and often have more parameters than lumped parameter models do. These models can be classified as Eulerian approaches or Lagrangian approaches (Leray et al., 2016). The Eulerian approach directly solves the partial differential equations (PDEs) derived from mass conservation with "age mass" as the primary variable (Goode, 1996; Ginn, 2000; Engdahl et al., 2016). The Lagrangian approach, including the smoothed particle hydrodynamics (SPH) approach and the random walk particle tracking (RWPT) approach, is numerically robust and less restrictive on time-step size in solving advection-dominated problems (Tompson and Gelhar, 1990). Consequently, Lagrangian methods are more promising in simulating complex real-world transport processes, as they avoid spurious mixing error in grid-fixed Eulerian methods (Benson et al., 2017). Therefore, the Lagrangian approach has been widely used to simulate large-scale reactive transport and biogeochemical problems (Park et al., 2008; de Rooij et al., 2013; Selle et al., 2013).

A reliable application of groundwater transport modeling is subject to many sources of uncertainty, including measurement, model structural, and parameter uncertainty (Beven, 1993). Specifically, the reliability of model prediction suffers from the uncertainty of external forcings, the uncertainty of internal hydraulic characteristics, and the interplay between them (Ajami et al., 2007). The spatially sparse measurements of recharge lead to a biased characterization of spatiotemporal patterns of recharge (Healy and Scanlon, 2010; Cheng et al., 2017). On the other hand, the spatial scarcity of hydrogeological data always hampers the right characterization of aquifer properties such as porosity and permeability, thus allowing a range of various realistic parameter values. The best-fit parameter may suffer from a fitting error caused by overparameterization and equifinality (Schoups et al., 2008). Such biased parameters cause uncertain predictions because parameter error may compensate for model structural defects (Doherty, 2015). Accordingly, predictive uncertainty can be hardly assessed in a precise way.

Biased characterization of the hydrodynamic system and oversimplified assumptions will lead to a problematic prediction of TTDs. Many past studies offer insights into the influence of recharge and hydrogeological configuration on the prediction of TTDs. For example, some research studies have been devoted to the development of analytical solutions for the idealized catchment (or aquifer) under some essential assumptions and simplifications (Neuman, 1972; Haitjema, 1995; Engdahl et al., 2016; Leray et al., 2016). Among them, Haitjema (1995) derived an analytical solution in an idealized groundwatershed under steady-state conditions and the Dupuit-Forchheimer assumption and found that the groundwater mean travel time appears to be only dependent on recharge, saturated aquifer thickness, and porosity, provided that the hydraulic conductivity is locally homogeneous. Basu et al. (2012) evaluated analytical, GIS, and numerical approaches for the prediction of groundwater TTDs and found that the simulated TTDs show a moderate difference. Many recent studies have reported the dependency of transient TTDs on the temporal variability of input forcings (Benettin et al., 2015; Yang et al., 2018; Remondi et al., 2018; Kaandorp et al., 2018), but the dependency of TTDs (as well as SAS functions) on the spatial variability of input forcings has rarely been studied.

Although studies on catchment-scale groundwater TTDs are numerous, comprehensive uncertainty analysis that aims to unveil the different roles of external forcing and internal hydrostratigraphic structure using both a numerical model and SAS functions is scarce. In this regard, two important questions are the following: (1) How does the uncertainty of recharge (including its spatial nonuniformity) and hydraulic conductivities affect the TTD predictions in a mesoscale agricultural catchment,

provided that the model is constrained to reality and groundwater head observations? (2) How does the uncertainty of inputs (forcings) and parameters influence the prediction of systematic preference for young/old water?

In this paper, we aim to answer these questions through a detailed (uncertainty) analysis of an example application in a mesoscale real-world catchment. In doing so, we establish a detailed groundwater model coupled to a random walk particle tracking system for predicting groundwater TTDs. The groundwater model OpenGeoSys (OGS) is used to simulate the groundwater flow, while the input forcing is fed by the mesoscale hydrologic model (mHM) via the coupling interface mHM-OGS (Jing et al., 2018). The numerical model follows the steady-state assumption of groundwater flow systems. This assumption is made because at the regional scale, the groundwater flow process has a much larger time scale than that of the high-frequency oscillation of recharge, which essentially dampens the effect of recharge oscillation (Leray et al., 2016). An ensemble of simulations using multiple recharge realizations and multiple equifinal hydraulic conductivity (K_s) fields is established. An analytical model is used as a reference for unveiling the mixing mechanism of the system. The StorAge Selection function is also used to interpret the simulation results of the numerical model, with an overall aim to quantify the predictive uncertainty of systematic preference for young/old water.

2   Site Description

The candidate site in this paper is the Nägelstedt catchment, located in central Germany (see Figure 1). With an area of approximately 850 km^2, the Nägelstedt catchment is a headwater catchment of the Unstrut river. The terrain elevation of this area varies from 164 m to 516 m a.m.s.l. (above mean sea level). It is a subcatchment of the Unstrut basin, one of the most intensively used agricultural regions in Germany. About 88% of the land in this site is marked as arable land, which is significantly higher than the average level of Thuringian (Wechsung et al., 2008). The agricultural nitrogen input has varied over the years and locations, from 5 - 24 kg/ha in the soils of the lowlands to 2 - 30 kg/ha in the feeding area (Wechsung et al., 2008). The mean annual precipitation is approximately 660 mm.

The dominating sediment in the study area is the Muschelkalk (Middle Triassic). The Muschelkalk has an overall thickness of about 220 m, and it has been divided into three subgroups according to mineral composition: Upper Muschelkalk (mo), Middle Muschelkalk (mm), and Lower Muschelkalk (mu). The Upper Muschelkalk (mo) is mainly composed of limestone, marlstone and claystone, and it forms fractured aquifers (Seidel, 2003; Jochen et al., 2014; Kohlhepp et al., 2017). The Middle Muschelkalk (mm) deposits are composed of evaporites, including dolomit marlstone, gypsum, dolomit limestone and eroded salt layers. The Lower Muschelkalk (mu) is composed of massive limestone (Seidel, 2003; McCann, 2008; Jochen et al., 2014). The Muschelkalk formation consists of limestone sediments, which may form fractured and karst aquifers. A recent study demonstrated that karstification and the development of conduits are limited at the base of the Upper Muschelkalk at the Hainich critical zone in the Nägelstedt catchment (Kohlhepp et al., 2017). In the middle of the study area, Keuper deposits, including Middle Keuper (km) and Lower Keuper (ku), overlay the Muschelkalk formation. The Lower Keuper formation forms the low permeable aquitard on top of the Upper Muschelkalk aquifer (Seidel, 2003; Kohlhepp et al., 2017). Lithologically, the
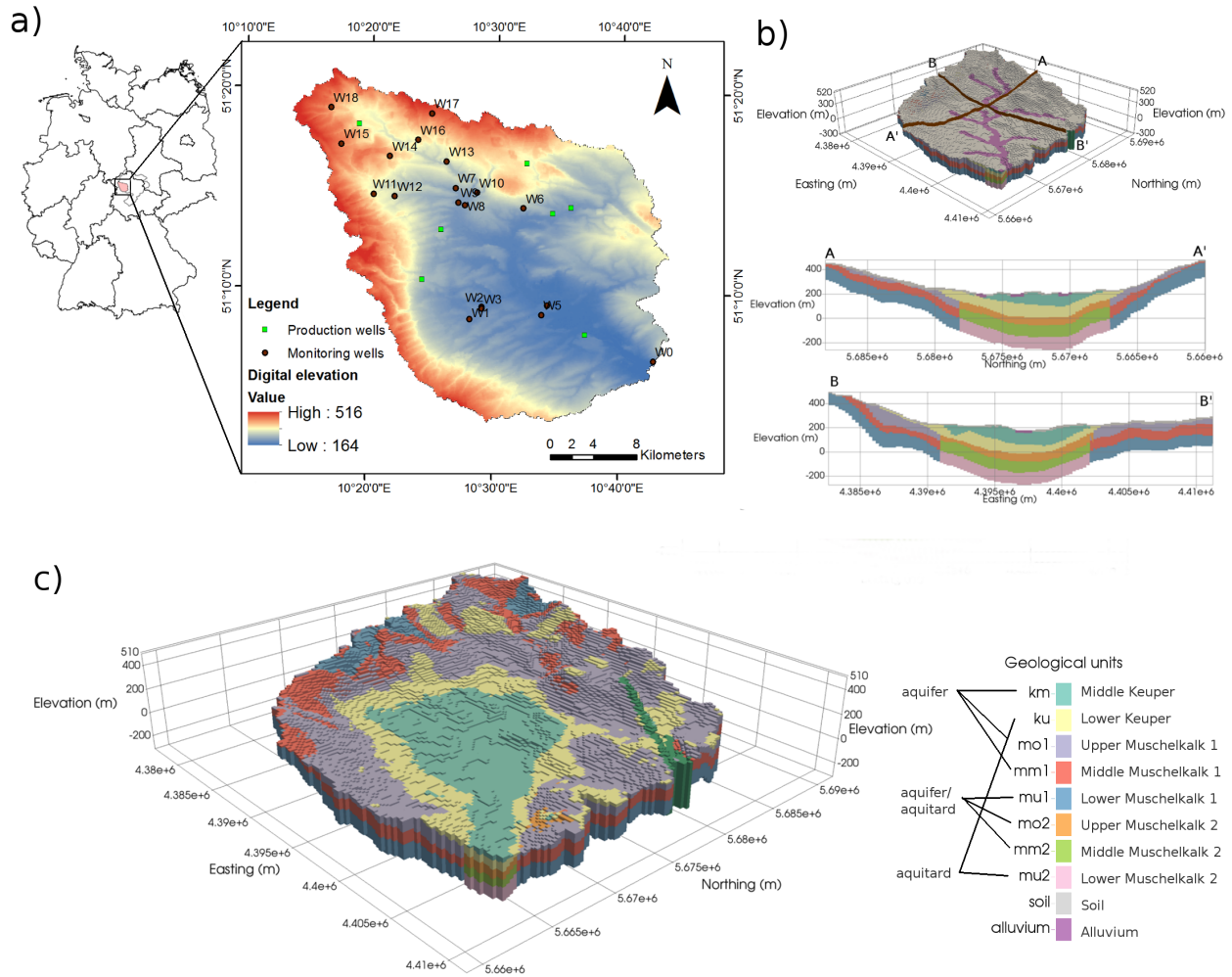
Figure 1. The Nägelstedt catchment used as the test catchment for this study (Jing et al., 2018). a) An overview of the Nägelstedt catchment and the locations of the monitoring wells used in this study. b) 3-D view highlighting the arrangement of alluvium and soil and cross-sectional view of the study area. c) 3-D view highlighting the zonation of the sedimentary aquifer-aquitard system. Note that the Muschelkalk layers (mo, mm and mu) are divided into more permeable subunits (mo1, mm1 and mu1) and less permeable subunits (mo2, mm2, and mu2).

Muschelkalk aquifer system is a "layer-cake" aquifer system that contains interbedded marlstone aquitards (Aigner, 1982; Merz, 1987; Kohlhepp et al., 2017).

Eighteen monitoring wells distributed in this area are used to calibrate the model (Figure 1a, in which the well W0 is abandoned in this study due to the proximity to the outer edge). The geological layers to which the wells belong are listed as follows: 5 wells in Middle Keuper (km), 4 in Lower Keuper (ku), 6 in Upper Muschelkalk (mo), 2 in Middle Muschelkalk (mm), and 1 in alluvium.

3 Methodology and materials

3.1 Numerical model

We use the coupled model mHM-OGS, proposed by Jing et al. (2018), to simulate terrestrial hydrological processes. This coupled model was developed for extending the predictive capability of mHM from land surface processes to the subsurface flow and transport processes. Specifically, mHM (Samaniego et al., 2010; Kumar et al., 2013) is used to partition water budget components, while the porous media simulator OGS (Kolditz et al., 2012) is used to compute groundwater flow and transport processes by using mHM-generated recharge as driving forces. For details on the coupled model mHM-OGS, please refer to Jing et al. (2018).

The catchment water storage is conceptually partitioned into soil zone storage and deep groundwater storage; the two corresponding components are computed by mHM and OGS, respectively. The soil zone dynamics of TTD has been well studied using mHM in a previous work (Heße et al., 2017). Hence, in this paper, we perform explicit forward modeling of the saturated-zone TTD through a 3-D OGS groundwater model by using the mHM-generated recharge as the external forcing.

In this study, we focus on the travel times in the saturated zone. Saturated groundwater flow is characterized by the continuity equation and Darcy's law:

$$S\frac{\partial \psi_p}{\partial t} = -\nabla \cdot \boldsymbol{q} + q_s \tag{1}$$

$$\boldsymbol{q} = -K_s \nabla(\psi_p - z) \tag{2}$$

where S is the specific storage coefficient in confined aquifers, or the specific yield in unconfined aquifers [1/L]; ψ_p is the pressure head in the porous medium [L]; t is the time [T]; $\boldsymbol{q}$ is the specific discharge or the Darcy velocity [LT^{-1}]; q_s is the volumetric source/sink term [T^{-1}]; K_s is the saturated hydraulic conductivity tensor [LT^{-1}]; and z is the vertical coordinate [L].

We use the RWPT method to track the particle movement. The RWPT method is embedded in the source codes of OGS (Kolditz et al., 2012; Park et al., 2008). Derived from stochastic physics, RWPT is under the assumption that the advection process is deterministic and the diffusion-dispersion process is stochastic. The theoretical background of the RWPT method is described in detail in Appendix A.

3.2 Numerical model setup

3.2.1 Boundary conditions

The steady-state model configuration is achieved using a temporally averaged recharge of the simulated daily recharges over a long period (1955–2005). The gridded recharges estimated by mHM are interpolated and then assigned to each grid node on the upper surface of the OGS mesh using a bilinear interpolation approach. No-flow boundaries are assumed at the outer edges that are defined by catchment divides, except for the northwestern and northeastern edges, where fixed-head boundaries are applied (Wechsung et al., 2008). The streams are assigned with fixed-head boundaries, wherein the heads are equal to the long-term averaged water levels. For the steady-state system and the one-way coupled model, baseflow component generated

Table 1. Adjustable ranges of the hydraulic parameters.

	Hydraulic conductivity K_s [m/s]						
	soil	alluvium	km	ku	mo1	mm1	mu1
Upper limit	9.0×10^{-4}	2.0×10^{-3}	9.0×10^{-4}	8.5×10^{-5}	8.0×10^{-4}	9.1×10^{-4}	2.0×10^{-5}
Lower limit	5.0×10^{-5}	4.5×10^{-6}	1.0×10^{-6}	9.6×10^{-7}	9.0×10^{-7}	3.1×10^{-7}	2.0×10^{-8}
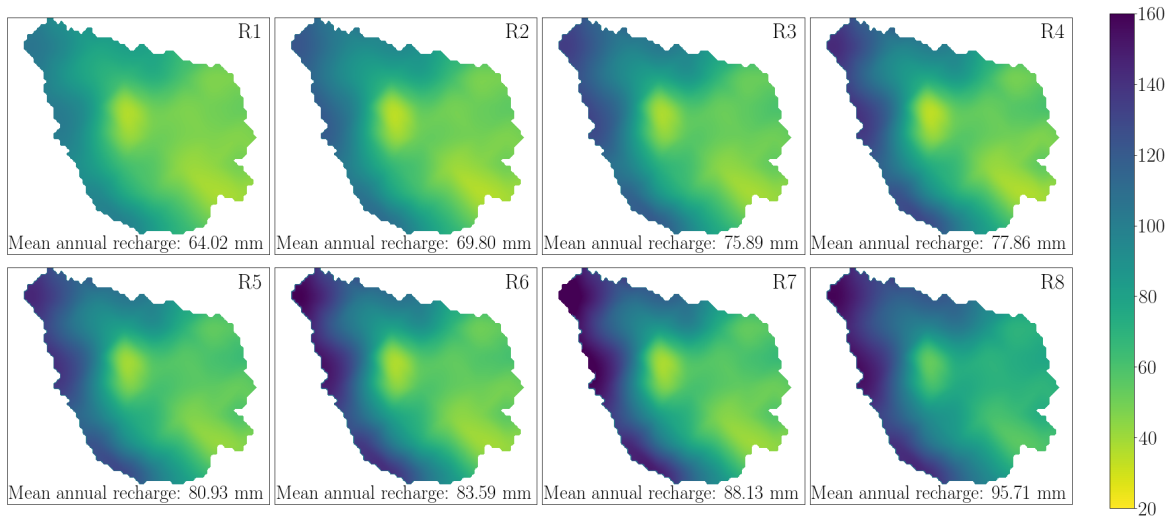
Figure 2. Recharge realizations used in this study (unit: mm). They were sampled from a high-resolution dataset of land surface fluxes for Germany (Zink et al., 2017).

by OGS proves to be consistent with the baseflow estimated by mHM, implying that the water budget in the subsurface system is essentially closed (Jing et al., 2018). Neumann boundaries are prescribed for 7 drinking water production wells. However, the amount of pumping makes up only around 3% of the total amount of outflow, and it therefore has a marginal influence on the water budget (Wechsung et al., 2008).

3.2.2 Modeling procedures

The numerical experiment to explore the uncertainty of TTDs is performed through the following workflow:

1. Eight spatially distributed recharge realizations are sampled from a high-resolution dataset of land surface fluxes for Germany, in which mHM is used to simulate land surface hydrological processes. The details of the dataset and the sampling method are described in the following section.

2. For each recharge realization, a series of equally probable realizations of K_s fields are generated using the null-space Monte Carlo (NSMC) method.

The NSMC method takes advantage of the hybrid Tikhonov-TSVD method in the parameter estimation code PEST to produce Monte Carlo realizations of parameters (Tonkin and Doherty, 2009; Doherty and Hunt, 2010). This approach is able to efficiently generate an ensemble of parameter fields that are conditioned to expert knowledge and measurements. Here, the observations of groundwater levels from 18 spatially distributed monitoring wells are used to calibrate the OGS model (the locations of monitoring wells are illustrated in Figure 1a). Before generating parameter sets, we calibrate the model to obtain the best-fit hydraulic conductivities, as well as a covariance matrix of the parameter probability distributions. On the basis of this information, many fully distributed K_s fields are randomly generated from a uniform distribution of hydraulic conductivity values (Doherty, 2015). As shown in Table 1, the range of hydraulic conductivities is predefined based on values obtained from a geological survey (Wechsung et al., 2008). As a result, a total of 400 parameter sets conditioned on both observations and reality are generated for the uncertainty analysis.

3. In each parameter realization, a large number of particles are injected through the top surface of the groundwater model. The spatial density of particles is proportional to the spatially distributed recharge rates.

To accurately interpret the travel time distribution, a large number of particles (e.g., approximately 80 000 particles in the case study) is released into the top surface of the groundwater model. The released particles serve as samples of water parcels for deriving their travel time distributions. In doing so, the density of particles is set proportional to the recharge at the corresponding grid cell (Figure 3). Each particle tracer represents a volumetric recharge rate of around 700 m^3/year.

4. An ensemble of forward simulations using the RWPT method is performed over all realizations of K_s fields.

In each realization of the ensemble parameter sets, forward simulations of particle tracking are performed. In this study, we focus on the predictive uncertainty within the convection process. Therefore, the molecular diffusion coefficients are universally set to 0 for all ensemble simulations. The porosity of the study domain is set to 0.2 universally. Through the above procedures, the flow paths and the corresponding residence times can be fully traced in the model at random times and locations, facilitating the detailed characterization of TTDs.

In parallel to this analysis, a sensitivity analysis for the spatial variability of recharge is also performed. Two different recharge scenarios are compared for this purpose: (1) the spatially distributed recharge generated by mHM and (2) the uniform recharge that is equal to the spatial average of the distributed recharge. Other parameters, including the porosity and the hydraulic conductivity, remain identical in these two recharge scenarios.

3.2.3 Recharge realizations

A high-resolution dataset of land surface fluxes and states across Germany is used for sampling recharge scenarios. This dataset was established on the basis of a daily simulation with 4 km spatial resolution using mHM for a time span of 60 years (1951–2010) (Zink et al., 2017). This dataset consists of an ensemble (100 realizations) of land surface variables, including evapotranspiration, groundwater recharge, soil moisture and discharge, with a spatial resolution of 4 km. A total of 100 re-
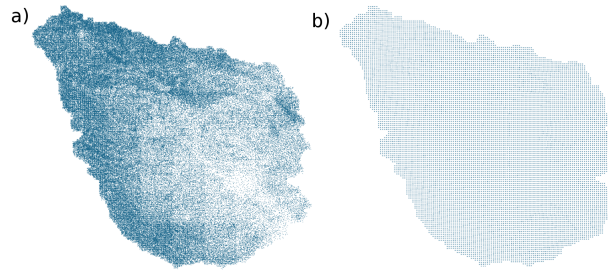
Figure 3. Two different spatial distributions of particle tracers for the RWPT method. a) The mass-weighted distribution of particles based on the recharge estimated by mHM. This is the default spatial pattern of particle tracers in this study. b) The uniformly distributed particle tracers used in the uniform recharge scenario.

alizations of land surface states are all conditioned on the observed daily discharge, and each of them has been derived by incorporating the uncertainty of parameterization caused by the heterogeneity of geometry, topography and geology. The modeled datasets are further validated against observation-based evapotranspiration and soil moisture data from eddy covariance stations (Heße et al., 2017). The derived recharges show a good correspondence with the estimation from the Hydrologic Atlas of Germany (Zink et al., 2017).

Eight representative recharge realizations (R1-R8) are sampled from 100 realizations for this study to save computational time. To enhance the representativeness of the samples, the 100 recharge realizations are sorted in an ascending order by their spatial averages. The selected recharge realizations are uniformly sampled from the sorted recharge realizations. In doing so, the maximum and minimum recharges are included in the samples such that the whole range of recharge realizations is fully covered.

3.2.4 3-D stratigraphic model

A 3-D stratigraphic mesh is established on the basis of hydrogeological characterizations elaborated in Section 2 (Figure 1). This mesh is based on data obtained from Thuringian State office for the Environment and Geology (TLUG) and its generation is described in Fischer et al. (2015). The structured mesh is composed of 310 599 nodes (132 rows, 140 columns, and 82 vertical layers). The 3-D cell size of 250 m, 250 m, and 10 m in the x-, y- and z-directions is used in this study. Based on the German stratigraphy (Menning, 2002), the Middle Muschelkalk, the Upper Muschelkalk, the Lower Keuper, and the Middle Keuper outcrop in the Nägelstedt catchment. Accordingly, a stratigraphic aquifer system with 10 geological units is set up. The uppermost 10 m of the mesh has been separated as a soil layer, while an alluvium layer consisting of high-permeability sandy gravel is set at the nodes beneath and near streams (Figure 1). Each of the Muschelkalk layers is further divided into two categories: the more permeable parts (mo1, mm1, and mu1) and the less permeable parts (mo2, mm2, and mu2) (see Figure 1). For each of the Muschelkalk units, the permeability of the less permeable part is tied to the corresponding more permeable part with a factor of 0.1. The equivalent porous medium approach is applied to characterize the karst aquifer of the Upper Muschelkalk (mo). This approach translates the parameters describing highly heterogeneous hydraulic properties at the point
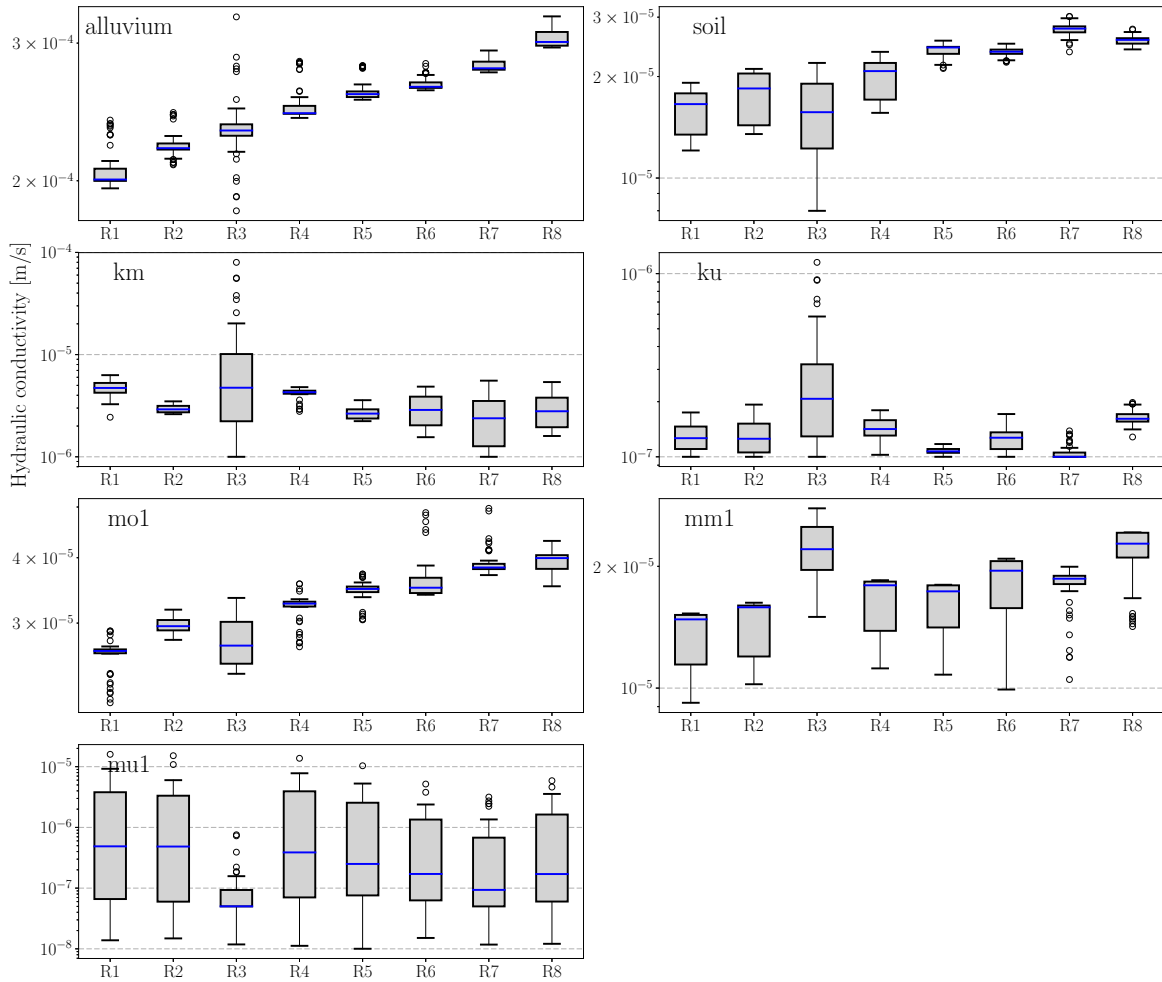
Figure 4. Box-plot of stochastically generated hydraulic conductivity (K_s) for each geological layer in 8 recharge realizations. Note that the parameters mo2, mm2, and mu2 are not shown in this figure because the less permeable subunits of the Muschelkalk (mo2, mm2, and mu2) are tied with the respective more permeable subunits (mo1, mm1, and mu1) with a factor of 0.1.

scale to the equivalent homogeneous medium at the regional scale to avoid adding redundant parameters and therefore avoid overfitting.

3.2.5 Parameter uncertainty

Multiple calibration-constrained K_s fields were stochastically generated for each recharge realization. Figure 4 shows the box-plot of generated hydraulic conductivities in all realizations categorized by geological unit. The hydraulic conductivity of the Lower Muschelkalk (mu) has the highest uncertainty (10^{-8} - 10^{-5} m/s) because the conductivity of mu is insensitive to groundwater head observations, given that it is the deepest geological layer and that no monitoring well is located in this
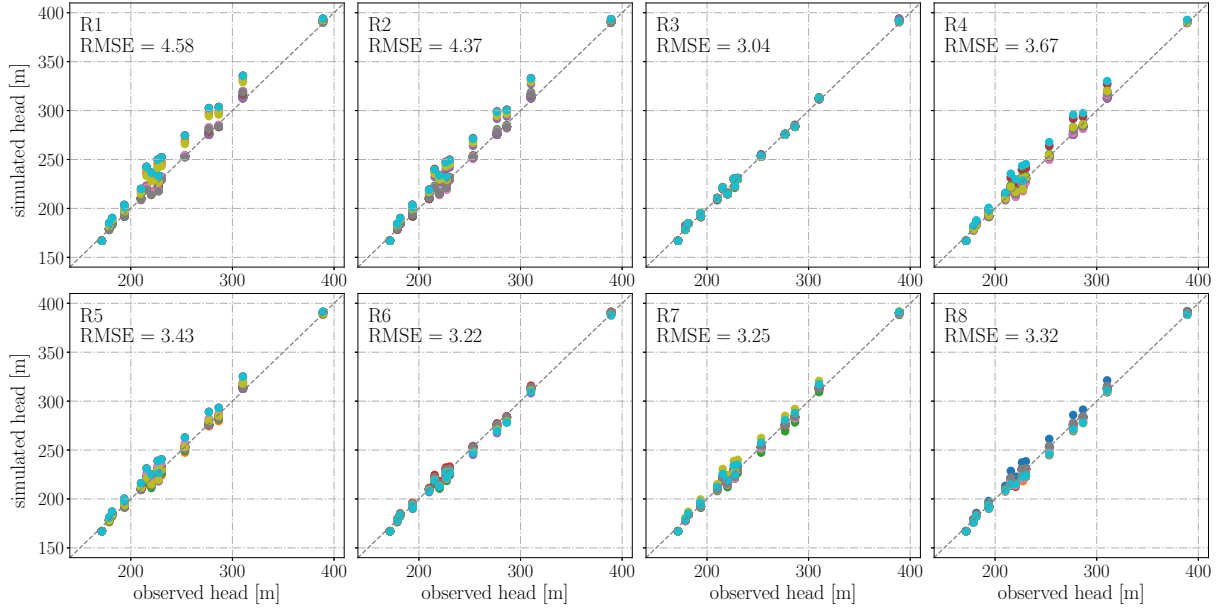
Figure 5. Observed and simulated groundwater heads for each parameter and recharge realization. The results of 400 realizations (R1K1 - R8K50) are categorized by recharge realization and shown in different panels.

layer (Table B1). The other parameters fluctuate moderately and are constrained within one order of magnitude in most of the recharge realizations. Hydraulic conductivities of several permeable layers (mo, mm, alluvium, and soil) increase from R1 to R8, which is not surprising because the hydraulic conductivity increases with increasing recharge and constant groundwater head. Moreover, the hydraulic conductivities of the above layers are roughly linearly correlated to the corresponding recharge

5 in each recharge realization. Figure 5 shows the simulated and observed groundwater heads for all 400 realizations. All of the 400 realizations are well constrained to observations, with the root mean square error (RMSE) of groundwater level residuals being lower than 4.6 m in all of the considered recharge realizations.

3.3 Theory of analytical StorAge Selection function

The travel time is defined as the time spent by a moving element (either a water particle or a solute) in a control volume of

10 a hydrologic system. In principle, the control volume can be defined at arbitrary spatial scales (i.e., from the molecular scale to the regional scale). Considering a hydrologic system in which the input flux (J) and the output fluxes ($Q_1, Q_2, ..., Q_n$) are known, each parcel of water within the system is tagged using its current age τ. The age-ranked storage $S_T = S_T(T, t)$ is defined as the mass of water in the system with age $\tau < T$. The backward form of the master equation (ME) for TTD in a control volume can be expressed as follows (Botter et al., 2011; Van Der Velde et al., 2012; Harman, 2015):

15 $$\frac{\partial S_T}{\partial t} = J(t) - \sum_{j=1}^{n} Q_j(t)\overleftarrow{P}_{Q_j}(T,t) - \frac{\partial S_T}{\partial T} \tag{3}$$

with boundary condition $S_T(0,t) = 0$, where T is the residence time of the oldest water parcel in storage S_T; t is the chronological time; $\overleftarrow{P}_{Q_j}(T,t)$ is the cdf of the backward travel time distribution of output flux Q_j; $J(t)$ is the input flux at time t; and $Q_j(t)$ is the output flux at time t. Specifically in this study, J is the groundwater recharge, and Q is composed of two components: the stream baseflow and the abstraction at production wells.

5   The StorAge Selection (SAS) function describes the fraction of water parcels leaving the control volume at time t, which is selected from the age-ranked storage S_T. Following the above definition, the SAS function can be linked with the backward travel time distribution $\overleftarrow{P}_Q(T,t)$ (Harman, 2015):

$$\Omega_Q(S_T,t) = \overleftarrow{P}_Q(T,t) \tag{4}$$

for $S_T = S_T(T,t)$. Ω_Q is the cumulative form of the StorAge Selection (SAS) function.

10   Three instances of SAS functions using gamma distribution are shown in Figure 9a. In case the age distribution of each outflow is uniformly selected from all water storages with various ages, the outflux TTDs turn into a random sample of the storage residence time distribution (RTD). The random sampling (RS) is equivalent to the uniform SAS function (Figure 9a). Many past studies have also considered the random sampling as a proper description of the sampling behavior for heterogeneous catchments (Cartwright and Morgenstern, 2015; Benettin et al., 2015; Heße et al., 2017). Eq. (4) in this case has the analytical

15 solution (Harman, 2015; Danesh-Yazdi et al., 2018):

$$p_S(T,t) = \overleftarrow{p}_Q(T,t) = \frac{J(t-T)}{S(t)} \exp\Big[-\int_{t-T}^{t} \frac{Q(\tau)}{S(\tau)} d\tau \Big] \tag{5}$$

where $p_S(T,t)$ is the pdf of the residence time distribution and $S(t)$ is the storage at time t. Specifically, in the case of a steady-state hydrodynamic system, Eq. (5) is further simplified into an exponential form:

$$\overleftarrow{p}_Q(T) = \frac{J}{S} \exp\Big(-\frac{J}{S} T \Big) \tag{6}$$

20 Eq. (6) is the analytical solution of backward TTD under the RS assumption.

     In the idealized saturated groundwater aquifer, Eq. 6 is equivalent to the analytical solution derived by Haitjema (1995). Based on Dupuit–Forcheimer's assumption, Haitjema (1995) derived a formula about the frequency distribution of residence time:

$$p_s(T) = \frac{1}{\overline{T}} \exp\Big(-\frac{T}{\overline{T}} \Big) \tag{7}$$

25    $$\overline{T} = \frac{nH}{J} \tag{8}$$

provided that nH/J is constant over the entire domain, the recharge is spatially uniform, and the aquifer is locally homogeneous, where n is the porosity, H is the saturated aquifer thickness, and $\overline{T}$ is the weighted mean travel time in the aquifer.

3.4   Linking the SAS functions to the physically based numerical model

Danesh-Yazdi et al. (2018) developed an approach to link the analytical SAS functions to the fully distributed numerical model.

30 Although differing in numerical model and particle tracking scheme, we apply the same approach to link the SAS functions

and the numerical model in this study. Eq. (3) under the steady-state condition can be further simplified as:

$$\frac{\partial S_T}{\partial T} = Q(1 - \Omega_Q(S_T)).$$

(9)

By combining Eq. (4) and Eq. (9), the age-ranked storage S_T can be calculated directly under the steady-state assumption:

$$S_T(T) = Q\Big(T - \int_0^T \overleftarrow{P}_Q(\tau)d\tau\Big).$$

(10)

Combining Eq. 10 with Eq. 4, the SAS function can be directly computed from the simulated backward TTD using numerical model.

3.5 Predictive uncertainty of TTDs

The theoretical framework of predictive uncertainty in this paper is based on Doherty (2015). As indicated in Bayes' theorem, the parameters of a model retain uncertainty, given that they have been adjusted to best-fit values achieved during calibration. Nevertheless, the uncertainty of parameters is subject to constraints. One of the constraints resides in the fixed adjustable range of parameters, in which expert knowledge must be respected. Another constraint is exerted by the parameterization process.

While the computationally expensive Bayesian approach offers a complete theoretical framework for predictive uncertainty evaluation, practical modeling efforts are often based on model calibration and a subsequent analysis of error or uncertainty in post-calibration predictions (Doherty, 2015). Ideally, the best-fit parameters achieved through calibration can reduce the predictive error to a minimum, with the minimum predictive error being the inherent uncertainty. However, the best-fit parameter is always biased from the true parameter because the essential imperfection of the model may facilitate or hamper the achievement of the minimum. Therefore, the motivation of uncertainty analysis in this study is to quantify and minimize the predictive uncertainty of travel time distributions, given that the parameters are plausible and that the model can reproduce well the groundwater heads.

4 Results

For the sake of clarity, we number the recharge realizations from R1 (with the lowest recharge rate) to R8 (with the highest recharge rate). For each recharge realization, 50 K_s fields are numbered from K1 to K50. Accordingly, R1K1 represents the K_s field K1 in the recharge realization R1.

4.1 Uncertainty of TTD predictions

Flow paths of particle tracers in a random parameter realization (R5K1) are displayed in Figure 6, serving as a visual reference for the regional groundwater flow pattern and the residence time distributions. In this realization, the deep low-permeability geological layers (e.g., mm2 and mu2) act as low-permeability aquitards. Therefore, the majority of streamlines do not enter these geological layers (Figure 6).
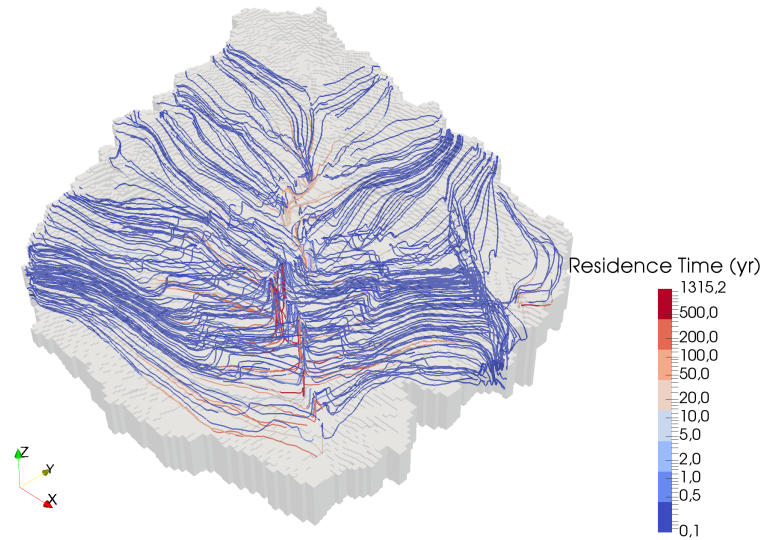
Figure 6. 3-D view of flow pathlines of some particles in realization R5K1. Note that only a limited number of particle pathlines are displayed here.

Figure 7 displays both the simulated TTDs using 50 K_s fields for 8 recharge realizations (orange solid lines) and the reference TTDs, represented by fitted blue dash-dot curves using the exponential model (Eq. 6). The ensemble average (μ) and the coefficient of variation (c_v) of MTTs for each recharge realization are also calculated and shown in Figure 7. Note that if the number of parameter realizations is sufficiently large, the ensemble average of MTTs (μ) will converge to the simulation result using the best-fit parameters achieved through model calibration (Doherty, 2015). Noticeable variability of TTDs can be observed with respect to different recharge realizations. Generally, the μ values show a decreasing trend from 166.5 yrs in recharge realization R1 to 110.9 yrs in recharge realization R8, with only two exceptions (R3 and R6), which is not surprising based on the inversely linear dependency between recharge J and μ, indicated by Eq. 8. In each recharge realization, the different realizations of K_s fields manipulate the mean travel time. c_v varies from 7.81% (for R5) to 15.56% (for R3), indicating a modest degree of uncertainty propagated from K_s estimation to TTD prediction.

The exponential model under the RS assumption is fitted to the ensemble averaged TTD of numerical solutions (see black lines in Figure 7) using Eq. 6. As shown in Figure 7, the shape of numerically simulated TTDs significantly deviate from the exponential distribution under the RS assumption, indicating a nonuniform sampling behavior of different water ages. The TTDs of numerical simulations are more right-skewed than the analytical TTDs under the RS assumption. This phenomenon reveals that the catchment TTD cannot be replicated by the single random sampling store.

Based on Eq. 8, we can approximate the "effective volume" of water involved in the transport process in the aquifer. The effective volume of groundwater storage related to the transport process is calculated and shown in Table 2. The effective volumes of storage (S_eff) estimated by the numerical solutions range from 9.8 m to 12.0 m, whereas the S_eff estimated by the analytical solution range from 6.8 m to 7.5 m. The groundwater storage that contributes to the streamflow is significantly

Table 2. Effective groundwater storages related to the transport process for each recharge realization.

	Effective volume of storage [m]								
	R1	R2	R3	R4	R5	R6	R7	R8	mean
S_{eff} (numerical)	10.7	10.6	12.0	10.5	9.8	10.2	10.2	10.6	10.6
S_{eff} (analytical)	6.9	6.9	7.5	7.1	6.8	6.9	6.9	7.2	7.0

smaller than the total groundwater storage (48.3 m). This difference is because most of the released particles only exist in the upper permeable layers rather than spread evenly over the whole aquifer/aquitard system. We are aware that this is only a first-order approximation because the analytical solution is only rigorously valid for the idealized homogeneous aquifer system (Haitjema, 1995).

5   Moreover, we assess the propagation to the MTT predictions from input and parameter uncertainty yielded by the 8 recharge realizations and the Monte Carlo realizations of hydraulic conductivities. Figure 8a depicts the distribution of MTTs of the ensemble simulations. The MTTs of the 400 realizations range from 87 to 212 yrs. At the same time, the ensemble average of MTTs over all realizations of recharges and K_s fields is 135.1 yrs, and the coefficient of variation is 18.93%. Figure 8b depicts the relationship between the ensemble average of MTTs and the spatially averaged recharge rates. We observe a roughly

10 inversely proportional relationship between the ensemble average of MTTs and the spatially averaged recharge rates (Figure 8b). The standard deviations (σ) of simulated MTTs range from 12.9 yrs (for R6) to 24.7 yrs (for R3).

4.2   Uncertainty of young/old water preference

Figure 9a provides an intuitive illustration of the relationship between the cumulative rank SAS functions and the systematic preference for discharging water with different ages. Figure 9b shows the cumulative rank SAS functions of all ensemble

15 simulations (obtained from 400 realizations of K_s fields in 8 recharge scenarios). The figure is categorized into 8 groups by different colors and line styles, each representing a recharge realization. Generally, the system has a weak preference to select younger water as discharge, despite different recharge realizations and K_s realizations. Nevertheless, a moderate variation in SAS functions for the ensemble simulations can be observed. The variation of interest is introduced by the spatial distribution and velocity of flow pathlines that are controlled by different K_s fields. For example, a more permeable shallow aquifer layer

20 will activate more flow pathways in this layer and thus will introduce a stronger preference for young water. Particularly, the significant variation in hydraulic conductivities in the deepest geological layer (e.g., Lower Muschelkalk) has a pronounced impact on the discharge of old water. With a thickness of up to 100 m, the hydraulic conductivity of this layer controls how many water parcels can enter into this layer and how deep the flow paths can develop. This effect can be evidenced by a large difference in the SAS functions related to old ages and a relatively smaller difference in those related to young ages in Figure

25 9b.
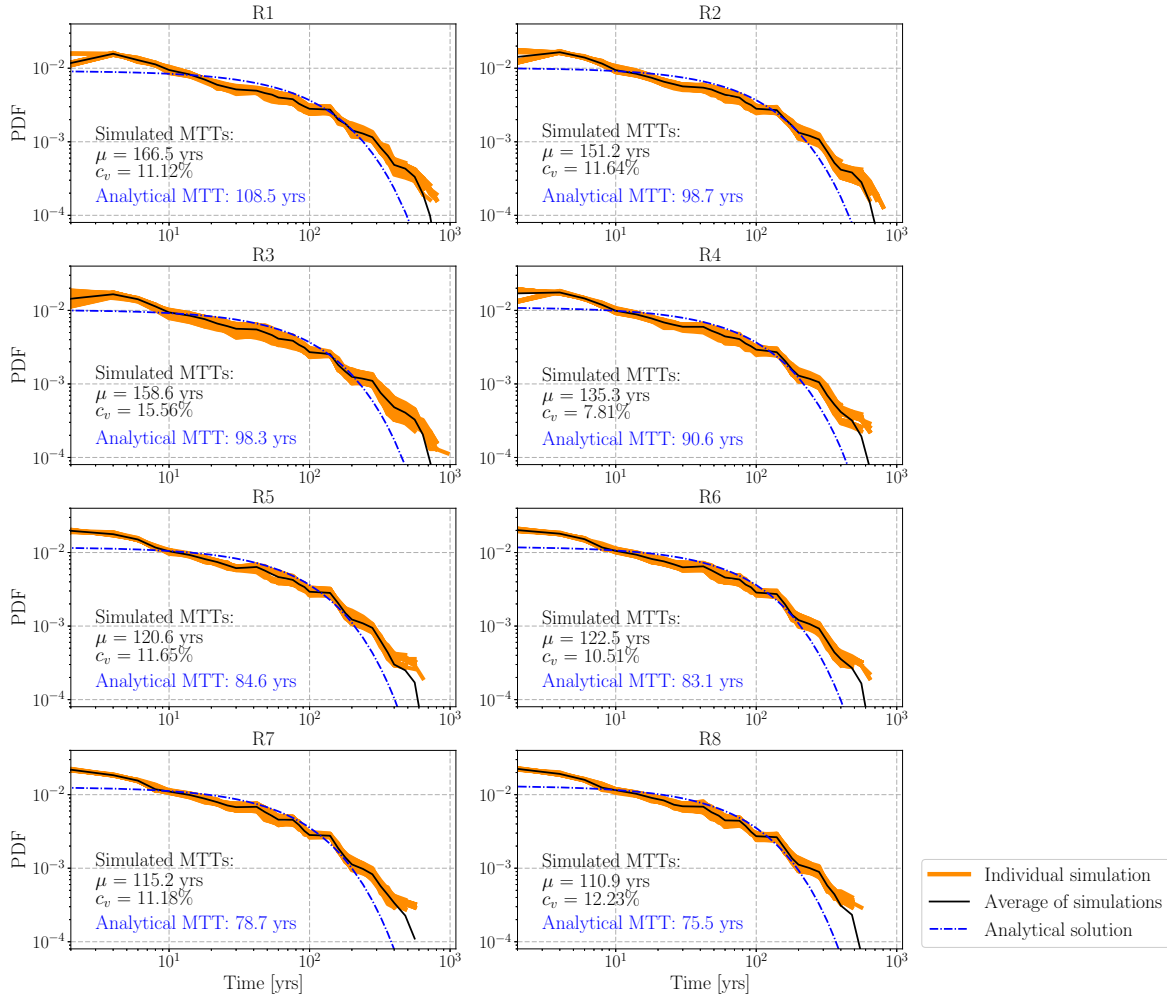
Figure 7. Travel time distributions of ensemble simulations and analytical solutions categorized by recharge realization. The orange lines show the simulated TTDs of all realizations of K_s fields for each recharge realization. The black lines denote the ensemble averaged TTDs of each recharge realization. The blue dash-dot line is the fitted analytical TTD under the RS assumption. The analytical MTT, the mean (μ), and the coefficient of variance (c_v) of the simulated MTTs are also shown in this figure.

4.3 Sensitivity to the spatial pattern of recharge

Figure 10a depicts the sensitivity of simulated TTDs and MTTs to the spatial distribution of recharge, while Figure 10b shows the sensitivity of the cumulative SAS function to the spatial pattern of recharge. The reference simulation is set up using the spatially uniform recharge that is equal to the spatial average of the spatially distributed recharge, while all of the other parameters in these two simulations are held identical. The different spatial distributions of recharge have a clear effect on the shape of TTDs. It appears that the most evident difference between the TTDs of the two recharge scenarios occurs at the
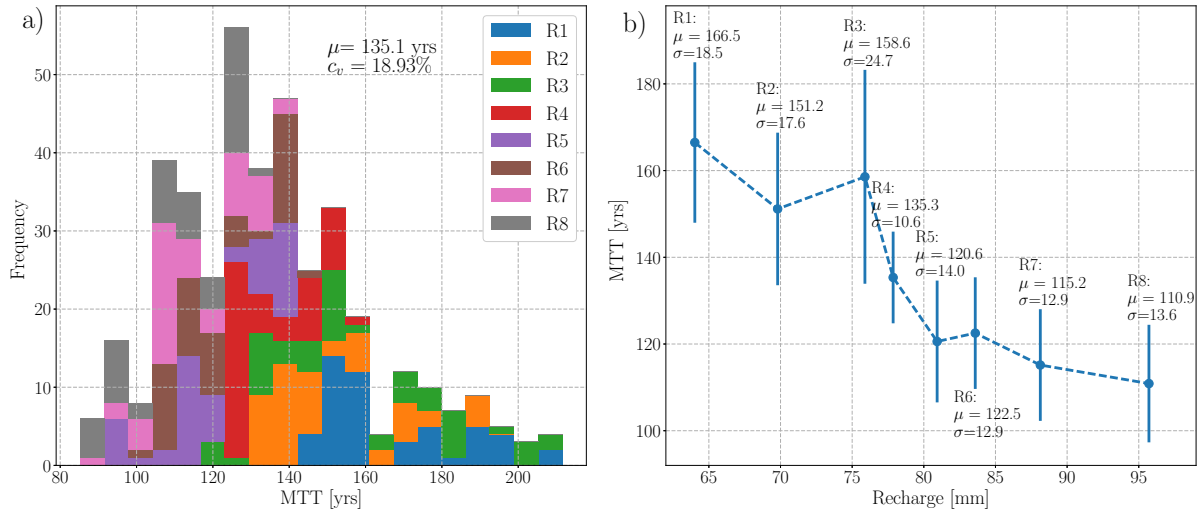
Figure 8. Uncertainty quantification: Monte Carlo simulations of MTT predictions categorized by recharge realization. Panel a) shows a histogram of MTT predictions. Panel b) shows the relationship between the ensemble averaged MTT (μ) and the spatially averaged recharge. The error bars represent the standard deviation of MTTs (σ) for each recharge realization.
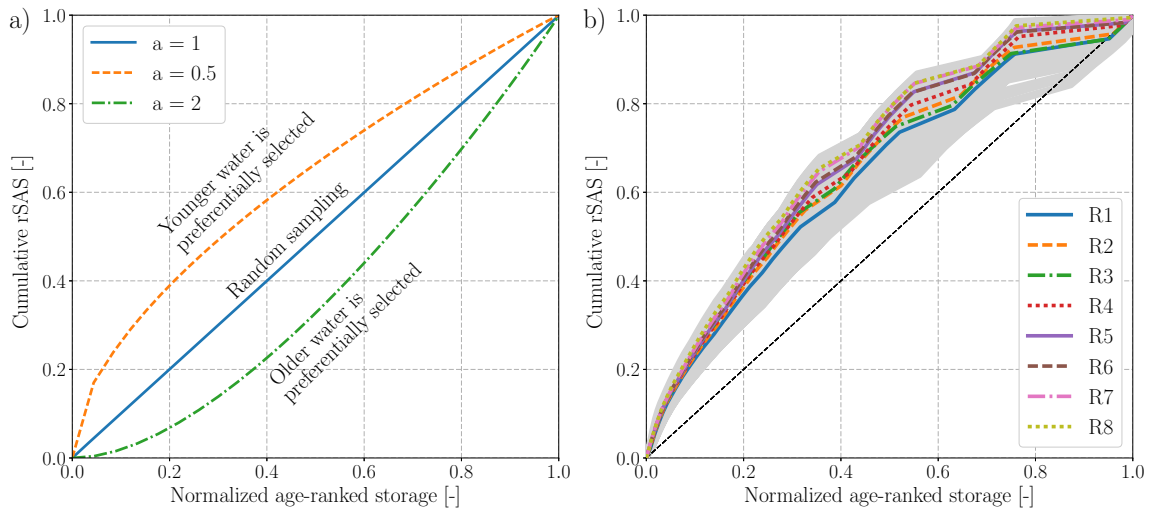
Figure 9. Cumulative rank SAS functions as a function of normalized age-ranked storage. (a) Schematic of cumulative rank SAS functions parameterized by gamma distribution with the shape parameter a = 0.5, 1, and 2. (b) Cumulative rank SAS functions of the ensemble simulations (light grey lines) and the ensemble average for each recharge realization.

early period. Additionally, the simulated MTT using the uniform recharge appears to be smaller than that using the spatially distributed recharge. Figure 10b indicates that the simulation using uniform recharge has a consistently stronger preference
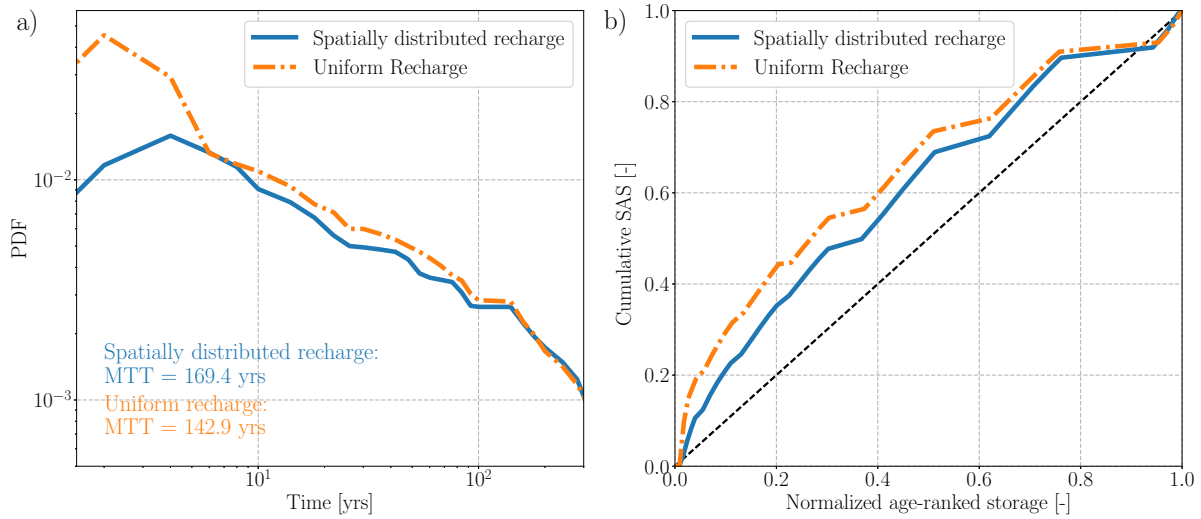
Figure 10. Sensitivity of a) TTDs and MTTs and b) SAS functions to the spatial pattern of recharge.

for sampling young water than the simulation using spatially distributed recharge. Nevertheless, both scenarios show a general preference for young water.

The difference in TTDs and SAS functions is not induced by the variability in internal hydraulic properties, since the two simulations share the same K_s field. Rather, it is mainly induced by the different flow paths of particle tracers in the two recharge scenarios. The spatially distributed recharge simulated by mHM reveals that the upstream mountainous area has higher recharge rates than those in the lowland plain. By contrast, the uniform recharge scenario neglects this spatial nonuniformity. This difference results in the following: (a) under a uniform recharge scenario, more particle tracers enter the system from locations near the streams at lowland plains (Figure 3b), indicating that more particle tracers are transported in the local flow system rather than in the regional flow system (Toth, 1963); and (b) higher recharge rates at lowland plains accelerate the particles' movement in this area and shorten their travel times. As such, local particle flow paths within the shallow aquifer layer at lowland plains are activated, leading to a stronger preference for sampling local flow paths in shallow aquifer layers and therefore a stronger preference for young ages. Our findings are in line with the observations by Kaandorp et al. (2018), wherein the authors found a relatively higher preference for the selection of older water in the upstream area than that in the downstream area of the Springendalse Beek catchment.

5 Discussion

5.1 Uncertainty of external forcing, internal property, and TTD predictions

In the idealized aquifers where groundwater flow is Dupuit-Forchheimer type, the recharge is uniform, and the aquifer is locally homogeneous. TTD is controlled by recharge, saturated aquifer thickness, and porosity, and it is independent of hydraulic

conductivity (Haitjema, 1995). In a real-world catchment with complex geometry and topography, stratigraphic aquifers, and nonuniform recharge, our numerical exploration demonstrates that the groundwater TTD is dependent on both recharge and the hydrostratigraphic K_s field.

The mechanisms behind the effects of the recharge rate and the K_s field are different. Given that the spatial pattern of recharge remains the same, a higher recharge rate intensifies the fluxes for the complete range of flow pathways. This process forces water downward in the recharge zones and upward to the discharge area. As such, flow rates through all flow pathways are increased equally, and the spatial distribution of flow pathways is not changed. The corresponding SAS function is also not changed (Botter et al., 2010; Cartwright and Morgenstern, 2015; Kaandorp et al., 2018). In contrast, a different K_s field activates flow pathways in more permeable layers, deactivates flow pathways in less permeable layers, changes the spatial distribution of flow pathways, and therefore changes the shape of the SAS function (Harman et al., 2016; Kim et al., 2016).

We also underline the value of observational data in reducing predictive uncertainty in simulated TTDs. In this study, the majority of model parameters can be adequately conditioned by spatially distributed groundwater head observations (Figure 4). Provided that most of the hydraulic conductivities are constrained to the model-to-measurement misfit and reality, the TTD predictions can also be effectively bounded. This is evidenced by Figure 7, from which moderate values of c_v ranging from 7.81% to 15.56% in different recharge realizations can be observed. The ensemble averaged MTTs for different recharge realizations also have a high c_v (15.70%), implying that the TTD prediction appears highly sensitive to recharge. Our findings are in line with Danesh-Yazdi et al. (2018), in which the interplay between recharge and subsurface heterogeneity was investigated and a strong dependency of TTDs on the recharge was observed.

5.2   Analytical model and SAS functions

The analytical solution of TTD, assuming a random sampling of water, cannot properly replicate the TTD of numerical simulation in the study domain. In the stratigraphic aquifer with complex topography and diffuse recharge, the analytical solution using Eq. 6 underestimates the MTT. Note that this conclusion holds when the simulated TTD has a relatively larger long-tail behavior than the exponential distribution. Such observations have also been reported for other real-world aquifers (Basu et al., 2012; Eberts et al., 2012; Kaandorp et al., 2018). This finding can be seen as an extension of Basu et al. (2012), in which the authors found a moderate discrepancy between the analytical solution using Eq. 6 and the numerical solution in a small catchment.

It is obvious that the analytical solution under the RS assumption cannot explicitly include the impact of the distributed hydraulic properties of stratigraphic aquifers and the spatially nonuniform recharge. The above limitations of analytical models may introduce a significant predictive error for the TTD predictions, as shown in Figure 7. Moreover, the effective volume of storage related to the transport process S_{eff} is calculated utilizing the numerical solutions of distributed models. As a first-order approximation of effective storage volume, it suggests that only a small fraction of total groundwater volume is involved in the active hydrologic cycle.

The SAS function provides a good interpretability for simulation results using the fully distributed model in terms of characterizing the preference for releasing water of different ages. We find that the SAS functions are weakly dependent on the K_s

fields in the stratigraphic aquifer system, but the overall preference for young water does not change. This weak dependency can be explained by the fact that different realizations of K_s fields modify the spatial distribution of particle flow paths. The overall tendency for young groundwater in the saturated aquifer has been reported by many past studies (Danesh-Yazdi et al., 2018; Kaandorp et al., 2018; Yang et al., 2018). Our study links the explicit simulations of travel times and the analytical SAS functions and offers original insights into the uncertainty propagated from recharge and the K_s fields to the SAS functions.

5.3 Dependency of TTDs and SAS functions on the spatial pattern of forcings

The sensitivity of the TTDs and SAS functions to the spatial pattern of recharge forcings can be explained mainly by the different flow paths of particle tracers, resulting primarily from the spatially heterogeneous fields of recharge across the study catchment. For the regional groundwater system, the spatial variation of recharge determines the distribution of starting points of the flow pathlines of tracer particles. For example, more particles will be injected from recharge zones that are typically located in high-elevation regions, resulting in a higher weight of flowlines starting from high-elevation regions. The pronounced spatial variability of recharge also controls the systematic (water age) preference for particles existing from the system (to river discharge) that originated from different regions and therefore exerts a strong control on the shape of the SAS function.

In the study catchment, an oversimplified spatially uniform recharge results in a smaller MTT and a stronger preference for discharging young water compared to those taking the spatial variability of recharge. Such observations are conditioned to site-specific features of the study catchment, which is noticed only when the following apply: (a) a site is located in a headwater catchment under a humid climate condition; (b) the recharge rate in areas close to the drainage network is generally lower than that in areas far away from the drainage network; and (c) the system is under (near) natural conditions, meaning that artificial drainage and pumping do not dominate the groundwater budget.

The assumption of spatially uniform input forcing has been widely applied in regional-scale subsurface hydrologic models (Zghibi et al., 2015; Yang et al., 2018). Our study indicates that a reasonable characterization of spatial pattern of recharge is crucial for reliable TTD prediction. Unfortunately, it is quite challenging to confidently quantify the groundwater recharge at the regional scale under today's technique due to the lack of reliable measurements (Healy and Scanlon, 2010; Cheng et al., 2017; Zink et al., 2017). Appropriate techniques should be chosen to estimate groundwater recharge according to the study goals and the spatial and temporal scales.

5.4 Implications for the applied groundwater modeling

Uncertainty limits the applicability of groundwater models. Most of the applied groundwater models are deterministic models that use direct values of inputs and parameters instead of probabilistic distributions of them. Specifically, both the model inputs and the inversion process are deterministic, leading to a deterministic best-fit parameter set achieved during model calibration. Our study reveals limitations of the above modeling procedure and suggests that the probabilistic distribution of inputs and parameters should be considered for the applied modeling. The main limitation is that the single exclusive assignment of recharge is inadequate for the simulation of transport processes because the error can be propagated from inputs to the conditioned parameters (e.g., hydraulic conductivities) through the calibration process (Figure 4). This accumulated

error will further lead to a seriously biased prediction of travel times. Additionally, the modeling workflow used in this study is computationally more efficient than the Bayesian approach and is suitable for complex real-world applications.

The degree of predictive uncertainty is highly dependent on the parameterization scheme. Some highly parameterized models are potentially ill-posed due to the paucity of data and therefore cannot be constrained by the available calibration dataset. In this case, the predictive uncertainty of TTDs is potentially high (Weissmann et al., 2002; Danesh-Yazdi et al., 2018). Stratigraphic aquifer models with zoned parameters are still widely used for applied groundwater modeling because the field representation of local-scale heterogeneity is difficult. Given that the aquifer model is stratigraphic and the number of parameters is less than the number of observations, most of the adjustable parameters can be effectively bounded. In this case, the uncertainty of input data (e.g., recharge) appears to have a primary influence on the TTD predictions. Note that here, we do not account for the error caused by model structural deficiency.

6 Conclusions

In this study, we explore the relationship between the uncertainty of recharge, calibration-constrained hydraulic conductivity realizations, and predictions of groundwater TTDs. Using both a physically based numerical model and a lumped analytical model, a comprehensive case study is performed in an agricultural catchment (the Nägelstedt catchment). The RWPT method is used to track the water samples through the modeling domain and compute their travel times. Moreover, the analytical model is fitted against the numerical solutions to provide a reference for the effective storage and sampling behavior of the system. Based on this study, the following conclusions are made:

1. In the Nägelstedt catchment model, the simulated MTTs are strongly dependent on the recharge rate and weakly dependent on the postcalibrated K_s field. We highlight the importance of recharge quantification and the worth of reliable data in reducing the predictive uncertainty of TTDs.

2. The framework of the SAS function provides a good interpretability of simulated TTDs in terms of characterizing the systematic preference for sampling young/old water as outflow. On the basis of this framework, we find that the ensemble simulations have a consistent preference for young water, despite the different recharge and hydraulic conductivity realizations. Our study provides a novel modeling framework to explore the effect of input uncertainty and parameter equifinality on TTDs and SAS functions through a combination of calibration-constrained Monte Carlo parameter generation, a numerical model, and a SAS function framework.

3. Both the shape and the breadth of catchment groundwater TTDs and SAS functions are sensitive to the spatial distribution of recharge. Therefore, a reasonable characterization of the spatial pattern of recharge is crucial for the reliable TTD prediction in the catchment-scale groundwater models.

Appendix A: Random walk particle tracking

Random walk particle tracking solves a diffusion equation at local Lagrangian coordinates rather than the classical advection-diffusion equation, which can be expressed as:

$$\mathbf{x}(t_i) = \mathbf{x}(t_{i-1}) + \mathbf{v}(\mathbf{x}(t_{i-1}))\Delta t + Z\sqrt{2\mathbf{D}(\mathbf{x}(t_{i-1})\Delta t} \tag{A1}$$

where $\mathbf{x}$ denotes the coordinates of the particle location, Δt denotes the time step size, and Z denotes a Gaussian random number, with the mean being zero and the variance being unity.

The velocity $\mathbf{v}$ in Eq. (A1) is replaced by $\mathbf{v}_i^*$ to maintain consistency with the classical advection-dispersion equation (Kinzelbach, 1986). The expressions of $\mathbf{v}_i^*$ and the hydrodynamic dispersion tensor $\mathbf{D}_{ij}$ are:

$$\mathbf{v}_i^* = \mathbf{v}_i + \sum_{i=1}^{3} \frac{\partial \mathbf{D}_{ij}}{\partial x_{ij}} \tag{A2}$$

$$\mathbf{D}_{ij} = \alpha_T |\mathbf{v}|\delta_{ij} + (\alpha_L - \alpha_T)\frac{\mathbf{v}_i\mathbf{v}_j}{|\mathbf{v}|} + \mathbf{D}_{ij}^d \tag{A3}$$

where δ_{ij} denotes the Kronecker symbol, α_L denotes the longitudinal dispersion length, α_T denotes the transverse dispersion length, $\mathbf{D}_{ij}^d$ denotes the tensor of the molecular diffusion coefficient, and $\mathbf{v}_i$ denotes the mean pore velocity component in the ith direction.

The stochastic governing equation of 3-D RWPT can therefore be expressed as:

$$x_{t+\Delta t} = x_t + \left(V_x(x_t, y_t, z_t, t) + \frac{\partial D_{xx}}{\partial x} + \frac{\partial D_{xy}}{\partial y} + \frac{\partial D_{xz}}{\partial z}\right)\Delta t + \sqrt{2D_{xx}\Delta t}Z_1 + \sqrt{2D_{xy}\Delta t}Z_2 + \sqrt{2D_{xz}\Delta t}Z_3 \tag{A4}$$

$$y_{t+\Delta t} = y_t + \left(V_y(x_t, y_t, z_t, t) + \frac{\partial D_{yx}}{\partial x} + \frac{\partial D_{yy}}{\partial y} + \frac{\partial D_{yz}}{\partial z}\right)\Delta t + \sqrt{2D_{yx}\Delta t}Z_1 + \sqrt{2D_{yy}\Delta t}Z_2 + \sqrt{2D_{yz}\Delta t}Z_3 \tag{A5}$$

$$z_{t+\Delta t} = z_t + \left(V_z(x_t, y_t, z_t, t) + \frac{\partial D_{zx}}{\partial x} + \frac{\partial D_{zy}}{\partial y} + \frac{\partial D_{zz}}{\partial z}\right)\Delta t + \sqrt{2D_{zx}\Delta t}Z_1 + \sqrt{2D_{zy}\Delta t}Z_2 + \sqrt{2D_{zz}\Delta t}Z_3 \tag{A6}$$

where x, y, z are the spatial coordinates of a particle, Δt is the time step, and Z_i is a random number with a mean of zero and a unit variance.

Appendix B: Composite parameter sensitivity

The PEST algorithm calculates the sensitivity with respect to each parameter of all observations (with the latter weighted as per user-assigned weights), namely the "composite sensitivity" (Doherty, 2015). The composite sensitivity of parameter i is defined as $csp_i = \frac{[\mathbf{J}^t\mathbf{Q}\mathbf{J}]_{ii}^{1/2}}{n}$, where $\mathbf{J}$ denotes the Jacobian matrix that includes the sensitivities of all predictions to all model parameters, $\mathbf{Q}$ is the weight matrix, and n is the number of observations with nonzero weights. In this study, all weights assigned to observations are equally set to 1. Table B1 displays the composite parameter sensitivities in each recharge realization. The mean composite parameter sensitivities of calibrations in all recharge realizations are also included in this table. The hydraulic conductivity of the Middle Muschelkalk (mm) is highly sensitive to groundwater head observations, whereas that of the Lower

Table B1. Composite parameter sensitivities to the groundwater head observations.

Recharge realizations	Parameter sensitivities [-]						
	soil	alluvium	km	ku	mo	mm	mu
R1	6.01	1.89	0.58	1.50	9.47	20.45	0.23
R2	6.15	1.93	0.49	1.52	9.62	20.61	0.35
R3	4.05	1.78	1.38	1.91	7.20	25.84	0.82
R4	5.97	1.91	0.39	1.56	9.41	20.99	0.19
R5	7.31	1.86	0.31	1.34	10.09	19.40	0.015
R6	5.87	1.90	0.39	1.67	9.50	21.03	0.23
R7	7.77	1.93	0.57	1.93	10.53	19.07	0.41
R8	5.03	1.87	0.48	1.81	9.07	22.64	0.11
mean	6.77	1.88	0.57	1.66	9.36	21.25	0.29

Muschelkalk (mu) is insensitive to groundwater head observations. The sensitivity of mu, however, varies widely between different recharge realizations, from the highest one in R3 (0.82) to the lowest one in R5 (0.015).

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

Acknowledgements. This study receives support from the Deutsche Forschungsgemeinschaft via Sonderforschungsbereich CRC 1076 AquaDiva.
We kindly thank Sabine Sattler from Thuringian State office for the Environment and Geology (TLUG) for providing basic geological data.
We acknowledge the EVE Linux Cluster team at UFZ for their support for this study. We also acknowledge the Chinese Scholarship Council
(CSC) for supporting Miao Jing's stay in Germany.

References

Aigner, T.: Calcareous Tempestites: Storm-dominated Stratification in Upper Muschelkalk Limestones (Middle Trias, SW-Germany), in: Cyclic and Event Stratification, edited by Einsele, G. and Seilacher, A., pp. 180–198, Springer Berlin Heidelberg, Berlin, Heidelberg, 1982.

5   Ajami, N. K., Duan, Q., and Sorooshian, S.: An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, Water Resources Research, 43, https://doi.org/10.1029/2005WR004745, http://dx.doi.org/10.1029/2005WR004745, 2007.

Basu, N. B., Jindal, P., Schilling, K. E., Wolter, C. F., and Takle, E. S.: Evaluation of analytical and numerical approaches for the estimation of groundwater travel time distribution, Journal of Hydrology, 475, 65 – 73, https://doi.org/https://doi.org/10.1016/j.jhydrol.2012.08.052,
10   http://www.sciencedirect.com/science/article/pii/S0022169412007317, 2012.

Benettin, P., Kirchner, J. W., Rinaldo, A., and Botter, G.: Modeling chloride transport using travel time distributions at Plynlimon, Wales, Water Resources Research, pp. 3259–3276, https://doi.org/10.1002/2014WR016600, 2015.

Benettin, P., Soulsby, C., Birkel, C., Tetzlaff, D., Botter, G., and Rinaldo, A.: Using SAS functions and high-resolution isotope data to unravel travel time distributions in headwater catchments, Water Resources Research, 53, 1864–1878, https://doi.org/10.1002/2016WR020117,
15   http://dx.doi.org/10.1002/2016WR020117, 2017.

Benson, D. A., Aquino, T., Bolster, D., Engdahl, N., Henri, C. V., and Fernandez-Garcia, D.: A comparison of Eulerian and Lagrangian transport and non-linear reaction algorithms, Advances in Water Resources, 99, 15–37, https://doi.org/10.1016/j.advwatres.2016.11.003, http://dx.doi.org/10.1016/j.advwatres.2016.11.003, 2017.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, Advances in water resources, 16, 41–51, 1993.

20   Böhlke, J. K.: Groundwater recharge and agricultural contamination, Hydrogeology Journal, 10, 153–179, https://doi.org/10.1007/s10040-001-0183-3, 2002.

Botter, G., Bertuzzo, E., and Rinaldo, A.: Transport in the hydrologic response: Travel time distributions, soil moisture dynamics, and the old water paradox, Water Resources Research, 46, 1–18, https://doi.org/10.1029/2009WR008371, 2010.

Botter, G., Bertuzzo, E., and Rinaldo, A.: Catchment residence and travel time distributions: The master equation, Geophysical Research
25   Letters, 38, 1–6, https://doi.org/10.1029/2011GL047666, 2011.

Böhlke, J. K. and Denver, J. M.: Combined Use of Groundwater Dating, Chemical, and Isotopic Analyses to Resolve the History and Fate of Nitrate Contamination in Two Agricultural Watersheds, Atlantic Coastal Plain, Maryland, Water Resources Research, 31, 2319–2339, https://doi.org/10.1029/95WR01584, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/95WR01584, 1995.

Cartwright, I. and Morgenstern, U.: Transit times from rainfall to baseflow in headwater catchments estimated using tritium: The Ovens
30   River, Australia, Hydrology and Earth System Sciences Discussions, 12, 5427–5463, https://doi.org/10.5194/hessd-12-5427-2015, 2015.

Cheng, Y., Zhan, H., Yang, W., Dang, H., and Li, W.: Is annual recharge coefficient a valid concept in arid and semi-arid regions?, Hydrology and Earth System Sciences, 21, 5031–5042, https://doi.org/10.5194/hess-21-5031-2017, https://www.hydrol-earth-syst-sci.net/21/5031/2017/, 2017.

Danesh-Yazdi, M., Klaus, J., Condon, L. E., and Maxwell, R. M.: Bridging the gap between numerical solutions of travel time distributions
35   and analytical storage selection functions, Hydrological Processes, https://doi.org/10.1002/hyp.11481, http://doi.wiley.com/10.1002/hyp.11481, 2018.

de Rooij, R., Graham, W., and Maxwell, R. M.: A particle-tracking scheme for simulating pathlines in coupled surface-subsurface flows, Advances in Water Resources, 52, 7–18, https://doi.org/10.1016/j.advwatres.2012.07.022, http://dx.doi.org/10.1016/j.advwatres.2012.07.022, 2013.

Doherty, J.: Calibration and uncertainty analysis for complex environmental models, Watermark Numerical Computing, 2015.

5   Doherty, J. and Hunt, R.: Approaches to highly parameterized inversion: a guide to using PEST for groundwater-model calibration, U. S. Geological Survey Scientific Investigations Report 2010-5169, p. 70, https://doi.org/2010-5211, http://pubs.usgs.gov/sir/2010/5169/, 2010.

Eberts, S. M., Böhlke, J. K., Kauffman, L. J., and Jurgens, B. C.: Comparison of particle-tracking and lumped-parameter age-distribution models for evaluating vulnerability of production wells to contamination, Hydrogeology Journal, 20, 263–282, https://doi.org/10.1007/s10040-011-0810-6, 2012.

10 Engdahl, N. B., McCallum, J. L., and Massoudieh, A.: Transient age distributions in subsurface hydrologic systems, Journal of Hydrology, 543, 88–100, https://doi.org/10.1016/j.jhydrol.2016.04.066, http://dx.doi.org/10.1016/j.jhydrol.2016.04.066, 2016.

Fischer, T., Naumov, D., Sattler, S., Kolditz, O., and Walther, M.: GO2OGS 1.0: A versatile workflow to integrate complex geological information with fault data into numerical simulation models, Geoscientific Model Development, 8, 3681–3694, https://doi.org/10.5194/gmd-8-3681-2015, 2015.

15 Ginn, T. R.: On the distribution of multicomponent mixtures over generalized exposure time in subsurface flow and reactive transport: Theory and formulations for residence-time-dependent sorption/desorption with memory, Water Resources Research, 36, 2885–2893, https://doi.org/10.1029/2000WR900170, 2000.

Goode, D. J.: Direct simulation of groundwater age, Water Resources Research, 32, 289–296, https://doi.org/10.1029/95WR03401, 1996.

20 Haitjema, H.: On the residence time distribution in idealized groundwater sheds, Journal of Hydrology, 172, 127–146, 1995.

Hale, V. C. and McDonnell, J. J.: Effect of bedrock permeability on stream base flow mean transit time scaling relations: 1. A multiscale catchment intercomparison, Water Resources Research, 52, 1358–1374, 2016.

Harman, C. J.: Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed, Water Resources Research, 51, 1–30, 2015.

25 Harman, C. J., Ward, A. S., and Ball, A.: How does reach-scale stream-hyporheic transport vary with discharge? Insights from rSAS analysis of sequential tracer injections in a headwater mountain stream, Water Resources Research, 52, 7130–7150, https://doi.org/10.1002/2016WR018832, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR018832, 2016.

Healy, R. W. and Scanlon, B. R.: Estimating Groundwater Recharge, Cambridge University Press, https://doi.org/10.1017/CBO9780511780745, 2010.

30 Heße, F., Zink, M., Kumar, R., Samaniego, L., and Attinger, S.: Spatially distributed characterization of soil-moisture dynamics using travel-time distributions, Hydrology and Earth System Sciences, 21, 549–570, https://doi.org/10.5194/hess-21-549-2017, http://www.hydrol-earth-syst-sci.net/21/549/2017/, 2017.

Howden, N. J. K., Burt, T. P., Worrall, F., Whelan, M. J., and Bieroza, M.: Nitrate concentrations and fluxes in the River Thames over 140 years (1868–2008): are increases irreversible?, Hydrological Processes, 24, 2657–2662, https://doi.org/10.1002/hyp.7835, http://dx.doi.org/10.1002/hyp.7835, 2010.

35 Jing, M., Heße, F., Kumar, R., Wang, W., Fischer, T., Walther, M., Zink, M., Zech, A., Samaniego, L., Kolditz, O., and Attinger, S.: Improved regional-scale groundwater representation by the coupling of the mesoscale Hydrologic Model (mHM v5.7) to the groundwa-

ter model OpenGeoSys (OGS), Geoscientific Model Development, 11, 1989–2007, https://doi.org/10.5194/gmd-11-1989-2018, https://www.geosci-model-dev.net/11/1989/2018/, 2018.

Jochen, L., , R. D., and Röhling, H.-G.: Lithostratigraphie des Buntsandstein in Deutschland, Schriftenreihe der Deutschen Gesellschaft für Geowissenschaften, 69, 69–149, https://doi.org/10.1127/sdgg/69/2014/69, http://dx.doi.org/10.1127/sdgg/69/2014/69, 2014.

5   Kaandorp, V. P., de Louw, P. G. B., van der Velde, Y., and Broers, H. P.: Transient Groundwater Travel Time Distributions and Age-Ranked Storage-Discharge Relationships of Three Lowland Catchments, Water Resources Research, pp. 1–18, https://doi.org/10.1029/2017WR022461, http://doi.wiley.com/10.1029/2017WR022461, 2018.

Kim, M., Pangle, L. A., Cardoso, C., Lora, M., Volkmann, T. H. M., Wang, Y., Harman, C. J., and Troch, P. A.: Transit time distributions and StorAge Selection functions in a sloping soil lysimeter with time-varying flow paths: Direct observation of internal and external transport

10 variability, Water Resources Research, 52, 7105–7129, https://doi.org/10.1002/2016WR018620, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR018620, 2016.

Kinzelbach, W.: Groundwater modelling: an introduction with sample programs in BASIC, vol. 25, Elsevier, 1986.

Kirchner, J. W.: Aggregation in environmental systems-Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments, Hydrology and Earth System Sciences, 20, 279–297, https://doi.org/10.5194/hess-20-279-

15 2016, 2016.

Kohlhepp, B., Lehmann, R., Seeber, P., Küsel, K., Trumbore, S. E., and Totsche, K. U.: Aquifer configuration and geostructural links control the groundwater quality in thin-bedded carbonate–siliciclastic alternations of the Hainich CZE, central Germany, Hydrology and Earth System Sciences, 21, 6091–6116, https://doi.org/10.5194/hess-21-6091-2017, https://www.hydrol-earth-syst-sci.net/21/6091/2017/, 2017.

20 Kolditz, O., Bauer, S., Bilke, L., Böttcher, N., Delfs, J. O., Fischer, T., Görke, U. J., Kalbacher, T., Kosakowski, G., McDermott, C. I., Park, C. H., Radu, F., Rink, K., Shao, H., Shao, H. B., Sun, F., Sun, Y. Y., Singh, A. K., Taron, J., Walther, M., Wang, W., Watanabe, N., Wu, Y., Xie, M., Xu, W., and Zehner, B.: OpenGeoSys: an open-source initiative for numerical simulation of thermo-hydro-mechanical/chemical (THM/C) processes in porous media, Environmental Earth Sciences, 67, 589–599, https://doi.org/10.1007/s12665-012-1546-x, http://link.springer.com/10.1007/s12665-012-1546-x, 2012.

25 Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resources Research, 49, 360–379, https://doi.org/10.1029/2012WR012195, http://dx.doi.org/10.1029/2012WR012195, 2013.

Leray, S., Engdahl, N. B., Massoudieh, A., Bresciani, E., and McCallum, J.: Residence time distributions for hydrologic systems: Mechanistic foundations and steady-state analytical solutions, Journal of Hydrology, 543, 67–87, https://doi.org/10.1016/j.jhydrol.2016.01.068, http://dx.doi.org/10.1016/j.jhydrol.2016.01.068, 2016.

30 McCallum, J. L., Engdahl, N. B., Ginn, T. R., and Cook, P. G.: Nonparametric estimation of groundwater residence time distributions: What can environmental tracer data tell us about groundwater residence time?, Water Resources Research, 50, 2022–2038, https://doi.org/10.1002/2013WR014974, 2014.

McCann, T.: The Geology of Central Europe Volume 2: Mesozoic and Cenozoic, Geological Society of London, https://doi.org/10.1144/CEV2P, https://doi.org/10.1144/CEV2P, 2008.

35 Menning, M.: Deutsche Stratigraphische Kommission (2002) Eine geologische Zeitskala 2002, Deutsche Stratigraphische Kommission (ed) Stratigraphische Tabelle von Deutschland, 2002.

Merz, G.: Zur Petrographie, Stratigraphie, Paläogeographie und Hydrogeologie des Muschelkalks (Trias) im Thüringer Becken, Zeitschrift der geologischen Wissenschaften, 15, 457–473, 1987.

Molnat, J. and Gascuel-Odoux, C.: Modelling flow and nitrate transport in groundwater for the prediction of water travel times and of consequences of land use evolution on water quality, Hydrological Processes, 16, 479–492, https://doi.org/10.1002/hyp.328, 2002.

5 Neuman, S. P.: Theory of flow in unconfined aquifers considering delayed response of the water table, Water Resources Research, 8, 1031–1045, https://doi.org/10.1029/WR008i004p01031, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/WR008i004p01031, 1972.

Park, C. H., Beyer, C., Bauer, S., and Kolditz, O.: Using global node-based velocity in random walk particle tracking in variably saturated porous media: Application to contaminant leaching from road constructions, Environmental Geology, 55, 1755–1766, https://doi.org/10.1007/s00254-007-1126-7, 2008.

10 Remondi, F., Kirchner, J. W., Burlando, P., and Fatichi, S.: Water Flux Tracking With a Distributed Hydrological Model to Quantify Controls on the Spatiotemporal Variability of Transit Time Distributions, Water Resources Research, 54, 3081–3099, https://doi.org/10.1002/2017WR021689, 2018.

Rinaldo, A., Beven, K. J., Bertuzzo, E., Nicotina, L., Davies, J., Fiori, A., Russo, D., and Botter, G.: Catchment travel time distributions and water flow in soils, Water Resources Research, 47, 1–13, https://doi.org/10.1029/2011WR010478, 2011.

15 Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, n/a–n/a, https://doi.org/10.1029/2008WR007327, http://doi.wiley.com/10.1029/2008WR007327, 2010.

Schoups, G., van de Giesen, N. C., and Savenije, H. H. G.: Model complexity control for hydrologic prediction, Water Resources Research, 44, n/a–n/a, https://doi.org/10.1029/2008WR006836, http://dx.doi.org/10.1029/2008WR006836, w00B03, 2008.

Seidel, H. G., ed.: Geologie von Thüringen, Schweizerbart Science Publishers, Stuttgart, Germany, http://www.schweizerbart.de/
20 /publications/detail/isbn/9783510652051/Geologie_von_Thuringen_herausg_v_G_Se, 2003.

Selle, B., Rink, K., and Kolditz, O.: Recharge and discharge controls on groundwater travel times and flow paths to production wells for the Ammer catchment in southwestern Germany, Environmental Earth Sciences, 69, 443–452, https://doi.org/10.1007/s12665-013-2333-z, 2013.

Stewart, M. K., Morgenstern, U., McDonnell, J. J., and Pfister, L.: The 'hidden streamflow' challenge in catchment hydrology: a call to action
25 for stream water transit time analysis, Hydrological Processes, 26, 2061–2066, 2012.

Stewart, M. K., Morgenstern, U., Gusyev, M. A., and Maloszewski, P.: Aggregation effects on tritium-based mean transit times and young water fractions in spatially heterogeneous catchments and groundwater systems, and implications for past and future applications of tritium, Hydrology and Earth System Sciences Discussions, pp. 1–26, https://doi.org/10.5194/hess-2016-532, http://www.hydrol-earth-syst-sci-discuss.net/hess-2016-532/, 2016.

30 Tetzlaff, D., Birkel, C., Dick, J., Geris, J., and Soulsby, C.: Storage dynamics in hydropedological units control hillslope connectivity, runoff generation, and the evolution of catchment transit time distributions, Water Resources Research, 50, 969–985, https://doi.org/10.1002/2013WR014147, http://dx.doi.org/10.1002/2013WR014147, 2014.

Tompson, A. F. and Gelhar, L. W.: Numerical simulation of solute transport in three-dimensional, randomly heterogeneous porous media, Water Resources Research, 26, 2541–2562, 1990.

35 Tonkin, M. and Doherty, J.: Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques, Water Resources Research, 45, 1–17, https://doi.org/10.1029/2007WR006678, 2009.

Toth, J.: A Theoretical Analysis of Groundwater Flow in Small Drainage Basins, Journal of Geophysical Research, 68, 4795–4812, https://doi.org/10.1029/JZ068i016p04795, 1963.

Van Der Velde, Y., Torfs, P. J. J. F., Van Der Zee, S. E. A. T. M., and Uijlenhoet, R.: Quantifying catchment-scale mixing and its effect on time-varying travel time distributions, Water Resources Research, 48, 1–13, https://doi.org/10.1029/2011WR011310, 2012.

van der Velde, Y., Heidbüchel, I., Lyon, S. W., Nyberg, L., Rodhe, A., Bishop, K., and Troch, P. A.: Consequences of mixing assumptions for time-variable travel time distributions, Hydrological Processes, 29, 3460–3474, https://doi.org/10.1002/hyp.10372, http://doi.wiley.com/10.1002/hyp.10372, 2015.

Van Meter, K., Basu, N., Veenstra, J., and Burras, C. L.: The nitrogen legacy: emerging evidence of nitrogen accumulation in anthropogenic landscapes, Environmental Research Letters, 11, 035 014, 2016.

Van Meter, K. J., Basu, N. B., and Van Cappellen, P.: Two centuries of nitrogen dynamics: Legacy sources and sinks in the Mississippi and Susquehanna River Basins, Global Biogeochemical Cycles, 31, 2–23, https://doi.org/10.1002/2016GB005498, 2017.

Wang, H., Richardson, C. J., Ho, M., and Flanagan, N.: Drained coastal peatlands: A potential nitrogen source to marine ecosystems under prolonged drought and heavy storm events—A microcosm experiment, Science of The Total Environment, 566-567, 621 – 626, https://doi.org/https://doi.org/10.1016/j.scitotenv.2016.04.211, http://www.sciencedirect.com/science/article/pii/S0048969716309573, 2016.

Wechsung, F., Kaden, S., Behrendt, H., and Klöcking, B., eds.: Integrated Analysis of the Impacts of Global Change on Environment and Society in the Elbe Basin, Schweizerbart Science Publishers, Stuttgart, Germany, http://www.schweizerbart.de//publications/detail/isbn/9783510653041/Wechsung_Integrated_Analysis_of_the_Imp, 2008.

Weissmann, G. S., Zhang, Y., LaBolle, E. M., and Fogg, G. E.: Dispersion of groundwater age in an alluvial aquifer system, Water Resources Research, 38, 16–1–16–13, https://doi.org/10.1029/2001WR000907, http://doi.wiley.com/10.1029/2001WR000907, 2002.

Yang, J., Heidbüchel, I., Musolff, A., Reinstorf, F., and Fleckenstein, J. H.: Exploring the Dynamics of Transit Times and Subsurface Mixing in a Small Agricultural Catchment, Water Resources Research, pp. 2317–2335, https://doi.org/10.1002/2017WR021896, http://doi.wiley.com/10.1002/2017WR021896, 2018.

Zghibi, A., Zouhri, L., Chenini, I., Merzougui, A., and Tarhouni, J.: Modelling of the groundwater flow and of tracer movement in the porous and fissured media: chalk aquifer (Northern part of Paris Basin, France), Hydrological Processes, pp. n/a–n/a, https://doi.org/10.1002/hyp.10746, http://doi.wiley.com/10.1002/hyp.10746, 2015.

Zink, M., Kumar, R., Cuntz, M., and Samaniego, L.: A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty, Hydrology and Earth System Sciences, 21, 1769–1790, https://doi.org/10.5194/hess-21-1769-2017, http://www.hydrol-earth-syst-sci.net/21/1769/2017/, 2017.