AGU
ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research®

Key Points:
- A flexible approach to explicitly model the transport of ^{222}Rn in saturated porous media using MODFLOW-NWT and MT3D-USGS is presented
- Spatially varying ^{222}Rn production rates, found to vary by a factor of two at relatively small spatial scales (<100 m), are considered
- A joint inversion of hydraulic heads, temperature, and ^{222}Rn reduced the high degree of uncertainty in the estimated groundwater travel times

Correspondence to:
J. L. Schaper,
jonas.schaper@uni-tuebingen.de

Spatial Variability of Radon Production Rates in an Alluvial Aquifer Affects Travel Time Estimates of Groundwater Originating From a Losing Stream

Jonas L. Schaper[1], Christiane Zarfl[1], Karin Meinikmann[2], Eddie W. Banks[3], Sandra Baron[1], Olaf A. Cirpka[1], and Joerg Lewandowski[4,5]

[1]Center for Applied Geoscience, Eberhard Karls University of Tübingen, Tübingen, Germany, [2]Julius Kühn Institute – Federal Research Centre for Cultivated Plants, Institute for Breeding Research on Agricultural Crops, Berlin, Germany, [3]National Centre for Groundwater Research and Training and College of Science and Engineering, Flinders University, Adelaide, SA, Australia, [4]Department Ecohydrology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany, [5]Geography Department, Humboldt University Berlin, Berlin, Germany

Abstract Assessing water travel times in hyporheic and adjacent alluvial sediments is important to quantify exchange rates, biogeochemical turnover, and pollution dynamics across groundwater-surface water interfaces and in floodplain aquifers. Heat and ^{222}Rn are useful natural tracers for this purpose. While heat transport is commonly simulated via the convection-conduction equation, the transport of ^{222}Rn is often simulated assuming purely advective transport and a homogenous distribution of ^{222}Rn production rates across the model domain. In the present study, we explicitly model the production, transport, and radioactive decay of ^{222}Rn after surface water infiltration into an alluvial aquifer with the numerical models MODFLOW-NWT and MT3D-USGS. Using field observations, numerical modeling, and laboratory experiments, we show that ^{222}Rn production rates vary substantially in the alluvial aquifer of a lowland river. Distributions of mean river-to-groundwater travel time in the alluvial aquifer, estimated via a Bayesian approach, shifted considerably depending on whether ^{222}Rn and time series of groundwater heads and temperature were jointly inverted and whether ^{222}Rn production rates were allowed to vary spatially. With distance to the river, differences between the median river-to-groundwater travel time and apparent ^{222}Rn ages increased by factors ranging from 1.4 at 4 m to 11.9 at 59 m, a finding that highlights the need to simulate ^{222}Rn transport explicitly. The joint inversion of ^{222}Rn, groundwater heads, and temperature reduced uncertainty associated with mean travel times, suggesting that the uncertainty introduced by spatially varying ^{222}Rn production rates can at least partly be compensated by using a combination of natural tracers.

Plain Language Summary Rivers and groundwater bodies exchange water back and forth. Knowing the time that water spends traveling from rivers to certain points in groundwater is important for water management, particularly in situations where drinking water is produced via bank filtration: It allows to predict what happens to natural and toxic compounds that come from the river and to assess the quantity of water moving from rivers to aquifers. Here, we present a modeling approach to estimate travel times of groundwater originating from a nearby river. The approach uses the open-source software packages MODFLOW-NWT and MT3D-USGS and a combination of natural tracers, such as the noble gas ^{222}Rn, which is produced in the subsurface with varying rates but decays with a constant rate constant. We demonstrate that by using both ^{222}Rn and other natural tracers, such as heat, we estimate travel times from rivers to points in groundwater with less uncertainty.

1. Introduction

Knowledge of water travel times in riverbeds of losing streams and in adjacent alluvial sediments is important for quantifying exchange rates across groundwater-surface water interfaces, assessing biogeochemical turnover rates from concentration differences and evaluating the impacts of anthropogenic pollution in floodplain aquifers. Transport characteristics along subsurface flow paths control biogeochemical processes, including redox zonation. They also impact the temperature distribution within hyporheic and alluvial sediments. Both influence the structure of the microbial community in the hyporheic zone (Peralta-Maraver et al., 2018), the biogeochemical cycling of nutrients and carbon (Zarnetske et al., 2011), and the removal rates of anthropogenic

compounds (Schaper et al., 2019). The distribution of travel times is also important for quantifying the rates at which compounds are transformed and cycled (Frei & Peiffer, 2016; Ginn, 1999). Furthermore, knowledge on travel times can be used to assess the characteristics of groundwater-surface water interactions, such as the magnitude and direction of exchange flows and the volume of subsurface transient storage zones (Cranswick & Cook, 2015). Travel times also influence the opportunity for transformation reactions to proceed and thus constrain the overall attenuation capacity of the riverbed and alluvial sediments (Ocampo et al., 2006). From an anthropocentric perspective, the transport time scales through hyporheic and riparian sediments to a particular location in adjacent alluvial aquifers are of critical importance where riverbank filtration is used for drinking water production (Hellauer et al., 2018; Henzler et al., 2014). The assessment of transport time scales, that is, water and solute travel times in both the riverbed and alluvial aquifers, is therefore crucial for freshwater ecology, river management, pollutant dynamics research, and drinking water supply.

In aquifers, water samples are often not characterized by a single travel time but by a distribution of travel times (Danckwerts, 1953; J. L. McCallum, Cook, et al., 2014). This is partially a sampling effect, but even at idealized points within an aquifer, the mixing by local dispersion along flow paths and even more so between flows paths of distinct ages deems travel times to be distributions (Engdahl et al., 2016; Leray et al., 2016; J. L. McCallum et al., 2015; Varni & Carrera, 1998).

In the present study, we refer to the mean travel time as the average time the water molecules within a water sample have spent within the subsurface, that is, we use the term synonymously to mean groundwater age (Bethke & Johnson, 2008). Water travel time distributions and mean travel times are typically estimated by combining models based on either parametric (Luo et al., 2006; Maloszewski & Zuber, 1993) or shape-free transfer functions (Cirpka et al., 2007; J. L. McCallum, Engdahl, et al., 2014; Z. Liao et al., 2014) or numerical groundwater flow and transport models (Engelhardt et al., 2013; Gilfedder et al., 2019; Varni & Carrera, 1998) with either artificial or natural tracers. In the context of groundwater-surface water interactions, the addition of artificial tracers to rivers requires injecting large masses of tracer, the majority of which remaining in the river, hampering the application of artificial-tracer tests in larger rivers. Moreover, artificial tracers are often injected over short time periods, which limits their ability to study transient hydrological processes that occur over longer periods in the natural environment. Environmental or natural tracers, such as heat, electrical conductivity (EC), and ^{222}Rn, have thus often been measured and interpreted to estimate travel times in alluvial floodplain and hyporheic sediments.

The use of EC to estimate travel times in alluvial aquifers and river sediments is often constrained to losing conditions and situations where EC fluctuations in surface water are high enough and travel times to observation wells short enough to allow a time-varying signal to be measured in the observation wells (Cirpka et al., 2007; Vieweg et al., 2016; Vogt et al., 2010). Heat, on the contrary, typically shows pronounced diurnal fluctuations in surface waters and because of the low costs involved in measuring temperature, is among the most widely used tracers to quantify groundwater-surface water exchange flows in riverbed sediments (Anderson, 2005; Rau et al., 2014) and alluvial aquifers (Hoehn & Cirpka, 2006). The main disadvantage of heat as a tracer to estimate alluvial travel times is that its transport depends on the effective heat capacity of the sediment determining the retardation of temperature in comparison to an ideal tracer. On the submeter to meter scale close to the sediment-water interface, thermal parameters can be inferred from daily temperature fluctuations (A. M. McCallum et al., 2012; Luce et al., 2013). On the decameter scales, however, seasonal temperature time series are required for the estimation of travel times. Seasonal temperature signals, however, cannot univocally be related to such signals in the river as the temperature at the land surface shows practically the same signal (Molina-Giraldo et al., 2011). Additionally, thermal parameters need to be measured or inferred from other measured sediment properties (Engelhardt et al., 2013). To reduce uncertainties stemming from sediment heterogeneity and conceptual uncertainty with respect to flow path geometry, heat as a tracer has been combined with other tracers, such as the artificial sweetener acesulfame (Engelhardt et al., 2013), the artificial tracer bromide (Ma et al., 2012), and the natural tracer ^{222}Rn (F. Liao et al., 2021; Hoehn & Cirpka, 2006).

Compared to EC and temperature, tracer methods that use the noble gas ^{222}Rn hardly depend on surface water signals. The radioactive ^{222}Rn has a half-life of 3.82 d and is produced from radioactive decay of ^{226}Ra as part of the ^{238}U decay chain. Due to gas exchange with the atmosphere, where natural ^{222}Rn concentrations are extremely low, ^{222}Rn concentrations in surface waters are also typically very low, except immediately downstream of strongly gaining river reaches (Cartwright & Gilfedder, 2015; Cook et al., 2006). Upon infiltration of surface water into the subsurface and below the capillary fringe, where groundwater is not in direct gas exchange with soil

air, ^{222}Rn concentrations increase as a result of ^{222}Rn emanation from ^{226}Ra bearing sediment materials resulting in a reaction rate $r_{222\,Rn}$ of:

$$r_{222\,Rn} = \gamma - \lambda_{222}\,^{222}Rn \tag{1}$$

where $r_{222\,Rn}$ is the rate of change of ^{222}Rn concentration due to radioactive processes (Bq L^{-1} s^{-1}), λ_{222} is the radioactive decay constant of ^{222}Rn [T^{-1}], ^{222}Rn is the concentration of ^{222}Rn (Bq L^{-1}), and γ denotes the effective production rate of ^{222}Rn into the groundwater. The latter rate is proportional to the ^{226}Ra content at the grain surface of the sediments and denotes the amount of ^{222}Rn released from the solid phase into the pore space per unit time (Bq L^{-1} s^{-1}) (Porcelli & Swarzenski, 2003). It differs from the emanation fraction, which is the proportion of total ^{222}Rn produced within the bulk aquifer to the total amount of ^{222}Rn released into the pore space, and the emanation rate E, which is the amount of ^{222}Rn released into the pore space per unit dry sediment (Cook et al., 2006). Because ^{226}Ra has a half-life of 1,600 years and is bound to the solids, the ^{222}Rn production rate γ can be viewed as constant over the time scales of interest considered in the present study. After 5 to 6 half-lives of ^{222}Rn a secular equilibrium is reached:

$$^{222}Rn_\infty = \frac{\gamma}{\lambda_{222}} \tag{2}$$

where ^{222}Rn$_\infty$ is the secular equilibrium concentration. In a system with constant production rate γ, the ^{222}Rn concentration of a water parcel that does not mix with other water parcels then evolves according to

$$^{222}Rn\,(t) = \,^{222}Rn_\infty + \left(^{222}Rn\,(t) - \,^{222}Rn_\infty\right)\exp\left(-\lambda_{222}t\right) \tag{3}$$

This expression was originally employed by Hoehn and von Gunten (1989) to estimate apparent Rn ages (τ_{app}) in alluvial aquifers recharged by surface water. While heat and solute transport are typically described via the advection-dispersion equation, the Hoehn and von Gunten (1989) model has been widely used to approximate groundwater ages by τ_{app} (Cranswick et al., 2014; Hoehn & Cirpka, 2006; Lamontagne & Cook, 2007; Pittroff et al., 2017; Popp et al., 2021). Similar to other apparent-ages models (Bethke & Johnson, 2008), the application of Equation 3 to estimate groundwater travel times assumes steady state and piston flow and therefore neglects the effects of dispersion and mixing between flow paths, which can lead to erroneous travel time estimates (Gilfedder et al., 2019; J. L. McCallum et al., 2015; Varni & Carrera, 1998). Moreover, in the derivation of Equation 3, the production rate γ is treated as a constant and thus the effects of heterogeneous production rates cannot be considered. In aquifers, however, γ does not only depend on the overall ^{226}Ra content of the sediment and the overall mineral composition, but also on the specific surface area of the sediments (Sakoda et al., 2011). Given these multiple factors on which γ depends and the heterogeneous distribution of sediment properties typically encountered at groundwater-surface water interfaces, it is reasonable to assume that ^{222}Rn production rates are not uniform in the subsurface.

While it has been demonstrated that ^{222}Rn production rates can vary considerably in sediments used in laboratory sand box experiments (Hoehn et al., 1992), along river reaches (Bouchez et al., 2021; Cook et al., 2006), and across lake bottoms (Corbett et al., 1997), the spatial variability of ^{222}Rn production rates at scales of typical groundwater-surface water interactions studies (<100 m) is poorly documented. Mullinger et al. (2009) showed that within an alluvial aquifer, ^{222}Rn emanation coefficients, that is, emanation rates normalized by sediment ^{226}Ra content, decreased with depth across spatial scales of <10 m, but did not investigate the effects of their findings on travel time estimates. By explicitly simulating ^{222}Rn concentrations using HydroGeoSphere (Therrien et al., 2010), Gilfedder et al. (2019) showed that neglecting dispersion leads to a systematic underestimation of mean river-to-groundwater travel times. The influence of heterogeneously distributed ^{222}Rn production rates on the estimation of travel times in alluvial aquifers, however, is unclear. While HydroGeoSphere and COMSOL Multiphysics (F. Liao et al., 2021) have been employed to explicitly simulate ^{222}Rn transport in the hyporheic zone and within alluvial aquifers, the applicability of open-source groundwater flow and transport software, such as MODFLOW-NWT (Niswonger et al., 2011) and MT3D-USGS (Bedekar et al., 2016), to ^{222}Rn transport has not been investigated. Explicit formulations of ^{222}Rn transport during aquifer recharge from surface waters allow the joint inversion of ^{222}Rn and other natural tracers, such as heat, to improve the precision of travel time estimates in alluvial aquifers.

In the present study, we evaluate how the transport of ^{222}Rn in saturated porous media can explicitly be simulated using MODFLOW-NWT and MT3D-USGS via the advection-dispersion equation. Specifically, we (a) investigate the influence of nonuniform ^{222}Rn production rates within an alluvial aquifer on the estimation of mean river-to-groundwater travel times and (b) assess whether the joint inversion of hydraulic heads, temperature time series, and ^{222}Rn concentrations can reduce uncertainty in the estimation of travel times in situations where surface waters infiltrate into alluvial aquifers. To this end, ^{222}Rn$_\infty$ concentrations of both riverbed and alluvial sediments were measured in incubation experiments and time series of hydraulic head observations, temperature, and ^{222}Rn concentrations were collected in the alluvial aquifer adjacent to a losing lowland river. Using a Bayesian framework, we estimate distributions of mean river-to-groundwater travel times in the alluvial aquifer based on model scenarios that consider different combinations of model constraints and ^{222}Rn transport formulation. We hypothesize that neglecting the effects of spatially nonconstant γ can lead to erroneous estimates of mean river-to-groundwater travel times and that the joint inversion of heads, temperature, and ^{222}Rn considerably reduces the uncertainty associated with estimating river-to-groundwater travel times.

2. Field Study and Laboratory Analysis

2.1. Site Description and Sediment Sampling

River Erpe is an urban lowland river east of Berlin, Germany. The study site is located close to Heidemühle (Latitude 52.478669°N and Longitude 13.635111°E) at a river section, which due to losing conditions and sandy sediments has been subject to previous plot-scale investigations on the fate of trace organic compounds in the hyporheic zone (Schaper et al., 2018, 2019). The present study focuses on river-to-groundwater travel time and ^{222}Rn transport along a transect in the floodplain aquifer southeast of River Erpe (Figure 1a).

Sediment characteristics of both the riverbed ($n = 20$) and the alluvial aquifer ($n = 27$) were determined from samples collected near the transect close to the southeastern riverbank of the River Erpe (see Supporting Information S1, for exact locations). Sediment samples were collected from the upper 10 cm of the riverbed using KSAT rings (METER Group, Germany, USA). In the alluvial aquifer, a hand auger (Eijkelkamp Soil & Water, The Netherlands) was used to drill a borehole until the saturated zone was reached. Sediment samples from the saturated zone were collected using a suction corer (Eijkelkamp) and transferred into KSAT rings. It should be noted that due to limitations of this sampling procedure, sediment samples were collected 1–2 m above the depth of the constructed piezometer screen intervals. Saturated hydraulic conductivities (K) were measured using a KSAT device (METER Group). Thermal conductivities and bulk volumetric heat capacities were assessed in saturated sediments using a KD2Pro device (METER Group).

2.2. Site Instrumentation

Seven piezometers (P1–P7) were installed in the southeastern alluvial aquifer in a transect perpendicular to the channel of the River Erpe (Figure 1a). P0 was installed at the northern end of the transect directly within the stream channel. Two stilling wells (SW) were installed; SW1 was installed in the surface water in close vicinity to P0 and SW2 was installed downstream of a rock ramp. In addition, three piezometers were installed northeast and southwest of the main transect (P8, P9, and P11). Piezometers were constructed of PVC tubes (inner diameter: 4 cm), which were installed using a hand auger for the unsaturated zone and a percussion hammer (Cobra TT, Eijkelkamp, The Netherlands) for the saturated zone. The 25-cm-long filter-screened section of each piezometer was located 4–6 m below the land surface and the riverbed, respectively (Figure 1c, Table S2 in Supporting Information S1). P3, P7, P9, and P11 were equipped with data loggers that measured temperature and absolute pressure, while SW1, SW2, P0, P1, and P8 were equipped with data loggers that measured temperature, absolute pressure, and EC (Micro Divers and CTD divers, respectively, van Essen Instruments, The Netherlands). Absolute pressure was converted into water pressure using air pressure data collected by a barologger (van Essen Instruments) installed in P1. Data were collected between 30 May 2019 and 5 November 2019 with the exception of P0, where data collection commenced on 5 June 2019.

Six multilevel temperature sticks (TS1 to TS6, UIT, Dresden, Germany (Munz et al., 2011)), were installed in the hyporheic zone of the River Erpe close to the southeastern bank (Figure 1a) and collected time series of temperature profiles from 2 June 2019 to 17 June 2019 at varying depths (Table S3 in Supporting Information S1). These time series data were used to calculate the magnitude and direction of vertical Darcy flows using the MATLAB
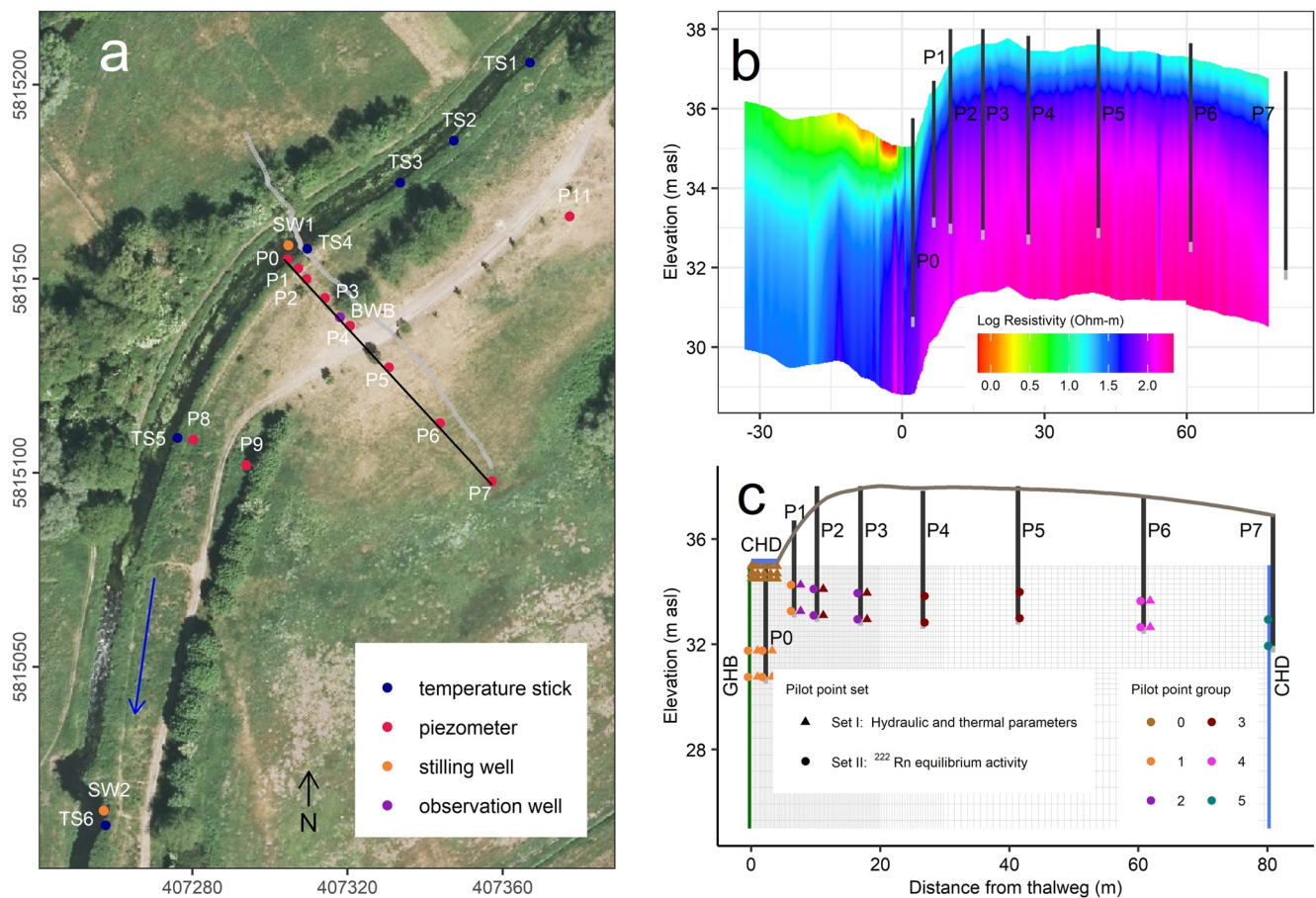
Figure 1. (a) Map of the field site showing the main piezometer transect (black line, P0–P7) and the location of a groundwater observation well of the Berlin Water Works (BWB) located between P3 and P4, two surface water stilling wells (SW1 and SW2), the surrounding piezometers not included in the groundwater flow and transport model (P8, P9, and P11), and 6 multilevel temperature sticks (TS) installed in the riverbed to obtain information on groundwater-surface water exchange flows. The black line shows the length of the model domain. The blue arrow indicates the flow direction of the River Erpe and is located next to a rock ramp. (b) Log resistivity of the subsurface measured along the transect (gray line in panel (a)) from northwest to southeast. (c) Two-dimensional vertical model domain from the thalweg of the River Erpe to piezometer P7, including boundary conditions (GHB = general head boundary, CHD = constant head boundary) and pilot point groups (colored symbols) of the two pilot point sets (I and II). Pilot point set I was used to prescribe hydraulic aquifer parameters (hydraulic conductivity, porosity, and bulk density) and thermal aquifer parameters (thermal conductivity and volumetric heat capacity) (shown as circles), while pilot set II was used to prescribe ^{222}Rn production rates (shown as triangles).

toolbox VFLUX 2.0 (Gordon et al., 2012; Irvine et al., 2015). Within VFLUX, dynamic harmonic regression (Young et al., 1999) is employed to extract phase angles and amplitudes from measured temperature time series, which are then used to calculate phase shifts and amplitude ratios of temperature time series at two different depths. In the present study, measured temperature time series were evaluated via both the combined phase-shift ($\Delta\varphi$) and amplitude ratio (Ar) method of A. M. McCallum et al. (2012) and the Ar method of Hatch et al. (2006). Sediment porosity, volumetric heat capacity, and thermal conductivity were obtained from sediment samples collected as described in Section 2.1. Uncertainty of flux estimates was assessed by a Monte Carlo simulation with 1,000 iterations using the VFLUX 2.0 function "vfluxmc.m." Additional information on flux calculations is provided in Supporting Information S1.

2.3. Near-Surface Geophysics

Ground conductivity measurements were collected using the Frequency Domain Electromagnetic Induction (FDEM) instrument CMD Explorer (GF Instruments, Brno, Czech Republic) to characterize the subsurface. The CMD Explorer is a multireceiver coil, electromagnetic conductivity meter. A transmitting coil creates a time-varying electric field at a frequency of 10,000 Hz. This frequency together with the coil spacings set the depth of

investigation for each receiver-transmitter pair with the depth ranging from ∼2.2 to ∼6.7 m. Further details of the FDEM method are described in Telford et al. (1990) and Reynolds (2011).

The data that the instrument records are converted to apparent conductivity using a low induction number approximation, which is acceptable if the subsurface material is not overly conductive (McNeil, 1980). To evaluate the results, the FDEM data sets were inverted to produce smooth-model inversions of "true" resistivity-depth sections (electrical resistivity, Ω-m where electrical conductivity, S m^{-1} is the inverse of resistivity) using the Aarhus GeoSoftware inversion code AarhusInv (Auken et al., 2015).

Ground conductivity data were collected at 1 s intervals along the transect. Accuracy is 4% at 50 mS m^{-1} with an apparent conductivity range of 0–1,000 mS m^{-1} (resolution 0.1 mS m^{-1}). Measurement locations were georeferenced with a handheld GPS connected to the CMD control unit, and the topographic elevation of each location was obtained from a high-resolution LIDAR digital elevation model of the study area.

2.4. Water Sampling and Analysis

An automated water sampler (3700 Teledyne ISCO, USA) was used to obtain hourly grab samples from the River Erpe, 100 m upstream of P0 between 30 May 2019 10:00 and 14 June 2019 11:00, which were combined in a time integrated manner to yield one surface water sample per day. Groundwater samples were collected from all piezometers over 4 days (5, 7, 11, and 15 June 2019) using a peristaltic pump (12 VDC, Eijkelkamp Soil & Water, The Netherlands) so that three-well volumes were removed prior to sampling. Additionally, in situ parameters (pH, temperature, and EC) had to show constant values before samples were collected.

Samples for chloride analysis were filtered through 0.2 μm cellulose acetate syringe filters (Sartorious, Germany) and analyzed via ion chromatography (Metrohm, Switzerland), following standard protocols (DIN EN ISO 10304-1). For ^{222}Rn analysis in water, water samples of 10 ml each were taken directly from the pump hose using a gas tight glass syringe and carefully injected beneath 10 ml of a liquid scintillation cocktail (MaxiLight®, Hidex Oy, Finland). Laboratory analysis did not commence for at least 3 hr to allow ^{222}Rn and its two progenies ^{218}Po and ^{214}Po to diffuse from the water into the cocktail phase. Afterward, ^{222}Rn concentrations in water samples were determined via liquid scintillation counting (LSC) using a Hidex 300 SL (Hidex Oy, Finland) LSC spectrometer. The spectrometer is equipped with pulse length index processing electronics to discriminate between alpha and beta decays and possesses a triple to double coincidence value of 1 (Eikenberg et al., 2014), that is, an alpha counting efficiency of 100%, and hence does not require calibration. The measurement time of each sample was set to 2 hr and the pulse length index to 10. Subsequently, overall groundwater ^{222}Rn concentrations (Bq L^{-1}) were calculated from overall alpha counts and corrected for decay after sampling via:

$$^{222}Rn = \frac{\text{CPM}_\alpha - \text{CPM}_0}{3\varepsilon V \cdot 60 \cdot e^{-t\lambda_{222}}} \qquad (4)$$

where CPM$_\alpha$ is counts per minute within the alpha (α) channel, CPM$_0$ is the counts per minute within the α channel of a ^{222}Rn free sample, V is the sample volume (10 ml), and λ_{222} is the radioactive decay constant of ^{222}Rn (0.1813 d^{-1}). ε is a transfer factor describing the fractionations of ^{222}Rn, ^{214}Po, and ^{218}Po between water and the LSC cocktail, which was assumed to be 0.9. Groundwater samples were collected and measured in duplicates and triplicates. For ^{222}Rn analysis in river water, two samples were collected 10 cm below the water surface on the 15 June 2019 using the same procedure as described above.

2.5. Determination of ^{222}Rn Equilibrium Concentrations in the Sediments

^{222}Rn equilibrium concentrations were determined in sediment samples collected from the riverbed close to P0 and P1, P3, P5, P6, P7, P8, and P11. Approximately 1.45 kg of saturated sediment was incubated in 700-ml gas-tight glass jars (height = 21 cm, inner diameter 6.5 cm). Sediment was filled to the top of the glass jar with a maximum of 40 ml of water above the sediment-water interface. Sets of duplicate or triplicate water samples were taken after incubation periods of at least 22 days. Ten ml of pore water were sampled from the bottom of the glass jars using a peristaltic pump connected to USGS Minipoint samplers (Duff et al., 1998) via silicon tubing and a pump rate of approximately 5 ml min^{-1}. To minimize ^{222}Rn degassing, sampled water was injected directly

beneath 10 ml of MaxiLight® scintillation cocktail. Exact incubation periods and sediment characteristics of the incubated materials are compiled in Table S4 in Supporting Information S1.

3. Numerical Modeling

MODFLOW-2005 (Harbaugh, 2005) was used to simulate groundwater flow, applying the Newton formulation of MODFLOW-NWT (Niswonger et al., 2011). MT3D-USGS (Bedekar et al., 2016) was employed to simulate heat transport, groundwater age, and the reactive transport of ^{222}Rn. Scripting of input files and post-processing of model results were conducted using the Python package FloPy (Bakker et al., 2016) and R (R Core Team, 2019), respectively. Conditional parameter distributions were inferred using the differential adaptive Metropolis algorithm implemented in DREAM (Vrugt, 2016).

3.1. Groundwater Flow Model

The model domain was a two-dimensional vertical cross section discretized in 46 layers, 1 column, and 176 rows with a bottom elevation of 17.5 m asl and a top elevation of 35 m asl. The layer thickness at the top of the model domain was set to 0.05 m and increased to 0.75 m with depth. The row width was set to 0.2 m at the northern end of the model domain and increased to 1 m at the southern end. The simulated time period was 62.5 days (30 May 2019 12:00 to 31 July 2019 23:55) and was discretized into 6-hr-long stress periods. Measured time series of river stage served as an inflow boundary condition, which was assigned to the first 20 rows in the uppermost layer as time varying, specified head using the constant head package (CHD; Figure 1c). At the southern end of the model domain, the outflow boundary was also defined as a time varying specified head boundary (CHD), using the groundwater levels measured in P7. A general head boundary (GHB) was implemented 5 m below the river boundary to simulate regional groundwater flow underneath River Erpe. Because the exact groundwater level of the GHB was not known, the assigned groundwater level was adjusted during parameter inference.

Measured time series of hydraulic heads in P0, P1, and P3 and manual head measurements in P4 to P6 were used as model constrains. Between 29 May 2019 and 31 July 2019, total precipitation measured at the weather station at IGB Berlin (Latitude 52.448788°N and Longitude. 13.64753°E) was 96.4 mm and the mean ambient air temperature measured by the baro datalogger installed 20 cm below the land surface in P1 was 20.6°C. During the same time period, water levels in both P11 and P7 decreased by 8.6 cm (Figure S5 in Supporting Information S1). Therefore, recharge was assumed to be negligible during the time period of the present study. The initial distribution of groundwater heads for the transient groundwater flow simulation was determined by running MODFLOW-NWT in a steady-state mode using the same hydraulic conductivity field and the measured groundwater levels of 29 May 2019.

3.2. Heat Transport Simulations

In the present study, the two-dimensional transport of heat was simulated via the convection-conduction equation. Similar to heat transport in MT3DMS (Zheng & Wang, 1999), heat transport in MT3D-USGS is based on the flow field provided by MODFLOW-NWT and is simulated as solute transport replacing concentration by temperature and applying thermal retardation via equilibrium "sorption" and a "linear isotherm" (compare Section S3 in Supporting Information S1 for details). Measured temperature time series in the River Erpe (SW1) and in P7 served as fixed-temperature in the inflow and outflow, respectively, which were implemented using the MT3D-USGS Source and Sink Mixing Package. Measured time series of temperature in P0, P1, and P3 were used as data to calibrate the model. An initial temperature field was obtained by linearly interpolating temperatures at the first simulation day (29 May 2019) via kriging. It should be noted that vertical conductive heat transport from the land surface was assumed to be negligible, because river-to-groundwater travel times to P0, P1, and P3 were relatively short compared to the time scales of seasonal land surface temperature fluctuations (compare Section 4.4).

3.3. Explicit Simulation of ^{222}Rn Using MT3D-USGS

In the present study, reactive transport of ^{222}Rn was explicitly simulated by substituting the rate expression of Equation 1 into the advection-dispersion equation:

$$\theta \frac{\partial (^{222}Rn)}{\partial t} = \nabla \left(\theta \boldsymbol{D}_h \nabla^{222} Rn \right) - \nabla (\boldsymbol{q}^{222} Rn) + \theta \left(\gamma - \lambda_{222}{}^{222} Rn \right) \tag{5}$$

where $\boldsymbol{D}_h$ is the hydrodynamic dispersion tensor (m^2 s^{-1}), parameterized by longitudinal and transverse dispersivities times the mean linear velocity of water, $\boldsymbol{q}$ is the specific discharge vector (m s^{-1}), ^{222}Rn denotes the ^{222}Rn concentration in groundwater (Bq L^{-1}), and γ is the quasi-zero-order production rate of ^{222}Rn (Bq L^{-1} s^{-1}). MT3D-USGS cannot simultaneously simulate zero-order production and first-order decay for the same species. We therefore simulate the production of ^{222}Rn by introducing a virtual dissolved mother compound ^{moth}Rn* that linearly decays to ^{222}Rn but hardly changes its concentration. The latter can be achieved by guaranteeing an extremely small Damköhler number ($Da = \lambda_{moth} x \theta / q$) of the mother compound within the domain, in which λ_{moth} is the first-order decay coefficient of this compound. Note that ^{moth}Rn* is a virtual solute with practically uniform concentration and spatially variable decay coefficient λ_{moth}, whereas the true mother compound ^{226}Ra is immobile, has a constant decay coefficient and a spatially variable concentration. We introduce an arbitrarily high inflow concentration ^{moth}Rn* at all inflow boundaries. Then, the first-order decay coefficient λ_{moth} of the mother compound must be set in the following way in order to meet the desired zero-order production rate of ^{222}Rn:

$$\lambda_{moth}(\mathbf{x}) = \frac{\gamma(\mathbf{x})}{^{moth}Rn^*} \tag{6}$$

Formally, the virtual mother compound undergoes advective-dispersive-reactive transport:

$$\theta \frac{\partial \left(^{moth}Rn^* \right)}{\partial t} = \nabla \left(\theta \boldsymbol{D}_h \nabla^{moth} Rn^* \right) - \nabla \left(\boldsymbol{q}^{moth} Rn^* \right) - \theta \lambda_{moth} \tag{7}$$

subject to uniform initial and inflow concentrations of 153,578 Bq L^{-1}. The coupled transport equations of ^{222}Rn and ^{moth}Rn* are implemented in MT3D-USGS via its linear-decay-chain capabilities. As pointed out above, ^{moth}Rn* does practically not vary within the domain. It is neither the true mother compound ^{226}Ra, nor is λ_{moth} its true decay coefficient. The only purpose of introducing this compound is to achieve a spatially variable zero-order production rate γ of ^{222}Rn, which is discussed in the following by its associated equilibrium concentration:

$$^{222}Rn_\infty(\mathbf{x}) = \frac{\gamma(\mathbf{x})}{\lambda_{222}} = \frac{\lambda_{moth}(\mathbf{x})}{\lambda_{222}}{}^{moth} Rn^* \tag{8}$$

In Section S5 in Supporting Information S1, we demonstrate that the chosen approach yields the analytical ^{222}Rn-age profile for the case of uniform flow. We have also checked that the concentrations of the virtual mother compound vary by less than 1% within the domain so that unintended spatial differences in the production rate γ of ^{222}Rn can be excluded. It should be noted that a spatially variable zero-order production rate γ of ^{222}Rn could also be expressed in terms of the sediment emanation rate (E, see Section S5 in Supporting Information S1 for details). The longitudinal dispersivity was allowed to vary between 0.01 and 1 m and the ratio of longitudinal dispersivity to vertical transverse dispersivity was set to 0.1. The ^{222}Rn concentration prescribed at the inflow boundary at the northwestern section of the model domain was set to 0 Bq L^{-1} because sampled water in P0 was chemically distinct from regional groundwater and considered to originate from infiltrated surface water (compare Section 4.1).

3.4. Parameterization, Parameter Inference, and Model Scenarios

Values for hydraulic conductivity (K), sediment porosity (θ), bulk density of the sediment (ρ_b), the thermal transport parameters, and equilibrium ^{222}Rn$_\infty$ concentrations were assumed to be time invariant during the simulation period. Model parameterization was achieved using two sets of pilot points (Doherty, 2003). To reduce the number of total parameters, both sets of pilot points were further grouped into pilot point groups (PPGs). At all pilot points of a PPG, the same parameter was prescribed during parameter inference and subsequently kriging was used as a spatial interpolation method to distribute parameter values across the model domain. The semivariogram parameters were chosen such that interpolated values were close to prescribed values (exponential semivariogram model with slope 190 and nugget 0.001). The first set of pilot points (set I) was used to distribute the values of K, θ and the thermal transport parameters across the model domain and consisted of 30 pilot points,

Table 1
Overview of the ^{222}Rn Transport Formulation and Data to Condition the Model Used in the Different Model Scenarios A to D

Model scenario	Groundwater heads	Temperature	^{222}Rn transport formulation	^{222}Rn production rate (γ)	Prior range for γ
A	x	x	–	–	–
B	x		Explicit	Variable	Lab and field data
C	x	x	Explicit	Constant	Lab and field data
D	x	x	Explicit	Variable	Lab and field data
τ_{app}	–	–	Implicit	Constant	Field data

Note. All model scenarios used hydraulic heads as model constraints. Additionally, scenario A used temperature as a model constraint, scenario B used ^{222}Rn concentrations as model constraints and was based on nonuniform ^{222}Rn production rates, scenario C used temperature and ^{222}Rn concentrations as model constrains and was based on a constant ^{222}Rn production rate, and scenario D used temperature and ^{222}Rn concentrations as model constraints and was based on nonuniform ^{222}Rn production rates.

which were assigned to five PPGs. PPGs 0, 1, 2, 3, and 4 of set I corresponded to sediments in the riverbed (Group 0) and to sediments in the alluvial aquifer in the vicinity of P0 (Group 1), P1 (Group 2), P2 and P3 (Group 3), and P6 (Group 4) and contained 14, 5, 2, 4, and 2 pilot points, respectively (Figure 1c). Values of thermal transport parameters and θ prescribed at PPG 3 were also prescribed at PPG 4 due to a lack of measured temperature time series in piezometers P4 to P6. Specific storage was set to 1×10^{-5} m^{-1} and was assigned homogeneously throughout the model domain. Specific yield was assumed to be a fraction of porosity and calculated from porosity using a specific yield to porosity ratio, which was prescribed for the entire model domain and inferred during parameter inference. To account for the heterogeneity of measured ^{222}Rn concentrations within the aquifer, particularly between P5 and P6, and for the data availability compared to hydraulic head and temperature time series, ^{222}Rn$_\infty$ concentrations were distributed across the model domain using a second set of pilot points (set II). Set II consisted of 32 pilot points grouped into six PPGs, which are located in the riverbed (Group 0), in the vicinity of P0 and P1 (Group 1), P2 and P3 (Group 2), P4 and P5 (Group 3), P6 (Group 4), and P7 (Group 5) and contained 14, 5, 2, 4, 2, and 2 pilot points, respectively (Figure 1c).

In the present study, parameter inference was achieved using the differential evolution adaptive Metropolis algorithm, DREAM (Vrugt, 2016; Vrugt et al., 2009). During each DREAM run, 28 Markov chains were run in parallel. Convergence of the chains to a limiting posterior distribution was declared after the $\hat{R}$-statistic of Gelman and Rubin (Gelman & Rubin, 1992) for each model parameter had fallen below 1.2. Following convergence, 28,000 additional iterations were performed to sample the posterior distribution.

Hydraulic and thermal parameters as well as distributions of mean river-to-groundwater travel times to each piezometer were estimated for four model scenarios (A–D). In all four scenarios, hydraulic heads were used to condition the parameters of the flow model, while temperature was used only in scenarios A, C, and D. Measured ^{222}Rn concentrations in groundwater were used as data for conditioning in scenarios B, C, D. In scenarios B and D, ^{222}Rn concentrations were modeled using spatially varying production rates as described in Section 3.3 (Equations 6–8), while in scenario C, a constant production rate was prescribed across the entire model domain (Table 1). The same set of boundary conditions and the same prior parameter distributions were used in all four model scenarios (Table 2).

All prior parameter ranges were set as uniform priors. The priors for the thermal sediment parameters, porosity, and hydraulic conductivity were in accordance with measured parameter ranges (Table 1, Figure 4), while the prior range for longitudinal dispersivity (α_L) was chosen according to values used in similar studies (Engelhardt et al., 2013; Gilfedder et al., 2019; Ma et al., 2012). Values for bulk densities and specific storage were not adjusted during parameter estimation. For specific storage, a literature value was chosen (Engelhardt et al., 2013), while bulk densities were set according to the mean of their measured values in riverbed and alluvial sediment samples. Because measured values for hydraulic conductivity and porosity were significantly different between riverbed and alluvial sediments (Section 3.1), prior parameter ranges were set differently for PPG 0 and PPGs 1

Table 2

Measured Values and Uniform Prior Ranges of Saturated Hydraulic Conductivity (K), Porosity (θ), Thermal Conductivity (κ₀), Bulk Volumetric Heat Capacity ($p_b c_b$), Specific Storage, Specific Yield, Bulk Density (ρ_b), ^{222}Rn Concentration at Secular Equilibrium ($^{222}Rn_\infty$), Longitudinal Dispersivity (α_L) Specific Storage, Specific Yield Ratio, and the Hydraulic Head Prescribed at the General Head Boundary of the Northwestern Model Domain Boundary Expressed as Head Difference to the Head Measured in P0 (Δh to P0) Used for Parameter Inference in the Joint MODFLOW and MT3D-USGS Model

| Parameter | Units | Measured values | | | Uniform prior range | | Pilot point set |
		Median, IQR	Min–max	n	Min–max	Pilot point group	
K riverbed	m s^{-1} 10^{-5}	2.5, 5.8	0.6–11.1	8	0.28–13.9	0	I
K all. aq.	m s^{-1} 10^{-4}	1.9, 3.2	0.44–13.8	18	0.28–13.9	1, 2, 3, 4	I
K NW all. aq.	m s^{-1} 10^{-8}	2.8, n.a.	n.a.	1	–	–	I
θ riverbed	m^3 m^{-3}	0.38, 0.05	0.33–0.42	10	0.3–0.45	0	I
θ all. aq.	m^3 m^{-3}	0.32, 0.03	0.23–0.37	18	0.2–0.4	1, 2, 3	I
κ_0 riverbed	W m^{-1} K^{-1}	2.6, 0.3	2.1–2.8	8	2.1–2.8	0, 1, 2, 3, 4	I
κ_0 all. aq.	W m^{-1} K^{-1}	2.8, 0.21	2.6–3.0	12	2.6–3.0	0, 1, 2,3, 4	I
$p_b c_b$ riverbed	MJ m^{-3} K^{-1}	2.9, 0.1	2.7–3.1	8	2.7–3.1	0, 1, 2, 3, 4	I
$p_b c_b$ all. aq.	MJ m^{-3} K^{-1}	2.7, 0.2	2.6–2.9	12	2.6–2.9	0, 1, 2, 3, 4	I
^{222}Rn$_\infty$ riverbed	Bq L^{-1}			2	1.0–2.0	0	II
^{222}Rn$_\infty$ all. aq.	Bq L^{-1}			9	1.3–7.0	1, 2, 3, 4, 5	II
ρ_b riverbed	g cm^{-3}	1.93, 0.08	1.81–2.02	9	1.93*	n.a.	n.a.
ρ_b all. aq.	g cm^{-3}	2.08, 0.06	1.95–2.17	18	2.07*	n.a.	n.a.
α_L	m	n.a.	n.a.	n.a.	0.01–1	e.d.	n.a.
Specific yield ratio	m^3 m^{-3}	n.a.	n.a.	n.a.	0.2–1.0	e.d.	n.a.
Specific storage	m^{-1}	n.a.	n.a.	n.a.	5.5×10^{-5}*	e.d.	n.a.
Δh to P0	m	n.a.	n.a.	n.a.	−0.1–0.1	e.d.	n.a.

Note. Pilot point set I comprises 5 pilot point groups, while set II comprises 6 pilot point groups. n = number of samples; n.a. = not available; e.d. = prescribed over entire model domain; * = values were not modified during parameter inference; all. aq. = alluvial aquifer; NW = northwestern; IQR = interquartile range.

to 4 in set I. Likewise, because measured ^{222}Rn concentrations differed between riverbed and alluvial sediments, prior parameter ranges were set differently for PPGs 0 and 1 to 5 within set II.

3.5. Simulation of River-to-Groundwater Travel Times

Similar to previous studies on travel times from rivers into alluvial aquifers (Engelhardt et al., 2013; Gilfedder et al., 2019), the mean travel time from the River Erpe through the model domain (alluvial aquifer), that is, the mean groundwater age, was approximated by solving the mean groundwater age equation of Goode (1996). The method treats time as "age mass" and simulates the mean groundwater travel time μ_τ using the advection-dispersion equation accounting for mixing along and between flow paths. Aging of water parcels is achieved by implementing a zero-order source term of unity across the entire model domain. The travel time in the inflow is zero. During posterior sampling, that is, after convergence of the Markov chains to a stable distribution, groundwater travel times were simulated in each realization using the same parameter set as used for heat and ^{222}Rn transport (Sections 3.3 and 3.4). This approach results in posterior probability density distributions of the mean travel time from the river to locations in the model domain arising from various model realizations during posterior sampling. The distribution of mean travel times among the accepted model realizations was further summarized by its ensemble median (med$_{\mu_\tau}$) and ensemble inter-quartile range (IQR$_{\mu_\tau}$). Note that IQR$_{\mu_\tau}$ expresses the uncertainty in identifying the mean travel time at a given point rather than the mean spread of the travel time distribution at this point within single realizations.

In addition to μ_τ, apparent ^{222}Rn ages τ_{app} were calculated using the analytical solution to Equation 1 of Cranswick et al. (2014), assuming a uniform spatial distribution of the ^{222}Rn production rate and thus of the corresponding equilibrium concentration ^{222}Rn$_\infty$:

$$\tau_{\mathrm{app}} = \frac{1}{\lambda_{222}} \ln \left[\frac{^{222}Rn_{\infty} - {}^{222}Rn_{riv}}{^{222}Rn_{\infty} - {}^{222}Rn_{p}} \right] \tag{9}$$

in which $^{222}Rn_{riv}$ and $^{222}Rn_{p}$ are the ^{222}Rn concentrations in the river and in the piezometer under study, respectively. To account for the uncertainty in ^{222}Rn concentration at secular equilibrium ($^{222}Rn_{\infty}$), a Monte Carlo simulation was conducted to estimate distributions of τ_{app}. The uniform range of possible $^{222}Rn_{\infty}$ concentrations was chosen according to measured ^{222}Rn concentrations in groundwater collected from P9, P7, and P11 and thus was 3.8–5.2 Bq L^{-1}.

4. Results and Discussion

4.1. Site Characteristics, Conceptual Model of Groundwater Flow, and Patterns of Groundwater-Surface Water Interactions

Sediments of the riverbed and northwestern and southeastern alluvial aquifer showed considerable differences (Table 2). While riverbed sediment and sediment samples in the southeastern alluvial aquifer mainly consisted of medium- to coarse-grained sand (d$_{50}$ median ± interquartile range = 0.48 ± 0.14 mm, $n = 10$) with varying fractions of organic material, the only sample collected from the northwestern alluvial aquifer (Figure S1 in Supporting Information S1) predominantly consisted of peat-like organic material. Saturated hydraulic conductivities (K) ranged between 1.7×10^{-6} and 1.4×10^{-3} m s^{-1} in samples collected from the riverbed and from the southeastern alluvial aquifer, while hydraulic conductivity of the peat-like materials found in the northwestern alluvial aquifer was considerably lower ($<1.0 \times 10^{-8}$ m s^{-1}). Moreover, log-transformed hydraulic conductivities in the southeastern alluvial aquifer (geometric mean, $\mu_g \pm \sigma_g = 1.8 \times 10^{-4} \pm 7.8 \times 10^{-4}$ m s^{-1}, $n = 18$) were significantly higher than in the riverbed around the main transect ($\mu_g \pm \sigma_g = 1.9 \times 10^{-5} \pm 7.2 \times 10^{-4}$ m s^{-1}, $n = 7$, Welch's t-test, $p < 0.01$, $t(11.3) = 4.5$). On the contrary, porosity (θ) determined from oven dried sediments (105°C, 72 hr) and bulk densities (ρ_b) were significantly lower (Welch's t-test, $p < 0.01$, t(14.4) = -5.6 and Welch's t-test, $p < 0.01$, t(14.6) = 5.0, respectively) in the riverbed around the main transect (0.38 ± 0.04, $n = 9$ and 1.93 ± 0.037 g cm^{-3}, $n = 9$, respectively) than in the southeastern alluvial aquifer (0.31 ± 0.03, $n = 18$ and 2.07 ± 0.06 g cm^{-3}, $n = 18$, respectively). Thermal conductivities and bulk volumetric heat capacities ranged between 2.1 and 3.0 W m^{-1} K^{-1} and 2.6 and 3.1 MJ m^{-3} K^{-1}, respectively.

The bulk resistivity profiles, obtained by FDEM inversion, chloride concentrations, hydraulic heads, and estimated vertical Darcy fluxes were used to conceptualize the general hydrogeological setting and infer the patterns of groundwater-surface water exchange flows at the field site (Figures 1b and 1c). The regional groundwater flow direction, inferred from groundwater contour lines provided by the Berlin Senate, indicates flow from north to south (Figure S2 in Supporting Information S1). Downward Darcy fluxes estimated via the combined McCallum $\Delta\varphi$Ar method ranged from 0.40 to 1.50 m d^{-1} at TS4 and from -0.03 to 0.21 m d^{-1} elsewhere (Figure S3 in Supporting Information S1). In comparison, the estimated fluxes using the Hatch Ar method, ranged from 0.43 to 1.27 m d^{-1} at TS4 and from -0.47 to 0.29 m d^{-1}, elsewhere. The fluxes estimated from TS4 agree with fluxes estimated previously at the same site (Schaper et al., 2019) and with negative hydraulic gradients (i) measured between P0 and SW1 ($i = -0.15 \pm 0.02$) and P1 and SW1 ($i = -0.22 \pm 0.02$). Although head gradients between P8 and SW1 ($i = -0.18 \pm 0.02$) were also negative, strong vertical fluxes were likely impeded by lower riverbed hydraulic conductivities measured around TS5 (Figure S1 and Table S1 in Supporting Information S1). In contrast, head gradients between P11 and SW1 ($i = -0.02 \pm 0.00$) were small and unlikely to create considerable downward fluxes upstream of TS4. It should be noted that because temperature time series recorded in the riverbed by TS1–TS6 were measured only to a maximum depth of 0.25 m, the downward Darcy fluxes estimated via VFLUX 2.0 were not used as model constrains in the numerical heat transport model presented in Sections 3.2 and Section S3 in Supporting Information S1.

Median (± interquartile range, IQR) chloride concentrations in the River Erpe (115 ± 22 mg L^{-1}, $n = 17$) and the southeastern alluvial aquifer (121 ± 16 mg L^{-1}, $n = 44$) were largely similar, but were considerably higher compared to chloride concentrations measured in the groundwater observation well (BWB, screened section -5 to -1 m above sea level, asl, 38.5–42.5 m below land surface, 37 mg L^{-1}, $n = 2$) of the Berlin Water Works located between P3 and P4. It is therefore likely that water sampled from the relatively deep BWB observation well represents regional groundwater, which did not mix with infiltrating surface water from the River Erpe.
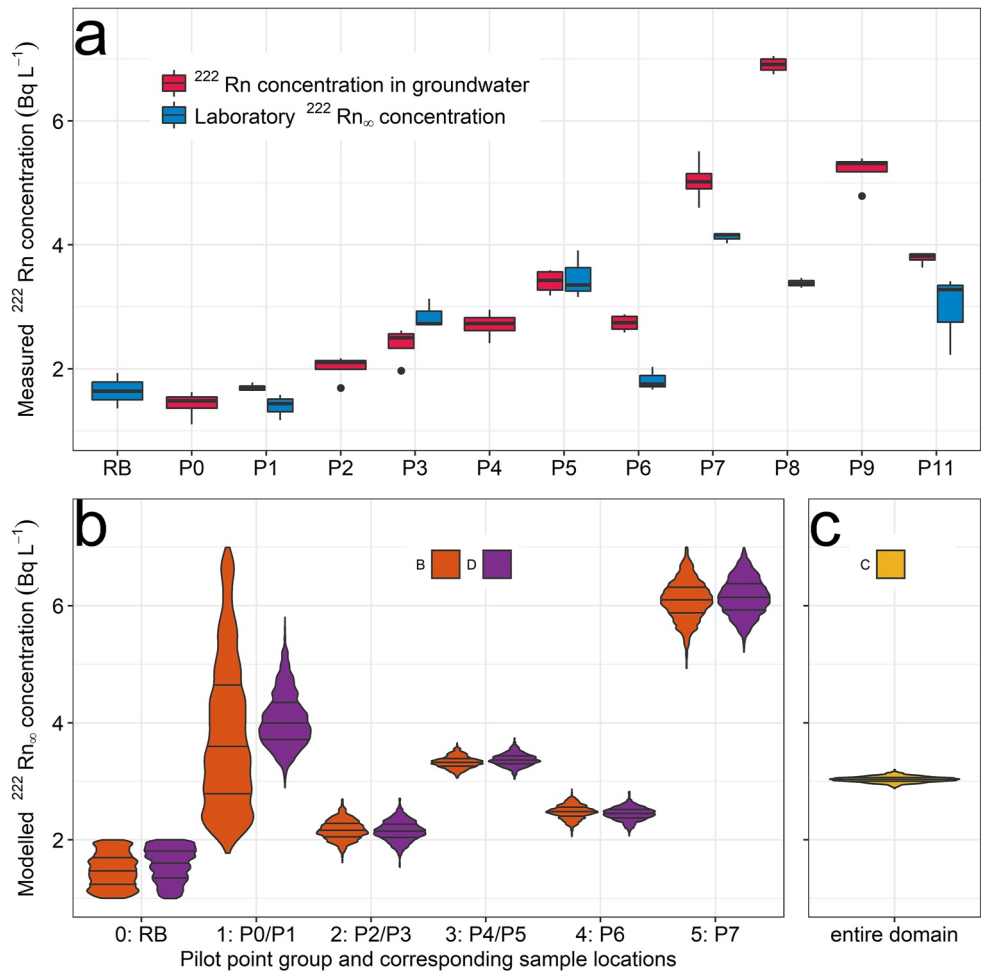
Figure 2. (a) Boxplots of measured ^{222}Rn concentrations (Bq L^{-1}) in the southeastern alluvial aquifer (red) and in the incubation experiments after 21 days (blue). (b) Violin plots of posterior distributions of estimated ^{222}Rn$_\infty$ concentrations at the five pilot point groups of set II for model scenarios B and D (Table 1). (c) Violin plot of the posterior distribution of estimated uniform ^{222}Rn$_\infty$ concentration for model scenario C (Table 1). In both panels, median concentrations and interquartile ranges are represented by the horizontal black line and the boxes, respectively. The most extreme data values, which are away from the box by less than 1.5 times the length of the box, are shown as whiskers and data values beyond the whiskers are plotted as black dots. RB = riverbed.

The electromagnetics-derived inverted bulk-resistivity data highlight the spatial heterogeneity and extent of the alluvial aquifer material along the piezometer transect and beneath the river. The medium- to coarse-grained sand of the alluvial aquifer on the southeastern side of the River Erpe extends from 5 to 95 m along the transect and has resistivity values that range from 1.8 to 3 Ω-m. Whereas for the more peat-rich, organic material on the northwestern side of the river, the resistivity values are lower (<1.25 Ω-m). Measured head gradients and vertical Darcy fluxes suggest that surface water infiltration into the southeastern alluvial aquifer was strongest around the location of the main piezometer transect (P0–P7). This observation agrees with the ^{222}Rn measurements and temperature time series collected in P11, P8, P1, and P3 (compare Sections 3.2 and 4.3). Consequently, a two-dimensional vertical cross-sectional model domain was set up along the main piezometer transect between P0 and P7, extending from the thalweg of the River Erpe and 80 m into the alluvial floodplain (Figure 1c).

4.2. ^{222}Rn Concentrations in the Alluvial Aquifer

Along the piezometer transect (P0 to P7), mean ($\pm$ one standard deviation, SD), ^{222}Rn concentrations increased from 0 Bq L^{-1} ($n = 2$) in the surface water of the River Erpe to 5.0 $\pm$ 0.4 Bq L^{-1} in P7 (Figure 2a) with the exception of P6, where concentrations were slightly lower compared to P5 and P7. In P11 and P9, ^{222}Rn concentrations

Table 3

Overview of Median Travel Times (Days, med$_{\mu_\tau}$) and Associated Interquartile Ranges (IQR$_{\mu_\tau}$) to Piezometers P0–P7 Calculated Using Implicit (Equation 9, τ_{app}) and Explicit (MT3D-USGS, Model Scenarios A to D) ^{222}Rn Transport Models, Including Calibration Constrains, Root-Mean-Square Errors (RMSE) of the Different Calibration Constrains and Posterior Probability Density Distributions for Longitudinal Dispersivity (α$_L$), Specific Yield to Porosity Ratio, and Δh Used to Prescribe the General Head Boundary at the Northeastern End of the Model Domain

Model scenario	Units	τ_{app}	A	B	C	D
^{222}Rn transport model		Implicit	–	Explicit	Explicit	Explicit
Calibration constrains		Rn	H, temp	H, Rn	H, temp, Rn	H, temp, Rn
^{222}Rn production rate (γ)		Constant	–	Variable	Constant	Variable
RMSE head	cm	–	0.71	0.68	0.70	0.72
RMSE temperature	°C	–	0.21	–	0.25	0.23
RMSE ^{222}Rn	Bq L^{-1}	–	–	0.12	0.75	0.12
Median (±IQR)						
α_L	m		0.6 ± 0.3*	0.7 ± 0.7*	0.8 ± 0.7*	0.9 ± 0.1*
Specific yield ratio	m^3 m^{-3}		0.6 ± 0.2*	0.6 ± 0.2*	0.6 ± 0.2*	0.6 ± 0.2*
$\Delta h = h_{boundary} - h_{P0}$	cm		0.06 ± 0.4	0.7 ± 1.0	0.4 ± 0.4	0.3 ± 0.3
Piezometer						
				med$_{\mu_\tau}$, IQR$_{\mu_\tau}$		
P0	days	2.1 ± 0.6	2.3 ± 0.8	3.3 ± 2.1	3.5 ± 0.5	3.0 ± 0.6
P1	days	2.5 ± 0.6	4.0 ± 1.0	5.0 ± 3.1	5.1 ± 0.3	4.8 ± 0.9
P2	days	3.3 ± 0.9	9.2 ± 1.6	8.9 ± 4.7	8.6 ± 0.6	8.4 ± 1.3
P3	days	4.3 ± 1.3	18.3 ± 4.2	16.8 ± 6.5	14.9 ± 1.1	15.3 ± 2.3
P4	days	5.1 ± 1.5	32.1 ± 8.6	28.9 ± 8.7	24.8 ± 2.0	25.7 ± 4.3
P5	days	7.8 ± 3.0	52.8 ± 14.9	46.8 ± 12.2	39.9 ± 3.6	41.6 ± 7.5
P6	days	5.2 ± 1.4	79.6 ± 23.3	69.5 ± 16.6	59.3 ± 5.6	61.9 ± 11.5
P7	days	n.e.	102.8 ± 30.2	91.5 ± 21.1	76.9 ± 7.3	80.1 ± 14.9

Note. Temp = temperature; H = hydraulic head; Rn = ^{222}Rn; * = posterior indifferent from uniform prior probability density distribution; IQR = interquartile range; n.e. = not estimated.

were similar to concentrations in P5 and P7, respectively. The highest ^{222}Rn concentrations encountered at the field site, however, were found in P8 (6.9 ± 0.1 Bq L^{-1}), despite its location in close vicinity to the River Erpe. No notable trend was detected within the four ^{222}Rn measurements conducted between 5 June 2019 and 15 June 2019, suggesting that with respect to ^{222}Rn, the groundwater system was in a quasi-steady state. This assumption is supported by the fact that hydraulic gradients across the field site changed very little between June and November 2019 (Figure S7 in Supporting Information S1).

Upon infiltration of river water into the subsurface, ^{222}Rn concentrations are expected to increase until secular equilibrium is reached, that is, final equilibrium concentration, where ^{222}Rn production from sediment emanation equals radioactive decay (Equation 1). Acknowledging measurement uncertainty of ^{222}Rn of 5%–10%, secular equilibrium in field settings can typically be assumed when mean travel times in the alluvial aquifer are larger than four to six times the ^{222}Rn half-life, corresponding to 15.2 days (0.93·^{222}Rn$_\infty$) and 23 days (0.98·^{222}Rn$_\infty$), respectively (Gilfedder et al., 2019; Hoehn & von Gunten, 1989). The reactive-transport simulations conducted in the present study suggest that along the groundwater flow path from P0 to P7, secular equilibrium was reached somewhere between P3 and P4 with corresponding travel times of 15.3 ± 2.3 d and 25.7 ± 4.3 d, respectively (compare Section 4.4 and Table 3).

Mean travel times from the river to piezometers P8, P9, and P11, which were not located within the model domain, were semiquantitatively estimated by comparing temperature time series recorded between June and November 2019 to temperature records in the piezometer transect (P0–P7). All temperature time series measured in the alluvial aquifer showed a distinct seasonal pattern (Figure S6 in Supporting Information S1). In P8, close to the river,

differences in mean arrival time estimated from EC time series recorded in August 2019 indicated a mean travel time of approximately 14 days (Figure S5 in Supporting Information S1). Temperature time series recorded in P3 were slightly delayed compared to temperature time series recorded in P8. As differences in thermal sediment parameters of the alluvial aquifer were small (compare Section 4.1), this finding is in agreement with calculated mean travel times to P3 of 15.3 $\pm$ 2.3 d (Table 3, Figure 5) and suggests that the relatively high ^{222}Rn concentrations measured in P8 were approaching ^{222}Rn$_\infty$. By contrast, temperature time series recorded in P9 and P11 resembled the temperature time series measured in P7. It is therefore likely that ^{222}Rn concentrations measured in P8, P9, and P11 and the concentrations measured in P5–P7 represent ^{222}Rn concentrations at or close to secular equilibrium. Assuming that secular equilibrium concentrations are linked to local sediment production rates, this finding suggests that the ^{222}Rn production rates vary by a factor of 2 across the field site.

The observation of non-homogenous ^{222}Rn production rates is qualitatively confirmed by the secular equilibrium concentrations measured in the incubation experiments (^{222}Rn$_{\infty,\,inc}$). ^{222}Rn$_{\infty,\,inc}$ ranged from 1.4 $\pm$ 0.4 Bq L^{-1} (mean $\pm$ 1 SD) in sediments collected in the riverbed to 3.9 $\pm$ 0.4 Bq L^{-1} in sediments collected close to P7 (Figure 2a). In the present study, correlations between ^{222}Rn$_{\infty,inc}$ and resulting emanation rates (E, Bq/kg dry sediment, see Section S7 in Supporting Information S1 for details) and sediment characteristics were assessed via Spearman rank correlation coefficients. ^{222}Rn$_{\infty,inc}$ concentrations were negatively correlated with porosity ($r_s = 0.77$, $n = 10$, $p < 0.05$), but both ^{222}Rn$_{\infty,inc}$ concentrations and resulting emanation rates E were positively correlated with the weight fraction of the sediment that was smaller than 0.2 mm (^{222}Rn$_{\infty,inc}$: $r_s = 056$, $n = 10$, $p < 0.05$, E: $r_s = 0.62$, $n = 10$, $p < 0.05$) and distance from the river thalweg (^{222}Rn$_{\infty,inc}$: $r_s = 0.71$, $n = 10$, $p < 0.05$, E: $r_s = 0.64$, $n = 10$, $p < 0.05$). This finding is in agreement with previous studies which found that ^{222}Rn production rates and the emanation fraction increase with decreasing grain size (Greeman & Rose, 1996; Hoehn et al., 1992; Sakoda et al., 2011).

Measured ^{222}Rn$_{\infty,inc}$ concentrations at P6, P7, P11, and P8 were lower than the direct field measurements of ^{222}Rn concentrations in groundwater samples with the largest differences encountered at P8 (Figure 2). This discrepancy may result from different sediment porosity induced during sediment sampling. Because only disturbed samples could be used, the ratio of pore water to sediment might have been increased compared to in situ conditions. This explanation would be in line with the observation that the posterior parameter distributions for porosity generally centered around lower values in the alluvial aquifer compared to measured values in the sediment samples (Figure 4). The differences could also be due to small-scale heterogeneity and the fact that sediments were not sampled from the piezometer screening depth but about approximately 1–2 m above the filter screen. In addition, it cannot fully be excluded that the observed discrepancy may have been due to methodological shortcomings (e.g., ^{222}Rn loss through the silicon tubing during sampling). Such effects, however, would have affected all sediment incubations equally.

4.3. Model Performance and Parameter Estimates

The modeled time series of groundwater levels and temperatures using the maximum-likelihood parameter set of the different model scenarios closely fit the observed data (Figure 3, Figures S8–S10 in Supporting Information S1) and all have similar root-mean-square errors (RMSE) for hydraulic head and temperature (Table 2). Model scenarios B and D, in which ^{222}Rn production rates vary across the domain, captured the measured ^{222}Rn concentrations relatively well, indicated by a low RMSE of ^{222}Rn concentrations, which falls within a measurement uncertainty of 10%–15%. However, in model scenario C, the RMSE of ^{222}Rn was considerably higher than the measurement uncertainty and also than in the model scenarios B and D. The explicit transport model, which assumed a constant ^{222}Rn production rate, was thus not capable to represent the measured spatial variability in ^{222}Rn concentration along the transect.

The posterior probability density distributions (posteriors) of hydraulic and thermal transport parameters were distinct from their prior probability density distributions (priors) in all model scenarios, suggesting that in general, the models were sensitive to these parameters (Figure 4) and the degree of equifinality across the different scenarios was low. An exception was porosity (θ), for which the posteriors of scenarios A and B were similar to their priors for all PPGs and for PPG 0 and 1, respectively. In all model scenarios, θ correlated with thermal retardation (R_{th}) and dispersion coefficients (D_T) as both thermal parameters are lumped parameters, including the effects of parameters related to heat transport in porous media, that is, porosity and the thermal conductivities and heat capacities of both the bulk sediment and the solid phase (Figures S11–S14 in Supporting Information S1).
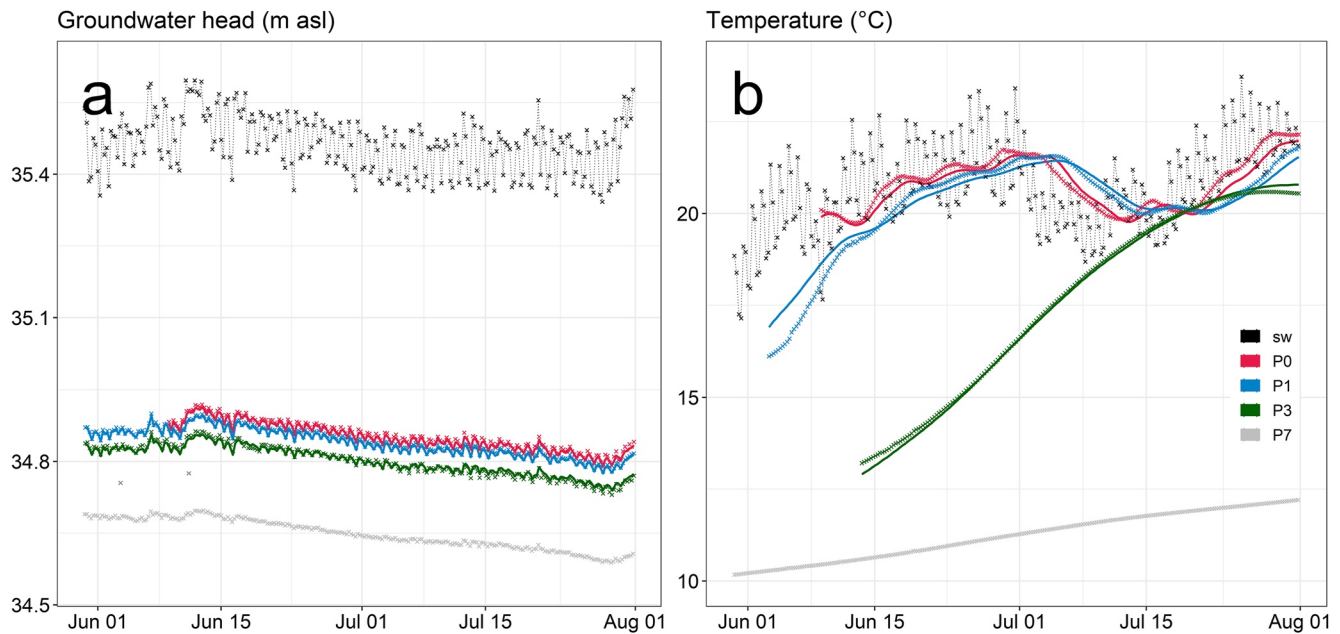
Figure 3. Measured (crosses) and modeled (lines, none for surface water) times series (scenario D, Table 1) of hydraulic heads (a) and temperature (b) in piezometers P0, P1, and P3. Also shown are measured time series in River Erpe (SW1) and in P7, which served as inflow and outflow boundary (CHD) for both groundwater flow and temperature transport simulations. Note that the measured and modeled curves for P0, P1, and P3 are hardly distinguishable because they lie on top of each other. Note further that data collection in P0 only commenced on 06/07/2019. For temperature transport, only data collected after one mean travel time to the respective piezometer are depicted. Modeled time series of scenario A to C are shown in Figures S6–S8 in Supporting Information S1; sw = surface water.

In scenario A, effects of θ were thus incorporated in R_{th} and D_T as no further constrain on transport was available. Accordingly, in scenarios C and D, where both ^{222}Rn and temperature measurements were used to constrain the model, the posterior variances of D_T and R_{th} decreased compared to scenario A, particularly for PPGs located in the alluvial aquifer. Likewise, the variance of ^{222}Rn production rates and θ in PPG 1 decreased from scenario B to D as temperature was added as data. Both the posteriors of specific yield to porosity ratio and of longitudinal dispersivity (α_L) were not distinct from their uniform priors, indicating that the model was not sensitive to specific yield and used a wide range of α_L (0.01–1) and transverse dispersivity (0.001–0.1) during transport simulations.

Median values of the posteriors for R_{th} and D_T in PPG 0 were in reasonable agreement with median measured values, while for PPG 1 and PPGs 2 and 3, the median values were slightly lower and higher compared to measured values, respectively. For hydraulic conductivity (K) and porosity (θ), median measured values in the riverbed were similar to modeled ranges, whereas median posterior values in the alluvial aquifer were, on average, considerably higher and lower than the median of measured values, respectively. In particular, K values of the last PPG close to P6 tend to cluster around the upper boundary of their prior range, while values of θ are grouped around their lower prior range. The question whether the discrepancy between measured and modeled parameters is due to the boundary condition prescribed at P7, a lack of transferability of laboratory measurements to in situ conditions or issues related to the conceptual groundwater flow model cannot fully be addressed within the scope of the present study. However, since prior parameter ranges were constrained by measured sediment characteristics and simulated heads matched measured heads closely in all model scenarios tested, all posterior model realizations are still considered to be plausible representations of groundwater flow and transport along the main piezometer transect.

The median of the posteriors of ^{222}Rn equilibrium concentrations for scenario C, for which a constant ^{222}Rn production rate was prescribed across the entire domain was 3.3 Bq L^{-1} and thus considerably lower compared to the concentrations measured in P7, P8, and P9 (Figure 2). In the two scenarios, in which ^{222}Rn production rates could vary across the domain (B and D), median posterior values ranged from 1.5 to 1.6 to 6.1 and 6.1 Bq L^{-1} in model scenarios B and D, respectively. For the southeastern section of the model domain (PPG 3 to PPG 5), posteriors of scenarios B and D were similar, whereas posteriors of PPGs 0 and 1, close to the River Erpe are more constrained in scenario D. This finding was expected because temperature time series were available only
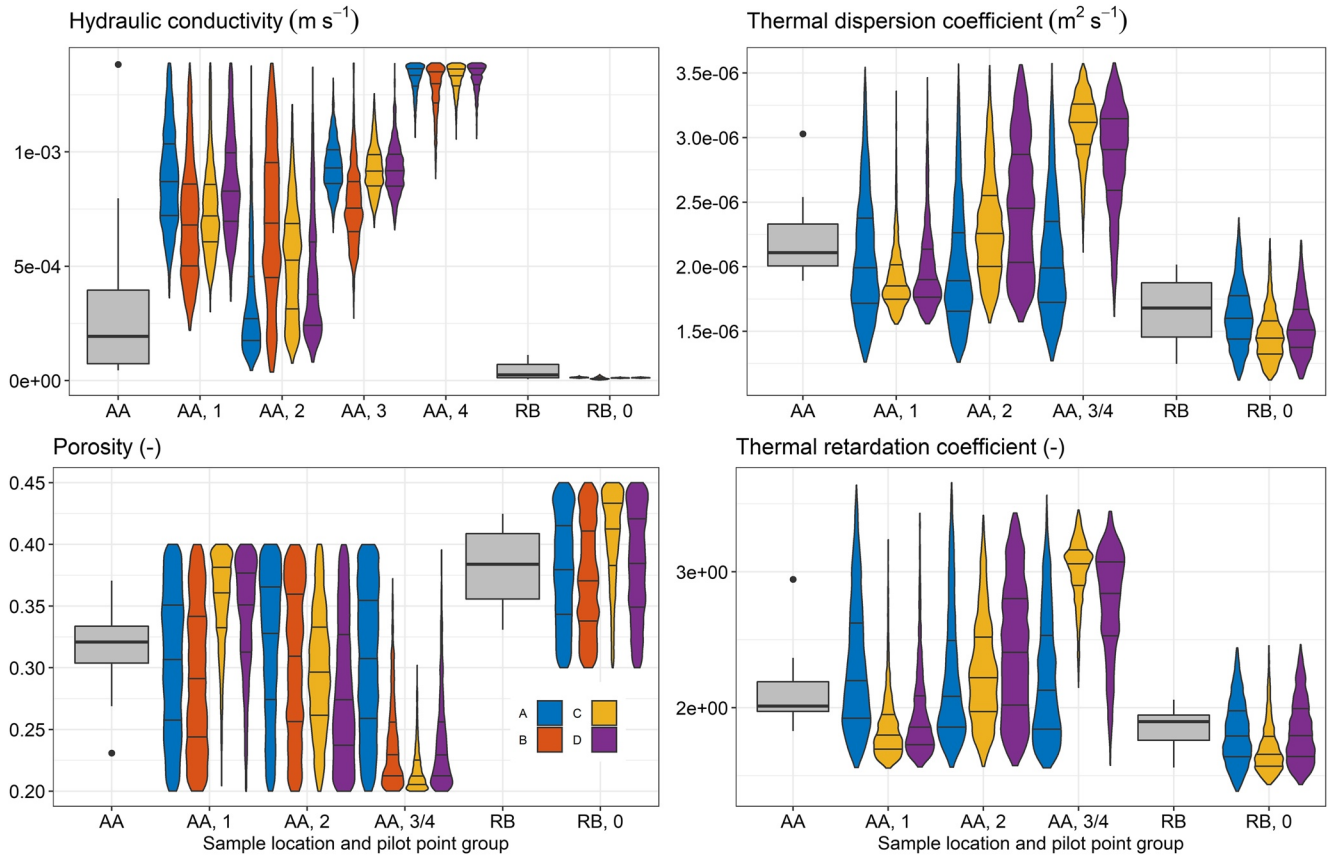
Figure 4. Measured values (boxplots) in the riverbed (RB) and the alluvial aquifer (AA) and posterior probability density distributions (violin plots) of saturated hydraulic conductivity (K), porosity (θ), thermal dispersion, and thermal retardation coefficients estimated by DREAM runs for the different pilot point groups and all four model scenarios (Table 1). In the boxplots and violin plots, median values and interquartile ranges are represented via the horizontal black line and the boxes and the three horizontal lines, respectively.

in P0, P1, and P3 and thus transport characteristics between P4 and P7 were only defined by measured ^{222}Rn concentrations and heads.

4.4. Estimation of Water Travel Times in the Alluvial Aquifer

The ensemble median of the mean travel times (med_{μ_τ}) estimated in the different model scenarios A (hydraulic heads and temperature), B (hydraulic heads, ^{222}Rn, and variable γ), C (hydraulic heads, temperature, ^{222}Rn, and constant γ), and D (hydraulic heads, temperature, ^{222}Rn, and variable γ) increased from P0 ($\mathrm{med}_{\mu_\tau,\mathrm{A}}$: 2.3 d, $\mathrm{med}_{\mu_\tau,\mathrm{D}}$: 3.0) to P7 ($\mathrm{med}_{\mu_\tau,\mathrm{A}}$: 102.8 d, $\mathrm{med}_{\mu_\tau,\mathrm{D}}$: 80.1, Table 3, Figure 5).

The travel times estimated by explicitly simulating ^{222}Rn transport were by factors ranging from 1.4 (at P0) to 11.9 (at P6) larger than the mean apparent ^{222}Rn ages (Table 3). This finding can be explained by dispersive mixing leading to a distribution $p(\tau)$ of travel times at an observation point. Already in the simplest case of one-dimensional advective-dispersive-reactive transport with uniform coefficients, the recovery of the decaying compound is affected by dispersion, namely $\exp\left(-2\lambda x/\left(v + \sqrt{v^2 + 4D\lambda}\right)\right)$, but interpreted in the apparent-age interpretation of the data by $\exp(-2\lambda x/v_{\mathrm{app}})$, so that the apparent velocity is $v_{app} = (v + \sqrt{v^2 + 4D\lambda})/2$, which is bigger than the true velocity v for any combination of positive D and λ. Any other distribution $p(\tau)$ of travel times at observation points, except for the Dirac delta function, would also lead to a ^{222}Rn-age $\tau_{\mathrm{app}} = -\ln\left(\int_0^\infty p(\tau)\exp(-\lambda\tau)d\tau\right)/\lambda$ that differs from the mean groundwater age $\int_0^\infty \tau p(\tau)d\tau$. To evaluate the effect of dispersive mixing on τ_{app}, a test case scenario was constructed in which, μ_τ, simulated using the median parameter values for hydraulic conductivity
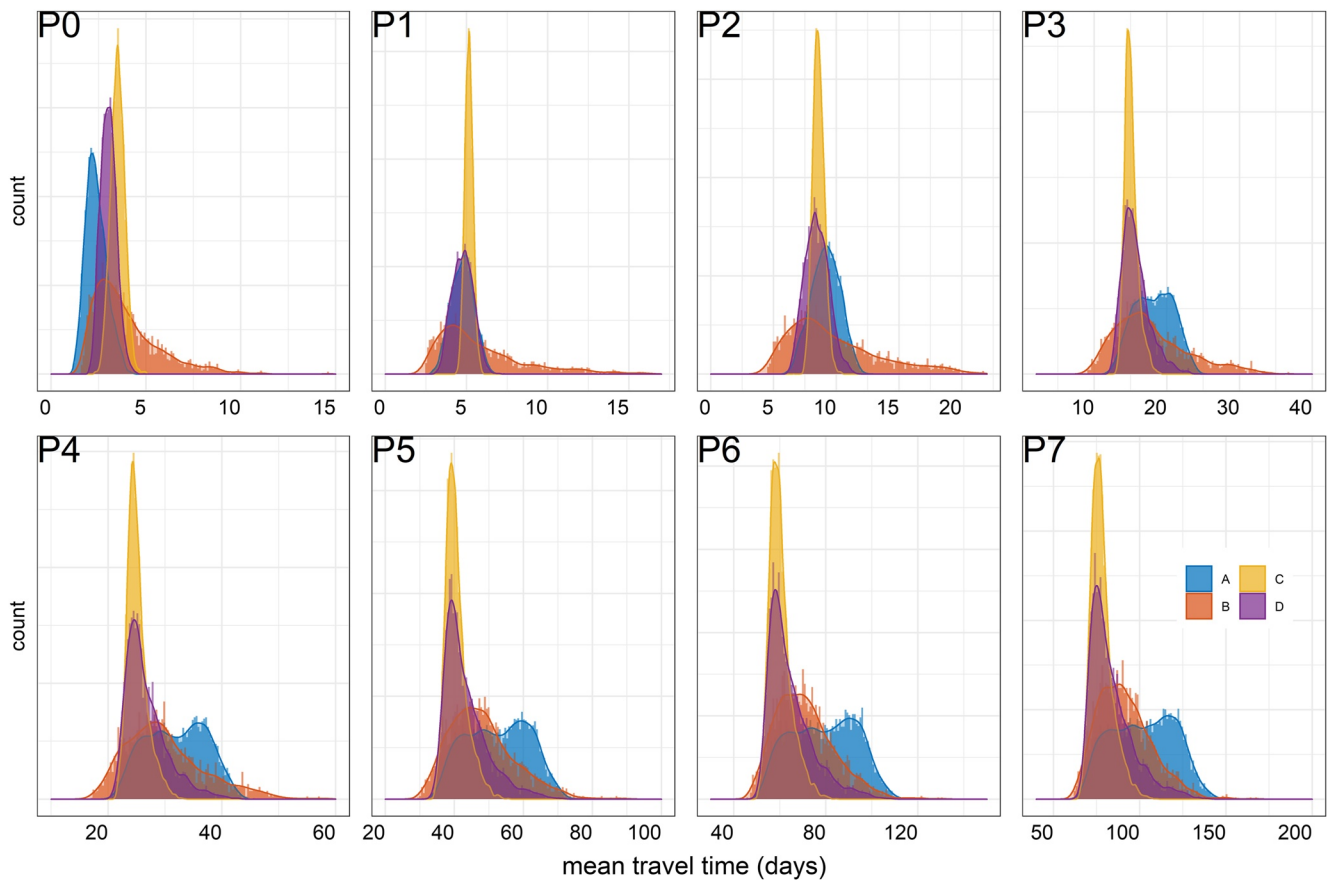
Figure 5. (a) Posterior probability density distributions of mean travel times to P0 to P7 estimated for scenarios A (based on measured hydraulic heads and temperature), B (based on measured hydraulic heads, ^{222}Rn, and variable γ), C (based on measured hydraulic heads, temperature, ^{222}Rn, and constant γ), and D (based on measured hydraulic heads, temperature, ^{222}Rn, and variable γ). Note the difference in *x*-axis scales.

and porosity of scenario D, was compared to τ_{app}, estimated from explicitly simulated ^{222}Rn concentrations and a uniform ^{222}Rn production rate. Apparent ^{222}Rn ages increasingly underestimate μ_τ as dispersivity increases (Figure S3 in Supporting Information S1), a finding that is in agreement with simulations conducted by Gilfedder et al. (2019). However, even for a longitudinal dispersivity of 5 m, μ_τ in the present study is larger than τ_{app} by factors that do not exceed 1.3. The deviation between τ_{app} and μ_τ can thus not be attributed to dispersive mixing alone, but is also caused by relatively low ^{222}Rn production rates in the riverbed sediments and at P1 compared to ^{222}Rn production rates in the alluvial aquifer.

It is reasonable to assume that the median groundwater travel times of scenario D are the most realistic estimates, because this scenario is conditioned on hydraulic heads, ^{222}Rn concentrations, and temperature measurements and relied on ^{222}Rn$_{\infty,inc}$ laboratory data to constrain prior ranges for the ^{222}Rn production rate. The median groundwater travel times in scenarios A and B were higher compared to those in scenario D with the exception of P0 in scenario A, where med$_{\mu_\tau,A}$ was lower than med$_{\mu_\tau,B}$ and P1, where median travel times were relatively similar for all scenarios (Table 3). Using only temperature or ^{222}Rn alone in combination with hydraulic heads as data would thus have resulted in an average overestimation of the travel time of 10%. For P2–P7, the median groundwater travel time in scenario C was similar to that of scenario D but was smaller for P0 and P1. The differences in travel time between scenarios D and C close to the river are attributed to an increase in ^{222}Rn production rates with distance from the river. Differences between scenarios D and A and D and B, however, are likely due to a lack of calibration data in scenarios A and B compared to scenario D, resulting in wider posterior distributions of μ_τ and thus larger med$_{\mu_\tau}$ values.

In the present study, uncertainty in identifying the mean travel time at a given point was estimated as the inter-quartile range (IQR_{μ_τ}) of the posterior distribution of mean groundwater travel times. In scenario D, the relative uncertainty ranged between 15% and 20% of the median travel time (Table 3). Compared to scenario D, the relative uncertainty was slightly lower in scenario C (6%–14% of $med_{\mu_\tau,C}$) but higher in scenarios A (17%–35% of $med_{\mu_\tau,A}$) and B (24%–62% of $med_{\mu_\tau,B}$). The uncertainty in identifying the mean travel time mainly arises from parameter equifinality, which is indicated by strong correlations among model parameters and relatively wide posterior parameter distributions. In scenario D, nonuniform ^{222}Rn production rates were partly correlated to sediment porosity (Figure S14 in Supporting Information S1), while no such correlations were observed in scenario C. The relatively low uncertainty in scenario C can thus be attributed to the prescription of one uniform ^{222}Rn production rate for the entire model domain. The relatively high uncertainty associated with med_{μ_τ} in scenarios A and B, however, reflects correlations among porosity and thermal sediment parameters and among hydraulic conductivity and ^{222}Rn production rates, respectively (Figures S11 and S12 in Supporting Information S1). This finding suggests that the joint inversion of ^{222}Rn concentrations and temperature can reduce the relative uncertainty in identifying the mean travel time by up to 42% and thus can at least partly compensate the uncertainty induced by nonuniform ^{222}Rn production rates.

It should be noted that the accuracy of our estimates of parameters and mean groundwater travel time is also influenced by conceptual model uncertainty and that our approach assumes time-invariant transport coefficients and accounts for the effects of mixing along and between distinct flow paths only via a uniform range of dispersivities. Mixing between flow paths of distinct ages, however, can not only result from transverse dispersion (Castro & Goblet, 2005), but also from exchange with aquitards (Cornaton & Perrochet, 2006; Sanford, 1997) and dual-domain transport (Neumann et al., 2008; Sanford et al., 2017). These effects can lead to erroneous estimates of groundwater travel time derived from concentrations of radiometric tracers as outlined above (see also J. L. McCallum et al., 2015). By setting up a cross-sectional model domain, the assumption was made that there was no lateral inflow of groundwater into the model domain and that the general groundwater flow direction in the south-astern alluvial aquifer was perpendicular to the River Erpe (i.e., within the transect of piezometers). The relatively high K values estimated for the southern part of the model domain could be indicative for lateral groundwater inflows, which would result in an underestimation of travel times, particularly in P6 and P7. The possibility that flow paths circumvent piezometers and are thus located outside of the model domain cannot fully be excluded. However, the geophysical data as well as the sediment analysis suggest that the aquifer materials at the field site were relatively homogenous. Vertical seepage fluxes and hydraulic gradients indicate that surface water infiltration into the alluvial aquifer was strongest around the location of the main transect, an observation that is further supported by temperature time series and ^{222}Rn concentrations. Furthermore, distances from the piezometers along the River Erpe are shortest along the course of the main transect and sampled water in the piezometers P0 to P7 was chemically distinct from regional groundwater. Despite the slight decrease in groundwater heads during the model simulation period, head gradients remained relatively constant and the measured ^{222}Rn concentrations showed no trend over time, thus justifying the steady-state assumption implied in the direct simulation of mean groundwater travel time. It is therefore reasonable to assume that there is a high level of congruence between the modeled transect and the actual groundwater flow field and that the median groundwater travel time simulated by scenario D, particularly the ones estimated for the northern part of the transect, is relatively close to the actual mean of the travel time distribution.

5. Conclusions and Implications

^{222}Rn is distinct from other radiogenic tracers, such as ^{3}H/^{3}He and ^{14}C, because its input concentration function is typically well known and it is produced in the sediment matrix. The present study, however, demonstrates that along alluvial groundwater flow paths, ^{222}Rn production rates may vary by a factor of 2 across relatively short distances (<100 m). As a consequence, differences between apparent ^{222}Rn ages and mean river-to-groundwater travel times can be substantial and likely exceed differences induced by dispersive mixing in two-dimensional flow fields. The fact that ^{222}Rn production rates can be nonuniform within alluvial aquifers and may differ between riverbed sediments and aquifer material presents a major disadvantage of the method, particularly in settings in which the flow paths under study span across sediments with different grain size distributions, porosity, or mineral composition. If the observed ^{222}Rn concentrations deviate from an exponential ingrowth model as predicted by Equation 3 or if field measurements of ^{222}Rn concentrations or other tracers suggest that ^{222}Rn$_\infty$

concentrations differ substantially, implicit ^{222}Rn transport formulations should be avoided. Under such conditions, the spatial distribution of ^{222}Rn production rates can be constrained by laboratory investigations on ^{222}Rn production rates or by employing additional tracers.

The present study further shows that ^{222}Rn transport can explicitly be simulated using the open-source software packages MODFLOW-NWT and MT3D-USGS. To reduce the uncertainty of travel times induced by sediment heterogeneity and to improve the accuracy of travel time estimates, we suggest to combine ^{222}Rn with other natural tracers, such as heat or electric-conductivity fluctuations, and jointly calibrate transport models of all tracers. Heat as a tracer has the particular advantage that temperature can easily be measured in riverbed sediments, where estimates of ^{222}Rn production rates can only be obtained from sediment incubations. In this context, Bayesian analysis provides a useful tool to quantify uncertainty originating from aquifer heterogeneity including sediment porosity, hydraulic conductivity, and the spatial variation in ^{222}Rn production rates. More research, however, is needed to investigate the effects of mixing between flow paths and transient flow on the estimation of travel time distributions from rivers to alluvial aquifers when heat and ^{222}Rn are used as tracers. Future research should also aim at comparing different methods to more reliably and accurately determine ^{222}Rn$_{\infty}$ concentrations in sediment samples.

Data Availability Statement

The data in this paper are published as Schaper et al. (2021) and can be found https://doi.org/10.5281/zenodo.4956781.

References

Anderson, M. P. (2005). Heat as a ground water tracer. *Ground Water*, *43*(6), 951–968. https://doi.org/10.1111/j.1745-6584.2005.00052.x

Auken, E., Christiansen, A. V., Kirkegaard, C., Fiandaca, G., Schamper, C., Behroozmand, A. A., et al. (2015). An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data. *Exploration Geophysics*, *46*(3), 223–235. https://doi.org/10.1071/EG13097

Bakker, M., Post, V., Langevin, C. D., Hughes, J. D., White, J. T., Starn, J. J., & Fienen, M. N. (2016). Scripting MODFLOW model development using Python and FloPy. *Ground Water*, *54*(5), 733–739. https://doi.org/10.1111/gwat.12413

Bedekar, V., Morway, E., Langevin, C., & Tonkin, M. (2016). MT3D-USGS version 1: A U.S. Geological Survey release of MT3DMS updated with new and expanded transport capabilities for use with MODFLOW. *Groundwater Resources Program*, *84*. https://doi.org/10.3133/tm6A53

Bethke, C. M., & Johnson, T. M. (2008). Groundwater age and groundwater age dating. *Annual Review of Earth and Planetary Sciences*, *36*, 121–152. https://doi.org/10.1146/annurev.earth.36.031207.124210

Bouchez, C., Cook, P. G., Partington, D., & Simmons, G. T. (2021). Comparison of surface water – Groundwater exchange fluxes derived from hydraulic and geochemical methods and a regional groundwater model. *Water Resources Research*. https://doi.org/10.1029/2020wr029137

Cartwright, I., & Gilfedder, B. (2015). Mapping and quantifying groundwater inflows to Deep Creek (Maribyrnong catchment, SE Australia) using ^{222}Rn, implications for protecting groundwater-dependant ecosystems. *Applied Geochemistry*, *52*, 118–129. https://doi.org/10.1016/j.apgeochem.2014.11.020

Castro, M. C., & Goblet, P. (2005). Calculation of ground water ages – A comparative analysis. *Ground Water*, *43*(3), 368–380. https://doi.org/10.1111/j.1745-6584.2005.0046.x

Cirpka, O. A., Fienen, M. N., Hofer, M., Hoehn, E., Tessarini, A., Kipfer, R., & Kitanidis, P. K. (2007). Analyzing bank filtration by deconvoluting time series of electric conductivity. *Ground Water*, *45*(3), 318–328. https://doi.org/10.1111/j.1745-6584.2006.00293.x

Cook, P. G., Lamontagne, S., Berhane, D., & Clark, J. F. (2006). Quantifying groundwater discharge to Cockburn River, southeastern Australia, using dissolved gas tracers ^{222}Rn and SF$_6$. *Water Resources Research*, *42*(10), 1–12. https://doi.org/10.1029/2006WR004921

Corbett, D. R., Burnett, W. C., Cable, P. H., & Clark, S. B. (1997). Radon tracing of groundwater input into Par Pond, Savannah River site. *Journal of Hydrology*, *203*(1–4), 209–227. https://doi.org/10.1016/S0022-1694(97)00103-0

Cornaton, F., & Perrochet, P. (2006). Groundwater age, life expectancy and transit time distributions in advective-dispersive systems: 1. Generalized reservoir theory. *Advances in Water Resources*, *29*(9), 1267–1291. https://doi.org/10.1016/j.advwatres.2005.10.009

Cranswick, R. H., & Cook, P. G. (2015). Scales and magnitude of hyporheic, river-aquifer and bank storage exchange fluxes. *Hydrological Processes*, *29*(14), 3084–3097. https://doi.org/10.1002/hyp.10421

Cranswick, R. H., Cook, P. G., & Lamontagne, S. (2014). Hyporheic zone exchange fluxes and residence times inferred from riverbed temperature and radon data. *Journal of Hydrology*, *519*(PB), 1870–1881. https://doi.org/10.1016/j.jhydrol.2014.09.059

Danckwerts, P. v. (1953). Continuous flow systems. Distribution of residence times. *Chemical Engineering Science*, *2*(24), 1–13. https://doi.org/10.1016/0009-2509(96)81811-2

Doherty, J. E. (2003). Ground water model calibration using pilot points and regularization. *Ground Water*, *41*(2), 170–177. https://doi.org/10.1111/j.1745-6584.2003.tb02580.x

Duff, J. H., Murphy, F., Fuller, C. C., Triska, F. J., Judson, W., & Jackman, A. P. (1998). A mini drivepoint for measuring pore water solute concentrations sampler zone of sand-bottom streams. *Limnology*, *43*(6), 1378–1383. https://doi.org/10.4319/lo.1998.43.6.1378

Eikenberg, J., Beer, H., & Jäggi, M. (2014). Determination of ^{210}Pb and ^{226}Ra/^{228}Ra in continental water using HIDEX 300SL LS-spectrometer with TDCR efficiency tracing and optimized αβ-discrimination. *Applied Radiation and Isotopes*, *93*, 64–69. https://doi.org/10.1016/j.apradiso.2014.02.021

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

Engelhardt, I., Prommer, H., Moore, C., Schulz, M., Schüth, C., & Ternes, T. A. (2013). Suitability of temperature, hydraulic heads, and acesulfame to quantify wastewater-related fluxes in the hyporheic and riparian zone. *Water Resources Research*, *49*(1), 426–440. https://doi.org/10.1029/2012WR012604

Frei, S., & Peiffer, S. (2016). Exposure times rather than residence times control redox transformation efficiencies in riparian wetlands. *Journal of Hydrology*, *543*, 182–196. https://doi.org/10.1016/j.jhydrol.2016.02.001

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 428–456. https://doi.org/10.1214/ss/1177011136

Gilfedder, B. S., Cartwright, I., Hofmann, H., & Frei, S. (2019). Explicit modeling of Radon-222 in HydroGeoSphere during steady state and dynamic transient storage. *Ground Water*, *57*(1), 36–47. https://doi.org/10.1111/gwat.12847

Ginn, T. R. (1999). On the distribution of multicomponent mixtures over generalized exposure time in subsurface flow and reactive transport: Foundations, and formulations for groundwater age, chemical heterogeneity, and biodegradation. *Water Resources Research*, *35*(5), 1395–1407. https://doi.org/10.1029/1999wr900013

Goode, D. J. (1996). Direct simulation of groundwater age. *Water Resources Research*, *32*(2), 289–296. https://doi.org/10.1029/95wr03401

Gordon, R. P., Lautz, L. K., Briggs, M. A., & McKenzie, J. M. (2012). Automated calculation of vertical pore-water flux from field temperature time series using the VFLUX method and computer program. *Journal of Hydrology*, *420*(421), 142–158. https://doi.org/10.1016/j.jhydrol.2011.11.053

Greeman, D. J., & Rose, A. W. (1996). Factors controlling the emanation of radon and thoron in soils of the eastern U.S.A. *Chemical Geology*, *129*(1–2), 1–14. https://doi.org/10.1016/0009-2541(95)00128-X

Harbaugh, A. W. (2005). MODFLOW-2005, The U.S. Geological Survey modular ground-water model — The ground-water flow process. *U.S. Geological Survey Techniques and Methods*, *253*.

Hatch, C. E., Fisher, A. T., Revenaugh, J. S., Constantz, J., & Ruehl, C. (2006). Quantifying surface water-groundwater interactions using time series analysis of streambed thermal records: Method development. *Water Resources Research*, *42*(10), 1–14. https://doi.org/10.1029/2005WR004787

Hellauer, K., Karakurt, S., Sperlich, A., Burke, V., Massmann, G., Hübner, U., & Drewes, J. E. (2018). Establishing sequential managed aquifer recharge technology (SMART) for enhanced removal of trace organic chemicals: Experiences from field studies in Berlin, Germany. *Journal of Hydrology*, *563*, 1161–1168. https://doi.org/10.1016/j.jhydrol.2017.09.044

Henzler, A. F., Greskowiak, J., & Massmann, G. (2014). Modeling the fate of organic micropollutants during river bank filtration (Berlin, Germany). *Journal of Contaminant Hydrology*, *156*, 78–92. https://doi.org/10.1016/j.jconhyd.2013.10.005

Hoehn, E., & Cirpka, O. A. (2006). Assessing residence times of hyporheic ground water in two alluvial flood plains of the Southern Alps using water temperature and tracers. *Hydrology and Earth System Sciences*, *10*(4), 553–563. https://doi.org/10.5194/hess-10-553-2006

Hoehn, E., & von Gunten, H. R. (1989). Radon in groundwater: A tool to assess infiltration from surface waters to aquifers. *Water Resources Research*, *25*(8), 1795–1803. https://doi.org/10.1029/wr025i008p01795

Hoehn, E., von Gunten, H. R., Fritz, S., & Themistocles, D. (1992). Radon-222 as a groundwater tracer. A laboratory study. *Environmental Science and Technology*, *26*(4), 734–738. https://doi.org/10.1021/es00028a010

Irvine, D. J., Lautz, L. K., Briggs, M. A., Gordon, R. P., & McKenzie, J. M. (2015). Experimental evaluation of the applicability of phase, amplitude, and combined methods to determine water flux and thermal diffusivity from temperature time series using VFLUX 2. *Journal of Hydrology*, *531*, 728–737. https://doi.org/10.1016/j.jhydrol.2015.10.054

Lamontagne, S., & Cook, P. G. (2007). Estimation of hyporheic water residence time in situ using 222Rn disequilibrium. *Limnology and Oceanography: Methods*, *5*(11), 407–416. https://doi.org/10.4319/lom.2007.5.407

Leray, S., Engdahl, N. B., Massoudieh, A., Bresciani, E., & McCallum, J. (2016). 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

Liao, F., Cardenas, M. B., Ferencz, S. B., Chen, X., & Wang, G. (2021). Tracing bank storage and hyporheic exchange dynamics using 222Rn: Virtual and field tests and comparison with other tracers. *Water Resources Research*, *57*(5), 1–20. https://doi.org/10.1029/2020WR028960

Liao, Z., Osenbrück, K., & Cirpka, O. A. (2014). Non-stationary nonparametric inference of river-to-groundwater travel-time distributions. *Journal of Hydrology*, *519*(PD), 3386–3399. https://doi.org/10.1016/j.jhydrol.2014.09.084

Luce, C. H., Tonina, D., Gariglio, F., & Applebee, R. (2013). Solutions for the diurnally forced advection-diffusion equation to estimate bulk fluid velocity and diffusivity in streambeds from temperature time series. *Water Resources Research*, *49*(1), 488–506. https://doi.org/10.1029/2012WR012380

Luo, J., Cirpka, O. A., Fienen, M. N., Wu, W. M., Mehlhorn, T. L., Carley, J., et al. (2006). A parametric transfer function methodology for analyzing reactive transport in nonuniform flow. *Journal of Contaminant Hydrology*, *83*(1–2), 27–41. https://doi.org/10.1016/j.jconhyd.2005.11.001

Ma, R., Zheng, C., Zachara, J. M., & Tonkin, M. (2012). Utility of bromide and heat tracers for aquifer characterization affected by highly transient flow conditions. *Water Resources Research*, *48*(8), 1–18. https://doi.org/10.1029/2011WR011281

Maloszewski, P., & Zuber, A. (1993). Principles and practice of calibration and validation of mathematical models for the interpretation of environmental tracer data in aquifers. *Advances in Water Resources*, *16*(3), 173–190. https://doi.org/10.1016/0309-1708(93)90036-F

McCallum, A. M., Andersen, M. S., Rau, G. C., & Acworth, R. I. (2012). A 1-D analytical method for estimating surface water-groundwater interactions and effective thermal diffusivity using temperature time series. *Water Resources Research*, *48*(11), 1–8. https://doi.org/10.1029/2012WR012007

McCallum, J. L., Cook, P. G., & Simmons, C. T. (2015). Limitations of the use of environmental tracers to infer groundwater age. *Ground Water*, *53*(S1), 56–70. https://doi.org/10.1111/gwat.12237

McCallum, J. L., Cook, P. G., Simmons, C. T., & Werner, A. D. (2014). Bias of apparent tracer ages in heterogeneous environments. *Ground Water*, *52*(2), 239–250. https://doi.org/10.1111/gwat.12052

McCallum, J. L., Engdahl, N. B., Ginn, T. R., & Cook, P. G. (2014). Nonparametric estimation of groundwater residence time distributions: What can environmental tracer data tell us about groundwater residence time? *Water Resources Research*, *50*(3), 2022–2038. https://doi.org/10.1002/2013WR014974

McNeil, J. D. (1980). Technical Note TN-6, Electromagnetic terrain conductivity measurement at low induction numbers. *Geophysics*.

Molina-Giraldo, N., Bayer, P., Blum, P., & Cirpka, O. A. (2011). Propagation of seasonal temperature signals into an aquifer upon bank infiltration. *Ground Water*, *49*(4), 491–502. https://doi.org/10.1111/j.1745-6584.2010.00745.x

Mullinger, N. J., Pates, J. M., Binley, A. M., & Crook, N. P. (2009). Controls on the spatial and temporal variability of 222Rn in riparian groundwater in a lowland Chalk catchment. *Journal of Hydrology*, *376*(1–2), 58–69. https://doi.org/10.1016/j.jhydrol.2009.07.015

Munz, M., Oswald, S. E., & Schmidt, C. (2011). Sand box experiments to evaluate the influence of subsurface temperature probe design on temperature based water flux calculation. *Hydrology and Earth System Sciences*, *15*(11), 3495–3510. https://doi.org/10.5194/hess-15-3495-2011

Neumann, R. B., Labolle, E. M., & Harvey, C. F. (2008). The effects of dual-domain mass transfer on the tritium-helium-3 dating method. *Environmental Science and Technology*, *42*(13), 4837–4843. https://doi.org/10.1021/es7025246

Niswonger, R. G., Panday, S., & Ibaraki, M. (2011). MODFLOW-NWT, A Newton Formulation for MODFLOW-2005: U.S. Geological Survey Techniques and Methods 6–A37. *Groundwater Book 6, Section A, Modeling Techniques, Book 6-A37*, 44.

Ocampo, C. J., Oldham, C. E., & Sivapalan, M. (2006). Nitrate attenuation in agricultural catchments: Shifting balances between transport and reaction. *Water Resources Research*, *42*(1), 1–16. https://doi.org/10.1029/2004WR003773

Peralta-Maraver, I., Galloway, J., Posselt, M., Arnon, S., Reiss, J., Lewandowski, J., & Robertson, A. L. (2018). Environmental filtering and community delineation in the streambed ecotone. *Scientific Reports*, *8*(October 2017), 1–11. https://doi.org/10.1038/s41598-018-34206-z

Pittroff, M., Frei, S., & Gilfedder, B. S. (2017). Quantifying nitrate and oxygen reduction rates in the hyporheic zone using ^{222}Rn to upscale biogeochemical turnover in rivers. *Water Resources Research*, *53*(3), 563–579. https://doi.org/10.1002/2016wr018917

Popp, A. L., Pardo-Álvarez, Á., Schilling, O. S., Scheidegger, A., Musy, S., Peel, M., et al. (2021). A framework for untangling transient ground-water mixing and travel times. *Water Resources Research*, *57*(4), 1–16. https://doi.org/10.1029/2020WR028362

Porcelli, D., & Swarzenski, P. W. (2003). The behavior of U- and Th-series nuclides in groundwater. *Uranium-Series Geochemistry*, *52*, 317–361. https://doi.org/10.2113/0520317

Rau, G. C., Andersen, M. S., McCallum, A. M., Roshan, H., & Acworth, R. I. (2014). Heat as a tracer to quantify water flow in near-surface sediments. *Earth-Science Reviews*, *129*, 40–58. https://doi.org/10.1016/j.earscirev.2013.10.015

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Reynolds, J. M. (2011). *An introduction to applied and environmental geophysics*. John Wiley & Sons.

Sakoda, A., Ishimori, Y., & Yamaoka, K. (2011). A comprehensive review of radon emanation measurements for mineral, rock, soil, mill tailing and fly ash. *Applied Radiation and Isotopes*, *69*(10), 1422–1435. https://doi.org/10.1016/j.apradiso.2011.06.009

Sanford, W. E. (1997). Correcting for diffusion in carbon-14 dating of ground water. *Ground Water*, *35*(2). https://doi.org/10.1111/j.1745-6584.1997.tb00093.x

Sanford, W. E., Plummer, L. N., Casile, G., Busenberg, E., Nelms, D. L., & Schlosser, P. (2017). Using dual-domain advective-transport simulation to reconcile multiple-tracer ages and estimate dual-porosity transport parameters. *Water Resources Research*, *53*, 5002–5016. https://doi.org/10.1002/2016wr019469

Schaper, J. L., Posselt, M., Bouchez, C., Jaeger, A., Nuetzmann, G., Putschew, A., et al. (2019). Fate of trace organic compounds in the hyporheic zone: Influence of retardation, the benthic biolayer, and organic carbon. *Environmental Science and Technology*, *53*(8), 4224–4234. https://doi.org/10.1021/acs.est.8b06231

Schaper, J. L., Seher, W., Nützmann, G., Putschew, A., Jekel, M., & Lewandowski, J. (2018). The fate of polar trace organic compounds in the hyporheic zone. *Water Research*, *140*, 158–166. https://doi.org/10.1016/j.watres.2018.04.040

Schaper, J. L., Zarfl, C., Meinikmann, K., Banks, E. W., Baron, S., & Lewandowski, J. (2021). *River Erpe (Berlin, Germany) groundwater levels, temperature and 222-radon data set, June 2019*. https://doi.org/10.5281/zenodo.4956781

Telford, W. M., Geldart, L. P., & Sheriff, R. E. (1990). *Applied geophysics*. Cambridge University Press.

Therrien, R., McLaren, R. G., Sudicky, E. A., & Panday, S. M. (2010). *HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transportGroundwater simulations group*. University of Waterloo.

Varni, M., & Carrera, J. (1998). Simulation of groundwater age distributions. *Water Resources Research*, *34*(12), 3271–3281. https://doi.org/10.1029/98wr02536

Vieweg, M., Kurz, M. J., Trauth, N., Fleckenstein, J. H., Musolff, A., & Schmidt, C. (2016). Estimating time-variable aerobic respiration in the streambed by combining electrical conductivity and dissolved oxygen time series. *Journal of Geophysical Research: Biogeosciences*, *121*, 2199–2215. https://doi.org/10.1002/2016JG003345.Received

Vogt, T., Hoehn, E., Schneider, P., Freund, A., Schirmer, M., & Cirpka, O. A. (2010). Fluctuations of electrical conductivity as a natural tracer for bank filtration in a losing stream. *Advances in Water Resources*, *33*(11), 1296–1308. https://doi.org/10.1016/j.advwatres.2010.02.007

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013

Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Stimulation*, *10*(3), 273–290. https://doi.org/10.1515/IJNSNS.2009.10.3.273

Young, P. C., Pedregal, D. J., & Tych, W. (1999). Dynamic harmonic regression. *Journal of Forecasting*, *18*(6), 369–394. https://doi.org/10.1002/(SICI)1099-131X(199911)18:6<369::AID-FOR748>3.0.CO;2-K

Zarnetske, J. P., Haggerty, R., Wondzell, S. M., & Baker, M. A. (2011). Dynamics of nitrate production and removal as a function of residence time in the hyporheic zone. *Journal of Geophysical Research*, *116*(1), 1–12. https://doi.org/10.1029/2010JG001356

Zheng, C., & Wang, P. (1999). *MT3DMS: A modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems; documentation and user's guide* (p. 220). US Army Corps of Engineers.

References From the Supporting Information

Anderson, M. P. (2005). Heat as a ground water tracer. *Ground Water*, *43*(6), 951–968. https://doi.org/10.1111/j.1745-6584.2005.00052.x

Cook, P. G., Lamontagne, S., Berhane, D., & Clark, J. F. (2006). Quantifying groundwater discharge to Cockburn River, southeastern Australia, using dissolved gas tracers ^{222}Rn and SF$_6$. *Water Resources Research*, *42*(10), 1–12. https://doi.org/10.1029/2006WR004921

de Marsily, G. (1986). *Quantitative hydrogeologyy: Groundwater hydrology for engineers* (p. 440). Academic Press.

Engelhardt, I., Prommer, H., Moore, C., Schulz, M., Schüth, C., & Ternes, T. A. (2013). Suitability of temperature, hydraulic heads, and acesulfame to quantify wastewater-related fluxes in the hyporheic and riparian zone. *Water Resources Research*, *49*(1), 426–440. https://doi.org/10.1029/2012WR012604

Ma, R., Zheng, C., Zachara, J. M., & Tonkin, M. (2012). Utility of bromide and heat tracers for aquifer characterization affected by highly transient flow conditions. *Water Resources Research*, *48*(8), 1–18. https://doi.org/10.1029/2011WR011281

Munz, M., Oswald, S. E., & Schmidt, C. (2011). Sand box experiments to evaluate the influence of subsurface temperature probe design on temperature based water flux calculation. *Hydrology and Earth System Sciences*, *15*(11), 3495–3510. https://doi.org/10.5194/hess-15-3495-2011

Palmer, D., Blowes, D. W., Frind, E. O., & Molson, J. W. (1992). Thermal energy storage in an unconfined aquifer: 1. Field injection experiment. *Water Resources Research*, *28*(10), 2845–2856. https://doi.org/10.1029/92WR01471

Rau, G. C., Andersen, M. S., & Acworth, R. I. (2012). Experimental investigation of the thermal dispersivity term and its significance in the heat transport equation for flow in sediments. *Water Resources Research*, *48*(3), 1–21. https://doi.org/10.1029/2011WR011038

Rau, G. C., Andersen, M. S., McCallum, A. M., Roshan, H., & Acworth, R. I. (2014b). Heat as a tracer to quantify water flow in near-surface sediments. *Earth-Science Reviews*, *129*, 40–58. https://doi.org/10.1016/j.earscirev.2013.10.015

Shook, G. M. (2001). Predicting thermal breakthrough in heterogeneous media from tracer tests. *Geothermics*, *30*, 573–589. https://doi.org/10.1016/S0375-6505(01)00015-3