Different tracer, different bias: using radon to reveal flow paths beyond the Window of Detection

Mortimer L. Bacher[1,2], Julian Klaus[2], Adam S. Ward[3], Jasmine Krause[3], Catalina Segura[4], Clarissa Glaser[2]

[1] Department of Geosciences, University of Tübingen, Tübingen, Germany
[2] Department of Geography, University of Bonn, Bonn, Germany
[3] Department of Biological and Ecological Engineering, Oregon State University, Corvallis, OR, USA
[4] Forest Engineering, Resources, and Management, Oregon State University, Corvallis, OR, USA

Correspondence to: Clarissa Glaser (cglaser@uni-bonn.de)

Abstract. Slug tracer experiments have greatly advanced our understanding of solute transport in streams. Breakthrough curves (BTCs) from these experiments are biased toward faster flow paths, highlighting the need for alternative tracers to cover longer timescales. The radioactive tracer radon (^{222}Rn) is increasingly used to quantify transit times in subsurface transient storage zones, ~~as it traces~~capturing durations ~~transit times~~ of up to 21 days. However, it remains unclear whether calibrating transient storage models (TSMs) with radon yields longer subsurface timescales of transit times ~~– and thus greater transient storage areas –~~ than calibrating them with slug tracers such as sodium chloride (NaCl). To address this, we conducted radon measurements and NaCl slug tracer experiments in Oak Creek (Oregon, USA) and jointly and individually calibrated TSM parameters with both tracers. We applied parameter identifiability analysis and ~~evaluated~~ used information theory to evaluate how the two tracers constrain model parameters.~~the information provided by both tracers in constraining model parameters~~. Our results show that ~~calibrating the~~ TSM calibration with both radon and chloride increases ~~information on model~~ parameter~~s~~ information compared to ~~calibrating the~~ TSM calibration with ~~each~~ either tracer ~~individually~~alone. This suggests that incorporating radon into calibration improves estimates of solute transport ~~estimates~~ in future studies. However, when calibrating the TSM with only radon measurements, all resulting parameters of the TSM were non-identifiable. This non-identifiability arises because radon activity in streams remains at steady-state and is highly sensitive to the location and amount of groundwater inflow, as well as contributions of flow paths from subsurface transient storage zones. ~~This non-identifiability arises from steady state activity of radon in streams and radon's high sensitivity to the amount and location of groundwater inflow, which is not explicitly accounted for in TSMs.~~ As a result, radon measurements are biased toward longer-timescale flow paths, limiting ~~its~~ their ~~usefulness~~ applicability to uniquely constrain~~for characterizing~~ solute transport parameters ~~in calibrating~~ in TSM~~s~~ calibration without complementary ~~chloride~~slug tracers.

1 Introduction

The time a water parcel spends in river corridors is a key variable controlling biogeochemical processes and the ecological functioning of streams (Harvey and Gooseff, 2015; Ward and Packman, 2019). In environmental systems, the ~~ensemble of~~ combined distribution of these timescales for many parcels of water ~~are termed~~ is called transit time distributions (TTDs). Empirical studies of TTDs in streams typically rely on solute tracer experiments (e.g., Stream Solute Workshop, 1990), which involve releasing a known mass of solute tracer into the stream and measuring its concentration over time, i.e., the breakthrough curve (BTC), at a downstream location (Day, 1976). Despite their widespread use, solute tracer experiments are biased toward measuring faster flow paths within TTDs due to the 'window of detection' (WoD). The WoD refers to the longest temporal scale of tracer-labelled flow paths that contribute to measurable tracer concentrations distinguishable from the background concentration (Harvey et al. 1996; Schmadel et al., 2013; Wagner and Harvey 1997; Ward et al. 2023), ultimately defining the longest timescales that can be observed in a given study. Despite decades of research (Harvey and Bencala, 1993; Wagner and Harvey, 1997), measuring flow paths with timescales beyond the WoD remains ~~a~~ challenging~~e,~~. This challenge leav~~ing~~es critical gaps in our understanding of solute transport in streams when relying solely on BTCs~~. This~~ and underscores the need for new approaches to capture these overlooked timescales.

Solute tracer studies are often evaluated by calibrating transient storage models (TSMs) to match empirical BTCs. In their simplest forms, TSMs assume a uniform, steady-state, one-dimensional flow, modeled using the advection-dispersion equation (ADE), while also accounting for first-order mass transfer between the advective flow and a storage zone ~~of effectively infinite dimensions~~ (Bencala and Walters, 1983; Gooseff et al., 2008). Extensions allow for implementing non-uniform groundwater inflow via lateral inflow terms (Runkel et al., 1998). Water in the storage zone is delayed relative to the main channel flow and is located in surface transient storage zones within the channel (Nordin and Troutman, 1980), either due to eddies and turbulence caused by in-stream obstructions (Jackson et al., 2013) or in subsurface transient storage zones (i.e., the hyporheic zone; Bencala and Walters, 1983; Cardenas and Wilson, 2007). The parameter values derived from TSMs provide a means of comparing solute transport within a single stream or across multiple streams (Runkel, 2002). Despite the widespread use of TSMs, their application often produces contradictory results (Ward and Packman, 2019) due to two fundamental issues. First, model parameters are frequently non-identifiable~~,~~. In streams where exchange between advective flow and storage zones is not minimal, non-identifiability may arise when multiple parameter combinations yield equivalent model performance. ~~meaning that multiple parameter combinations can yield equivalent model performance.~~ One approach to addressing this issue has been the application of parameter identifiability analysis. Previous studies highlighted the importance of incorporating identifiability analysis when calibrating TSMs with BTCs to enhance certainty ~~of~~ in model parameters (Bonanno et al., 2022; Camacho and González, 2008; Kelleher et al., 2013; Wagner and Harvey, 1997; Wagener et al., 2002). In addition to identifiability analysis, adding observations is another commonly used strategy for reducing parameter uncertainty and improving model constraints (e.g., Nearing and Gupta, 2015). Research has demonstrated that incorporating additional tracer observations in TSM applications enhances the accuracy of solute transport estimations (Briggs et al., 2009; Neilson et al., 2010a; 2010b).

The second fundamental issue leading to contradictory results from TSMs is that they can only fit observed solute tracer data, meaning they do not account for flow beyond the WoD. This limitation is critical, as a growing body of research highlights the presence of flow paths that exceed the duration of experiments using instantaneous tracer injections (hereafter 'slug tracer experiments'; ~~(~~e.g., Ward et al., 2023). Specifically, tracer mass that is released but remains unrecovered (i.e., 'lost') within the WoD may either follow flow paths that exceed the duration of slug tracer experiments or bypass the downstream sampling location entirely by traveling through subsurface pathways (e.g., Covino et al., 2007; Payn et al., 2009). These subsurface flow paths can occur at multiple scales, exhibiting a wide range of transport times and distances (Cardenas, 2008). ~~They~~ These subsurface flow paths play a crucial role in buffering temperature signals before returning to the channel (Briggs et al., 2022; Wu et al., 2020) and serve as reservoirs of exchange for shorter hyporheic flow paths that may mix with this water before re-entering the main channel (Payn et al., 2009). Some studies suggest that adapting study designs can effectively trace the entire continuum of subsurface flow paths, including large-scale exchange along the river corridor (Covino et al., 2011; Mallard et al., 2014; Ward et al., 2023). However, measuring flow paths beyond the WoD at the reach scale remains challenging when calibrating TSMs with 'traditional' measurements of solute tracer concentration.

The naturally occurring radon (^{222}Rn) may present an opportunity as a tracer to enhance our measurements of flow paths longer than the ~~duration of slug tracer experiments~~WoD. Radon has frequently been used to estimate transit times in subsurface transient storage zones (Cranswick et al., 2014; Frei et al., 2019; Gilfedder et al., 2019; Lamontagne and Cook, 2006; Pittroff et al., 2016) and to quantify groundwater inflows into streams (Cook et al., 2006; Cook, 2013). Radon is a radioactive noble gas that is produced through the decay of radium-226 (^{226}Ra), a parent isotope found in radium-bearing minerals in streambeds (Sakoda et al., 2011). As ^{226}Ra decays, radon activity increases exponentially until secular equilibrium, which occurs when radon production equals its decay. This equilibrium also defines the maximum achievable radon activity based on the availability of radium-bearing minerals. For radon, secular equilibrium is established after approximately 21 days, which is about ~~five~~ seven times its half-life of 3.18 days (Krishnaswami et al., 1982). Secular equilibrium is maintained in aquifers because they are typically closed systems, preventing radon from readily escaping into the atmosphere. Radon activity in groundwater can exceed 100,000 Bq m^{-3} (Cecil and Green, 2000), whereas surface water activity is usually several orders of magnitude lower due to atmospheric degassing. When surface water exchanges with subsurface transient storage zones and contacts radium-bearing minerals in the streambed, radon activity increases as a function of the time spent in the ~~hyporheic~~ subsurface transient storage zone. As a result, radon activity in streams offers insights into the duration that water parcels remain in the subsurface (i.e., in contact with radium-bearing minerals), particularly for transit times of less than 21 days. Subsurface transit times of up to 21 days exceed those measured in slug tracer experiments, where transit times usually range from minutes to hours.

The ~~overarching~~ goal of this study is to quantify flow paths of different timescales at the reach scale using measurements of solute tracer and naturally occurring radon. We expect that calibrating the TSM with radon and determining the model parameters will result in longer timescales of flow paths and, in turn, larger transient storage areas compared to calibration with ‘traditional’ slug tracer data. To test this expectation, we address the following questions:

- How do the values of model parameters and their identifiably differ when calibrating a TSM with only radon or chloride for the same study reach?
- How do parametric values and parameter identifiably change when jointly calibrating a TSM with radon and chloride, compared to calibrating each tracer ~~individually~~alone?

To answer these questions, we apply a coherent mathematical framework to radon and slug ~~tracer~~ injections of sodium chloride (NaCl). ~~Applying a coherent framework~~This approach ~~wa~~is motivated by ~~findings from~~ previous catchment-scale studies~~, which~~ ~~showed~~ showing that applying the same model~~s~~ framework to different tracers yields ~~a similar understanding of~~comparable insights into hydrological transport processes (e.g., Rodriguez et al., 2021; Wang et al., 2023). To ensure a coherent mathematical framework for both NaCl and radon, we adapt the transient storage model OTIS (‘One-Dimensional Transport with Inflow and Storage model,’ Runkel, 1998) by incorporating radon-specific processes ~~(e.g.,~~such as degassing~~)~~. We then jointly and individually calibrate the~~is~~ model with slug tracer ~~data~~ (chloride) and ~~with~~ radon data. We ~~apply~~ perform a global sensitivity analysis approach to ~~ensure~~ assess ~~the~~ parameter identifiability ~~of model parameters in our calibration approach~~. ~~Subsequently~~Finally, we apply~~use~~ information theory to ~~calculate~~ quantify the information gained from~~ the~~ joint and individual calibration of the TSM with these tracers.

2 Materials and Methods

2.1 Field site and experiments

We carried out slug tracer experiments and radon measurements in August 2023 in a 578-m long segment of Oak Creek (44°36'29.16"N, 123°19'56.05"W) in Oregon (USA) (Cargill et al., 2021; Katz et al., 2018; Milhous, 1973). Oak Creek mainly features basaltic lithology and stream sediments consist of cobble to gravel-sized weathering products of these basalts. We subdivided the selected Oak Creek segment into five reaches with lengths between 67 and 140 m (Fig. 1). Each reach length was at least 20 times the Wetted Channel Width. This was done to ~~control~~ensure thorough mixing of the tracer and solutes across both the width and depth of the channel, thereby minimizing spatial variability and potential biases due to reach selection ~~for expected variations in solute transport that occur as a function of reach selection~~ (Anderson et al., 2005; Becker et al., 2023; Day, 1977). We equipped the upstream and downstream location of each reach with conductivity loggers (CTD Diver from Eijkelkamp Soil & Water and Levelogger of Solinst, both with an accuracy of ±1% for conductivity, ±0.1°C for temperature, and ±0.5 cm for the water pressure). We prepared NaCl solutions using one gallon (3.8 L) of stream water and studied each reach by conducting a slug tracer injection both upstream and downstream. The amount of NaCl varied between

500 and 2500 g and was adapted for each injection location to ensure that the tracer signal at the reach's downstream location was elevated by at least 50 µS cm^{-1} above the background concentration. We determined the injection locations to ensure lateral and vertical mixing of stream water with the injected solution of solute tracers by the time the tracer entered each study reach (Payn et al., 2009; Ward et al., 2013). After tracer injection, we measured electrical conductivity (EC) every 5 seconds and normalized it to 25°C. We then corrected for background EC of the stream and corrected EC to chloride concentrations based on EC-concentration regression lines (R^2 = 0.99). Discharge was calculated from the resulting BTCs using dilution gauging (Kilpatrick and Cobb, 1985). Additionally, we determined the mass recovery of the injected tracer from the BTCs of the upstream and the downstream injection to quantify the amount of NaCl tracer mass lost during the experiment (Payn et al., 2009).

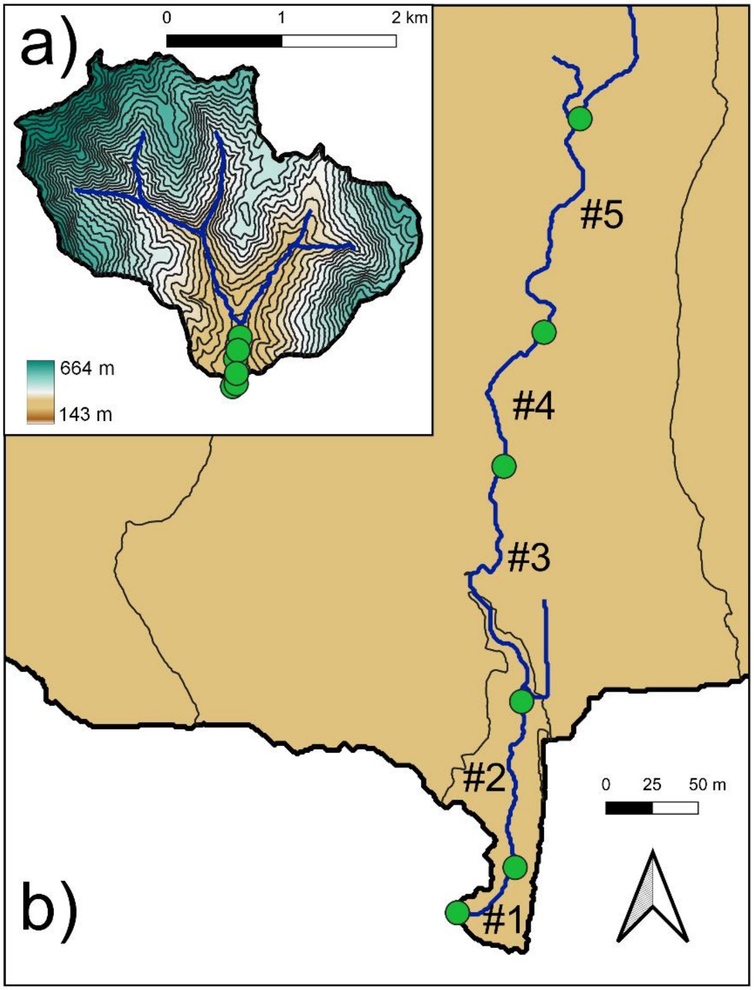

Figure 1: a) Map of the Oak Creek catchment, with colors indicating elevation. Green markers denote tracer measurement locations. (b) Close-up view, where labels (#1–#5) represent reaches between these measurement locations.

Radon sampling sites were co-located with BTCs observations, collected one day before the slug tracer injections. We filled one-liter amber glass bottles with stream water in the thalweg and tightly closed the bottles beneath the water surface to prevent degassing during sampling. The water samples were analyzed using a large dome, large detector RAD7 device (Durridge Company, Inc.). We employed a closed air loop approach as outlined by Lee and Kim (2006). Each one-liter water sample underwent degassing with the integrated pump of the RAD7 to strip the dissolved radon into the air phase. Subsequently, radon in the air of the closed-loop was counted up to six times over 30 minutes with the integrated detector of the RAD7. The resulting average value of the repeated radon measurements was corrected for the time between sampling and measurement to account for radioactive decay. Radon measurements were then multiplied by an empirical correction factor to adjust for differences in degassing between one-liter samples and the reference volume of the RAD H_2O method (250 mL).

We determined the maximum radon activity that can be achieved based on the available radium-bearing minerals of Oak Creek (i.e., the radon activity reached at secular equilibrium). For this purpose, we selected two locations at Oak Creek and collected five sediment samples from each location. Subsequently, these sediment samples were merged into two bulk samples to reduce the potential for small-scale spatial heterogeneity in radon activity in the stream sediment, and to ensure representative sediment samples. We then conducted incubation experiments with stream sediment and radon-free water (after Corbett et al. 1997; 1998; Peel et al., 2022). After incubating the bulk sediment samples for at least 21 days inside gas-tight two-liter containers, we analyzed the water from these samples using the closed-loop approach with the RAD7, as described previously for the surface water samples. We assumed the same mineralogical composition of the aquifer and the streambed sediment. This means the radon activity at secular equilibrium represents both the radon activity of the groundwater and the highest achievable activity along subsurface flow paths.

2.2 Transient storage modelling

Solute transport parameters are ~~is~~ commonly de~~rived~~termined ~~from~~by calibrating ~~the~~ TSMs ~~to~~ against measured BTCs. ~~The~~ TSMs describe~~s~~ the combined effect of flow velocity and dispersion on solute transport in a one-dimensional steady flow domain (Taylor 1922; 1954), dilution (or enrichment) of solutes from lateral groundwater inflow, and additionally considers a first-order mass exchange of solutes between the surface and a finite-size, ~~well~~completely-mixed transient storage zone. The partial differential equations of the TSM are (Bencala and Walters 1983):

$$\frac{\partial C}{\partial t} = -\frac{Q}{A}\frac{\partial C}{\partial x} + \frac{1}{A}\frac{\partial}{\partial x}\left(AD\frac{\partial C}{\partial x}\right) + \frac{q_I}{A}(C_I - C) + \alpha(C_{TS} - C),$$

$$\frac{d\cancel{\partial}C_{TS}}{d\cancel{\partial}t} = -\alpha\frac{A}{A_{TS}}(C_{TS} - C) \tag{1}$$

where C is the observed tracer concentration above the background concentration [M L^{-3}], t is time [T], Q the discharge in the stream channel [L^3 T^{-1}], A the channel's cross-sectional area [L~~m~~^{2}], x the distance [L], D the longitudinal dispersion coefficient [L^2 T^{-1}], q_I the groundwater inflow into the stream channel [L^3 L^{-1} T^{-1}], α the transient storage exchange coefficient [T^{-1}], C_{TS} the solute concentration in the transient storage zone [M L^{-3}], and A_{TS} the cross-sectional area of the transient storage zone [L^{-2}]. ~~In case of a lack of exchange where α = 0, eq. 1 reduces to the advection-dispersion equation.~~

The model formulation above ~~is not suited~~does not account for key processes affecting radon activity, as radon activity changes due to radioactive decay, degassing, and production in the transient storage zone (Cook 2013; Frei and Gilfedder 2015). Therefore, we implemented additional radon specific processes in the one-zone TSM as follows:

$$\frac{\partial C}{\partial t}=-\frac{Q}{A}\frac{\partial C}{\partial x}+\frac{1}{A}\frac{\partial}{\partial x}\left(AD\frac{\partial C}{\partial x}\right)+\frac{q_I}{A}(C_I-C)-\lambda C-\frac{k}{d}C+\alpha(C_{TS}-C),$$
$$\frac{d\cancel{\partial}C_{TS}}{\cancel{\partial}dt}=-\alpha\frac{A}{A_{TS}}(C_{TS}-C)-\lambda C_{TS}+\gamma\cancel{;}$$

(2)

where λ [T^{-1}] is the radioactive decay rate [~~T^{-1}; ~~(0.18 d^{-1} for radon)~~]~~, k [L T^{-1}] the gas exchange velocity, d [L] the stream depth, and γ [M L^{-3} T^{-1}] the production of radon in the transient storage zone. In the absence of radon-specific processes (k = 0, λ = 0 and $\gamma = 0$), eq. 2 reduces to the TSM described in eq. 1.

2.3 Numerical implementation of radon-specific processes in OTIS

The 'One Dimensional Transport with Inflow and Storage' (OTIS) model (Runkel, 1998) is one of the most commonly used implementations of the TSM. OTIS uses a Crank-Nicolson numerical scheme to solve the TSM. We adapted the existing code of OTIS (written in FORTRAN 77) to simulate radon activity. Hereafter, we will refer to the implementation of the TSM that considers radon-specific processes as OTIS-R (R for radon; Bacher et al., 2025).

2.4 Model calibration

We used measured chloride concentrations and radon activity to calibrate the model parameters of the TSM. The calibration was done in a Monte Carlo approach to assess the model performance for different combinations of parameter values (after Kelleher et al, 2013; Ward et al, 2017; 2018). We refer to a single combination of calibrated parameter values as a 'parameter set.' The same parameter sets were tested for both tracers to evaluate model performance for each tracer. The model performance was evaluated using the normalized root mean squared error (nRMSE). We performed this normalization to enable a relative comparison of both tracers. We conducted three different calibration approaches, each with 200,000 iterations. The model parameters were sampled using Latin Hypercube Sampling (LHS), a method that employs stratified sampling while retaining the simplicity and objectivity of fully random sampling (Helton and Davis, 2003). In all three calibration approaches, we calibrated D, α, and A_{TS} by sampling them from a defined parameter range (Table 1). We ~~assumed a uniform parameter distribution~~ sampled parameter values~~for~~ of D, α, and A_{TS}~~,~~ ~~sampling~~uniformly from a log10 transformed ~~distribution~~ space to ensure approximately equal representation for each order of magnitude within the parameter space (Kelleher et al. 2013; Ward et al. 2017). We extracted the 1% and 10% with the lowest values for the nRMSE of the parameter sets and considered them as behavioral parameter~~s~~ sets (Beven and Binley, 1992). We selected these behavioral thresholds to ensure consistency with previous solute transport studies (e.g., Bonanno et al., 2022; Kelleher et al., 2019; Wagener et al., 2002; Ward et al., 2013, 2017; Wlostowski et al., 2013). Since we tested the same combinations of parameter values in the TSM for both tracers, the intersection of the behavioral parameter sets from both tracers reflects the parameter sets obtained when the model is calibrated with both tracers together. This indicates that when the behavioral parameter sets for both tracers are ~~the same~~identical, the choice of tracer does not affect~~impact the parametric information related to~~ estimates of solute transport. We refer to this intersecting set as the 'joint calibration'. ~~Still~~In contrast, other parameter sets were included in the behavioral set for only one tracer but not the other~~,~~. These sets represent~~ing~~ parameter~~s~~izations that result in~~ with~~ acceptable model performance for a single tracer, but are not robust in ~~describing~~ simulating both tracers simultaneously. We refer to these as the 'individual calibration'. The behavioral parameter sets were used for all ~~further~~subsequent analysis and calculations.

Table 1: ~~Model~~ Calibration parameters used in OTIS-R. Parameter ranges are shown for those that were calibrated.

Model parameter	Range
D [m² s^{-1}]	1e-5 to 10
α~~A~~ [~~m~~s^{-1}~~²~~]	1e-5 to 0.1
A_{TS} [m²]	1e-5 to 100

Model parameters other than those calibrated - including the stream velocity (v), the cross-sectional area (A), the production term of radon in the storage zone (γ), and the gas exchange velocity (k) - were calculated before calibration. This reduces potential issues of equifinality with TSMs (Knapp and Kelleher 2020). We calculated v by dividing the stream length by the arrival time of the concentration peak of the downstream BTC, and calculated A from ~~these two parameters~~ v and Q after the calibration approach. This choice was motivated by findings from Bonanno et al. (2022), who showed that A_{TS} and α are often not identifiable when v is calibrated instead of calculating v by dividing the stream length by the arrival time of the concentration peak of the downstream BTC. We calculated the radon production term γ as the product of the decay constant (0.18 d^{-1}) and the measured equilibrium radon activity (Gilfedder et al. 2019). This approach assumes that radon production occurs only in the subsurface transient storage zone of the stream. However, radon may also increase when stream water interacts directly with the streambed surface. For the gas exchange velocity, we relied on gas tracer experiments previously conducted at the same stream section as our study at Oak Creek (Cargill et al. 2021). We scaled the gas exchange coefficients for SF_6 reported by Cargill et al. (2021) to radon (Jähne et al., 1987; Raymond et al., 2012). We tested two different gas exchange velocities for each run of our three calibration approaches to quantify the uncertainty of degassing in calibration of the TSM. The gas tracer experiments by Cargill et al. (2021) were conducted at three different discharge conditions ($0.05\ m^3s^{-1}$, $0.1\ m^3s^{-1}$, $1.07\ m^3s^{-1}$). We used values from the experiments conducted during the lowest and highest discharges for parameterization. This resulted in one model setup with a low gas exchange value and another with a high gas exchange value. Hereafter, we will refer to these different values used for parameterization as k_{low} (k_{600} = 206 d^{-1}) and k_{high} (k_{600} = 290 d^{-1}).

2.5 Evaluating parameter sensitivity, certainty, and interactions

After model calibration, we evaluated the parameter identifiability of the behavioral parameters through sensitivity, certainty, and interactions analysis. We refer to a parameter set as 'identifiable' when the values of the model parameters are ~~certain,~~sensitive, ~~sensitive~~certain, and do not have any parameter interactions. The~~se~~ parameter identifiability analyses include the visual inspection of I) nRMSE vs. parameter plots (Wagener et al., 2003), II) cumulative parameter distribution plots (Kelleher et al., 2019), III) posterior distribution plots (Wagener et al., 2002; Ward et al., 2017), ~~and~~ IV) scatter plots of the behavioral parameters, and V) calculation of the Shannon entropy for the posterior distribution of model parameters as metric for certainty (sensu Rodriguez et al. 2021, section 2.6). In the nRMSE vs. parameter plots, parameters needed to exhibit a distinct peak of performance in nRMSE ~~vs. parameter plots~~ to be categorized as sensitive~~ and certain~~. ~~The~~ Furthermore, the cumulative distribution functions (CDFs) of the top 1% or 10% of results (behavioral parameters) had to visibly deviate from the 1:1 line (representing a uniform distribution) to be categorized as sensitive. The probability density functions (posterior distributions) had to be peaked and narrow to categorize parameters as certain.

The posterior distribution of parameters is represented by the histogram of behavioral parameter sets and their performance (nRMSE). This histogram was created by dividing the parameter values into 15 equally sized bins, where the bar height illustrates the likelihood of a parameter falling within a specific bin. ~~Sensitive~~ Certain parameters exhibit higher variation in likelihood across different parameter values. In the scatter plots, narrow and constrained values of two model parameters indicate identifiable parameters. Parameter interactions are visible through changes in ~~one parameter's~~ the value of one parameter relative to changes in another within the parameter space. These interactions are visually depicted as a curve in the parameter space, suggesting that variations in parameter 1 result in corresponding variations in parameter 2.

For a quantitative measure of the parameter sensitivity that underpins the visual inspections, we applied the two-sample Kolmogorov–Smirnov (K-S) test that calculates the maximum distance K and the corresponding *p-value* between two cumulative distribution functions:

$$[K, p] = max\,|F(P_{behavioral}) - F(P_{non\text{-}behavioral})|$$

(3)

where $F(P_{behavioural})$ and $F(P_{nonbehavioral})$ are the cumulative distribution functions of a parameter P for the behavioral and non-behavioral parameter sets, respectively. The K-S test thus expresses the degree of sensitivity ~~of~~ for a parameter. We grouped parameter sensitivity into four different categories following the approach of Ouyang et al. (2014): highly ~~identifiable~~ sensitive (K > *0.2*; *p-value* ≤ 0.05), moderately sensitive~~identifiable~~ 0.1 ≤ *K* ≤ 0.2; *p-value* ≤ 0.05), poorly ~~identifiable~~ sensitive *(K <* 0.1*; p-value* ≤ 0.05) and non-~~identifiable~~ sensitive *(p-value > 0.05)*. Moreover, we calculated the non-parametric Spearman rank correlation coefficient~~s~~ ($\rho_{spearman}$) using a significance threshold of 0.05 to quantitatively assess the non-linear interactions between different model parameters observed~~for a quantitative measure of the visual inspection of the~~ in the scatter plots.~~ of model parameters. Given the non-linear nature of these interactions, we used the non-parametric Spearman rank correlation coefficient, with a *p-value* of 0.05 for determining statistical significance.~~

2.6 Evaluating information content of model parameters

We ~~used the information content to~~ evaluated ~~the~~parameter certainty using the information content,~~of the model parameters by calculating~~ calculated as the Shannon entropy of the posterior ~~parameter~~ distributions of ~~these~~ model parameters (Cover and Thomas, 2005; Loritz et al., 2018). Rodriguez et al. (2021) applied this approach in a catchment-scale study. The posterior ~~parameter~~ distribution is the probability density function of the behavioral parameters sets. The Shannon entropy reads:

$$H(X|T) = -\sum_{k=1}^{n_I} f(I_k) log_2 f(I_k)$$

(4)

Where H describes the Shannon entropy and X the parameter of interest. T is the tracer, for which behavioral parameter sets were extracted. The tracer could either be radon, chloride or a combination of both ($H(X|(radon \cap chloride))$). We binned the parameter values into 15 bins of equal size, similar to visual inspection of the posterior distribution of the parameter certainty. The rationale for choosing 15 bins was that the resulting histograms visually revealed the underlying structure of the parameter values without introducing uneven features, such as spiky histograms. The height of each bin describes the likelihood of the parameter being located in this specific bin. n_I [-] signifies the number of intervals (bins), and $f(I_k)$ [-] describes the probability of the parameter X falling within the interval I_k. $f(I_k)$ describes the probability of the parameter X to take a value in an interval I_k for the posterior distribution (either radon~~ or~~, chloride, or a combination of both)~~, a combination of both (posterior distribution),~~ or ~~none of those (~~the prior distribution (none of those). Smaller values of H show that the posterior distribution is not flat and that it is more certain than a uniform prior distribution.

Furthermore, we evaluated the information gain from the prior to the posterior distribution of model parameters when the TSM was calibrated with radon and chloride separately, as well as for both tracers together. In this context, the prior distribution describes the uniform distribution of all parameters prior to parameter calibration. The minimum and maximum values for this distribution are defined through the parameter range from which these parameters were sampled (Table 1). The information gain quantifies how much information the tracers add to the model parameters when calibration of the model was conducted with these tracers separately and with both tracers together. We evaluated the information gain from prior to posterior distributions for each model parameter using the Kullback-Leibler divergence D_{KL} (Rodriguez 2021):

$$D_{KL}(X|T,X) = \sum_{k=1}^{n_I} f(I_k) log_2 \frac{f(I_k)}{g(I_k)} \quad (5)$$

where $g(I_k)$ [-] is the probability of the parameter X to fall in the interval I_k in the prior distribution. Higher values of D_{KL} (in bits) show a higher information gain from prior to posterior parameter distribution during calibrating the model (Rodriguez 2021). Summing values of the Kullback-Leibler divergence for all TSM parameters yields the total information on solute transport from that tracer.

2.7 ~~Considering~~ Accounting for groundwater inflow ~~for~~ in ~~calibrating the~~ TSM calibration

Radon activity in streams varies with the amount of inflowing groundwater, as radon activity differs significantly between groundwater and surface water (Cook, 2013). Small changes in the amount of inflowing groundwater may lead to differences in model performance. To account for this, we either calibrated or calculated groundwater inflow within three different calibration approaches, in addition to calibrating D, α, and A_{TS}. These approaches to handling groundwater inflow were selected to assess how varying values of groundwater inflows affect model performance. For the first calibration approach (Q_{fix}), we calculated the groundwater inflow by dividing the difference of discharge between the upstream and downstream BTCs by the reach length ~~(Table 1)~~. In the second calibration approach (Q_{LHS}), ~~we calibrated~~ discharge ~~(Q_{LHS})~~ was calibrated as model parameter, like D, α, and A_{TS}, ~~and subsequently~~ before calculat~~ed~~ing groundwater inflow. For each reach, ~~D~~ discharge was sampled from a normal distribution, because Schmadel et al. (2010) reported that discharge measurement errors follow a normal distribution. We used the calculated discharge as the mean of the normal distribution and assumed that its standard deviation represents the uncertainty of dilution gauging. In the third calibration approach (Q_{out}), we directly calibrated groundwater inflow, whereas discharge was not calibrated ~~we calculated groundwater inflow from the calibrated discharge~~. ~~This~~ The direct calibration of groundwater inflow allows us to calculate gross water fluxes along reaches. This is because

OTIS, and by extension OTIS-R, accounts for water mass balance under steady state by parameterizing groundwater inflow using the following equation:

$$\frac{\partial Q}{\partial x} = \cancel{q_I\ q_{out}}\, I - O \tag{6}$$

where ~~q_I~~I [$L^2\ T^{-1}$] is the gross water inflow and ~~q_{out}~~O [$L^2\ T^{-1}$] the gross water outflow from the stream into the groundwater. The discretized form of this equation, with ΔQ describing the difference of discharge between the upstream and downstream BTCs and Δx the reach length, can be expressed in terms of q_{out} [$L^3\ T^{-1} L^{-1}$] rather than O. In the first and second calibration approaches, gross water outflow was assumed to be zero. These gross water fluxes are commonly derived from calculating the mass loss of BTCs relative to the injected tracer mass (i.e., 'channel water balance'; Payn et al., 2009). For all three calibration approaches, the measured equilibrium radon activity was used as the activity of the groundwater inflow.

Additionally, radon activity in streams depends on the location of the groundwater flow (Cook 2013). Assuming groundwater inflow as either discrete or linear in the TSM might therefore lead to different simulated radon activity and, in turn, to different calibrated parameter values to achieve a good fit between simulated and measured radon activity. To test this, we calibrated the TSM with radon activity ~~and chloride concentrations~~ for the downstream-most reach in three different model setups. In each setup, we assumed different locations of groundwater inflow along the selected reach (upstream-most point, mid-point, and downstream-most point in the study reach). The sub-reach with the groundwater inflow attributed to a 1 m long subdivision and at a magnitude equal to the total net increase in stream flow along the study reach (i.e., all groundwater inflow in a single location, representing a focused discharge of groundwater into the stream, like that attributable to a geologic discontinuity or fracture). For each of the three model setups, we ran the model 200,000 times and calibrated the model parameters (D, α, A_{TS}, discharge) following the same procedures as the prior model fitting. We then compared calibrated model parameters among the different model setups and applied identifiability analysis to the behavioral parameter sets (1% and 10%). Subsequently, we applied the Levene test for equality of variance to compare the distributions of the model parameters from the different model setups (upstream, middle, downstream), using a *p-value* of 0.05 for determining statistical significance. Although we expect groundwater inflow to primarily affect radon activity in streams, we also calibrated TSM parameters in three model setups that varied in groundwater inflow locations using chloride concentrations. This was motivated by the assumption that chloride-free groundwater, as commonly assumed in TSMs, dilutes chloride concentrations in streams.

3 Results

3.1 Tracer concentration in Oak Creek

BTCs (Fig. S1) showed a distinct peak concentration at both the upstream and the downstream locations of the study reaches, thereby meeting a key requirement for the calibration of the TSM.~~and measured radon activity revealed spatial variability across the study reaches.~~ Radon activity in groundwater was 23 times higher than in surface water~~, reaching 6765 (± 841) Bq m^{-3}.~~, providing the necessary contrast for quantifying groundwater inflows into the stream. ~~R~~Stream radon activity ~~in the stream~~ ranged from 285 (± 22) Bq m^{-3} to 337 (± 26) Bq m^{-3}~~.~~ with ~~T~~the highest activity was observed at reach #2 (Table S1). ~~Radon activity in groundwater was 23 times higher than in surface water, reaching 6765 (± 841) Bq m^{-3}.~~

3.2 Information content and information gain for model parameters

~~Calibrating the~~ TSM calibration with both tracers resulted in higher values of the Kullback-Leibler divergence and thus more information on model parameters compared to calibration with chloride or radon ~~individually~~ alone ($D_{KL}(X|(radon \cap chloride),X) > D_{KL}(X|chloride,X)$ and $D_{KL}(X|(radon \cap chloride),X) > D_{KL}(X|radon,X)$) (Table 2). When calibrating the TSM with each tracer ~~individually~~alone, ~~calibration with~~ chloride ~~yielded~~ provided more information on model parameters than ~~calibrating with~~ radon ($D_{KL}(X|chloride,X) > D_{KL}(X|radon,X)$). The only exception occurred in the calibration approach with fixed groundwater inflow (Q_{fix}). Values of the Kullback-Leibler divergence of individual parameters varied depending on the tracer used for calibration. In general, chloride provided more information on ~~dispersion~~ D but little information on groundwater inflow q_I, whereas radon provided more information on groundwater inflow q_I but less on D~~dispersion~~.

Calibrating the TSM with both tracers increased certainty in model parameters compared to using chloride and radon ~~individually~~alone. This is evident in values of the Shannon entropy of the model parameters, which show that the posterior distributions became narrower ($H(X|(radon \cap chloride)) < H(X)$; (Table 2)). Similarly, ~~calibrating the~~ TSM calibration with each tracer ~~individually~~ alone increased certainty in the model parameters ($H(X|radon) < H(X)$ and $D_{KL}(X|(radon), X) > 0$ and ($H(X|chloride) < H(X)$ and $D_{KL}(X|(chloride), X) > 0$).

Table 2: Shannon entropy H and Kullback-Leibler divergence D_{KL} for the prior and posterior distributions of model parameters (D, A_{TS}, α and the groundwater inflow q_I) resulting from joint ($H(X|radon \cap chloride)$; $D_{KL}(X|(radon \cap chloride),X)$) and individual ~~calibration~~ ($H(X|chloride)$; $H(X|radon)$; $D_{KL}(X|chloride)$; $D_{KL}(X|radon)$) calibration of the TSM with chloride and radon. Results from all three calibration approaches are shown here, which differ in how groundwater inflow was calibrated (Q_{fix}, Q_{LHS}, and Q_{out}). For simplicity, only the results of the top 10% behavioural parameter sets from the low-degassing model setup (k_{low}) with radon are shown here for reach #1. Results for reaches #2 - #5 can be found in the supporting information (Table S2~~1~~).

Calibration approach	Q_{fix}					Q_{LHS}					Q_{out}				
Parameter	D	α	A_{TS}	q_I	SUM	D	α	A_{TS}	q_I	SUM	D	α	A_{TS}	q_I	SUM
Unit	[m² s⁻¹]	[s⁻¹]	[m²]	[m³ s⁻¹ m⁻¹]		[m² s⁻¹]	[s⁻¹]	[m²]	[m³ s⁻¹ m⁻¹]		[m² s⁻¹]	[s⁻¹]	[m²]	[m³ s⁻¹ m⁻¹]	
$H(X)^a$	3.91	3.91	3.91	-		3.91	3.91	3.91	~~0~~-		3.91	3.91	3.91	3.90	
$H(X\|chloride)$	2.55	3.81	3.53	-		2.14	3.86	3.65	~~0~~-		2.16	3.86	3.70	3.87	
$H(X\|radon)$	3.88	3.74	1.65	-		3.89	3.88	3.39	~~0~~-		3.90	3.90	3.75	2.49	
$H(X\|(radon \cap chloride))$	2.55	2.40	1.51	-		2.00	3.72	3.72	~~0~~-		1.99	3.81	3.83	1.93	
$D_{KL}(X\|chloride,X)$	1.36	0.10	0.38	-	1.84	1.77	0.06	0.27	~~0~~-	2.1	1.76	0.04	0.21	0.04	2.05
$D_{KL}(X\|radon,X)$	0.03	0.17	2.27	-	2.47	0.01	0.03	0.52	~~0~~-	0.56	0.01	0.02	0.16	1.42	1.61
$D_{KL}(X\|(radon \cap chloride),X)$	1.95	1.50	2.40	-	5.85	1.91	0.23	0.19	~~0~~-	2.33	1.92	0.10	0.08	1.98	4.08

3.3 Parameter sensitivity and certainty

Parameter sensitivity and certainty depended on the tracer used for calibration. When the TSM was calibrated with chloride, *D*~~dispersion~~ showed high sensitivity, whereas groundwater inflow was largely insensitive (Table 3). Calibrating the TSM with radon resulted in no sensitivity to *D*~~dispersion~~ but high sensitivity to groundwater inflow. The sensitivity and certainty of A_{TS} and α in simulations depended both on the tracer used for calibration and the approach used to calibrate groundwater inflow (Fig. 2, Fig. 3, Fig. 4, Table 3, Table S3~~2~~). Calibration with chloride consistently resulted in high parameter certainty and moderate to high sensitivity for A_{TS} and α. In contrast, calibration with radon yielded moderate to high sensitivity of A_{TS} and α when groundwater inflow was fixed (Q_{fix}). A_{TS} and α became insensitive when groundwater inflow was included as a direct calibration parameter (~~Q_{LHS}~~Q_{Out}) or calculated from calibrated discharge (~~Q_{out}~~Q_{LHS}). A comparison of model performance showed that the nRMSE values of behavioral parameter sets were lower when calibrating with radon than with chloride (Fig. 2).

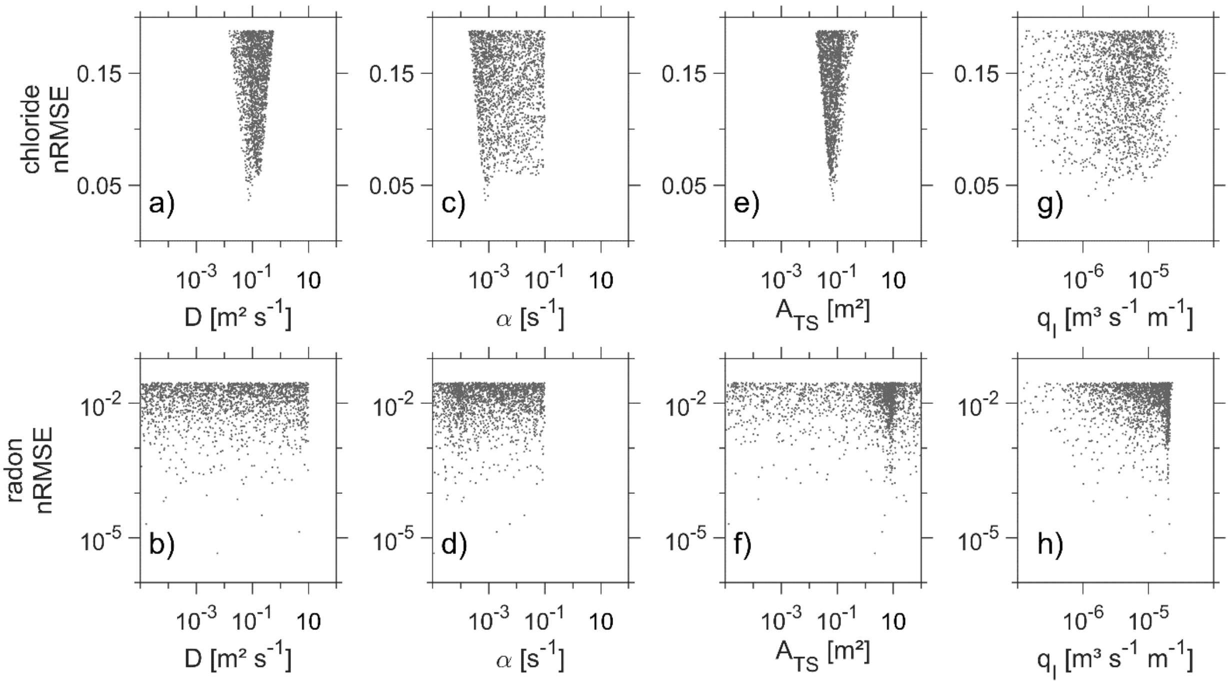

Figure 2: Parameter values (*D*, *α*, A_{TS}, and the groundwater inflow (q_I)) plotted against the corresponding normalized root mean square error (nRMSE) values. The figure shows the best 1% of model runs from calibrating the TSM with chloride (top row) and radon (bottom row) using the calibration approach in which groundwater inflow q_I was calculated from calibrated discharge (Q_{LHS}). The model runs shown are from reach #1. For radon, we present the k_{low} model setup only.

Table 3: Overview of the sensitivity based on the K-S test for all model parameters (D, α, A_{TS}, and the groundwater inflow (q_I)) ~~for calibrating the TSM with chloride or radon.~~from the individual calibration of the TSM with chloride and radon. The terms 'high', 'moderate', 'poor', and 'non-sensitive' refer to the classification of parameter sensitivity based on the K-value and *p-value* (see Section 2.6), following the methodology of Ouyang et al. (2014). Results from all three calibration approaches are shown here, which differ in how groundwater inflow was calibrated (Q_{fix}, Q_{LHS}, and Q_{out}). Results are presented for all reaches, but only for the 1% behavioral parameters. For simplicity, we show the k_{low} model setup for calibrating the TSM with radon only.

		Q_{fix}		Q_{LHS}		Q_{out}	
	parameter	chloride	radon	chloride	Radon	chloride	Radon
Reach #1	D	high	~~slight~~poor	high	~~Slight~~poor	high	~~Slight~~poor
	α	high	high	high	poor~~slight~~	high	poor~~slight~~
	A_{TS}	high	high	high	High	high	moderate
	q_I					~~slight~~poor	high
Reach #2	D	high	~~slight~~poor	high	~~Slight~~poor	high	moderate
	α	high	moderate	high	poor~~slight~~	high	non-sensitive~~insensitive~~
	A_{TS}	high	high	high	moderate	high	high
	q_I					high	high
Reach #3	D	high	non-sensitive~~insensitive~~	high	~~insensitive~~ non-sensitive	high	~~Slight~~poor
	α	high	high	high	slight	high	poor~~slight~~
	A_{TS}	high	high	high	High	high	moderate
	q_I					moderate	high
Reach #4	D	high	poor~~slight~~	high	poor~~slight~~	high	~~insensitive~~non-sensitive
	α	high	high	high	moderate	high	poor~~slight~~
	A_{TS}	high	high	high	High	high	moderate
	q_I					slight	high
Reach #5	D	high	~~insensitive~~non-sensitive	high	poor~~slight~~	high	moderate
	α	high	high	high	poor~~slight~~	moderate	~~insensitive~~ non-sensitive
	A_{TS}	high	high	high	moderate	high	high
	q_I					high	high

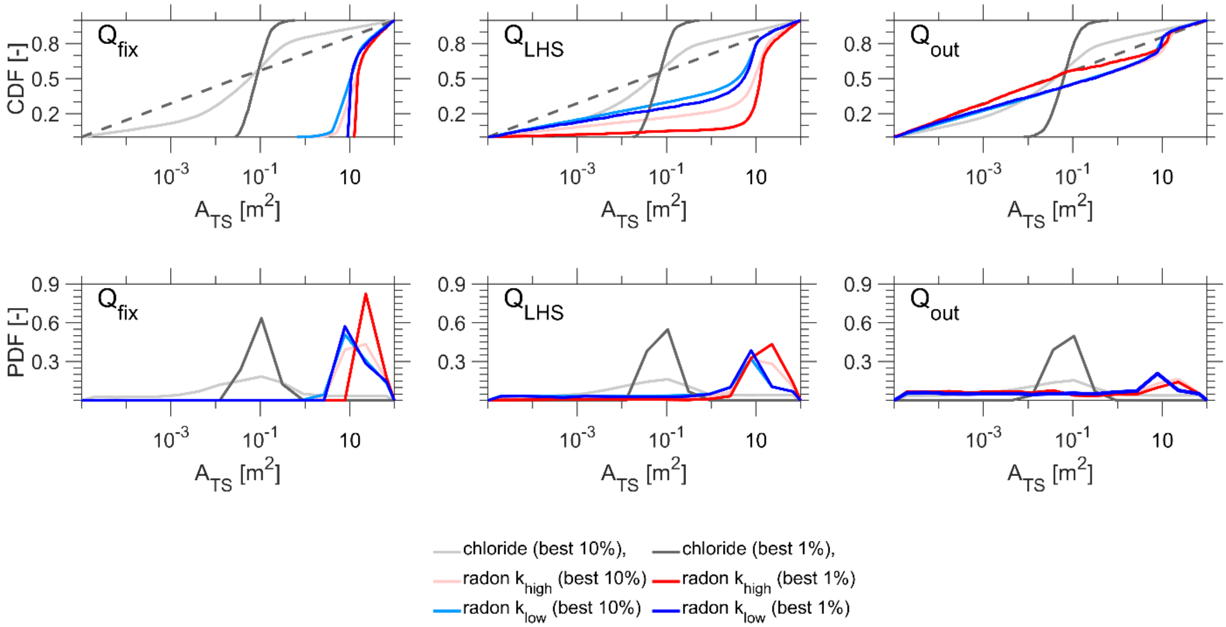

Figure 3: Cumulative distribution plots (upper row) and posterior distribution plots (lower row)<u>. Each plot shows</u> ~~for~~ *A_{TS}* <u>results</u> based on the best 1% and best 10% parameter sets <u>from the individual calibration</u>. We present results from the calibration approach where groundwater inflow was fixed (Q_{fix}), calculated from calibrated discharge (Q_{LHS}), and directly calibrated (Q_{out}). Non-behavioral A_{TS} parameters in the top row are shown with a grey dashed line. For simplicity, we show the results from reach #1 only.

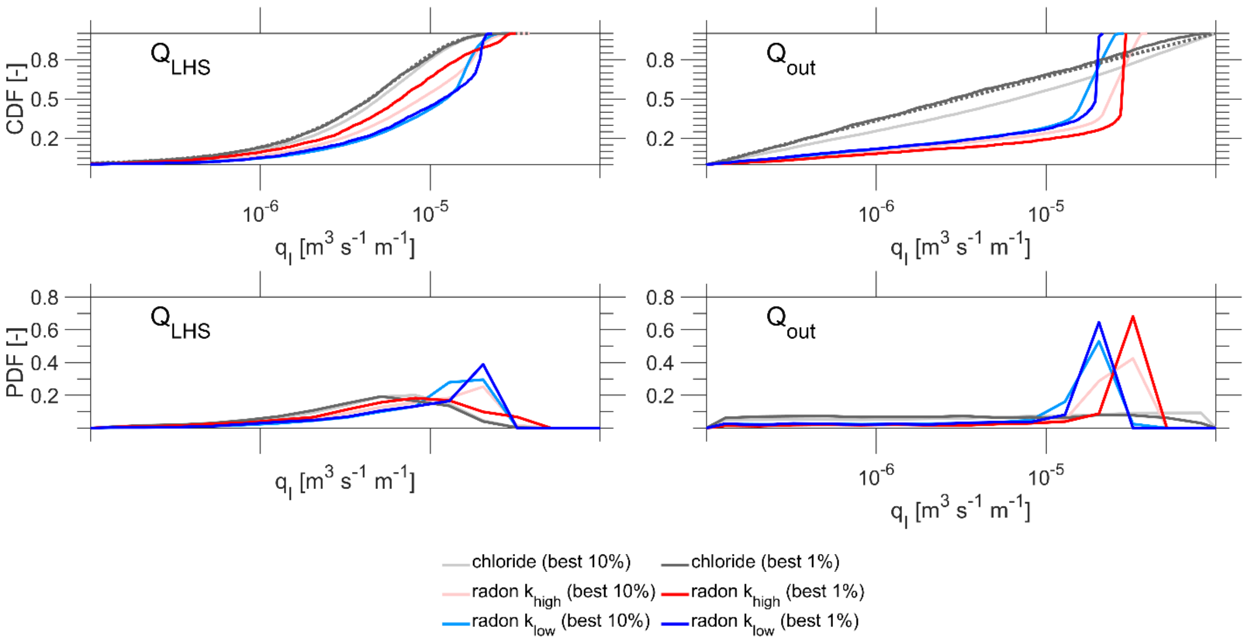

Figure 4: Cumulative distribution plots (upper row) and posterior distribution plots (lower row). Each plot shows ~~for the~~ groundwater inflow (q_I) results based on~~for~~ the best 1% and best 10% parameter sets from the individual calibration. We present results from the calibration approach where groundwater inflow was fixed (Q_{fix}), calculated from calibrated discharge (Q_{LHS}), and directly calibrated (Q_{out}). Non-behavioral q_I parameters in the top row are shown with a grey dashed line. For simplicity, the results from reach #1 are shown only.

3.4 Parameter interactions

We found no parameter interactions when the TSM was calibrated with chloride, with only a few exceptions (Fig. 5, Table ~~S3~~S4). Calibration with chloride resulted in a narrower range between the minimum and maximum values of the behavioral parameters, thereby better constraining parameter values. In contrast, when the TSM was calibrated with radon, parameters were tightly constrained only when others in the same set were less constrained (Fig. 5). For example, groundwater inflow was tightly constrained at higher values within the behavioral parameter range, but A_{TS} values for these groundwater inflow values remained unconstrained. Conversely, when A_{TS} values were tightly constrained at higher values, groundwater inflow values associated with them were unconstrained.

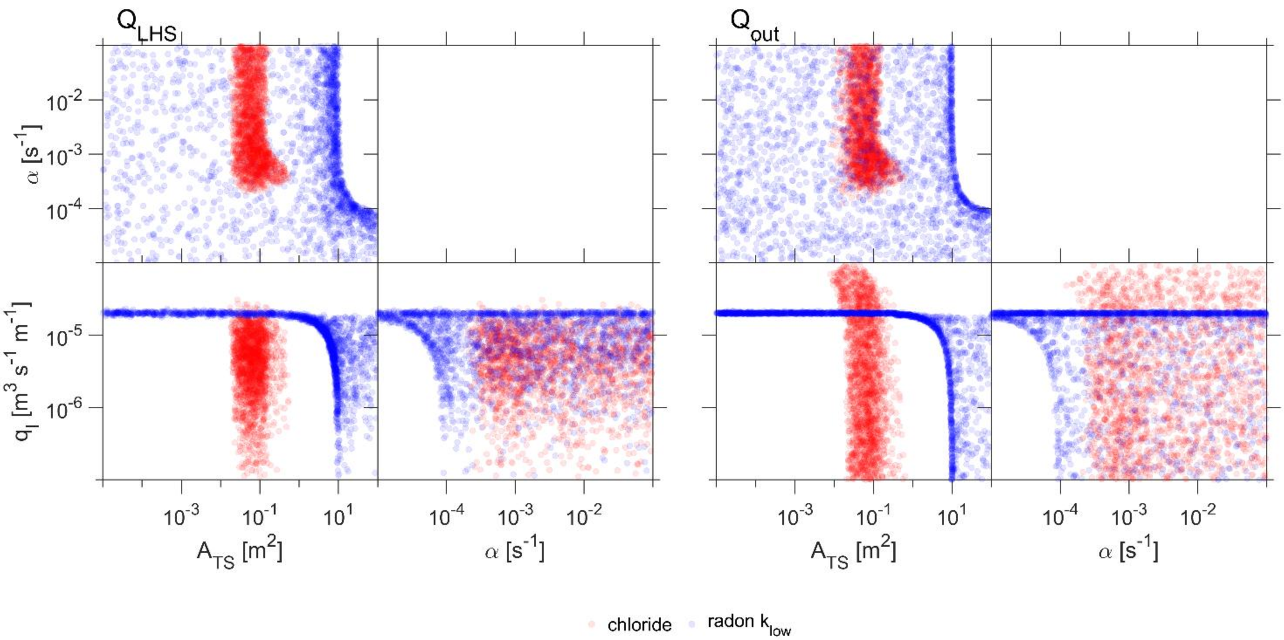

Figure 5: Scatter plots of the best 1% model parameters (A_{TS}, α and q_l) from calibrating the TSM with radon (blue) and chloride (red) ~~individually~~alone. Calibration with chloride resulted in constrained parameter values. When the TSM was calibrated with radon, parameter values were tightly constrained only when others in the same set were less constrained. We show the calibration approach where the groundwater inflow was calculated from calibrated discharge (Q_{LHS}) and the calibration approach where the groundwater inflow was directly calibrated (Q_{out}). For simplicity, only the results from calibrating the TSM with radon in the k_{low} model setup are shown. The parameter *D* is not included, as it was neither certain nor sensitive in calibrating the TSM with radon.

3.5 The effect of different locations of groundwater inflow on parameter interactions

We found significantly different distributions of the behavioral parameters dependent~~d~~ on the different locations of groundwater inflow (upstream-most point, mid-point, downstream-most point model setups) independently of which tracer was used for calibration (Fig. 6, Fig. 7). Calibrating the TSM with chloride resulted in constrained parameters when the inflow was located at the upstream-most point or mid-point (Fig. 7). In contrast, parameter interactions became evident when inflow was at the downstream-most point. When the TSM was calibrated with radon and groundwater inflow was set at the upstream-most point, A_{TS} was tightly constrained at higher values within the behavioral parameter range. However, the groundwater inflow values remained unconstrained for these higher A_{TS} values. In contrast, parameter values were less constrained when inflow was set at the downstream-most point.

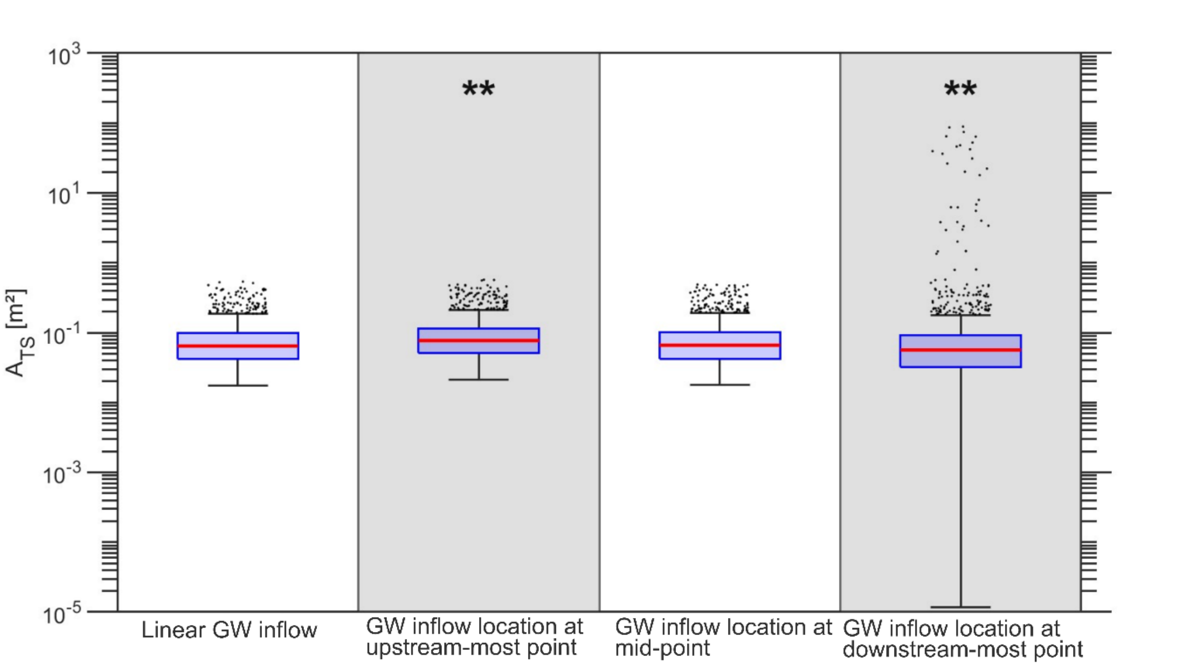

Figure 6: Distributions of behavioral A_{TS} values from model setups with~~that~~ varying ~~in~~ groundwater inflow locations, ~~and~~ each calibrated with chloride only. The model setup labelled 'linear groundwater (GW) inflow' refers to the calibration approach where the groundwater inflow was calculated from the calibrated discharge (Q_{LHS}). The red line is the median of the distributions, while black dots highlight outliers. Asterisks and the grey areas show a significant difference between the variance of the parameter distributions compared to the setup with linear groundwater inflow. Results are shown for the best 1% behavioral model parameters and reach #1 only.

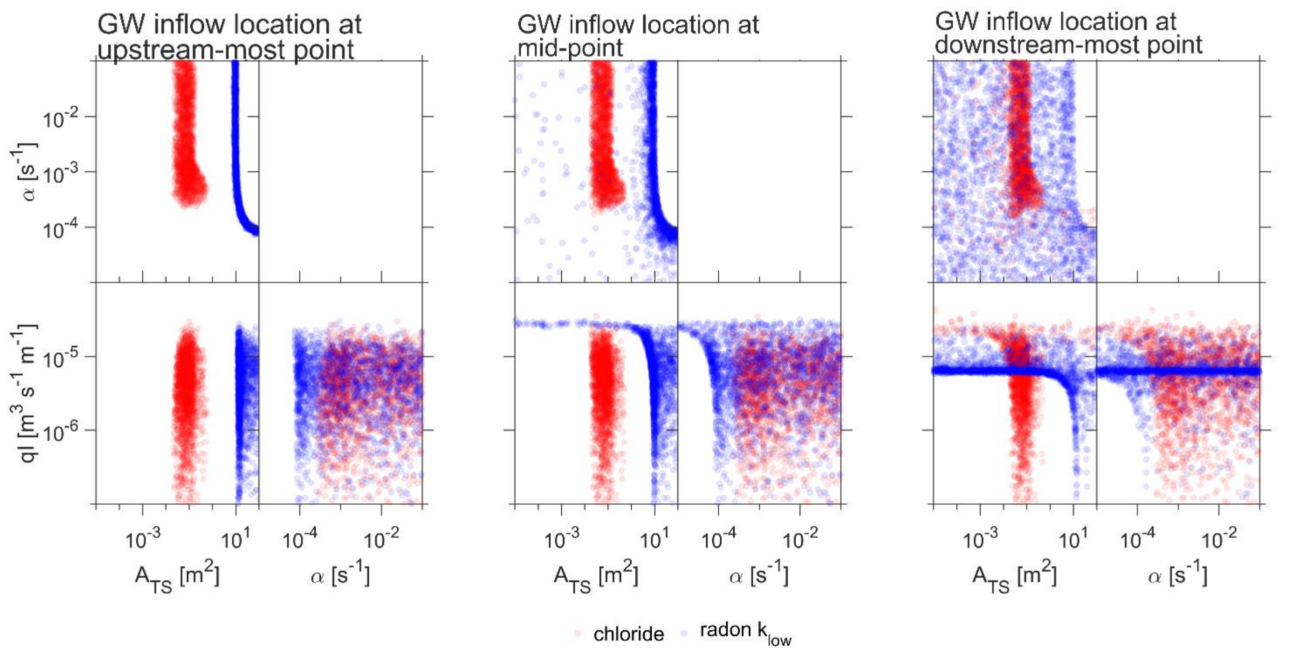

Figure 7: Scatter plots of the best 1% behavioral model parameters (A_{TS}, α and the groundwater inflow q_I) from calibrating the TSM with radon and chloride ~~individually~~alone. Three different model setups are presented that differ in the location of the groundwater inflow along the reach (upstream-most point, mid-point, downstream-most point model setups)~~location upstream, middle and downstream).~~ For simplicity, results from calibrating the TSM with radon in the k_{low} model setup are shown only.

3.6 Gross water fluxes from ~~calibrating the~~ the TSM calibrated with radon and chloride

Discharge gradually increased in the downstream direction and ranged from 9.5 to 12 L s^{-1} across all reaches (Fig. 8a). However, not all reaches exhibited this increase; specifically, discharge decreased in reaches #1 and #4. Gross loss and gain revealed spatial variability across reaches. Reaches #1 and #4 were characterized by higher gross losses compared to remaining reaches (Fig. 8b). Notably, gross water flux could only be derived from calibrating the TSM with radon and not with chloride.

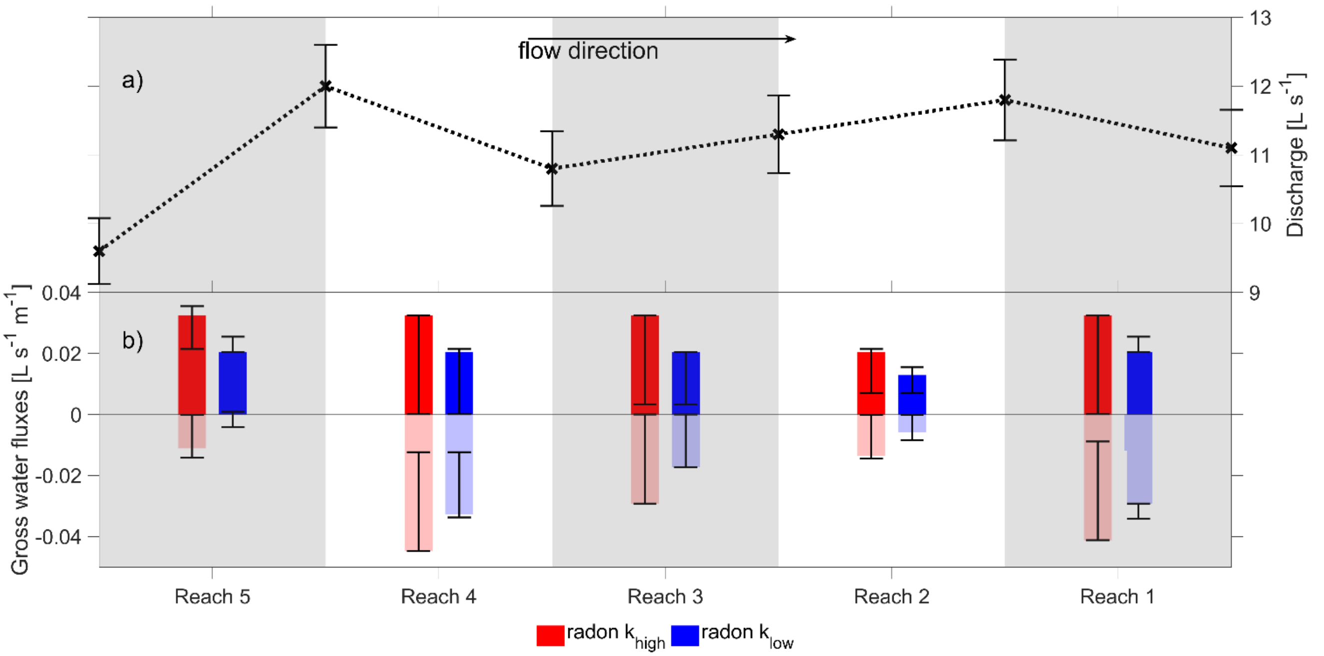

Figure 8: a) Discharge values upstream and downstream of each reach [L s^{-1}] with 95% confidence intervals. Discharge was calculated from BTCs using dilution gauging (Q_{out}). b) Gross water fluxes [L s^{-1} m^{-1}] for each reach, derived from calibrating the TSM with radon measurements. Both degassing model setups are shown (blue and red) and transparent bars highlight negative ~~gross~~ water fluxes. Bar heights correspond to the mode of the posterior distribution for the calibrated q_I and calculated q_{out} values. Error bars show the 95% confidence interval of the posterior distribution.

4 Discussion

4.1 Calibrating TSMs with multiple tracers better constrains model parameters

Calibrating the TSM with chloride provides more information on solute transport~~ (i.e., the sum of the Kullback-Leibler divergence for all TSM parameters)~~ than calibrating ~~it~~ with radon~~,~~ because~~as~~ chloride ~~provides more information~~is particularly informative on ~~the~~ *D* ~~dispersion parameter~~ (Table 2). Previous studies have shown that *D*~~dispersion~~ mainly affects the rising limb of BTCs (e.g., Kelleher et al., 2013; Scott et al., 2003; Wlostowski et al., 2013~~),~~). This highlights ~~suggesting~~ ~~that tracers~~that a tracer with a distinct rising concentration limb, such as chloride, ~~are~~is necessary to identify *D*~~dispersion~~. ~~Radon activity remains steady throughout the experiment, which limits its ability to identify dispersion. Although radon yields less information on dispersion than chloride, radon~~ Radon, by contrast, provides more information on groundwater inflow and A_{TS} than chloride (Table 2)~~. This is due to the higher radon~~ due to its higher activity in groundwater compared to surface water~~, with a 23-fold difference at Oak Creek, which increases radon activity in streams~~. ~~In contrast,~~Because chloride concentrations are ~~smaller~~ lower in groundwater compared to surface water~~.~~ ~~Groundwater~~ , groundwater inflow ~~therefore~~ simply dilutes chloride concentrations in the stream without ~~providing~~ adding additional information on ~~the~~ inflow ~~itself~~. In summary, the most information on solute transport is obtained by using both tracers jointly because each tracer contributions uniquely. ~~Ultimately, the~~ ~~unique information each tracer provides about different~~ Joint calibration with radon and chloride improves ~~model parameters enhances the~~ certainty ~~of~~ in solute transport estimates ~~when~~ compared to ~~the TSM is jointly~~ calibrat~~ed~~ing ~~with radon and chloride compared to~~ with either tracer ~~individually~~alone. We therefore recommend calibrating TSMs with multiple tracers to improve estimates of solute transport in future studies.

This recommendation aligns with recent calls for joint calibration of hydrological models with multiple tracers. For example, Neilson et al. (2010b) demonstrated that calibrating a TSM with both temperature and slug tracer data provides more insights into solute transport and exchange compared to calibrating the TSM with temperature alone~~individually~~. At the catchment scale, Rodriguez et al. (2021) demonstrated that jointly calibrating a storage selection function with deuterium and tritium ~~improved certainty~~reduced uncertainty ~~o~~in model parameters compared to calibrating ~~the storage selection function~~ with either tracer ~~individually~~alone. Notably, the authors ~~employed~~ used a different quality criterion for behavioral parameter selection than we did. Rodriguez et al. (2021) used distinct threshold values for each tracer to obtain a comparable number of behavioral parameter sets~~, as the~~ due to differences in sampling frequency, ~~—~~ and thus ~~the~~ dataset length, ~~– differed~~ between deuterium and tritium. In contrast, w~~W~~e selected the best 1% and 10% of parameter sets as behavioral to ~~maintain consistency~~remain consistent with previous solute transport studies (e.g., Kelleher et al., 2019; Wagener et al., 2002; Ward et al., 2013, 2017; Wlostowski et al., 2013). This~~S~~ selecti~~ng~~on ~~the best 1% and 10% of parameter sets led~~ resulted in~~to~~ a lower nRMSE for radon compared to chloride (Fig. 2), thus, ~~meaning that calibrating the TSM with radon in our study carried~~ carrying more weight in the joint calibration. Therefore, the conclusion that ~~jointly calibrating the~~ joint calibration of the TSM with radon and chloride yields more information than ~~calibrating~~ calibration~~it~~ with either tracer ~~individually~~ alone is partly influenced by the quality criterion for parameter sets~~,~~ and, ultimately, ~~thus~~ by subjective modeling decisions - a well-known challenge in hydrology (e.g., Beven and Binley, 1992).

4.2 The role of groundwater inflow for parameter identifiability

The sensitivity of radon to the amount and location of groundwater inflow ~~hinders the derivation~~makes it difficult to obtain ~~of~~ narrow, well-constrained estimates for groundwater inflow, A_{TS}, and α, when calibrating TSMs with radon ~~individually~~alone. The sensitivity to the amount of groundwater inflow is evident in the calibration approach where groundwater inflow was calculated from calibrated discharge (Q_{LHS}; Fig. 5). Discharge was sampled within the uncertainty range of measurements (95% confidence interval from dilution gauging). Even within this ~~uncertainty~~ range, ~~different~~ varying ~~values for the~~ groundwater inflow values led to wide, unconstrained values of A_{TS} and α. This suggests that ~~obtaining narrow, well-constrained estimates for~~constraining groundwater inflow, A_{TS}, or α ~~from calibrating the~~with ~~TSM with~~ radon calibration alone will remain challenging unless at least one of these parameters is ~~further~~ independently constrained. Deriving narrow, well-constrained TSM parameters is also restricted due to the sensitivity of radon to the location of groundwater inflow. This is

~~evident in~~shown by different parameter interactions across model setups that vary in the location of groundwater inflow (Fig. 7). For example, when groundwater inflow occurs at the downstream-most point in the study reach, ~~increases~~ radon activity ~~at this point~~increases there, where measured radon ~~values~~ activity ~~are~~is used for calibration. A shorter distance between ~~the~~ inflow and ~~the downstream-most~~this point ~~allows~~ means less time for radon ~~to~~ degassing. With less ~~time for~~ degassing time, only smaller~~,~~ and better constrained ~~values of~~ groundwater inflow values can close the radon mass balance and ~~result in~~achieve a good fit between simulated and measured radon activity. ~~Constrained~~ However, these constrained groundwater inflow values lead to less-constrained ~~estimates of~~ A_{TS} and α~~,~~ estimates, ~~as~~ since different ~~values~~ combinations of these parameters ~~A_{TS} and α~~ can still close the radon mass balance. ~~Thus~~Therefore, ~~the~~ spatial variability ~~of~~ in groundwater inflow~~s hinders better~~hampers precise constraints ~~on parameter estimates for~~ A_{TS} and α when using radon ~~individually~~ alone for model calibration.

Spatially heterogeneous groundwater inflows have been documented across various streams and attributed to transitions in valley structure (Mallard et al., 2014; Cartwright and Gilfedder, 2015; Payn et al., 2009; Pittroff et al., 2017; Somers et al., 2016), geological fractures (Genereux et al., 1993; Glaser et al., 2020), or ~~textural heterogeneities of the~~ subsurface textural heterogeneities (Fleckenstein et al., 2006). Groundwater inflows at discrete stream locations ~~of streams~~ are common (Sophocleous 2002). However, ~~yet~~ their ~~impact~~ effect on the identifiability of TSM parameters ~~with radon and chloride~~ has not been explored. ~~Instead, p~~Previous research has instead highlighted the critical role of degassing when simulating radon activity (e.g., Atkinson et al., 2015; Gilfedder et al., 2019). For example, Schubert et al. (2020) incorporated degassing tests alongside radon measurements to quantify groundwater inflow ~~through~~ using a numerical mass balance approach with transient storage parameters. ~~(Frei and Gilfedder, 2015). Schubert et al. (2020)~~They found that calibrated groundwater inflows exceeded the net increase in discharge, attributing this outcome to spatial variability in degassing rates. ~~The~~ In our study, the two degassing parameterizations ~~in our study~~ did not affect model performance (e.g., Fig. 3 and 4~~);~~). ~~rather~~Instead, differences in groundwater inflow locations along the reach resulted in different model parameters in behavioral parameter sets. ~~Despite differences in our study site and that~~Although our study site differs from ~~of~~ Schubert et al.'s (2020), both models assume linear groundwater inflow along the reach. Therefore, the ~~observed~~ overestimation of calibrated groundwater inflows by Schubert et al. (2020) may also result from spatial variability in groundwater inflows along their study reach, ~~leading to radon~~ which causes radon increases that ~~required~~ require ~~calibrating~~ higher calibrated groundwater inflow values. Similarly, Cook et al. (2006) ~~showed~~ reported an overestimation of groundwater inflow to the Cockburn River, Australia, by almost 70% compared to actual flow measurements. The authors concluded that including exchange with subsurface transient storage zones is essential when simulating radon activity. ~~. Cook et al. (2006) assumed~~ An underlying assumption of their model approach was linear inflow along the reach, similar to Schubert et al. (2020). In light of our findings, we suggest that the overestimation found by Cook et al. (2006) may ~~also~~ instead be due to spatial variability in groundwater inflows, rather than solely the omission of exchange with subsurface transient storage zones in their radon mass balance. Therefore, our findings emphasize that the spatial variability of groundwater inflow location should be explicitly accounted for in future radon-based studies. This consideration may challenge the common perception that radon mass balances can be fully closed solely by including exchange with subsurface transient storage zones. ~~The insights of our study are therefore critical for contextualizing results in future radon-based studies in river corridors, especially given the prevailing perception in the radon community — based on Cook et al.'s (2006) findings nearly twenty years ago — that exchange with subsurface transient storage zones should be included in radon mass balances to close them.~~

Notably, our results show that spatially heterogeneous groundwater inflows also affect the identifiability of α and A_{TS} when calibrating the TSM with chloride concentrations (Fig. 6). Previous studies ~~have noted~~reported uncertainty ~~of the~~in α estimates from the behavioral parameters when calibrating TSMs with chloride concentrations (Kelleher et al., 2013; Wagener et al., 2002; Wlostowski et al., 2013). These studies ~~attributed the~~linked the uncertainty ~~of α~~ to differences in stream-specific characteristics. Our findings ~~offer~~ provide a more detailed explanation~~,~~: ~~namely~~ the spatial variability in groundwater inflow along streams. We ~~therefore~~ thus suggest ~~that~~ further research ~~is needed~~ to understand how groundwater inflow ~~influences~~ affects chloride concentrations and, ~~in turn~~consequently, the identifiability of parameters ~~when calibrating the~~in TSM calibration. Future studies on solute transport with chloride should first identify groundwater inflow locations before selecting a reach for slug tracer injections. This could be ~~achieved~~ done by incorporating spatially resolved piezometers and hydraulic head measurements along streams, as implemented in the study design of Harvey (1993) and Bonanno et al. (2023). Alternatively, temperature surveys (distributed temperature sensing; 'DTS'), which provide high spatial and temporal resolution data on longitudinal groundwater inflows (Krause et al., 2012), could be used.

4.3 Radon is biased toward~~s~~ large spatial-scale subsurface flow paths

Spatially variable gross water gains exceeding net discharge from ~~calibrating the~~ TSM calibration with radon suggest that the water balance of Oak Creek~~'s water balance~~ is affected~~influenced~~ by stream–subsurface water exchange~~s~~ that is occurring at multiple spatial scales. Previous studies have shown that large spatial-scale subsurface flow paths, that are subsurface flow paths originating from further upstream of the stream reach, ~~-~~play a critical role in explaining water mass balances in streams (e.g., Payn et al., 2009; Stanford and Ward, 1993; Ward et al., 2023). ~~The inability to detect these gross water gains with chloride due to nearly complete BTC mass recovery highlights radon's superior sensitivity in revealing large-scale subsurface flow paths.~~Unlike chloride, radon uniquely labels these flow paths (Fig. 9, arrow C), which would be undetectable if chloride were used. This unique labeling ~~is~~ underpins the superior sensitivity of radon in revealing subsurface flow paths and highlights its novel contribution to understanding the role of these flow paths in stream water balances.~~because~~ . ~~radon activity was in~~ Conceptually, this labeling capacity of radon is reflected in measured steady state activity at both the upstream and downstream ends of the reach before the slug tracer experiment began.~~. Steady-state radon activity indicates~~ Steady state activity indicates that ~~that the ensemble of timescales for many parcels of water in the~~ the measured radon activity includes subsurface water parcels ~~is~~ -pre-labelled with radon prior to the experiment. Thus,~~-~~ ~~Subsurface flow paths~~large spatial-scale subsurface flow paths ~~originating~~ from further upstream also contribute to measured radon activity at the downstream end of the reach (Fig. 9, arrow C~~D~~), along~~side~~ with ~~contributions from~~ groundwater inflow and temporally shorter flow paths (Fig. 9, arrows A, B, and E; Cook et al., 2013; McCallum et al., 2012). ~~In contrast, chloride is injected as a distinct input signal (Dirac injection) at the upstream end of the reach, thus only revealing flow paths between locations of injection and measurement (Fig. 9, arrows A and B). This finding underscores radon's value as a tracer for detecting large spatial-scale subsurface flow paths that slug injections of chloride may overlook.~~ In contrast, chloride is injected as a distinct input signal (Dirac injection) at the upstream end of the reach and thus only reveals flow paths between locations of injection and measurement (Fig. 9, arrows A and B). This inherent bias in chloride injections toward labeling timescales within the WoD complements the bias of radon toward large spatial-scale subsurface flow paths, underscoring the value of using both tracers together. However, despite these novel insights on large spatial-scale flow paths from calibrating TSM with radon, ~~large spatial-scale subsurface flow~~such paths are not explicitly represented in the TSM. Instead, they~~the flow paths~~ are indirectly accounted for through the calibration of ~~the~~ groundwater inflow. This indirect consideration may lead to an overestimation of groundwater inflow values during calibration~~, ~~. Such overestimation ~~which~~ reduces~~es~~ ~~the~~ certainty ~~of~~in the A_{TS} and α parameters and ~~introduces~~ causes interactions ~~between the~~among groundwater inflow, A_{TS}, and α. Although radon's sensitivity to large spatial-scale subsurface flow paths provides valuable insights, it therefore also introduces bias in constraining transient storage parameters in TSMs. ~~Thus, radon traces subsurface timescales longer than the WoD, which cannot be identified through calibrating the TSM with radon.~~

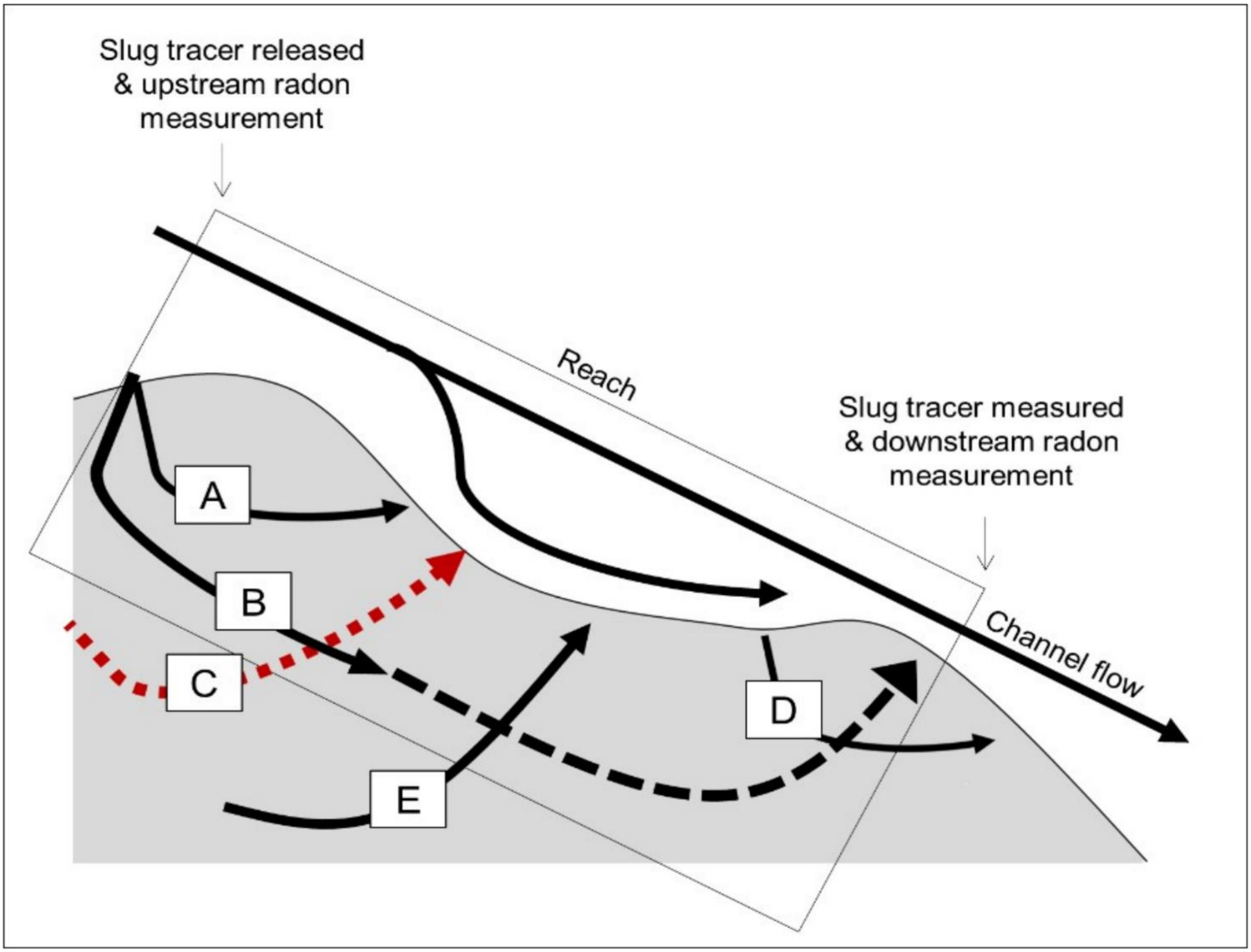

Figure 9: Conceptual figure of flow paths in streams. The box shows the reach of investigation. A is a flow path that is labelled by the tracer at the upstream location and returns within the WoD. B is a flow path with timescales longer than the WoD, where transit times exceed the duration of the tracer experiment, as indicated by the dashed arrow. ~~C is a flow path that bypasses the downstream end of the reach.~~ C is a subsurface flow path entering the hyporheic zone upstream of the reach. This flow path is characterized by subsurface transit times shorter than 21 days and radon activity that has not yet reached secular equilibrium. This flow path can be identified by radon but not by chloride concentration, which explains the red coloration. ~~D is a subsurface flow path entering the hyporheic zone upstream of the reach. This path is characterized by transit times shorter than 21 days and radon activity that has not yet reached secular equilibrium.~~ D~~C~~ is a flow path that bypasses the downstream end of the reach. E is a groundwater flow path with transit times longer than 21 days and radon activity at secular equilibrium. Flow path D is conceptually excluded from the TSM, which is why it is highlighted with a red dashed arrow. (Figure adapted from Payn et al. (2009))

4.4 Implications and future research

The improved parameter identifiability through joint TSM calibration with radon and chloride may be a critical step toward enhancing the physical realism of TSMs and providing more reliable estimates of solute transport. This is important because past studies found that the relationships between TSM parameters and hydrologic drivers are often contradictory (Ward and Packman, 2019; Bonanno et al., 2022). We envision that future studies will derive model parameters by jointly calibrating TSMs across diverse environments and hydrologic conditions. This approach can help clarify how model parameters vary with hydrologic drivers. To advance this goal, studies should compare streams across varying scales and geological settings to determine how these differences affect calibration outcomes when radon is included. Through such comparative studies, researchers could test our expectation that streams with higher radon activity, typically smaller-order streams with granite-rich geology, will yield greater certainty for groundwater inflow q_I and improved model performance when TSMs are jointly calibrated with radon and chloride. This greater certainty for q_I might result from the larger difference in radon concentration activity between groundwater and stream water. Furthermore, future research could jointly calibrate TSMs with radon and tracer data from constant rate injections. This approach would test whether radon captures temporally longer flow paths than those detected by constant rate but not by slug tracer injections.

For future research with radon on gross exchange fluxes, we recommend constraining the transient storage parameter by considering surface (e.g., pools) and subsurface transient storage zones (e.g., hyporheic zones) in TSM evaluations (e.g., Choi et al., 2000). This distinction~~step~~ might be critical because~~, given~~ ~~that~~ only surface storage contributes to radon degassing~~.~~, whereas radon activity increases mainly in the subsurface storage and, to a lesser extent, in surface storage.~~OTIS-R currently considers a single storage zone for the A_{TS} parameter, which actually includes both surface storage and subsurface storage.~~ These storage-specific processes are not fully captured by using a single storage zone, as implemented in OTIS-R, instead of two. We selected the one-zone storage model because this setup aligns conceptually with most established radon models (e.g., Cook et al., 2006; Frei and Gilfedder, 2015) and many TSM calibration with slug tracers (e.g., Bonanno et al., 2023). To our knowledge, no prior radon study has considered two storage zones, highlighting a promising opportunity for future research. ~~Moreover, w~~We also recommend ~~that future studies~~ incorporat~~e~~ing additional discharge measurements to reduce measurement errors. This is important~~necessary~~ because the calibrated gross exchange fluxes for Oak Creek fall within the confidence interval of changes in net discharge (Fig. 8). Therefore, caution is needed when interpreting these fluxes. By following~~addressing~~ these recommendations, future studies may unlock radon's full potential as a tracer. This will~~,~~ enabl~~ing~~e deeper insights into the continuum of subsurface flow paths beyond the WoD of slug tracer injections. Experimental setups could for instance measure radon activity alongside BTCs at several downstream locations from the point of injection (e.g., Ward et al., 2023), thereby supporting the assessment of large spatial-scale subsurface flow paths from slug tracer injections. Studies focusing on understanding the turnover of chemical compounds along large spatial-scale subsurface flow paths could benefit from incorporating radon as an additional tracer. This is especially relevant for slower chemical degradation processes, such as denitrification, which are more likely to occur along flow paths with longer timescales (sensu Jimenez-Fernandez et al., 2022) which radon may help resolve.

5 Conclusions

The ~~overarching~~ goal of this study was to quantify flow paths of different timescales at the reach scale using measurements of ~~solute tracer~~chloride concentration and naturally occurring radon. To achieve this, we calibrated a transient storage model (TSM) with ~~both~~ each tracer~~s~~ individually and with both jointly. Our results show that ~~calibrating the~~ TSM calibration with chloride yields identifiable, narrower parameter values, while calibration with radon does not result in identifiable parameters. The lack of identifiability with radon is due to steady-state activity of radon in the stream, the tracer's sensitivity to the amount and location of groundwater inflow, and to large-scale subsurface flow paths, demonstrating radon’s bias toward flow paths with longer temporal timescales. We found that joint calibration of a TSM with both tracers provided the most information on parameters of the TSM compared to calibrating it with either tracer ~~individually~~alone. Based on these findings, we recommend that future studies incorporate both radon and chloride when calibrating the TSM to improve estimates of solute transport. Furthermore, with an adapted sampling and modeling strategy, we see significant potential in incorporating radon in future studies to unlock its full potential in revealing flow paths with longer timescales than those traced by slug tracer injections.

Code availability

The code to obtain the identifiability of the transient storage parameter and OTIS-R are freely available at Bacher et al. (2025).

Data availability

Solute breakthrough curve and radon activity can be obtained at Bacher et al. (2025).

6 Author contribution

All authors developed the concept of the manuscript and discussed the content of this manuscript. Mortimer Bacher took care of the formal analysis, programming the software and writing the first draft. Julian Klaus contributed to the original idea and revised and edited the manuscript. Adam Ward and Catalina Segura revised and edited the manuscript. Jasmine Krause conducted experiments and revised and edited the manuscript. Clarissa Glaser had the original idea for this manuscript, conducted the experiments, took care of writing the first draft, implemented the revisions and administration of the project.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

The authors would like to thank Jaime Ortega, Madelyn Maffia, Keira Johnson and Stephen Arthur Fitzgerald for supporting field work and Stephen Good for providing lab space for conducting our lab analysis.

Financial support

This research was supported by the Argelander Starter Kit of the University of Bonn, the Klaus Tschira Stifung, and the Open Access Publication Fund of the University of Bonn.

References

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

Atkinson, A. P., Cartwright, I., Gilfedder, B. S., Hofmann, H., Unland, N. P., Cendón, D. I., Chisari, R.: A multi-tracer approach to quantifying groundwater inflows to an upland river; assessing the influence of variable groundwater chemistry, Hydrol. Process.29 (1), 1–12, https://doi.org/10.1002/hyp.10122, 2015.

Bacher, M. L., Klaus, J., Glaser, C.: OTIS-R, https://github.com/mortimerluka/OTIS-R, 2025.

Becker, P. S., Ward, A. S., Herzog, S. P., Wondzell, S. M.: Testing hidden assumptions of representativeness in reach-scale studies of hyporheic exchange, Water Resour. Res. 59 (1), https://doi.org/10.1029/2022WR032718, 2023.

Bencala, K. E.; Walters, R. A.: Simulation of solute transport in a mountain pool-and-riffle stream: A transient storage model, Water Resour. Res. 19 (3), 718–724, https://doi.org/10.1029/WR019i003p00718, 1983.

Bencala, K. E., Gooseff, M. N., Kimball, B. A.: Rethinking hyporheic flow and transient storage to advance understanding of stream-catchment connections, Water Resour. Res. 47 (3), 44, https://doi.org/10.1029/2010WR010066, 2011.

Beven, Keith; Binley, Andrew: The future of distributed models: Model calibration and uncertainty prediction, Hydrol. Process. 6 (3), 279–298, https://doi.org/10.1002/hyp.3360060305, 1992.

Bonanno, E., Blöschl, G., Klaus, J.: Exploring tracer information in a small stream to improve parameter identifiability and enhance the process interpretation in transient storage models, Hydrol. Earth Syst. Sci. 26 (23), S. 6003–6028, https://doi.org/10.5194/hess-26-6003-2022, 2022.

Bonanno, E., Blöschl, G., Klaus, J.: Discharge, groundwater gradients, and streambed micro-topography control the temporal dynamics of transient storage in a headwater reach, Water Resour. Res., e2022WR034053, https://doi.org/10.1029/2022WR034053, 2023.

Briggs, M. A., Gooseff, M. N., Arp, C. D., Baker, M. A.: A method for estimating surface transient storage parameters for streams with concurrent hyporheic storage, Water Resour. Res. 45(4), https://doi.org/10.1029/2008WR006959, 2009.

Briggs, M. A., Goodling, P., Johnson, Z. C., Rogers, K.M., Hitt, N. P., Fair, J. B., Snyder, C. D.: Bedrock depth influences spatial patterns of summer baseflow, temperature and flow disconnection for mountainous headwater streams, Hydrol. Earth Syst. Sci. 26 (15), 3989–4011, https://doi.org/10.5194/hess-26-3989-2022, 2022.

Camacho, L. A., González, R. A.: Calibration and predictive ability analysis of longitudinal solute transport models in mountain streams, Environ Fluid Mech 8 (5-6), 597–604, https://doi.org/10.1007/s10652-008-9109-0, 2008.

Cardenas, M. B., Wilson, J. L.: Exchange across a sediment–water interface with ambient groundwater discharge, J Hydrol 346 (3-4), 69–80, https://doi.org/10.1016/j.jhydrol.2007.08.019, 2007.

Cardenas, M. B.: Surface water-groundwater interface geomorphology leads to scaling of residence times, Geophys. Res. Lett 35 (8), https://doi.org/10.1029/2008GL033753, 2008.

Cargill, S. K., Segura, C., Villamizar, S. R., Warren, D. R.: The influence of lithology on stream metabolism in headwater systems, Ecohydrology 14 (4), e2284, https://doi.org/10.1002/eco.2284, 2021.

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

Cecil, L. D., Green, J. R.: Radon-222, in: Environmental tracers in subsurface hydrology. edited by Cook, P. G., Herczeg, A. L., Springer New York, NY, USA, 175–194, https://doi.org/10.1007/978-1-4615-4557-6, 2000.

Choi, J., Harvey, J. W., Conklin, M. H.: Characterizing multiple timescales of stream and storage zone interaction that affect solute fate and transport in streams, Water Resour. Res. 36 (6), 1511–1518, https://doi.org/10.1029/2000WR900051, 2000.

Cook, P. G., Lamontagne, S., Berhane, D., Clark, J. F.: Quantifying groundwater discharge to Cockburn River, southeastern Australia, using dissolved gas tracers ^{222}Rn and SF_6, Water Resour. Res. 42 (10), 718, https://doi.org/10.1029/2006WR004921, 2006.

Cook, P. G.: Estimating groundwater discharge to rivers from river chemistry surveys, Hydrol. Process. 27 (25), 3694–3707, https://doi.org/10.1002/hyp.9493, 2013.

Corbett, D. R., Burnett, W. C., Cable, P. H., Clark, S. B.: A multiple approach to the determination of radon fluxes from sediments, J Radioanal Nucl Chem 236 (1-2), 247–253, https://doi.org/10.1007/BF02386351, 1998.

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

Cover, T. M., Thomas, J. A.: Elements of information theory. 2nd ed., Wiley Online Books, https://doi.org/10.1002/047174882X, 2006.

Covino, T., McGlynn, B., Mallard, J.: Stream-groundwater exchange and hydrologic turnover at the network scale, Water Resour. Res. 47 (12), 41, https://doi.org/10.1029/2011WR010942, 2011.

Covino, T. P., McGlynn, B. L.: Stream gains and losses across a mountain-to-valley transition: Impacts on watershed hydrology and stream water chemistry, Water Resour. Res. 43 (10), https://doi.org/10.1029/2006WR005544, 2007.

Cranswick, R. H., Cook, P. G., Lamontagne, S.: Hyporheic zone exchange fluxes and residence times inferred from riverbed temperature and radon data, J Hydrol 519, 1870–1881, https://doi.org/10.1016/j.jhydrol.2014.09.059, 2014.

Day, T. J.: On the precision of salt dilution gauging, J Hydrol 31 (3-4), 293–306, https://doi.org/10.1016/0022-1694(76)90130-X, 1976.

Day, T. J.: Observed mixing lengths in mountain streams, J Hydrol 35 (1-2), 125–136, https://doi.org/10.1016/0022-1694(77)90081-6, 1977.

Fleckenstein, J. H., Niswonger, R. G., Fogg, G. E.: River-aquifer interactions, geologic heterogeneity, and low-flow management, Ground water 44 (6), 837–852, https://doi.org/10.1111/j.1745-6584.2006.00190.x, 2006.

Frei, S., Gilfedder, B. S.: FINIFLUX. An implicit finite element model for quantification of groundwater fluxes and hyporheic exchange in streams and rivers using radon, Water Resour. Res. 51 (8), 6776–6786, https://doi.org/10.1002/2015WR017212, 2015.

Frei, S., Durejka, S., Le Lay, H., Thomas, Z., Gilfedder, B. S.: Quantification of hyporheic nitrate removal at the reach scale. Exposure times versus residence times, Water Resour. Res. 55 (11), 9808–9825, https://doi.org/10.1029/2019WR025540, 2019.

Genereux, D. P., Hemond, H. F., Mulholland, P. J.: Spatial and temporal variability in streamflow generation on the West Fork of Walker Branch Watershed, J Hydrol 142 (1-4), 137–166, https://doi.org/10.1016/0022-1694(93)90009-X, 1993.

Gilfedder, B. S., Cartwright, I., Hofmann, H., Frei, S.: 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, 2019.

Glaser, C., Schwientek, Marc, Junginger, T., Gilfedder, B. S., Frei, S., Werneburg, M., Zwiener, C., Zarfl, C.: Comparison of environmental tracers including organic micropollutants as groundwater exfiltration indicators into a small river of a karstic catchment, Hydrol. Process. 34 (24), 4712–4726, https://doi.org/10.1002/hyp.13909, 2020.

Gooseff, M. N., Payn, R. A., Zarnetske, J. P., Bowden, W. B., McNamara, J. P., Bradford, J. H.: Comparison of in-channel mobile–immobile zone exchange during instantaneous and constant rate stream tracer additions: Implications for design and interpretation of non-conservative tracer experiments, J Hydrol 357 (1-2), 112–124, https://doi.org/10.1016/j.jhydrol.2008.05.006, 2008.

Harvey, J. W., Bencala, K. E.: The effect of streambed topography on surface-subsurface water exchange in mountain catchments, Water Resour. Res. 29 (1), 89–98, https://doi.org/10.1029/92WR01960, 1993.

Harvey, J. W., Wagner, B. J., Bencala, K. E.: Evaluating the reliability of the stream tracer approach to characterize stream-subsurface water exchange, Water Resour. Res. 32 (8), 2441–2451, https://doi.org/10.1029/96WR01268, 1996.

Harvey, J., Gooseff, M.: River corridor science: Hydrologic exchange and ecological consequences from bedforms to basins, Water Resour. Res. 51 (9), 6893–6922, https://doi.org/10.1002/2015WR017617, 2015.

Helton, J. C., Davis, F. J.: Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, Reliab. Eng. Syst. Saf 81 (1), 23–69, https://doi.org/10.1016/S0951-8320(03)00058-9, 2003.

Jackson, T. R., Haggerty, R., Apte, S. V.: A fluid-mechanics based classification scheme for surface transient storage in riverine environments: quantitatively separating surface from hyporheic transient storage, Hydrol. Earth Syst. Sci. 17 (7), 2747–2779, https://doi.org/10.5194/hess-17-2747-2013, 2013.

Jähne, B., Münnich, K. O., Bösinger, R., Dutzi, A., Huber, W., Libner, P.: On the parameters influencing air-water gas exchange, J. Geophys. Res. 92 (C2), 1937, https://doi.org/10.1029/jc092ic02p01937, 1987.

Jimenez-Fernandez, O., Schwientek, M., Osenbrück, K., Glaser, C., Schmidt, C., Fleckenstein, J. H.: Groundwater-surface water exchange as key control for instream and groundwater nitrate concentrations along a first-order agricultural stream, Hydrol. Process. 36 (2), https://doi.org/10.1002/hyp.14507, 2022.

Katz, S. B., Segura, C., Warren, D. R.: The influence of channel bed disturbance on benthic Chlorophyll a: A high resolution perspective, Geomorphology 305, 141–153, https://doi.org/10.1016/j.geomorph.2017.11.010, 2018.

Kelleher, C., Wagener, T., McGlynn, B., Ward, A. S., Gooseff, M. N., Payn, R. A.: Identifiability of transient storage model parameters along a mountain stream, Water Resour. Res. 49 (9), 5290–5306, https://doi.org/10.1002/wrcr.20413, 2013.

Kelleher, C., Ward, A., Knapp, J. L. A., Blaen, P. J., Kurz, M. J., Drummond, J. D., Zarnetske, J. P., Hannah, D. M., Mendoza-Lera, C., Schmadel, N. M., Datry, T., Lewandowski, J., Milner, A. M., Krause, S.: Exploring tracer information and model framework trade-offs to improve estimation of stream transient storage processes, Water Resour. Res. 55 (4), 3481–3501, https://doi.org/10.1029/2018WR023585, 2019.

Kilpatrick, F. A., Cobb, E. D.: Measurement of discharge using tracers: U.S. Geological survey techniques of water-resources investigations, book 3, chap A16, 52, https://pubs.usgs.gov/twri/twri3-a16/, 1985.

Knapp, J. L. A., Kelleher, C.: A Perspective on the future of transient storage modeling: Let's stop chasing our tails, Water Resour. Res. 56 (3), https://doi.org/10.1029/2019WR026257, 2020.

Krause, S., Blume, T.: Impact of seasonal variability and monitoring mode on the adequacy of fiber-optic distributed temperature sensing at aquifer-river interfaces, Water Resour. Res. 49 (5), 2408-2423, https://doi.org/10.1002/wrcr.20232, 2012.

Krishnaswami, S., Graustein, William C., Turekian, Karl K., Dowd, John F.: Radium, thorium and radioactive lead isotopes in groundwaters: Application to the in situ determination of adsorption-desorption rate constants and retardation factors, Water Resour. Res. 18 (6), 1663–1675, https://doi.org/10.1029/WR018i006p01663, 1982.

Lamontagne, S., Cook, P. G.: Estimation of hyporheic water residence time in situ using ^{222}Rn disequilibrium, Limnol. Oceanogr. Methods 5 (11), 407–416, https://doi.org/10.4319/lom.2007.5.407, 2007.

Lee, J.-M., Kim, G.: A simple and rapid method for analyzing radon in coastal and ground waters using a radon-in-air monitor, J Environ Radioact 89 (3), 219–228, https://doi.org/10.1016/j.jenvrad.2006.05.006, 2006.

Loritz, R., Gupta, H., Jackisch, C., Westhoff, M., Kleidon, A., Ehret, U., Zehe, E.: On the dynamic nature of hydrological similarity, Hydrol. Earth Syst. Sci. 22 (7), 3663 – 3684, https://doi.org/10.5194/hess-22-3663-2018, 2018.

Mallard, J., McGlynn, B., Covino, T.: Lateral inflows, stream-groundwater exchange, and network geometry influence stream water composition, Water Resour. Res. 50 (6), 4603–4623, https://doi.org/10.1002/2013WR014944, 2014.

McCallum, J. L., Cook, P. G., Berhane, D., Rumpf, C., McMahon, G. A.: Quantifying groundwater flows to streams using differential flow gaugings and water chemistry, J Hydrol 416-417, 118–132, https://doi.org/10.1016/j.jhydrol.2011.11.040, 2012.

Milhous, R.: Sediment transport in a gravel bottomed stream. PhD. Oregon State University (232 pp.), 1973.

Nearing, G. S., Gupta, H. V.: The quantity and quality of information in hydrologic models, Water Resour. Res. 51(1). 524-538, https://doi.org/10.1002/2014WR015895, 2015.

Neilson, B. T., Chapra, S. C., Stevens, D. K., Bandaragoda, C.: Two-zone transient storage modeling using temperature and solute data with multiobjective calibration: 1. Temperature, Water Resour. Res. 46(12), 2009WR008756, https://doi.org/10.1029/2009WR008756, 2010a.

Neilson, B. T., Stevens, D. K., Chapra, S. C., Bandaragoda, C.: Two-zone transient storage modeling using temperature and solute data with multiobjective calibration: 2. Temperature and solute, Water Resour. Res. 46(12), https://doi.org/10.1029/2009WR008759, 2010b.

Nordin, C. F., Troutman, B. M.: Longitudinal dispersion in rivers: The persistence of skewness in observed data, Water Resour. Res. 16 (1), 123–128, https://doi.org/10.1029/WR016i001p00123, 1980.

Ouyang, S., Puhlmann, H., Wang, S., Wilpert, K. von; Sun, O. J.: Parameter uncertainty and identifiability of a conceptual semi-distributed model to simulate hydrological processes in a small headwater catchment in Northwest China, Ecol Process 3 (1), https://doi.org/10.1186/s13717-014-0014-9, 2014.

Payn, R. A., Gooseff, M. N., McGlynn, B. L., Bencala, K. E., Wondzell, S. M.: Channel water balance and exchange with subsurface flow along a mountain headwater stream in Montana, United States, Water Resour. Res. 45 (11), 1470, https://doi.org/10.1029/2008WR007644, 2009.

Peel, M., Kipfer, R., Hunkeler, D., Brunner, P.: Variable ^{222}Rn emanation rates in an alluvial aquifer: Limits on using ^{222}Rn as a tracer of surface water – Groundwater interactions, Chem Geol 599, 120829, https://doi.org/10.1016/j.chemgeo.2022.120829, 2022.

Pittroff, M., Frei, S., Gilfedder, B. S.: Quantifying nitrate and oxygen reduction rates in the hyporheic zone using ^{222}Rn to upscale biogeochemical turnover in rivers, Water Resour. Res. 53 (1), 563–579, https://doi.org/10.1002/2016WR018917, 2017.

Raymond, P. A., Zappa, C. J., Butman, D., Bott, T. L., Potter, J., Mulholland, P., Laursen, A. E., McDowell, W. H., Newbold, D.: Scaling the gas transfer velocity and hydraulic geometry in streams and small rivers, Limnol. Oceanogr. 2 (1), 41–53, https://doi.org/10.1215/21573689-1597669, 2012.

Rodriguez, N. B., Pfister, L., Zehe, E., Klaus, J.: A comparison of catchment travel times and storage deduced from deuterium and tritium tracers using StorAge Selection functions, Hydrol. Earth Syst. Sci. 25 (1), 401–428, https://doi.org/10.5194/hess-25-401-2021, 2021.

Runkel, L. R.: One dimensional transport with inflow and storage (OTIS): A solute transport model for streams and rivers. US Geol. Surv. Water Resour. Invest. Rep. 98-4018, University of Michigan Library, Denver, https://doi.org/10.1002/wrcr.20277#wrcr20277-bib-0038, 1998.

Runkel, R.L.: A new metric for determining the importance of transient storage, J. North American Benthological Society, 21(4), 529–543, 10.2307/1468428, 2002.

Runkel, R.L., McKnight, D.M., and Andrews, E.D.: Analysis of transient storage subject to unsteady flow: Diel flow variation in an Antarctic stream, J North American Benthological Society, 17(2), 143–154, 10.2307/1467958, 1998.

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

Schmadel, N. M., Neilson, B. T., Stevens, D. K.: Approaches to estimate uncertainty in longitudinal channel water balances, J Hydrol 394 (3-4), 357–369, https://doi.org/10.1016/j.jhydrol.2010.09.011, 2010.

Schmadel, N. M., Neilson, B. T., Kasahara, T.: Deducing the spatial variability of exchange within a longitudinal channel water balance, Hydrol. Process. 28 (7), 3088 – 3103, https://doi.org./ 10.1002/hyp.9854, 2014.

Schubert, M., Knoeller, K., Mueller, C., Gilfedder, B.: Investigating river water/groundwater interaction along a rivulet section by ^{222}Rn mass balancing, Water 12 (11), 3027, https://doi.org/10.3390/w12113027, 2020.

Somers, L. D.; Gordon, R. P.; McKenzie, J. M.; Lautz, L. K.; Wigmore, O., Glose, A. M., Glaser, R., Aubry-Wake, C., Mark, B., Baraer, M., Condom, T..: Quantifying groundwater–surface water interactions in a proglacial valley, Cordillera Blanca, Peru, Hydrol. Process. 30 (17), 2915–2929, https://doi.org/10.1002/hyp.10912, 2016.

Sophocleous, M.: Interactions between groundwater and surface water. The state of the science, Hydrogeol J 10 (1), 52–67, https://doi.org/10.1007/s10040-001-0170-8, 2002.

Stanford, J. A., Ward, J. V.: An ecosystem perspective of alluvial rivers: Connectivity and the hyporheic corridor, J N Am Benthol Soc 12 (1), 48–60, https://doi.org/10.2307/1467685, 1993.

Stream Solute Workshop: Concepts and methods for assessing solute dynamics in stream ecosystems, JNABS 9 (2), 95–119, https://doi.org/10.2307/1467445, 1990.

Taylor, G.: Diffusion by continuous movements, P Lond. Math. Soc 2-20 (1), 196–212, https://doi.org/10.1112/plms/s2-20.1.196, 1922.

Taylor, G.: The dispersion of matter in turbulent flow through a pipe, Proc. R. Soc. Lond. A, 223, 446–468, https://doi.org/10.1098/rspa.1954.0130, 1954.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., Gupta, H. V.: Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis, Hydrol. Process. 17 (2), 455–476, https://doi.org/10.1002/hyp.1135, 2003.

Wagener, T., Camacho, L. A., Wheater, H. S.: Dynamic identifiability analysis of the transient storage model for solute transport in rivers, J Hydroinf 4 (3), 199–211, https://doi.org/10.2166/hydro.2002.0019, 2002.

Wagner, B. J., Harvey, J. W.: Experimental design for estimating parameters of rate-limited mass transfer: Analysis of stream tracer studies, Water Resour. Res. 33 (7), 1731–1741, https://doi.org/10.1029/97WR01067, 1997.

Wang, S., Hrachowitz, M., Schoups, G., Stumpp, C.: Stable water isotopes and tritium tracers tell the same tale: no evidence for underestimation of catchment transit times inferred by stable isotopes in StorAge Selection (SAS)-function models, Hydrol. Earth Syst. Sci. 27 (16), 3083–3114, https://doi.org/10.5194/hess-27-3083-2023, 2023.

Ward, A. S., Payn, R. A., Gooseff, M. N., McGlynn, B. L., Bencala, K. E., Kelleher, C. A., Wondzell, S. A., Wagener, T.: Variations in surface water-ground water interactions along a headwater mountain stream: Comparisons between transient storage and water balance analyses, Water Resour. Res. 49 (6), 3359–3374, https://doi.org/10.1002/wrcr.20148, 2013.

Ward, A. S., Kelleher, C. A., Mason, S. J. K., Wagener, T., McIntyre, N., McGlynn, B., Runkel, R. L., Payn, R. A.: A software tool to assess uncertainty in transient-storage model parameters using Monte Carlo simulations, Freshwater Science 36 (1), 195–217, https://doi.org/10.1086/690444, 2017.

Ward, A. S., Packman, A. I.: Advancing our predictive understanding of river corridor exchange, WIREs Water 6 (1), https://doi.org/10.1002/wat2.1327, 2019.

Ward, A. S., Wondzell, S. M., Gooseff, M. N., Covino, T., Herzog, S., McGlynn, B., Payn, R. A.: Breaking the Window of Detection: Using multi-scale solute tracer studies to assess mass recovery at the detection limit, Water Resour. Res. 59 (3), e2022WR032736, https://doi.org/10.1029/2022WR032736, 2023.

Wlostowski, A. N., Gooseff, M. N.; Wagener, Thorsten: Influence of constant rate versus slug injection experiment type on parameter identifiability in a 1-D transient storage model for stream solute transport, Water Resour. Res. 49 (2), 1184–1188, https://doi.org/10.1002/wrcr.20103, 2013.

Wu, L., Gomez-Velez, J. D., Krause, S., Singh, T., Wörman, A., Lewandowski, J.: Impact of flow alteration and temperature variability on hyporheic exchange, Water Resour. Res. 56 (3), e2019WR026225, https://doi.org/10.1029/2019WR026225, 2020.