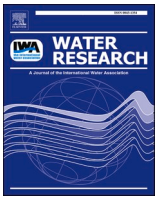
Water Research 235 (2023) 119880

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres



Exploring the reliability of ^{222}Rn as a tracer of groundwater age in alluvial aquifers: Insights from the explicit simulation of variable ^{222}Rn production

Morgan Peel [a,*], Hugo Delottier [a], Rolf Kipfer [b,c], Daniel Hunkeler [a], Philip Brunner [a]

[a] Centre d'hydrogéologie et de géothermie (CHYN), Université de Neuchâtel, Emile Argand 11, Neuchâtel 2000, Switzerland
[b] Department Water Resources and Drinking Water, Eawag, Überlandstrasse 133, Dübendorf 8600, Switzerland
[c] Institute of Biogeochemistry and Pollutant Dynamics and Institute for Geochemistry and Petrology, ETH Zürich, Zürich 8092, Switzerland

A B S T R A C T

Knowledge of groundwater residence times (GRT; the time elapsed since surface water infiltration) between losing rivers and pumping wells is crucial for management of water resources in alluvial aquifers. The radioactive noble gas radon-222 (^{222}Rn) has been used for decades as a natural indicator of surface water infiltration, as it can provide quantitative information on GRT. However, models using ^{222}Rn as a tracer of GRT are often based on a set of highly simplifying assumptions, including spatially homogenous ^{222}Rn production and exclusively advective mass transport within the aquifer. In this paper, we use the integrated surface-subsurface hydrological model HydroGeoSphere (HGS) to simulate ^{222}Rn transport, production, and decay in a bank filtration context. Spatially variable ^{222}Rn production, based on experimental data, is explicitly considered. We show that variable ^{222}Rn production rates, coupled with hydrodispersive mixing of groundwater, may lead to large biases in GRT estimates. Under certain transient conditions however, changes in tracer-derived GRTs correlate well with changes in mean groundwater age. Whereas ^{222}Rn-derived GRTs may only be reliable under a narrow range of field conditions, ^{222}Rn may serve as a powerful tracer of changes in mean GRT even in complex and heterogenous environments.

1. Introduction

Unconsolidated alluvial aquifers are often targeted for drinking water production given their high productivity and the convenience of shallow groundwater (GW) exploitation (Hiscock and Grischek, 2002; Margat and van der Gun, 2013). These systems can be described as natural water-treatment sites, as induced bank filtration, or the pumping of GW near a surface water (SW) body, is known to improve water quality through a range of chemical, biological and physical processes during underground flow (Hiscock and Grischek, 2002; Sprenger et al., 2011). Hydraulic gradients induced by pumping wells placed near streams coupled to high transmissivities can lead to significant infiltration fluxes and GW flow velocities; therefore, a large fraction of the abstracted water is composed of freshly infiltrated surface water with residence times of days to weeks (Frei and Gilfedder, 2021; Tufenkji et al., 2002). Knowledge of both the origin and the residence time (i.e. time from infiltration) of pumped water is critical for sustainable water supply, and is required to definition of capture zones, assessment of GW resources sustainability, and susceptibility to contamination and pollution.

In such contexts, environmental and/or artificial tracers provide observable information on sources, pathways, and travel times of GW, as well as groundwater-surface water (GW-SW) mixing ratios (Brunner et al., 2017; Cook and Herczeg, 2000). The radioactive noble gas radon-222 (^{222}Rn) is one of few environmental tracers that is sensitive to processes of GW-SW exchange at timescales of hours to ~10–15 days, which are often those most relevant for water management in alluvial aquifer contexts. This environmental gas tracer, which is naturally produced in the subsurface, is a powerful indicator of GW-SW interactions, as concentrations in GW are usually much higher (often orders of magnitude) than those in SW. This tracer has been extensively used over the last three decades to characterize SW infiltration into adjacent alluvial aquifers in terms of infiltration fluxes and GW residence times (e.g. Bertin and Bourg 1994, Frei and Gilfedder 2021, Hoehn and Von Gunten 1989, Popp et al. 2021, Stellato et al. 2013, Vogt et al. 2010).

Estimation of fluxes and residence times of bank filtrate with ^{222}Rn usually relies on the apparent age model of Hoehn and Von Gunten (1989), which is based on the definition of a single ^{222}Rn end-member activity representative of GW at radioactive equilibrium with the aquifer matrix. This is usually undertaken by sampling GW at a location where ^{222}Rn activities are expected to be at secular equilibrium with the aquifer matrix, or by measuring ^{222}Rn production rates directly from recovered sediment samples. However, the definition of a single ^{222}Rn

* Corresponding author.
E-mail address: morgan.peel@unine.ch (M. Peel).

https://doi.org/10.1016/j.watres.2023.119880
Received 30 September 2022; Received in revised form 9 March 2023; Accepted 14 March 2023
Available online 15 March 2023

end-member activity is only meaningful if ^{222}Rn production rates are spatially homogenous at the scale of investigation. This assumption contrasts with results from several field studies, which have shown that ^{222}Rn production rates can vary considerably, sometimes by over one order of magnitude, at small scales within alluvial aquifers (e.g., Mullinger et al. 2009, Peel et al. 2022, Schaper et al. 2022).

Recent studies put forward the potential of explicitly simulating ^{222}Rn activities with mass transport models rather than relying on simple apparent age (or implicit) models (e.g., Gilfedder et al. 2019, Peel et al. 2022, Schaper et al. 2022). The latter even recommend avoiding the apparent ^{222}Rn age model if field measurements suggest "substantial" variations in ^{222}Rn production rates. This follows a more general trend in hydrological sciences in which the explicit simulation of tracer concentrations in mass transport models is considered to extract the information content of tracer measurements (e.g. Schilling et al. 2019, Thiros et al. 2021, Turnadge and Smerdon 2014).

However, the explicit simulation of tracer concentrations might not be suited when the time or the ability to incorporate tracer measurements into mass transport models coupled with sophisticated calibration approaches is not available. In such conditions, simple mathematical models (e.g. apparent age models or lumped-parameter models) are usually preferred. It is therefore critical to understand how site- and tracer-specific properties (e.g. aquifer hydraulic properties, transient hydraulic conditions, spatiotemporal variations in tracer sources/sinks, etc..) may lead to bias in tracer-derived apparent GW ages. Systematic exploration of the performance of the apparent ^{222}Rn age model simultaneously accounting for different aquifer properties, variable hydraulic conditions and spatially heterogenous ^{222}Rn production has not yet been undertaken, although it is of clear interest for water resources management. Accordingly, the goal of the present study is to test the sensitivity of apparent (or radiometric) ^{222}Rn ages to (i) aquifer hydraulic and transport parameters, (ii) transient hydraulic conditions, and (iii) the magnitude and spatial scale of variability of ^{222}Rn production into GW. We quantify the bias between mean and apparent GW ages and illustrate some limits of the apparent ^{222}Rn age model. We also identify cases in which ^{222}Rn can provide reliable information on GW residence time without having to resort to complex numerical models. To this end, we constructed a synthetic 2-D mass transport model explicitly simulating mean GW age and dissolved ^{222}Rn, with both spatially homogenous and heterogenous ^{222}Rn production. The model simulates subsurface flow and transport along a transect downgradient of a losing river, with variable ^{222}Rn production rates based on high-resolution sediment incubation experiments (Peel et al., 2022). Specifically, we seek to disentangle the relative importance of different sources of apparent age bias arising from (i) mixing of water of different ages, (ii) transient hydraulic conditions, and (iii) variable ^{222}Rn production.

1.1. ^{222}Rn as a tracer of GW age

^{222}Rn (half-life ~3.8d) is an intermediary product in the ^{238}U decay-chain, and occurs naturally in GW mainly as a result of alpha-decay of matrix-bound ^{226}Ra (Cecil and Green, 2000). The magnitude of ^{222}Rn production in GW is controlled by several factors, including total ^{226}Ra activity of aquifer material, the location of ^{226}Ra in or around mineral grains, sediment specific surface area, pore geometry, as well as water saturation (Cecil and Green, 2000).

As there is no significant atmospheric source of ^{222}Rn, most surface water bodies exhibit ^{222}Rn activities generally much lower (often orders of magnitude) than those measured in GW (Cecil and Green, 2000). Only in and downstream of strongly gaining river reaches do ^{222}Rn activities in streams reach significant levels (Cartwright and Gilfedder, 2015).

1.2. ^{222}Rn dating method

GW dating with ^{222}Rn is based on gradual increase of ^{222}Rn activities in recently infiltrated water as a function of residence time (i.e. time since infiltration). Assuming spatially constant production of ^{222}Rn in GW, the apparent age of a water sample a_{app} [T] can be computed with the following equation (Cranswick et al., 2014; Hoehn and Von Gunten, 1989):

$$a_{app} = -\frac{1}{\lambda_{Rn}} ln\left(\frac{A_{Rn,eq} - A_{Rn,meas}}{A_{Rn,eq} - A_{Rn,0}}\right) \quad (1)$$

Where $A_{Rn,meas}$ [$T^{-1}L^{-3}$] is the activity of ^{222}Rn in a given GW sample, $A_{Rn,eq}$ [$T^{-1}L^{-3}$] is the activity of ^{222}Rn at equilibrium with the aquifer matrix (^{222}Rn end-member activity), $A_{Rn,0}$ [$T^{-1}L^{-3}$] is the ^{222}Rn activity of surface water at the time of infiltration, and λ_{Rn} [T^{-1}] is the ^{222}Rn decay constant (~0.18 d^{-1}). The equilibrium ^{222}Rn concentration of GW $A_{Rn,eq}$ is the ratio of the aquifer ^{222}Rn production rate γ_{Rn} [$T^{-2}L^{-3}$] and the ^{222}Rn decay constant:

$$A_{Rn,eq} = \frac{\gamma_{Rn}}{\lambda_{Rn}} \quad (2)$$

In practice, GW may be considered at equilibrium after approximately four to five half-lives of ^{222}Rn (~15–20 days). Eq. (1) has regularly been used to date young (i.e. < 15–20 days) GW in bank filtration contexts (e.g. Cranswick et al. 2014, Frei and Gilfedder 2021, Pittroff et al. 2017, Schilling et al. 2017). As is the case for most tracer-based apparent age models, Eq. (1) is only valid in the restrictive case of 1-D piston flow, meaning if there is no mixing of water of different ages in a GW sample (e.g. due to intra-aquifer mixing or sampling of multiple flowlines). Moreover, Eq. (1) also relies on the definition of a single ^{222}Rn end-member activity ($A_{Rn,eq}$) representative of GW at equilibrium with the host aquifer matrix, which inherently supposes homogenous ^{222}Rn production (γ_{Rn}) at the scale of investigation. Therefore, in cases where ^{222}Rn production is not homogenous, definition of a single or a range of ^{222}Rn end-member(s) for the computation of apparent GW ages is not trivial and is limited by the fragmentary knowledge of the distribution of ^{222}Rn activities and production rates at a given study site (e.g. Peel et al. 2022).

1.2.1. Apparent age bias and GW mixing

When tracer concentrations vary non-linearly with respect to time, any process that leads to mixing and spreading of a solute will systematically result in biases of apparent ages relative to mean ages McCallum et al., 2013; Varni and Carrera, 1998; Waugh et al., 2003). On top of spatially heterogenous production in the case of ^{222}Rn, sources of apparent age bias include intra-aquifer mixing due to dispersion and diffusion, as well as mixing of waters following different flow paths in piezometers or wells with long screens (Bethke and Johnson, 2008; Manning et al., 2005). Even in the simple case of 1D, homogenous steady-state flow along a hypothetical flowline, hydrodynamic dispersion will lead to mixing of water of different ages, and any GW sample will exhibit at least an inverse-Gaussian distribution of ages, with variance proportional to the dispersion coefficient (Ginn et al., 2009). Assuming homogenous ^{222}Rn production, the relationship between mean and apparent ^{222}Rn ages in such 1-D systems is given by the following functional relationship (derived from Massoudieh and Ginn 2011, Eq. ((7)):

$$\frac{a_{app}(\mathbf{x})}{a_{mean}(\mathbf{x})} = \frac{v_0}{2\lambda_{Rn}\alpha_L}\left(\sqrt{1 + \frac{4\alpha_L\lambda_{Rn}}{v_0}} - 1\right) \quad (3)$$

Where $a_{app}(\mathbf{x})$ [T] and $a_{mean}(\mathbf{x})$ [T] are apparent and mean ages at sampling location $\mathbf{x}$ [L], v_0 [LT^{-1}] the steady state GW flow velocity (or Darcy velocity), and α_L [L] the longitudinal dispersivity. Under such conditions, the apparent age bias is a function of the ratio $v_0/\lambda_{Rn}\alpha_L$, akin to the Péclet number, and is independent of the magnitude of the ^{222}Rn production rate γ_{Rn}. This relationship highlights how apparent ^{222}Rn ages are always smaller than mean ages in the case of homogenous ^{222}Rn production. Only in the extreme case of purely advective transport (high Péclet number) or minute tracer decay both ages are equal. In more

general terms, when the *rate* of tracer accumulation decreases with increasing residence time (such as in the case of ^{222}Rn), mixing of waters of different ages will lead to an apparent age younger than the mean age (McCallum et al., 2013).

2. Materials and methods

2.1. Synthetic 2-D numerical model

A generic 2-D mass transport model was constructed to analyse the effects of aquifer hydraulic and mass transport parameters, transient hydraulic conditions, and spatially variable ^{222}Rn production rates on the distribution of ^{222}Rn activities and apparent ^{222}Rn ages in the subsurface. The synthetic model simulates infiltration of river-water into an aquifer, and GW flow and transport in variably-saturated conditions. The numerical code HydroGeoSphere (Aquanty, 2023) was used, as it is capable of simulating variably-saturated flow and solute transport with both source and decay terms, corresponding respectively to ^{222}Rn ingrowth and disintegration. The scope of the 2-D model is to explicitly simulate ^{222}Rn signatures in groundwater along multiple flowlines downgradient from the site of infiltration (i.e. a stream or river) in both steady-state and transient hydraulic conditions.

Model dimensions are 400 m x 50 m in the horizontal (X) and vertical (Z) directions respectively. A channel representing a simplified river-bed and bank was included at the top-left of the model domain (channel depth: 4 m; width: 10 m; bank slope: 30°). The model was discretized into approximately 40′000 triangular elements, with element edges ranging from about 15 cm near the top of the model and stream/aquifer interface to 2.5 m at depth near the model outlet (X = 400 m).

The river was simulated by a constant head boundary condition (BC) along all channel nodes. To ensure continuously losing river conditions (i.e. infiltration of river water into the aquifer), the outlet (X = 400 m) was assigned a constant head at least 2 m lower than that at the inlet. To simulate the effect of transient hydraulic conditions on ^{222}Rn signatures (for example as a consequence of increased GW pumping), a variable-head BC was applied to the model outlet, which simulated two incremental 1-m decreases in hydraulic head followed by a gradual return to initial conditions (see Fig. 1). This variable-head boundary leads to transient 50% and 100% increase in overall hydraulic gradients, and is chosen to highlight the effect of significant variations in GW flow velocity on the distribution of ^{222}Rn throughout the model domain.

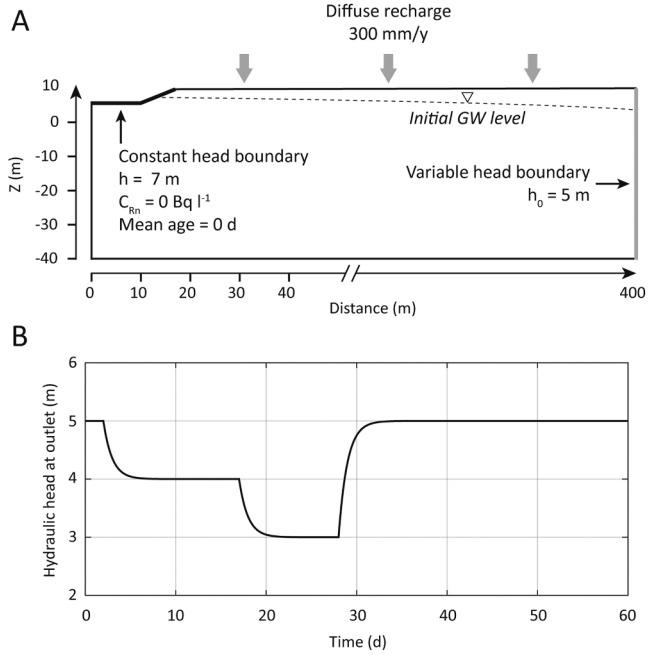


Fig. 1. 2D model setup. A: Geometry and boundary conditions; B: Variable head boundary condition at model outlet.

No-flow BCs were applied to the left (X = 0 m) and bottom (Z = -40 m) boundaries. The no-flow condition at the X = 0 boundary represents a hypothetical water divide, whereas the lower boundary represents the lower limit of the alluvial aquifer. A diffuse and spatially constant recharge of 300 mm/y was applied to the top boundary.

The effect of GW flow velocity and aquifer dispersivity on ^{222}Rn signatures in GW was explored by simulating various combinations of homogenous and isotropic hydraulic conductivity (K [L T^{-1}]) and longitudinal and transverse vertical dispersivity (α_L, α_{TV} [L]). Three spatially homogenous values of K (100, 250 and 500 m/d) and two of longitudinal dispersivity α_L (5 and 20 m) were selected. Transverse vertical dispersivity was set to 10% of longitudinal dispersivity. The values of K are representative of sand and gravel aquifers (Bear, 1972), while those of dispersivity are consistent with measurement scales between 1 and 1000 m (Gelhar et al., 1992; Schulze-Makuch, 2005). No clogging layer was considered at the stream/aquifer interface, and the hydraulic conductivity was the same as within the aquifer. The van Genuchten – Mualem model was used to simulate variably-saturated flow with parameters α and β set to 4.1 m^{-1} and 2.2 (-) respectively. A constant porosity 0.2 was assumed for all simulations, which is representative of unconsolidated sandy gravel aquifers (e.g. Fetter 2001).

The solute transport model simulated both ^{222}Rn concentrations and mean GW age. Mean age was simulated by applying a zero-age BC at the inlet and top boundaries, and all elements were assigned an age source term simulating a theoretical tracer with a growth rate of unity in the subsurface (i.e., 1 day per day, Goode (1996). Mean GW age is therefore directly simulated through the advection-dispersion transport equation, and treated as a conservative solute with zero age at model inlets, and a subsurface ageing term γ_{Age} [-] of 1 day per day (Goode, 1996).

A zero-concentration BC was applied to the inlet and top boundaries for ^{222}Rn. This boundary condition reflects the fact that ^{222}Rn activities in surface waters tend to be negligible compared to those in groundwater, as there are no atmospheric sources of ^{222}Rn, and any dissolved ^{222}Rn is quickly lost to the atmosphere through gas transfer (Cecil and Green, 2000). ^{222}Rn production by the aquifer matrix was simulated by applying a (spatially-variable) zero-order source BC to all model elements. Note that ^{222}Rn production and emanation rates are related by the following expression (e.g. Cook et al. (2006), Eq. (14)):

$$\gamma_{Rn} = E_m \rho_{grain} \lambda_{Rn} \frac{1-\varphi}{\varphi} \tag{4}$$

Where E_m [Bq kg^{-1}] is the ^{222}Rn emanation rate, ρ_{grain} [kg l^{-1}] is the density of the mineral phase, and φ [-] is the sediment porosity. A homogenous grain density of 2.65 kg l^{-1} was assumed. ^{222}Rn production rates in the unsaturated zone were scaled to account for variable water content and air/water partitioning of ^{222}Rn, as described in Delottier et al. (2022). ^{222}Rn production rates in the unsaturated zone are effectively decreased by simulating instant equilibrium of produced ^{222}Rn between air and water phases according to ^{222}Rn solubility and water saturation (Delottier et al., (2022); Eq. (7)):

$$\gamma_{Rn,eff}(S_w) = S_w \frac{H^{cc} \gamma_{Rn,sat}}{H^{cc} S_w + [1 - S_w]} \tag{5}$$

Where $\gamma_{Rn,eff}(S_w)$ [$T^{-2}L^{-3}$] is the effective (or scaled) ^{222}Rn production rate in water as a function of water saturation S_w [-], $\gamma_{Rn,sat}$ [$T^{-2}L^{-3}$] is the ^{222}Rn production rate in saturated conditions (i.e., S_w=1), and H^{cc} [-] is the dimensionless air-water partitioning coefficient of ^{222}Rn $\left(H^{cc} = \frac{C_{water}}{C_{air}}\right)$. A constant air-water partitioning coefficient equal to that at 10 °C was used for ^{222}Rn ($H^{cc} \approx 0.35$), calculated after Weigel (1978). Free-solution diffusion coefficients for both ^{222}Rn and mean age were set to 10^{-9} m^2/s (Goode, 1996; Ishimori et al., 2013). For each

transient simulation, initial ^{222}Rn concentrations were determined from steady-state conditions with outlet hydraulic head $h_0 = 5$ m. Finally, the transient model was run for a total simulation time of 60 days (Fig. 1B).

^{222}Rn production rates are based on a high-resolution dataset of ^{222}Rn emanation rates in a bank filtration setting (Peel et al., 2022). These production rates provide a realistic basis upon which a variety of spatial distributions of ^{222}Rn production could be constructed. A simplified geostatistical framework was employed to create spatially-variable distributions of ^{222}Rn production rates in the 2-D model. Three synthetic variogram models, representing small-, intermediate- and large-scale variability of ^{222}Rn production, were employed to capture a substantial range of spatial distributions of ^{222}Rn production rates. For all scenarios, an exponential variogram model was used, with ranges from 20 × 10 m up to 200 × 30 m in the X and Z directions respectively (see Table 1). Synthetic conditioning data, based on the statistical properties of measured ^{222}Rn emanation rates, were placed at regular intervals within the model grid with spacing equal to half the variogram ranges in both the X and Z directions. For the three variogram models, fifty sets of synthetic conditioning data were created. For each of these datasets, ten equiprobable realizations of the spatial distribution of emanation rates (total 500 realizations per variogram model) were generated with a Sequential Gaussian Simulation (SGS) algorithm with the GEONE python software package (Straubhaar, 2020). A full description of the geostatistical workflow can be found in Appendix A. This methodology led to a coefficient of variation (CV) of modelled ^{222}Rn production rates of over 50% for each realization, with a mean ^{222}Rn production rate (± 1-σ) of 3.0 ± 1.6 Bq l^{-1} s^{-1} (see Table 1).

2.2. Quantification of apparent age bias

The bias – or relative deviation - between apparent ^{222}Rn and mean ages at a given location can be defined as:

$$bias(\mathbf{x},t) = \frac{a_{app}(\mathbf{x},t) - a_{mean}(\mathbf{x},t)}{a_{mean}(\mathbf{x},t)} \quad (6)$$

Where $a_{mean}(\mathbf{x},t)$ [T] and $a_{app}(\mathbf{x},t)$ [T] are the mean (directly simulated) and apparent (computed from simulated ^{222}Rn activities) ages of groundwater at sampling location $\mathbf{x}$ and simulation time t. In the considered 2-D system, bias can arise from (i) mixing of water of different ages (hydrodynamic dispersion) and (ii) variable ^{222}Rn production rates (γ_{Rn}) i.e., a non-unique ^{222}Rn end-member activity ($A_{Rn,eq}$).

Table 1
Aquifer and geostatistical parameters used for the 2-D flow and transport model. An exponential variogram model representing the spatial correlation of ^{222}Rn production rates was selected for all cases (see Appendix A for details). K: hydraulic conductivity;.α_L: longitudinal dispersivity; φ: porosity; ρ_{grain}: aquifer grain density; λ_{Rn}: ^{222}Rn decay constant; $\gamma_{Rn,sat}$: zero-order source for ^{222}Rn (in saturated conditions); γ_{Age}: unit source term for age; $D_0^{Rn,Age}$: free-solution diffusion coefficient respectively for ^{222}Rn and mean GW age.

	Property	Modelled values/range
Aquifer properties	K (m d^{-1})	100
		250
		500
	α_L (m)	5
		20
	φ (-)	0.2
	ρ_{grain} (g cm^{-3})	2.65
Tracer properties	λ_{Rn} (d^{-1})	0.1814
	$\gamma_{Rn,sat}$ (Bq l^{-1} d^{-1})	3.0 ± 1.6 (1-σ)
	γ_{Age} (d^{-1})	1
	D_0^{Rn} (m^2 s^{-1})	10^{-9}
	D_0^{Age} (m^2 s^{-1})	10^{-9}
Variogram range for ^{222}Rn production	Range (X – Z)	20 m – 10 m
		50 m – 15 m
		200 m – 30 m

Mixing due to sampling of multiple flowlines is not considered in the present study, although this may be an issue in many real-world contexts due to long screens in piezometers.

To disentangle the contributions of mixing and variable ^{222}Rn production to the age bias (Eq. (6)), transient simulations were run for the six combinations of hydraulic conductivity and dispersivity (see Table 1) with a constant value of γ_{Rn} assigned to all model elements. In these scenarios, any age bias results from mixing of GW with different ages. Synthetic linear observation wells were placed throughout the model domain at regular intervals of 2.5, 6.25, and 12.5 m in the horizontal direction for cases where K was equal to respectively 100, 250, and 500 m/d (well placement shown in Appendix B). Each well is placed from the surface to a depth of 20 m. Output including mean GW age and ^{222}Rn activities was produced for all nodes located within 10 cm of these synthetic observation wells. This allowed the calculation of apparent age bias at various locations within the model domain at every timestep for cases where ^{222}Rn production is spatially constant. A generic value of γ_{Rn} equal to porosity (i.e. 0.2) was selected for these simulations. Note that apparent age bias does not depend on the choice of γ_{Rn} in the case of constant ^{222}Rn input, as all simulated ^{222}Rn activities are normalized by the end-member activity $A_{Rn,eq}$ for calculation of apparent age. For the calculation of apparent ages, $A_{Rn,eq}$ was constant and defined as the ratio of γ_{Rn} and λ_{Rn} (Eq. (2)).

Subsequently, simulations were run by assuming spatially variable ^{222}Rn production. The ^{222}Rn apparent age bias, that is the amount of bias attributed solely to the spatial variability of ^{222}Rn input, was defined from Eq. (6) as:

$$bias_{Rn}(\mathbf{x},t) = bias_{\gamma_{Rn,var}}(\mathbf{x},t) - bias_{\gamma_{Rn,cst}}(\mathbf{x},t) = \frac{a_{app}^{\gamma_{Rn,var}}(\mathbf{x},t) - a_{app}^{\gamma_{Rn,cst}}(\mathbf{x},t)}{a_{mean}(\mathbf{x},t)} \quad (7)$$

Where $bias_{\gamma_{Rn,var}}(\mathbf{x},t)$ and $bias_{\gamma_{Rn,cst}}(\mathbf{x},t)$ [-] are the apparent age biases in the case of spatially variable and constant ^{222}Rn production (γ_{Rn}) respectively, and $a_{app}^{\gamma_{Rn,var}}(\mathbf{x},t)$ [T] and $a_{app}^{\gamma_{Rn,cst}}(\mathbf{x},t)$ [T] are similarly the apparent ^{222}Rn ages in the two cases.

The definition of a single ^{222}Rn end-member equilibrium activity for the computation of apparent age is not straightforward in cases where the ^{222}Rn input is spatially variable. Indeed, local changes in production rates will lead to commensurate changes in ^{222}Rn activities, especially for "old" groundwater components with mean ages > 25 days. In real-world settings, groundwater may be sampled at one or several locations where it is expected to be at secular equilibrium with the aquifer matrix. If ^{222}Rn production is spatially variable, measured end-member activities may strongly depend on sampling location. To account for this non-uniqueness in the definition of ^{222}Rn end-member equilibrium activities, and to encompass all possible apparent ages based on simulation results, three ^{222}Rn end-member activities were defined, representing respectively the (1) *minimum,* (2) *mean,* and (3) *maximum* simulated ^{222}Rn concentrations at nodes where the mean age in steady-state flow and transport conditions is > 25 days. These end-member activities therefore represent integrated ^{222}Rn input signals from model inlets to sampling points. For each simulation, end-member activities were computed from simulated ^{222}Rn activities at over 150 nodes where mean GW age was above 25 days.

To simultaneously account for the influence of mass-transport processes (advection-dispersion), the rate of ^{222}Rn accumulation, and the scale of variability of ^{222}Rn production in relation to the sampling scale, we introduce a dimensionless Damköhler-like number N_{Rn}^0:

$$N_{Rn}^0 = \frac{\lambda_{Rn} L^2}{\alpha_L v} \frac{S}{L} \quad (8)$$

Where L [L] is the sampling scale (e.g., the distance from the inlet to the sampling point), S [L] is the scale of variability of ^{222}Rn production rates (e.g., geostatistical correlation length), and v [L T^{-1}] the representative GW flow velocity (Darcy velocity). The first set of terms $\lambda_{Rn}L^2/$

$\alpha_L v$ is the Damköhler number for a reaction with first-order coefficient λ_{Rn} (e.g., Oldham et al. 2013), and represents the relative dispersive mass-transport and radiochemical (i.e., ^{222}Rn accumulation) timescales. The ratio S/L is a measure of the relative timescale a given water sample will have spent in zones of distinct ^{222}Rn production (potentially high or low). Here the radiochemical timescale is given by λ_{Rn} instead of the ^{222}Rn production rates γ_{Rn}, as the rate of ^{222}Rn *accumulation* is one of first order increase, modulated by the ^{222}Rn decay constant. If we assume the ratio L/v approximates the mean age of a water sample, we can simplify Eq. (8):

$$N^0_{Rn} \approx N_{Rn} = a_{mean}\lambda_{Rn}\frac{S}{\alpha_L} \tag{9}$$

The amount of ^{222}Rn bias (Eq. (7)) as a function of N_{Rn} may offer some insight on the combined effects of mass-transport, spatial correlation of ^{222}Rn production rates, and the residence time of GW on the performance of the apparent ^{222}Rn age model. Indeed, low values of N_{Rn} are associated smaller-scale spatial variations in ^{222}Rn production rates (as given by S), lower GW residence times (a_{mean}) and therefore (on average) lower ^{222}Rn activities, and higher smoothing of ^{222}Rn activities due to hydrodynamic dispersion (α_L); resultant spatial distributions of ^{222}Rn activities tend to be smoother, and computed apparent ages less influenced by the asymptotic behaviour of the ^{222}Rn age model when activities approach those of the end-member. Taken together, these effects likely dampen the effect of variable ^{222}Rn production on ^{222}Rn bias. The converse is true for large values of N_{Rn}.

3. Results

3.1. Modelled ^{222}Rn end-member equilibrium activities

As described in Section 2.2, three different ^{222}Rn end-member equilibrium activities were defined for each simulation, representing respectively the (1) *minimum,* (2) *mean,* and (3) *maximum* simulated steady-state ^{222}Rn concentrations at nodes where the mean age is > 25 days. Each of these end-member activities vary from one simulation to the next, depending on the given spatial distribution of ^{222}Rn production rates. The range of end-member ^{222}Rn activities for each of the eighteen illustrative cases described in Section 2.1 is shown in Fig. 2.

For the six combinations of K and α_L modelled, ^{222}Rn end-member activities vary, on average, around a value representative of the mean production rate (~16.5 Bq l^{-1}, see Table 1). This results from the lack of a systematic trend in the spatial distribution of ^{222}Rn production rates (i.e. spatially constant mean production rate within the model domain). However, the spread of ^{222}Rn equilibrium activities, both within a given ^{222}Rn end-member class (*min, mean,* or *max*) and between these end-member activities, depends most strongly on hydraulic conductivity (K), albeit with significant influence of dispersivity (α_L) and the spatial distribution of production rates as given by the variogram models. Indeed, the range of end-member activities and differences between end-member classes are largest at low values of K and α_L, and large variogram ranges. Coefficients of variation of the mean ^{222}Rn end-member activities range from 0.04 for the case $\alpha_L = 20$ m, $K = 500$ m/d and smallest spatial correlation ("Vario 1″) to 0.09 for the case $\alpha_L = 5$ m, $K = 100$ m/d and largest variogram range ("Vario 3″). The ratio of mean *minimum* to *maximum* end-member activities ranges from 1.1 to 2.3 for the same cases.

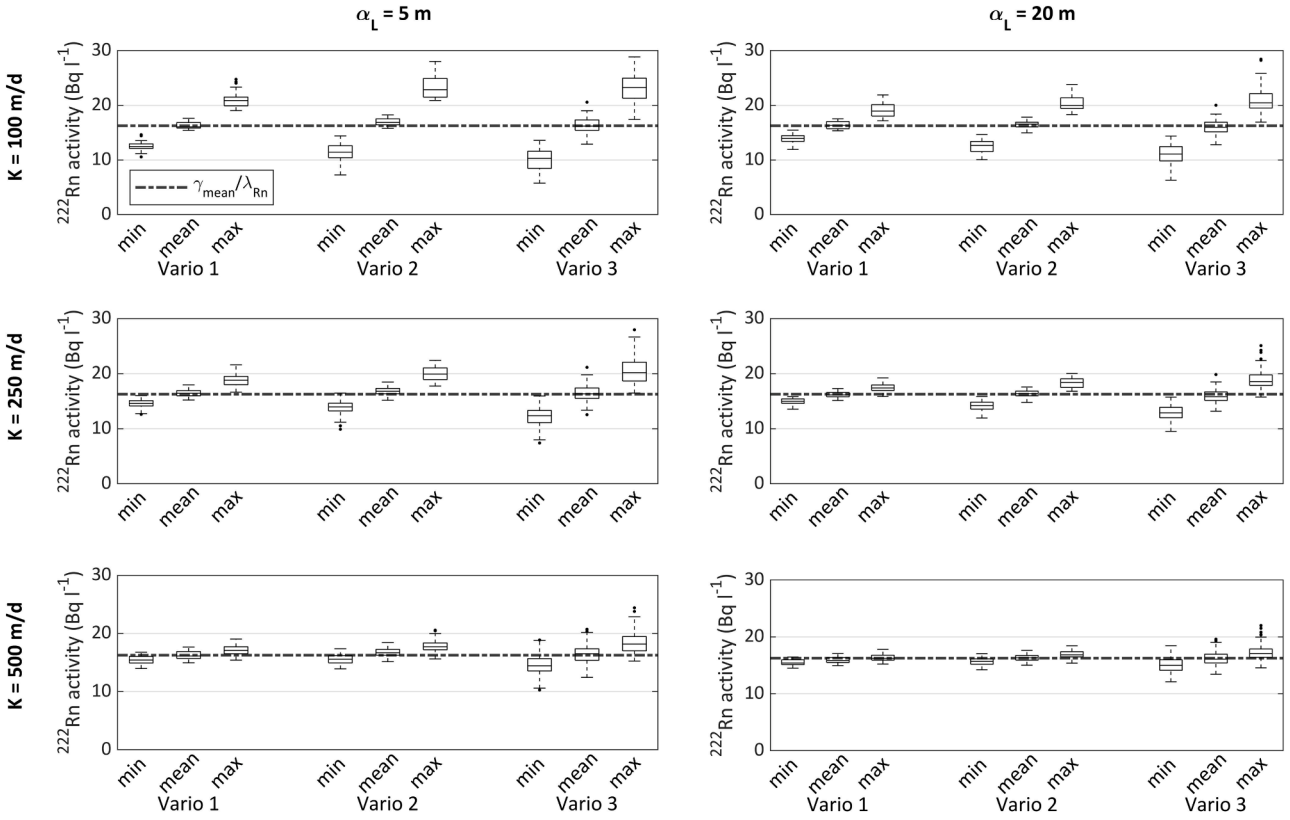


Fig. 2. Boxplots of ^{222}Rn end-member activities for different combinations of hydraulic conductivity (K), longitudinal dispersivity (α_L), and variogram models for the distribution of ^{222}Rn production. Aggregated statistics from all simulations are shown here. Vario 1, 2, and 3 represent variogram models with ranges in the X and Z directions of 20 × 10 m, 50 × 15 m, and 200 × 30 m respectively (see Table 1). The min, mean, and max parameters are the minimum, mean, and maximum end-member activities as defined in Section 2.2. The grey dashed line represents theoretical equilibrium ^{222}Rn activities assuming a spatially constant production rate equal to $\gamma_{mean} \approx 3.0$ Bq l^{-1} s^{-1}, leading to $A_{Rn,eq} \approx 16.5$ Bq l^{-1} (see Table 1).

These observations can be linked to the smoothing effect of both GW flow velocity and dispersion on observed ^{222}Rn activities. This is readily seen in Fig. 3, which shows the distribution of ^{222}Rn activities within the model domain for a single realization of variable ^{222}Rn production rates. At low velocities and dispersivities, ^{222}Rn activities are very sensitive to small-scale changes in production rates. Indeed, ^{222}Rn concentrations will be highly influenced by the local magnitude of ^{222}Rn input. As velocities and dispersivities increase, ^{222}Rn activities are averaged out as they integrate ^{222}Rn input signals from an increasing portion of the aquifer, and tend to be representative of the mean ^{222}Rn production rate.

This averaging effect is modulated by the range of spatial correlation of ^{222}Rn production rates; indeed, large zones of below- or above-average ^{222}Rn input will appear in the model domain when variogram ranges are largest. The core of these zones will retain distinct ^{222}Rn signatures even when flow velocities or dispersivity are large, as GW residence times within these zones allows ^{222}Rn activities to equilibrate with the local aquifer ^{222}Rn production rates.

3.2. Apparent age bias in steady-state conditions

3.2.1. Sensitivity to aquifer parameters (K, $\alpha_{L,TV}$) and scale of variability of ^{222}Rn production

The range of simulated ^{222}Rn apparent age biases in steady-state conditions (i.e., incurred only by the spatial variability of ^{222}Rn production, Eq. (7)) as a function of mean age at all synthetic observation points is shown in Fig. 4 for each of combination of hydraulic conductivity, dispersivity, and choice of ^{222}Rn end-member activity (Sections 2.1 and 2.2). Apparent ages were truncated at 16.5 days (i.e. at ~ 95% end-member activities). This apparent age is practically an upper limit for the ^{222}Rn dating technique (see Section 1.1); in these instances, additional age bias was therefore not computed. For readability, only results for one variogram model are shown in Fig. 4A ("Variogram 3″ with the largest correlation length for ^{222}Rn production); Results from other variogram models exhibit the same overall trends, and are shown in Appendix C. Note that the age bias incurred by mixing of water of different ages through hydrodynamic dispersion has been subtracted from the total apparent age bias (see Eq. (7)). In cases simulated here, the effect of mixing is well approximated by Eq. (3), and ranges from ~-0.05 for the case $K = 500$ m/d and $\alpha_L = 5$ m to approximately -0.5 for the case $K = 100$ m/d and $\alpha_L = 20$ m (see Appendix C).

As shown in Fig. 4A, the average ^{222}Rn apparent age bias is neutral, negative, and positive when computing apparent ages respectively with the *mean, maximum*, and *minimum* ^{222}Rn end-members. This is a result of the sensitivity of the apparent age equation (Eq. (1)) to end-member definition; indeed, using an end-member with large ^{222}Rn activities will lead on average to an underestimation of GW age, and vice-versa. In all cases presented here, age bias will tend to be slightly shifted towards positive values, as a result of the higher production rates modelled near infiltration zones (see Appendix A), leading on average to higher ^{222}Rn activities and thus apparent ages. This is most clearly visible at lower values of K and α_L (Fig. 4A), as lower GW velocities will lead to increased residence time in zones of high ^{222}Rn production, and the smoothing effect of dispersivity remains modest.

The magnitude and range of additional age bias is very sensitive to both GW flow velocity (controlled by hydraulic conductivity (K)), dispersivity (α_L), and the scale of variability of ^{222}Rn production (as given by the variogram models). Indeed, the range of ^{222}Rn apparent age bias increases when both velocities and dispersivities are low, and the scale of variability of ^{222}Rn input is large. For example, in the case where $K = 100$ m/d, $\alpha_L = 5$ m, and the scale of variability is largest ("Variogram 3″), the 25/75% interval for additional bias ranges between -0.3 and +1 at mean age = 5 days, and between -0.4 and +1 at mean age = 10 days. Conversely, when $K = 500$ m/d, $\alpha_L = 20$ m, and the scale of variability is lowest ("Variogram 1″, see Fig. 4B), additional bias is much more modest, with a 25/75% interval between 0 and +0.17 at mean age = 5 days, and between 0 and +0.25 at mean age = 10 days. This observation is linked to the smoothing effects of both GW flow velocity and dispersivity as described in Section 3.1. Indeed, even though ^{222}Rn activities oscillate around a value representative of the mean production rate, large variations in observed ^{222}Rn concentrations at greater distances

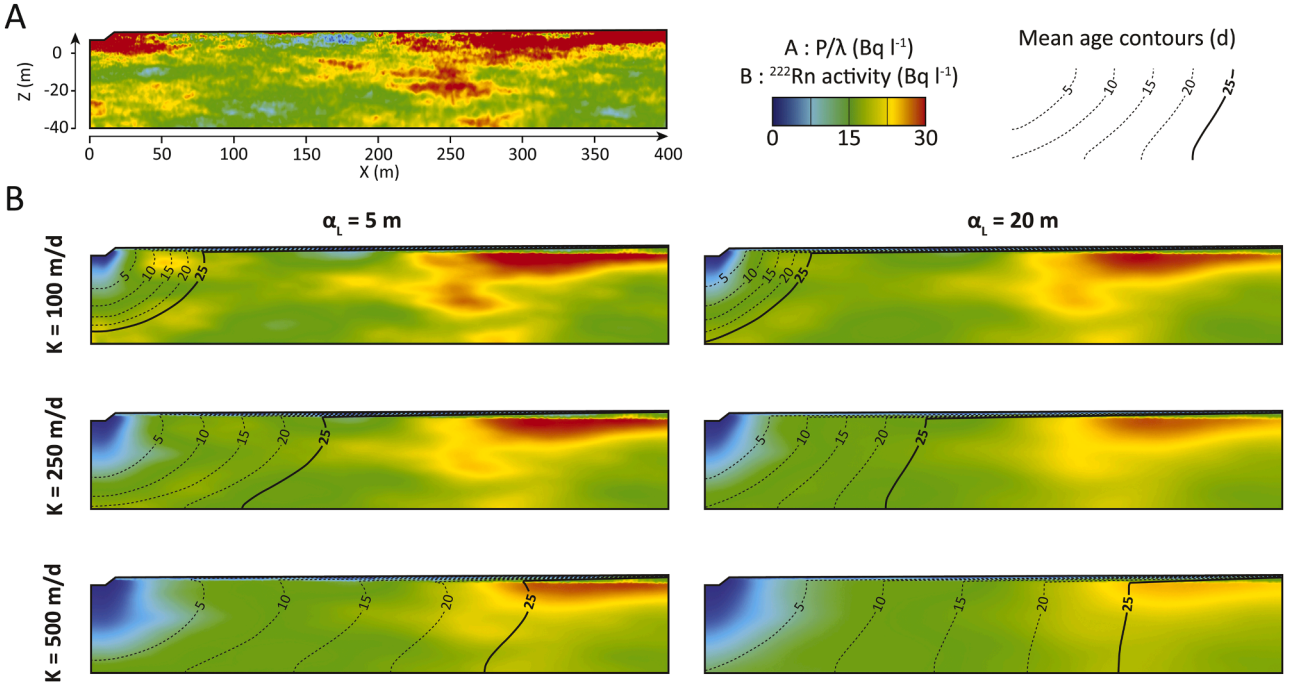
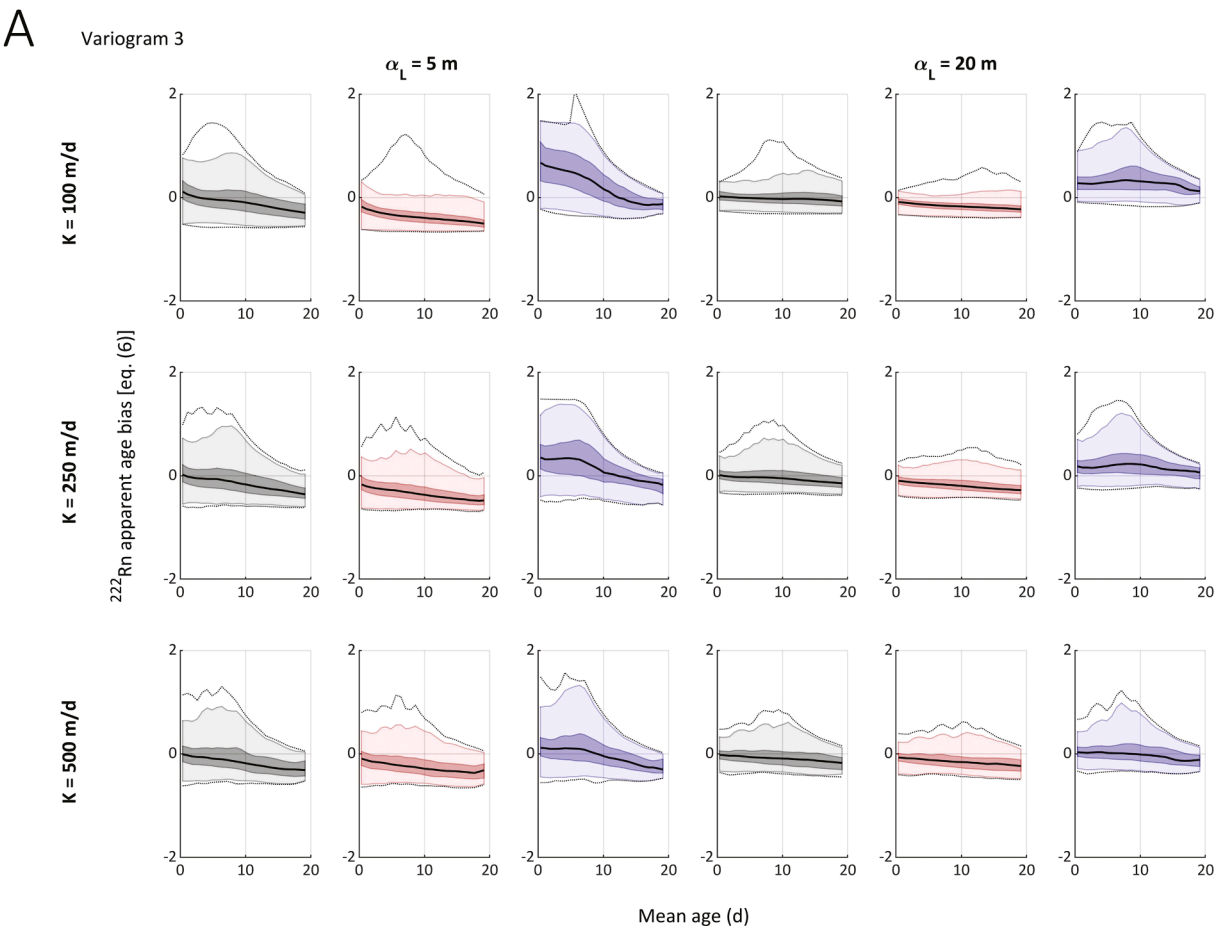


Fig. 3. Effect of hydraulic conductivity (K) and longitudinal dispersivity (α_L) on simulated steady-state ^{222}Rn concentrations. Results from a single realization of the distribution of ^{222}Rn production rates and the six parameter combinations are shown for illustrative purposes. A: Equivalent equilibrium ^{222}Rn concentrations ($=\gamma_{Rn}/\lambda_{Rn}$, Eq. (2)) in the model domain, shown here as a proxy for ^{222}Rn production rates. They represent theoretical equilibrium ^{222}Rn activities in saturated conditions without flow or molecular diffusion; B: ^{222}Rn activities in steady-state conditions. Black contour lines represent mean groundwater age.



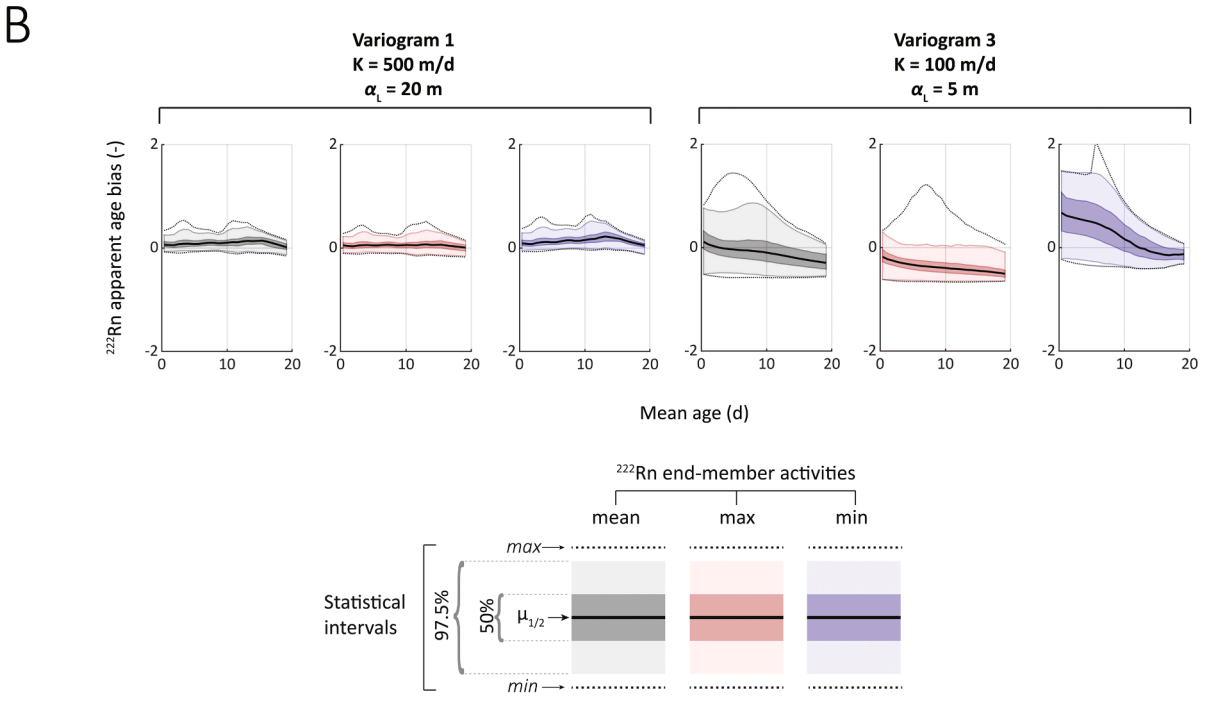


Fig. 4. ^{222}Rn age bias (Eq. (7)) in steady-state conditions as a function of mean age computed with the mean (grey), maximum (red), and minimum (blue) ^{222}Rn end-member activities (see Section 2.2 for the definition of these end-member activities). The median ($\mu_{1/2}$), min and max biases, as well as the 50% and 97.5% intervals are shown. A: Simulated ^{222}Rn bias as a function on mean age with variogram model 3 ("Variogram 3", highest correlation length for ^{222}Rn production). B: Comparison of two scenarios ((i) $\boldsymbol{K} = 500$ m/d, $\boldsymbol{\alpha_L} = 20$, lowest correlation length; (ii) $\boldsymbol{K} = 100$ m/d, $\boldsymbol{\alpha_L} = 5$, highest correlation length) with respectively the lowest and largest range of simulated ^{222}Rn apparent age bias. Results from 500 simulations are presented for each combination of $\boldsymbol{K}$ and $\boldsymbol{\alpha_L}$.

(or mean GW ages) will lead to a large range of possible end-member activities, and thus to highly non-unique apparent age estimates (see Fig. 2).

At low mean ages (< 5–10 days), the possible range of *positive* apparent age bias (i.e., overestimation of GW mean age) tends to increase with mean age, as the apparent age equation becomes more sensitive to small changes in ^{222}Rn activities when the latter start approaching end-member activities (see Eq. (1)). Conversely, the range of *negative* ^{222}Rn age bias remains mostly constant as a function of mean age. Cases where the ^{222}Rn age bias is lowest represents those where the ^{222}Rn production rates are lowest between the inlet and sampling points; ^{222}Rn activities will not near those of the end-member, and computed apparent ages remain low and are not affected by the asymptotic behaviour of the apparent age equation (Eq. (1)) at high ^{222}Rn activities.

The decrease in both the values and the spread of ^{222}Rn age bias at higher mean ages (> 5–10 days) results from two overlapping effects: (i) the smoothing effect from integrating signals from an increasing portion of the aquifer, and (ii) the truncation of apparent ages at 16.5 days. The former is related to the increasing distance between the inlet and sampling points at higher mean ages, and ^{222}Rn signals being increasingly representative of the mean production rate (i.e., lower sensitivity to small-scale variations in ^{222}Rn production rates). The latter effect is significant for scenarios where the *mean* and *minimum* ^{222}Rn end-members are used in apparent age computations (respectively grey and blue in Fig. 4). In such cases, ^{222}Rn activities will more rapidly approach end-member activities in instances where the mean production rate between the inlet and the sampling point is comparatively high; when ^{222}Rn activities exceed 95% of those of the end-member (i.e., apparent age > 16.5 days), apparent ages and thus ^{222}Rn age bias are not computed. Therefore, at higher mean ages, computable apparent ages will disproportionately reflect cases where the mean production rates between the inlet and the sampling point are generally lower. This leads to lower apparent ages, and therefore to a more negative ^{222}Rn age bias. This effect is even more pronounced when using the *minimum* end-member, as on average ^{222}Rn activities will more tend to approach even more rapidly those of the end-member. Therefore, lower ^{222}Rn age bias at higher values of mean age is not always synonymous with apparent age estimates being better constrained at higher mean ages; but rather that apparent ages can be computed at greater distances from the inlet only for a subset of simulations.

Finally, the decrease in apparent age bias at higher mean ages in cases where the *maximum* end-member is used is related to the existence of an apparent age threshold. In many cases, apparent ages will remain low at even high mean ages, as ^{222}Rn activities may remain significantly lower than those of the *maximum* end-member.

The combined influences of mean GW age, spatial correlation of ^{222}Rn production rates, as well as hydro-dispersive mixing on ^{222}Rn age bias can be seen in Fig. 5, which shows the ^{222}Rn age bias as a function of N_{Rn} (see Eq. (9)). The spread of ^{222}Rn age bias is lowest at low values of N_{Rn}, and tends to increase as N_{Rn} increases. For example, at values of $N_{Rn} = 0.1$, the 25/75% interval of ^{222}Rn age bias ranges from 0 to 0.2, whereas it ranges from -0.25 to +0.5% at $N_{Rn} = 10$. At the highest values of N_{Rn} (and thus of mean GW age), the ^{222}Rn bias tends to decrease for the same reasons noted in the previous paragraphs.

When $N_{Rn}>>1$ (generally higher spatial correlation of ^{222}Rn production, lower dispersivity, higher mean GW ages), measured ^{222}Rn activities are likely to be strongly affected by zones of high or low ^{222}Rn production, which lead to high variability in ^{222}Rn activities (which will often approach those of the end-member(s)) and a large range of ^{222}Rn end-members, leading in turn to high potential bias in computed apparent ^{222}Rn apparent ages. Conversely, when $N_{Rn}<< 1$, ^{222}Rn activities are more representative of the mean ^{222}Rn production rate; small differences between end-members will exist, and only moderate ^{222}Rn bias will arise when computing apparent ^{222}Rn ages.

3.3. Apparent age bias in transient conditions

Fig. 6 shows the range of transient simulated mean and apparent ages

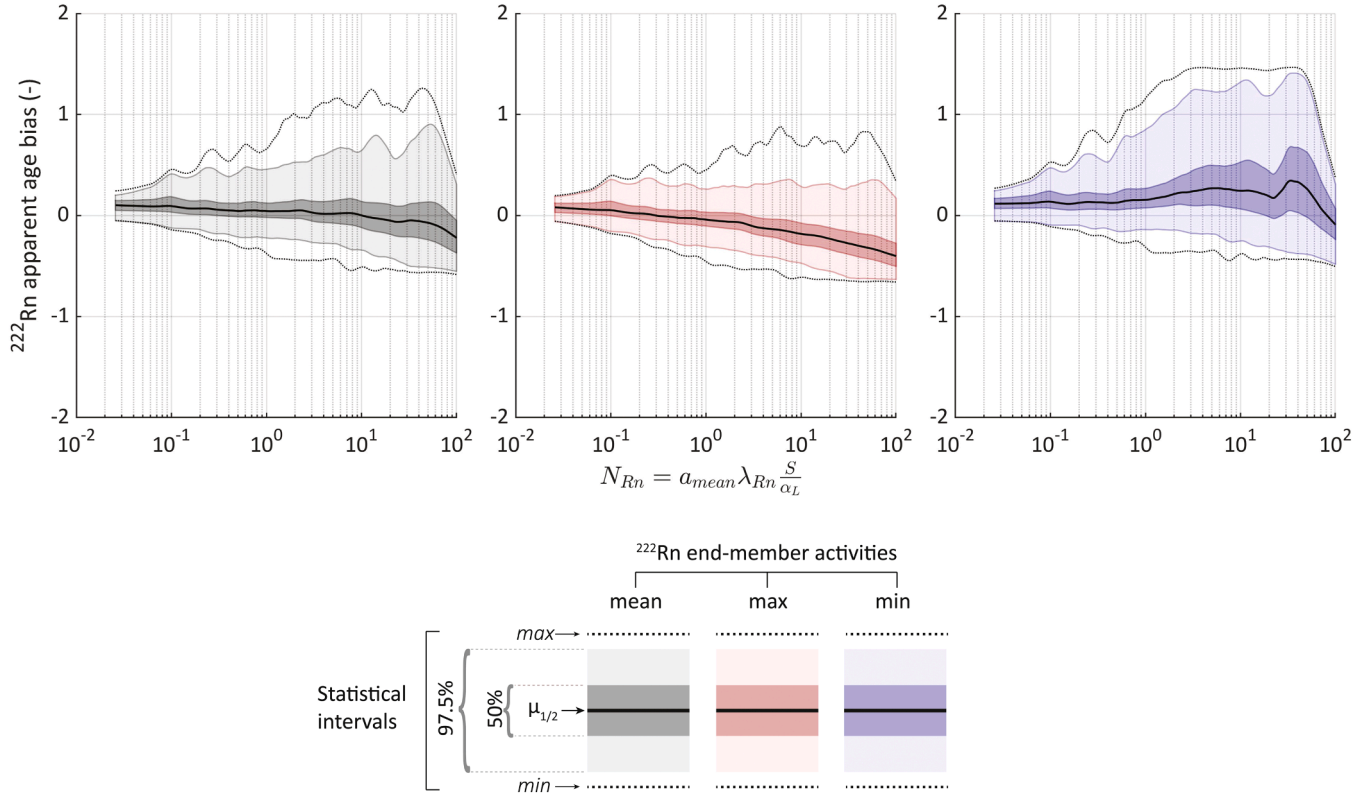


Fig. 5. Additional apparent ^{222}Rn age bias in steady-state conditions as a function of N_{Rn} (Eq. (9)). $\mu_{1/2}$ is the simulated median ^{222}Rn bias.

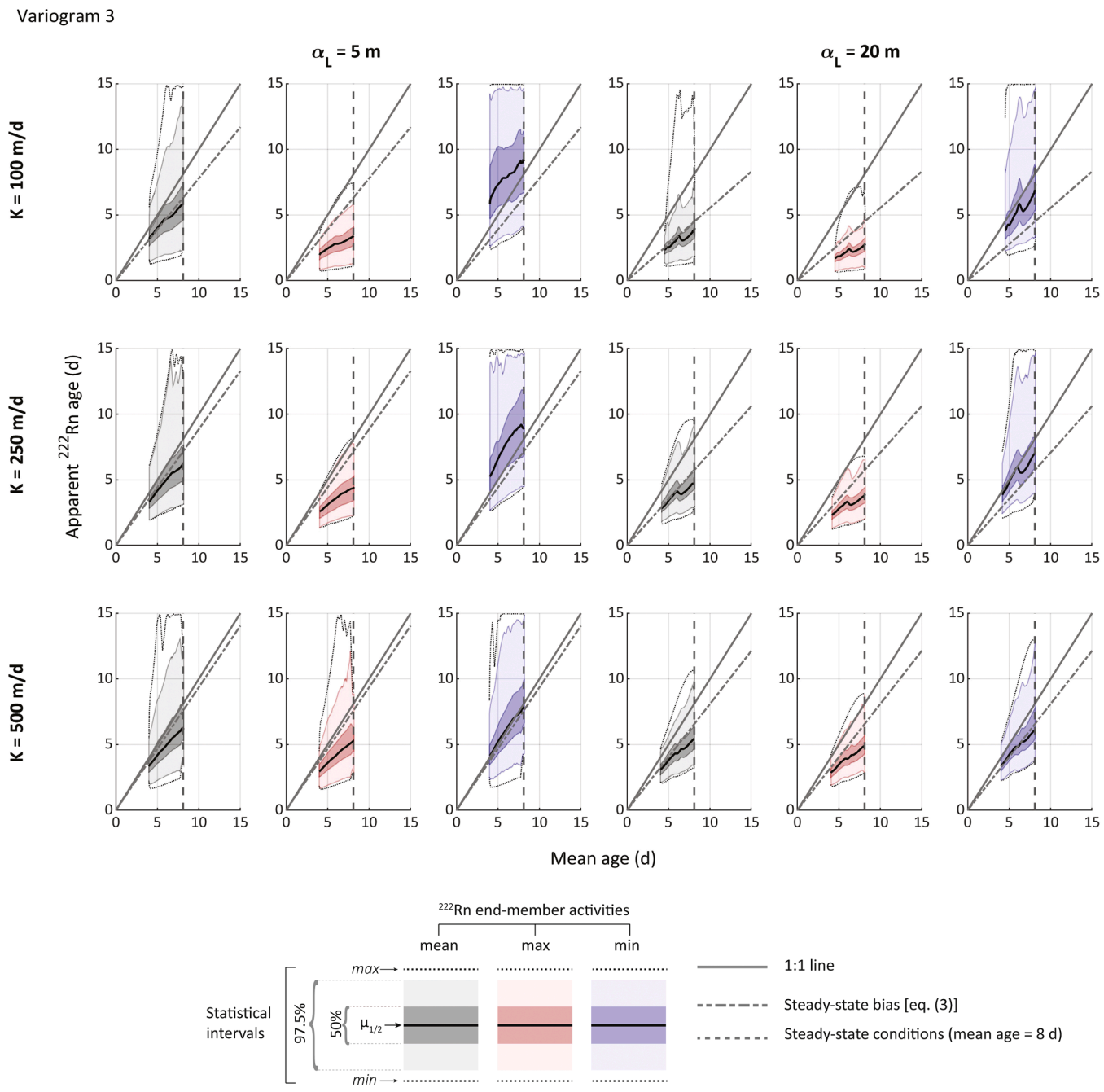


Fig. 6. Aggregated apparent ^{222}Rn ages vs mean ages for the 500 transient simulations, for all nodes where the mean age in steady-state conditions is $a_{mean} = 8 \pm 0.1$ d Apparent ages along the mean age = 8 d line (vertical dashed line) represent steady-state conditions. The steady-state bias line represents the 1-D apparent age bias with constant ^{222}Rn production and the mean GW flow velocity in steady-state conditions.

for the six combinations of hydraulic conductivity and longitudinal dispersivity at nodes where mean age in steady-state conditions is equal to 8.0 ± 0.1 d; results follow similar trends for different mean ages and variogram models (see Appendix B). All data points not located on the mean age = 8 d line result from transient hydraulic conditions; increased SW infiltration and GW flow velocity caused by the variable head boundary at the outlet (see Fig. 1) will lead to lower mean ages at any given observation point.

As already highlighted in Section 1.1.2, apparent ages in steady-state conditions tend to be lower than mean ages owing to hydrodynamic dispersion, and are clustered, *on average*, near (although slightly below) the 1-D bias line (see Fig. 6), which represents the steady-state relationship between mean and apparent ages in 1-D systems with constant ^{222}Rn input (Eq. (3)). This is because ^{222}Rn activities integrate information on the distribution of ^{222}Rn production between the inlet and a given sampling point, and tend to be, *on average*, representative of the mean production rate. Additional negative age bias is due to the additional effect of transverse dispersion which is not considered in 1-D systems. As explained in the previous section, apparent ages are consistently under- and overestimated in relation to the bias line when using the *maximum* and *minimum* ^{222}Rn endmembers respectively. Moreover, owing to the same smoothing mechanisms described in the previous sections, the range of computed apparent ages in steady-state conditions is greatest at low velocities and dispersivities, and when the scale of variability in ^{222}Rn input is highest. This is evident in Fig. 6 by the spread of apparent ages at mean age = 8 days for the different

combinations of K and α_L. In transient conditions, the twofold increase in GW flow velocity due to the transient head boundary at the outlet leads to a decrease of mean age up to 50% (from 8 to ~4 days), a decrease of apparent ^{222}Rn ages, as well as to an overall decrease in the spread of apparent ages. In other words, the potential for additional apparent age bias tends to decrease when GW flow velocities increase, once again due to the smoothing effects described previously.

In the conditions simulated here, changes in mean and apparent ages tend to follow a mostly linear pattern, with varying slopes depending on the choice of the ^{222}Rn endmember. In other words, in transient conditions, variations in apparent ages tend to be associated with proportional variations in mean ages, regardless of the choice of the ^{222}Rn endmember. To better visualize this relationship, we define $a^{rel}_{app,Rn}(\mathbf{x},t)$ and $a^{rel}_{mean}(\mathbf{x},t)$ as respectively the relative apparent ^{222}Rn and relative mean GW ages:

$$a^{rel}_{app,Rn}(\mathbf{x},t) = \frac{a_{app,Rn}(\mathbf{x},t)}{a_{app,Rn}(\mathbf{x},0)} \tag{10}$$

$$a^{rel}_{mean}(\mathbf{x},t) = \frac{a_{mean}(\mathbf{x},t)}{a_{mean}(\mathbf{x},0)} \tag{11}$$

where $a_{app,Rn}(\mathbf{x},t)$ [T] and $a_{mean}(\mathbf{x},t)$ [T] are the relative apparent ^{222}Rn and mean GW ages at location $\mathbf{x}$ and time t, and $a_{app,Rn}(\mathbf{x},0)$ [T] and $a_{mean}(\mathbf{x},0)$ [T] are similarly apparent ^{222}Rn and mean GW ages at the start of the simulations (t=0). $a^{rel}_{app,Rn}(\mathbf{x},t)$ and $a^{rel}_{mean}(\mathbf{x},t)$ as a function of simulation time are shown in Fig. 7. In cases where advection dominates over dispersive processes (i.e. high velocities and low dispersivities), relative mean and apparent ages behave in a very consistent manner;

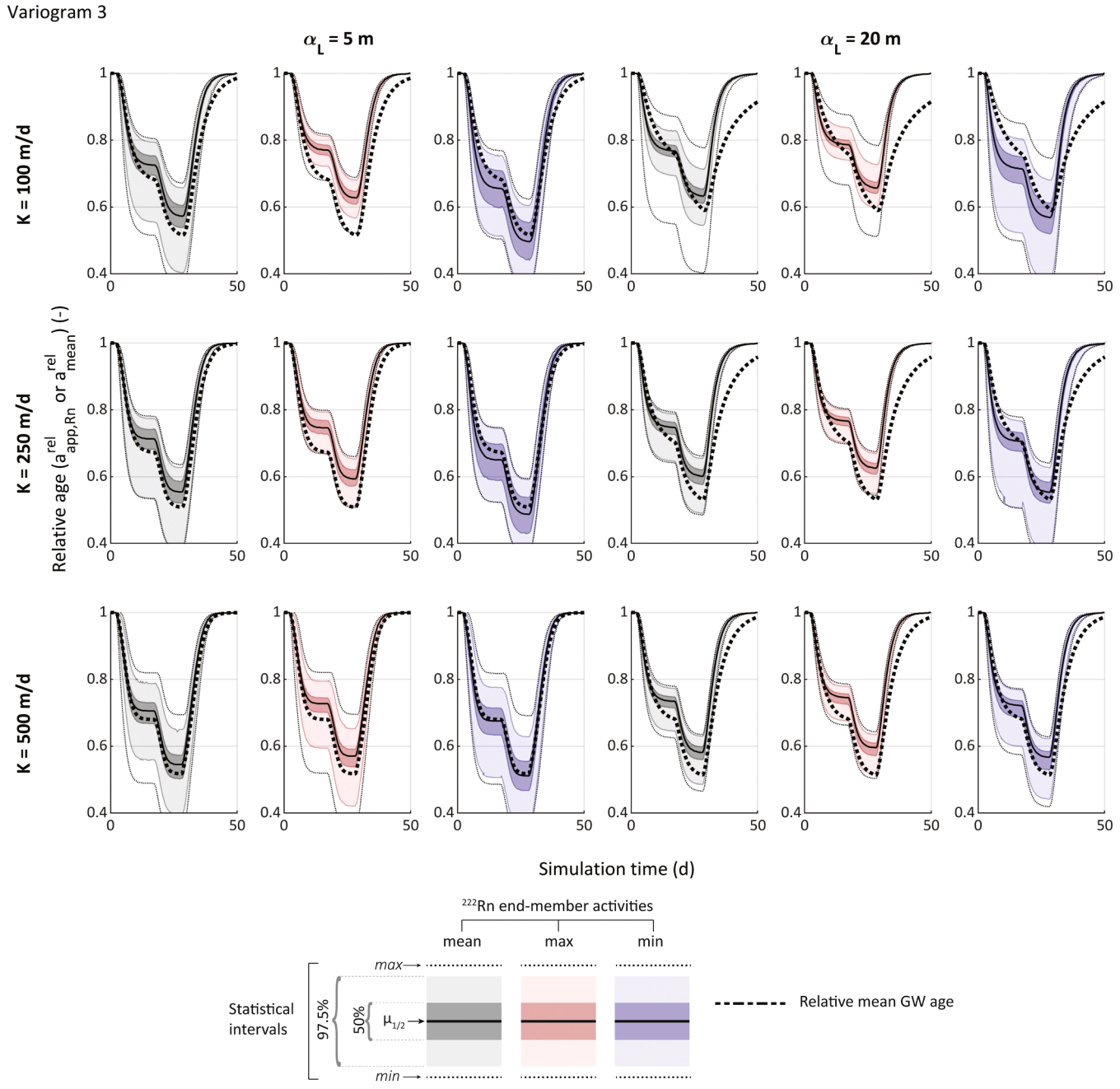


Fig. 7. Relative apparent ^{222}Rn ($a^{rel}_{app,Rn}(x,t)$) and relative mean ages ($a^{rel}_{mean}(x,t)$) as a function of simulation time for all nodes where the mean age in steady-state conditions is a_{mean} = 8.0 ± 0.1 d Apparent ages were computed using the mean (grey), maximum (red) and minimum (blue) ^{222}Rn end-member activities.

deviations between the two are mainly controlled by non-uniqueness in the definition of a ^{222}Rn end-member. Conversely, when dispersivity is large in relation to the flow velocity, changes in relative mean age are delayed in comparison to relative apparent ages. This most evident at the latter portion of the simulations, where relative mean ages return more slowly to steady-state conditions. This is a consequence of the relatively low half-life of ^{222}Rn (~3.8 d), which results in GW ^{222}Rn activities being quickly in local equilibrium with the aquifer matrix even in transient hydraulic conditions, and the persistence of a proportion of "younger" GW during the return to steady-state conditions will only marginally affect ^{222}Rn activities and apparent ^{222}Rn ages.

Despite some degree of bias arising from dispersion as well as from the non-uniqueness of the ^{222}Rn end-member, relative apparent ages track relative mean ages within ± 5–15% in most simulations. In contrast, apparent ages are much more poorly constrained, most notably at low GW flow velocities. In cases where mass transport is dominated by advective rather that dispersive processes, relative changes in mean apparent ^{222}Rn show excellent correlation with changes in mean GW age.

4. Discussion

4.1. Bias of apparent ^{222}Rn ages

Results from the synthetic models illustrate that, even in idealized homogenous 2-D flow systems, the ^{222}Rn apparent age model may not provide reliable estimates of mean GW age in steady-state conditions. On top of the negative age bias incurred by mixing of GW with different ages through hydrodispersive mixing and multiple flowline sampling, spatial heterogeneity of ^{222}Rn production rates could lead to large additional biases. Put together, these effects lead to apparent ^{222}Rn ages regularly differing from mean GW ages by a factor of two or more in the conditions simulated here. The magnitude of ^{222}Rn apparent age bias, that is the age bias incurred solely by variable ^{222}Rn production rates, can be captured by the dimensionless number N_{Rn}, with low values ($<<$ 1 in cases simulated here) correlated to low additional ^{222}Rn age bias, and vice versa. We show that the magnitude of age bias incurred by heterogenous ^{222}Rn production is highest (i) at low flow velocities, (ii) in cases where hydrodispersive mixing is limited, (iii) at high mean GW ages, and (iv) if the scale of spatial correlation of ^{222}Rn production rates is high compared to the sampling scale. Under such circumstances, ^{222}Rn signals are very sensitive to local variations in ^{222}Rn production rates, leading to high spatial variability in the distribution of ^{222}Rn activities, and ambiguity in the definition of a representative ^{222}Rn end-member activity. Conversely, when flow velocities and dispersion increase, ^{222}Rn activities are smoothed and tend to reflect mean ^{222}Rn production rates. Reliable estimates of GW age with apparent ^{222}Rn age model are therefore limited to cases where advective processes dominate (i.e., high Péclet number; little mixing of GW of different ages), and where there is little additional bias from spatially-variable ^{222}Rn production rates (i.e., low N_{Rn}). For both conditions to simultaneously hold, hydrodispersive processes must remain marginal in relation to advective transport, which requires the spatial scale of variability of ^{222}Rn production rates to be significantly smaller than the sampling scale, and/or mean GW age to remain relatively low (see Eq. (9)).

Results from transient simulations illustrate how time series of ^{222}Rn activities can provide reliable information on *changes* in mean GW age, even in cases where dispersive processes are significant and the ^{222}Rn end-members and apparent ^{222}Rn ages are poorly constrained. In the contexts modelled here, relative apparent and mean ages show very similar trends, although some deviations are noticeable in highly transient conditions. Indeed, ^{222}Rn signatures are more sensitive than mean GW age to abrupt changes in hydraulic conditions, especially in contexts where dispersive processes are significant. This results from the differing sensitivities of apparent and mean ages to the presence of GW components with residence times above 15–20 days. Whereas mean GW ages are determined by the entire age distribution of GW ages at any point in space and time, apparent ^{222}Rn ages are insensitive to the distribution of ages above the 15–20 day threshold, as these components are effectively at local radioactive equilibrium with the aquifer matrix. Therefore, any transient perturbation in the GW age distribution at ages above this threshold will not affect the computation of apparent ^{222}Rn ages, but will nevertheless have a lasting effect on mean GW ages until the distribution of ages reaches pseudo-steady-state conditions. Transient deviations between both ages are more evident when dispersion becomes a dominant mass-transport process, as GW age distributions tend to be broad and only slowly stabilize to equilibrium conditions. Conversely, when hydraulic conditions vary more smoothly, both apparent and mean GW ages remain in excellent agreement. The coherence of the relative apparent and mean ages can be attributed to the fact that throughout the simulation period, ^{222}Rn activities at all observation points integrate production rate signals from the same portions of the aquifer upgradient of the sampling points. In other terms, transient ^{222}Rn signals in GW near infiltration zones remain representative of an average ^{222}Rn production rate between infiltration and sampling points. Therefore, in the 2-D settings presented here, temporal variations in ^{222}Rn activities are almost exclusively attributed to changes in time since infiltration (i.e., GW age), and not to changes in GW flow paths through zones of differing ^{222}Rn input.

It must however be noted that a 2-D model will lead to obvious simplifications over complex 3-D flow and mass transport processes encountered in real-world environments. Changes in hydraulic conditions due to precipitation, variable river discharge, changes in river-bed permeability, GW pumping, etc. may lead to changes in GW flow and mass transport patterns not captured by the models presented in this study. Moreover, system transience is simulated exclusively through changes in hydraulic head at the model outlet, downgradient of observations points. In natural systems, changes in river stage (from storm events, seasonal patterns, etc..) will also control hydraulic gradients between infiltration zones and observation points. Abrupt variations in river stage may lead to complex flow conditions not modelled here, such as temporary flow reversal (e.g., bank storage return flow). Nevertheless, under less highly-transient conditions, flow patterns will generally reflect overall hydraulic gradients, whether they result from downgradient changes in GW head (e.g. through changes in GW pumping) or from variations in river stage. We therefore expect similar ^{222}Rn and mean GW age signatures to emerge in both cases.

Any significant changes if flowline geometry upstream of a sampling point will add an additional layer of complexity in interpreting ^{222}Rn signals, as sampled GW will be influenced by portions of the aquifer where ^{222}Rn production rates may differ. In field settings, the validity of assuming consistent flowline geometry under transient conditions must be justified by site-specific properties. This assumption is most likely to hold in circumstances where GW flow and mass transport are controlled by the existence of preferential flow pathways (i.e., highly conductive zones). Such preferential flow structures are abundant in many alluvial aquifers, and are known to strongly affect GW flow, solute transport, and SW-GW exchange fluxes (e.g. Boano et al. 2014, Huggenberger et al. 1998). In conditions where GW flow is consistently channelized through large-scale, hydraulically connected flow structures, we expect the correlation between mean and apparent ages to be less controlled by the heterogeneity of hydraulic parameters than by the distribution of ^{222}Rn production rates upgradient of sampling points, the degree of system transience, and the relative influence of advective and dispersive mass-transport processes.

4.2. Implications for use of ^{222}Rn as a tracer of GW age

The assumptions underpinning the apparent ^{222}Rn age model are quite restrictive and can often not be justified by field conditions. We've illustrated how apparent ^{222}Rn ages provide reliable estimates of mean GW age in losing river contexts only if each the following conditions are

met:

1 The scale of variation (or spatial correlation) of ^{222}Rn production rates is small in comparison to the sampling scale and/or mean GW age is much lower than the upper limit (~15 days) of the apparent ^{222}Rn age model (i.e., low N_{Rn});
2 There is no systematic trend in the spatial distribution of ^{222}Rn production rates;
3 Mass transport is advection-dominated (i.e., high Péclet number);
4 The ^{222}Rn endmember can be reasonably well constrained;
5 The assumption of sampling a single flowline is reasonable (e.g., short-screened intervals);
6 There is no mixing of different GW components (e.g., recently infiltrated SW and regional GW), or this mixing can explicitly be accounted for (e.g., via tracer-based approaches).

Conditions 3 to 6 also need to hold in conditions where the ^{222}Rn production rate is constant, as any mixing of water of different ages will lead to a (negative in the case of ^{222}Rn) bias of apparent ages in relation to mean GW ages.

Ideally, data for verification of conditions 1 and 2 can be obtained through high-resolution measurements of ^{222}Rn production rates from recovered aquifer material (e.g., through ^{222}Rn incubation experiments on sediment cores). As this is often practically infeasible, high-resolution measurements of ^{222}Rn activities in GW near radioactive equilibrium with the aquifer matrix can provide an estimate of the degree of heterogeneity of ^{222}Rn production. This requires a dense piezometer network at locations not influenced by recently infiltrated SW with ages $<$ 15–20 days. However, these measurements will not provide any direct information on the ^{222}Rn production rates at locations of active SW infiltration, i.e., where apparent ^{222}Rn age estimates are relevant.

In cases where field conditions and data are inconsistent with the above requirements, calculation of absolute ^{222}Rn ages should be done with caution or avoided altogether. Indeed, we have shown that even in simple systems such as those simulated in the present study, modest variability in the spatial distribution of ^{222}Rn production rates can lead to large discrepancies between mean and apparent ^{222}Rn ages. In conditions modelled here, mean and apparent ^{222}Rn ages may regularly differ by a factor of two or more, even for very young GW (mean age $<$ 5 days); such levels of uncertainty need to be factored into any estimates of GW age made with ^{222}Rn.

Additionally, the upper age limit of the ^{222}Rn dating method is modulated by the magnitude of spatial variability in ^{222}Rn production rates. Considering the variability in ^{222}Rn production modelled in this study, apparent age estimates for GW with mean ages greater than 10 days may become problematic, as ^{222}Rn activities may be influenced by zones of high ^{222}Rn production not captured by the chosen ^{222}Rn endmember activity. Indeed, as the sensitivity of ^{222}Rn age equation increases as ^{222}Rn activities approach those of the end-member activity, large positive biases in GW age estimates may result. This upper age limit of ~10 days is lower than the ~15 days put forward in other studies (e.g. Cranswick et al. 2014, Hoehn and Von Gunten 1989), which only applies to conditions where ^{222}Rn input is spatially constant. These conclusions build on those of recent studies of apparent ^{222}Rn ages in alluvial contexts (e.g., Gilfedder et al. 2019, Schaper et al. 2022), which have highlighted the limitations of the apparent ^{222}Rn age model in both homogenous and heterogenous environments.

In transient hydraulic conditions however, valuable information GW age can be gained from time series of ^{222}Rn activities. Indeed, in losing river contexts, variations in ^{222}Rn activities and computed apparent ^{222}Rn ages may be intimately linked to changes in mean GW age. Indeed, if the following criteria are met, our simulations have shown that relative changes in apparent ^{222}Rn ages can be reliably used as a proxy for variations in mean GW ages:

1 The system is not in a state of acute disequilibrium (i.e., system transience in not extreme at the time of sampling);
2 Flowline geometry of sampled water remains stable through transient conditions, i.e., sampled water integrates ^{222}Rn production information from the same zones of the aquifer;
3 The GW age distribution of sampled water is dominated by water with ages significantly lower than the upper age limit for the ^{222}Rn dating technique (~10 days in our simulations);
4 The ^{222}Rn end-member is significantly higher (at least 15%) than measured ^{222}Rn activities;
5 There is no mixing of different GW components (e.g., recently infiltrated SW and regional GW), or this mixing can explicitly be quantified (e.g., via tracer-based approaches).

Interestingly, these observations equally hold in cases the spatial distribution of ^{222}Rn production is highly heterogenous. Moreover, the computation of relative apparent ages is only moderately sensitive to the definition of the ^{222}Rn end-member, insomuch as its magnitude is significantly larger than measured activities and consistent with production rates at a given field site. Although absolute mean ages may not be adequately estimated from apparent ^{222}Rn ages, temporal variations in ^{222}Rn activities can be used as a tool to investigate changes in mean GW age at a given sampling location, assuming as the conditions listed above hold. These assumptions must be justified in individual field settings. If estimates of mean GW age can be made at different points in time through an alternate method (e.g., artificial tracer test), apparent ^{222}Rn and mean ages can be tied, allowing the computation of a time series of (absolute) mean ages from changes in apparent ^{222}Rn ages. This would apply to both continuous and occasional measurements of ^{222}Rn activities. We suggest, whenever possible, comparing several separate measurements of mean and apparent GW ages. This allows independent verification of the equivalence between variations in mean and apparent ^{222}Rn ages in transient conditions.

4.3. Limitations

Our analysis has been limited to simplified 2-D systems, which are unable to reproduce complex 3-D flow and mass transport phenomena. Moreover, GW mixing caused by lateral inflows, or water exchange with aquitards and/or low-permeability lenses was not considered, which may lead to strong underestimations of mean GW age. Furthermore, we only consider a simplified system with spatially homogenous hydraulic properties (conductivity, porosity, dispersivity). It is well established that bias between tracer-derived and mean GW ages increases with increasing heterogeneity of hydraulic parameters (e.g., Gardner et al. 2015, McCallum et al. 2013). As noted in the preceding section, the effect of hydraulic parameter heterogeneity on the performance of the ^{222}Rn age model may not be exceedingly detrimental in conditions were channelized flow and transport are dominant processes. However, the impact of hydraulic parameter heterogeneity has not been quantified in the present work, and should be systematically addressed in future studies. Furthermore, we suspect that the conclusions concerning transient ^{222}Rn signals do not hold for sampling locations where convergent flow occurs, e.g., at pumping wells. Indeed, disentangling contributions from all GW components with ages $>$ 10 days to the overall ^{222}Rn signal is practically impossible without resorting to additional tracer methods and/or numerical modelling tools. Additionally, the results presented in this study are valid only for the magnitude of variability of ^{222}Rn production rates modelled in the present study (i.e., CV of 50%, see Table 1, Section 2.1); we expect the amount of ^{222}Rn age bias to increase with increasing variability in ^{222}Rn production rates, although this should be investigated in future studies. Our simulations do not include systematic trends in the spatial distribution of ^{222}Rn production rates, which could lead to additional bias in apparent ^{222}Rn ages. This approach was chosen as no observable spatial trend could be discerned from the experimental data used in the simulations. It is unclear whether this is a peculiarity of

the selected field site or a feature common to many alluvial aquifers. The latter may be true at other similar sites, as alluvial deposits within a small region of an aquifer often share many features, such as common depositional histories, as well as rheological and geochemical properties.

5. Conclusions

The objective of the present study was to explore the reliability of ^{222}Rn as a tracer of GW age in bank filtration contexts where ^{222}Rn production rates are spatially variable. By including laboratory measurements of ^{222}Rn emanation rates in a physically-based model, we illustrated how spatially-variable ^{222}Rn input as well as hydrodispersive mixing may lead to biased estimates of ^{222}Rn-derived GW age.

We show that the combined effects of GW mixing and spatially-variable ^{222}Rn production rates may lead to high biases between apparent ^{222}Rn ages and mean GW age, limiting the reliability of the ^{222}Rn dating technique to contexts where mass-transport is advection-dominated, sampling scale is much larger than the scale of variability of ^{222}Rn production rates, and GW is sampled from shortly-screened intervals.

The deleterious effects of mixing and variable ^{222}Rn input on the apparent ^{222}Rn age model are somewhat dampened when interpreting time series of ^{222}Rn activities at a given observation point. Inasmuch as GW flow paths remain essentially stable in transient conditions, ^{222}Rn activities can provide reliable information on relative changes in mean GW age, even in environments where the spatial distribution of ^{222}Rn production is highly heterogenous.

The scope of application for the (relative) apparent ^{222}Rn age model remains quite restrictive, and information justifying its use in field settings may often be lacking. The sampling and continuous monitoring of additional tracers (e.g., electrical conductivity, stable isotopes of water, GW temperature) may provide further information on GW residence times at temporal scales relevant to bank filtration contexts. However, these tracers do not systematically exhibit distinct concentrations in river water and sampled GW or show measurable temporal variations in SW that can be linked to variations in GW. In such cases, ^{222}Rn often remains one of few practical tracers that can provide information on GW residence times at timescales of days to weeks.

The simulation of ^{222}Rn (and any other studied tracer) in fully-coupled, physically-based 3-D mass-transport models offers a more robust conceptual framework through which the information content of tracer measurements can be better extracted (e.g. Turnadge and Smerdon 2014). With such models, the uncertainty in tracer-relevant parameters (such as the spatial distribution of tracer sources and sinks, hydraulic & mass-transport parameters, etc.) may be explicitly accounted for. Parameterization of these models can be improved through the inclusion of tracer measurements (e.g. Schaper et al. 2022, Thiros et al. 2021), which in turn leads to increased predictive accuracy of estimates of management interest, such as GW-SW exchange fluxes, mixing ratios, and GW residence times.

However, when access to additional such complex modelling and calibration tools is not possible, monitoring changes in ^{222}Rn activities can be helpful to gain a better conceptual understanding of the sensitivity of the GW-SW system and GW age to changes in hydraulic conditions, although absolute age estimates should always be interpreted with caution.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We would like to thank the Editor as well as two anonymous reviewers for their constructive comments which greatly improved the quality of the manuscript. All data used in this study can be found in Supplement S1. This study was funded by the Swiss National Science Foundation (grant number 200021_179017).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2023.119880.

References

Aquanty, 2023. HGS Reference Manual. Aquanty, Waterloo, ON, p. 366.
Bear, J., 1972. Dynamics of Fluids in Porous Media. American Elsevier, New York.
Bertin, C., Bourg, A.C., 1994. Radon-222 and chloride as natural tracers of the infiltration of river water into an alluvial aquifer in which there is significant river/groundwater mixing. Environ. Sci. Technol. 28 (5), 794–798.
Bethke, C.M., Johnson, T.M., 2008. Groundwater Age and Groundwater Age Dating. Annu. Rev. Earth Planet Science, 36, pp. 121–152.
Boano, F., Harvey, J.W., Marion, A., Packman, A.I., Revelli, R., Ridolfi, L., Wörman, A., 2014. Hyporheic flow and transport processes: mechanisms, models, and biogeochemical implications. Rev. Geophys. 52 (4), 603–679.
Brunner, P., Therrien, R., Renard, P., Simmons, C.T., Franssen, H.J.H., 2017. Advances in understanding river-groundwater interactions. Rev. Geophys. 55 (3), 818–854.
Cartwright, I., Gilfedder, B., 2015. Mapping and quantifying groundwater inflows to Deep Creek (Maribyrnong catchment, SE Australia) using ^{222}Rn, implications for protecting groundwater-dependant ecosystems. Appl. Geochem. 52, 118–129.
Cecil, L.D. and Green, J.R. (2000) Environmental tracers in subsurface hydrology, pp. 175–194.
Cook, P.G., Herczeg, A.L., 2000. Environmental Tracers in Subsurface Hydrology. Springer, Boston, MA.
Cook, P.G., Lamontagne, S., Berhane, D., Clark, J.F., 2006. Quantifying groundwater discharge to Cockburn River, southeastern Australia, using dissolved gas tracers ^{222}Rn and SF6. Water Resour. Res. 42 (10), 1–12.
Cranswick, R.H., Cook, P.G., Lamontagne, S., 2014. Hyporheic zone exchange fluxes and residence times inferred from riverbed temperature and radon data. J. Hydrol. 519, 1870–1881.
Delottier, H., Peel, M., Musy, S., Schilling, O.S., Purtschert, R., Brunner, P., 2022. Explicit simulation of environmental gas tracers with integrated surface and subsurface hydrological models. Front. Water 4, 1–12.
Fetter, C.W., 2001. Applied Hydrogeology. Prentice Hall, Upper Saddle Creek, NJ.
Frei, S., Gilfedder, B.S., 2021. Quantifying residence times of bank filtrate: a novel framework using radon as a natural tracer. Water Res. 201, 117376.
Gardner, W.P., Hammond, G., Lichtner, P., 2015. High performance simulation of environmental tracers in heterogeneous domains. Ground Water 53 (Suppl 1), 71–80.
Gelhar, L.W., Welty, C., Rehfeldt, K.R., 1992. A critical review of data on field-scale dispersion in aquifers. Water Resour. Res. 28 (7), 1955–1974.
Gilfedder, B.S., Cartwright, I., Hofmann, H., Frei, S., 2019. Explicit modeling of Radon-222 in hydrogeosphere during steady state and dynamic transient storage. Ground Water 57 (1), 36–47.
Ginn, T.R., Haeri, H., Massoudieh, A., Foglia, L., 2009. Notes on groundwater age in forward and inverse modeling. Transp. Porous Media 79 (1), 117–134.
Goode, D.J., 1996. Direct simulation of groundwater age. Water Resour. Res. 32 (2), 289–296.
Hiscock, K.M., Grischek, T., 2002. Attenuation of groundwater pollution by bank filtration. J. Hydrol. 266 (3–4), 139–144.
Hoehn, E., Von Gunten, H.R., 1989. Radon in groundwater: a tool to assess infiltration from surface waters to aquifers. Water Resour. Res. 25 (8), 1795–1803.
Huggenberger, P., Hoehn, E., Beschta, R., Woessner, W., 1998. Abiotic aspects of channels and floodplains in riparian ecology. Freshw. Biol. 40 (3), 407–425.
Ishimori, Y., Lange, K., Martin, P., Mayya, Y.S., Phaneuf, M., 2013. Measurement and Calculation of Radon Releases from NORM Residues. Vienna, Austria, p. 103.
Manning, A.H., Solomon, K.D., Thiros, S.A., 2005. 3H/3He age data in assessing the susceptibility of wells to contamination. Ground Water 43 (3), 353–367.
Margat, J., van der Gun, J., 2013. Groundwater Around the World - A Geographic Synopsis. CRC Press, Boca Raton, FL.
Massoudieh, A., Ginn, T.R., 2011. The theoretical relation between unstable solutes and groundwater age. Water Resour. Res. 47 (10), 1–6.
McCallum, J.L., Cook, P.G., Simmons, C.T., Werner, A.D., 2013. Bias of apparent tracer ages in heterogeneous environments. Ground Water 52 (2), 239–250.

Mullinger, N.J., Pates, J.M., Binley, A.M., Crook, N.P., 2009. Controls on the spatial and temporal variability of ^{222}Rn in riparian groundwater in a lowland Chalk catchment. J. Hydrol. 376 (1–2), 58–69.
Oldham, C.E., Farrow, D.E., Peiffer, S., 2013. A generalized Damköhler number for classifying material processing in hydrological systems. Hydrol. Earth Syst. Sci. 17 (3), 1133–1148.
Peel, M., Kipfer, R., Hunkeler, D., Brunner, P., 2022. 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, 1–12.
Pittroff, M., Frei, S., Gilfedder, B.S., 2017. Quantifying nitrate and oxygen reduction rates in the hyporheic zone using ^{222}Rn to upscale biogeochemical turnover in rivers. Water Resour. Res. 53 (1), 563–579.
Popp, A.L., Pardo-Álvarez, Á., Schilling, O.S., Scheidegger, A., Musy, S., Peel, M., Brunner, P., Purtschert, R., Hunkeler, D., Kipfer, R., 2021. A framework for untangling transient groundwater mixing and travel times. Water Resour. Res. 57 (4), 1–16.
Schaper, J.L., Zarfl, C., Meinikmann, K., Banks, E.W., Baron, S., Cirpka, O.A., Lewandowski, J., 2022. Spatial variability of radon production rates in an alluvial aquifer affects travel time estimates of groundwater originating from a losing stream. Water Resour. Res. 58, 1-–22. https://doi.org/10.1029/2021WR03063.
Schilling, O.S., Cook, P.G., Brunner, P., 2019. Beyond classical observations in hydrogeology: the advantages of including exchange flux, temperature, tracer concentration, residence time, and soil moisture observations in groundwater model calibration. Rev. Geophys. 57 (1), 146–182.
Schilling, O.S., Gerber, C., Partington, D.J., Purtschert, R., Brennwald, M.S., Kipfer, R., Hunkeler, D., Brunner, P., 2017. Advancing physically-based flow simulations of alluvial systems through atmospheric noble gases and the Novel37Ar tracer method. Water Resour. Res. 53 (12), 10465–10490.
Schulze-Makuch, D., 2005. Longitudinal dispersivity data and implications for scaling behavior. Ground Water 43 (3), 443–456.
Sprenger, C., Lorenzen, G., Hulshoff, I., Grutzmacher, G., Ronghang, M., Pekdeger, A., 2011. Vulnerability of bank filtration systems to climate change. Sci. Total Environ. 409 (4), 655–663.
Stellato, L., Terrasi, F., Marzaioli, F., Belli, M., Sansone, U., Celico, F., 2013. Is ^{222}Rn a suitable tracer of stream–groundwater interactions? A case study in central Italy. Appl. Geochem. 32, 108–117.
Straubhaar, J., 2020. GEONE, Python Package for Geostatistical and Multiple-Point Statistics Modeling. University of Neuchâtel.
Thiros, N.E., Gardner, W.P., Kuhlman, K.L., 2021. Utilizing environmental tracers to reduce groundwater flow and transport model parameter uncertainties. Water Resour. Res. 57 (7), 1–27.
Tufenkji, N., Ryan, J.N., Elimelech, M., 2002. Peer Reviewed: The Promise of Bank Filtration. ACS Publications, pp. 423–428A.
Turnadge, C., Smerdon, B.D., 2014. A review of methods for modelling environmental tracers in groundwater: advantages of tracer concentration simulation. J. Hydrol. 519, 3674–3689.
Varni, M., Carrera, J., 1998. Simulation of groundwater age distributions. Water Resour. Res. 34 (12), 3271–3281.
Vogt, T., Hoehn, E., Schneider, P., Freund, A., Schirmer, M., Cirpka, O.A., 2010. Fluctuations of electrical conductivity as a natural tracer for bank filtration in a losing stream. Adv. Water Resour. 33 (11), 1296–1308.
Waugh, D.W., Hall, T.M., Haine, T.W.N., 2003. Relationships among tracer ages. J. Geophys. Res. 108 (C5) 1-16.
Weigel, F., 1978. Radon. Chem. Ztg. 102 (9), 287–299.