https://doi.org/10.35662/unine-thesis-3032

University of Neuchâtel, Switzerland

Faculty of Sciences

Centre for Hydrogeology and Geothermics (CHYN)

Noble gases as tracers of surface water – groundwater interactions: insights from novel field and modelling approaches

A thesis presented for the degree of Doctor of Sciences

by

Morgan PEEL

Accepted on the recommendation of

Prof. Philip Brunner, University of Neuchâtel, Switzerland (Thesis director)
Prof. Daniel Hunkeler, University of Neuchâtel, Switzerland (Rapporteur)
Prof. Peter Cook, Flinders University, Australia (Rapporteur)
Prof. Olaf Cirpka, University of Tübingen, Germany (Rapporteur)
Dr. Roland Purtschert, University of Bern, Switzerland (Rapporteur)

Defended on May 17, 2023

unine
Université de Neuchâtel
Faculté des sciences

Rue Emile-Argand 11
CH-2000 Neuchâtel
doctorat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Monsieur Morgan PEEL

Titre :

"Noble gases as tracers of surface water – groundwater interactions: insights from novel field and modelling approaches"

sur le rapport des membres du jury composé comme suit :

- Prof. Philip Brunner, directeur de thèse, Université de Neuchâtel, Suisse
- Prof. Daniel Hunkeler, Université de Neuchâtel, Suisse
- Prof. Peter Cook, Flinders University, Australie
- Prof. Olaf Cirpka, Tübingen Universität, Allemagne
- Dr Roland Purtschert, Universität Bern, Suisse

Neuchâtel, le 12 juillet 2023

Le Doyen, Prof. R. Bshary

Summary

Understanding the interactions between surface water (SW) and groundwater (GW) in alluvial aquifer systems is of critical importance for the sustainable management of water resources. Advances in real-time and continuous measurement of a range of hydrological tracers provide new opportunities for the characterization of SW-GW dynamics at unprecedented spatiotemporal resolutions. Amongst several promising tracer methods, noble gases are particularly well-adapted to the study of SW-GW interactions, and provide an integrated signal on the flow paths and travel times of water. Capitalizing on the insights offered by novel measurement technologies requires tracer interpretation methods that appropriately capture tracer transport processes in dynamic environmental conditions. However, recourse to highly-simplified tracer interpretation methods, conceptually detached from the complexity of natural systems, is still widespread. In such cases, the potentially rich information content of tracer measurements may be underutilized.

This thesis aims at investigating how established and emerging noble gas tracer methods can be optimally used - and when they should be avoided - for the study of SW-GW interactions in alluvial aquifer systems. To this end, a range of novel laboratory, field, and modelling approaches are employed to systematically assess the applicability, limitations, and potential of some gas tracer methods toward informing SW-GW exchange processes.

The first gas tracer method examined is the ^{222}Rn apparent age model, which is widely used to estimate the ages of very young GW (days to weeks). High-resolution measurements of the spatial distribution of ^{222}Rn emanation rates in an alluvial aquifer reveal significant spatial heterogeneity in ^{222}Rn production. The explicit simulation of ^{222}Rn in synthetic mass-transport models shows that this level of heterogeneity, combined with mixing of GW, can result in strongly biased estimates of GW age, effectively limiting the applicability of the ^{222}Rn apparent age method to a limited range of field conditions. Although temporal changes in ^{222}Rn concentrations may reveal insights into GW age dynamics, the information content of ^{222}Rn measurements may be best extracted through the integration of ^{222}Rn observations in the calibration process of physics-based flow and transport models.

Indeed, the second part of this thesis is devoted to exploring how the explicit simulation of tracer concentrations and the assimilation of untransformed tracer data in highly parameterized, physics-based models may inform model parameters and ultimately predictions of management interest. Within this framework, the joint assimilation of hydraulic and noble gas data (^{222}Rn and helium), acquired over the course of a novel tracer injection experiment, is shown to strongly inform model parameters and reduce predictive uncertainty of several important water management quantities,

such as GW age, SW-GW mixing ratios, and SW infiltration fluxes, far beyond what is achieved with "traditional" hydraulic data alone. These results build upon mounting evidence as to the benefits of explicitly simulating and assimilating diverse observation types with physically-based flow and transport models, avoiding the layer of conceptual simplification and potential bias accrued with simplified tracer interpretation models.

Finally, the successful combination of novel gas injection methods, developed over the course of this project, and the assimilation of high-resolution gas tracer measurements in an explicit tracer simulation framework strongly support further developments of (noble) gas tracer methods and tracer-numerical model synergies.

Keywords

Alluvial aquifers; Surface water – groundwater interactions; Environmental tracers; Artificial tracers; Tracer injection; Explicit tracer simulation; Parameter estimation; Integrated surface subsurface hydrological models; Noble gases; Radon-222; Groundwater; Groundwater age; Data assimilation; Mass transport; Water resources management

Résumé

La compréhension des interactions entre les eaux de surface (ESF) et les eaux souterraines (EST) dans les systèmes d'aquifères alluviaux est d'importance capitale pour la gestion durable des ressources en eau. De récents progrès réalisés dans le développement de techniques de mesure en continu d'une gamme de traceurs hydrologiques offrent de nouvelles possibilités pour la caractérisation de la dynamique des interactions ESF - EST à des résolutions spatio-temporelles sans précédent. Parmi de nombreuses classes de traceurs désormais disponibles, les gaz nobles sont particulièrement bien adaptés à l'étude des interactions entre les eaux souterraines et les eaux de surface, et fournissent un signal intégré sur le trajet et les temps de parcours de l'eau souterraine. L'exploitation du potentiel offert par ces nouvelles technologies de mesure est sujette à l'utilisation de méthodes d'interprétation des données de traceurs qui tiennent compte des processus de transport de masse dans des conditions environnementales complexes. Cependant, le recours à des méthodes d'interprétation très simplifiées, conceptuellement détachées de la complexité des systèmes naturels, est encore très répandu. Dans de tels cas, l'information contenue dans les données de traceurs est potentiellement diluée.

Cette thèse vise à étudier comment différentes méthodes de traceurs basées sur les gaz nobles peuvent être utilisées de manière optimale (et quand elles doivent être évitées) pour l'étude des interactions ESF-EST dans les systèmes d'aquifères alluviaux. A cette fin, une série de nouvelles approches en laboratoire, de terrain, et de modélisation numérique sont développées afin d'évaluer systématiquement l'applicabilité, les limites, et le potentiel de certaines méthodes de traceurs avec les gaz nobles pour informer les processus d'échange entre les ESF et EST.

La première méthode de traceur examinée est le ^{222}Rn, qui est communément mesuré afin d'estimer l'âge des très jeunes EST (infiltrées depuis quelques jours à semaines). Des mesures à haute résolution de la distribution spatiale des taux d'émanation du ^{222}Rn dans un aquifère alluvial révèlent une hétérogénéité spatiale significative dans la production de ^{222}Rn. La simulation explicite du ^{222}Rn avec des modèles numériques de transport de masse montre que ce niveau d'hétérogénéité, combiné au mélange des EST, peut donner lieu à des estimations fortement biaisées de l'âge des EST, ce qui limite l'applicabilité du modèle d'âge ^{222}Rn à une gamme étroite de conditions de terrain. Bien que les variations temporelles des concentrations de ^{222}Rn puissent révéler des informations sur la dynamique de l'âge des EST, l'information contenue dans les mesures de ^{222}Rn pourrait être mieux extraite par l'intégration des observations de ^{222}Rn dans le processus de calibration de modèles numériques d'écoulement et de transport basés sur la physique.

En effet, la deuxième partie de cette thèse explore comment la simulation explicite des concentrations de traceurs, couplée à l'assimilation des données de traceurs non transformées dans des modèles hautement paramétrés, peut contraindre les paramètres du modèle et ainsi diminuer l'incertitude de plusieurs prévisions pertinentes pour la gestion de l'eau. L'assimilation conjointe de données hydrauliques et de gaz nobles (^{222}Rn et hélium), acquises au cours d'une expérience novatrice d'injection de gaz nobles, permet de contraindre les paramètres d'un modèle numérique d'un site de filtration sur berge, et de fortement diminuer l'incertitude prédictive de plusieurs quantités importantes pour la gestion de l'eau, telles que l'âge des EST, des rapports de mélange ESF-EST, et les flux d'infiltration des ESF, bien au-delà de ce qui est envisageable avec seule l'assimilation de données hydrauliques "traditionnelles". Ces résultats s'ajoutent à un nombre croissant de travaux qui illustrent les avantages de la simulation explicite des traceurs et de l'assimilation de différents types d'observations avec des modèles numériques d'écoulement et de transport basés sur la physique. Cette méthodologie permet d'éviter la simplification conceptuelle et les biais potentiels découlant de l'utilisation de modèles simplifiés d'interprétation des traceurs.

La combinaison de nouvelles méthodes d'injection de gaz, développées au cours de ce projet, et de l'assimilation de mesures de gaz nobles dans un cadre de simulation explicite de traceurs soutiennent à la fois le déploiement de synergies entre les méthodes de traçage aux gaz et les modèles numériques, ainsi que le développement de nouvelles méthodes de traçage aux gaz (nobles).

Mots-clés

Aquifères alluviaux; Interactions eaux de surface – eaux souterraines; Traceurs environnementaux; Traceurs artificiels; Injection de traceurs; Simulation explicite des traceurs; Estimation de paramètres ; Modèles hydrologiques intégrés de surface et de subsurface; Gaz nobles; Radon-222; Eaux souterraine; Ages des eaux souterraines; Assimilation de données; Transport de masse; Gestion des ressources en eau

Acknowledgements

These few lines can only express a fraction of the gratitude towards the many colleagues, friends and family members who have provided support over the last four years (and more); this thesis would never have been possible without you all.

First and foremost, my deepest thanks go to my advisor Philip Brunner, for giving me the opportunity to participate in this amazing project devoted to hydrological gas tracers. Your keen eye for new research avenues, your endless enthusiasm, and your unflinching support have been inspiring and have kept me motivated from day one. To say it's been a pleasure working with you is an understatement.

Secondly, I would like to thank Daniel Hunkeler, my unofficial co-adviser, whose ideas were always a source of inspiration and in large part shaped this thesis. The projects on which we worked together were some of the most interesting during my time at the CHYN, and without a doubt helped uncover a hidden love for work in the field (and, to a somewhat lesser extent, in the lab…).

I would also like to thank thesis committee members Roland Purtschert, Olaf Cirpka, and Peter Cook for carefully reading and providing feedback on this thesis.

Of course, many thanks go to my "noble gas" colleagues, Rolf Kipfer and Matthias Brennwald from Eawag, as well as Théo Blanc from the CHYN, with whom we could explore plenty of fascinating ways to employ gases as artificial tracers. What started as a few fun ideas turned into something which has the potential to move forward the science of artificial tracer hydrology. I did not envisage this work would bring us so far, and it's been incredibly rewarding.

My gratitude also goes towards the organisations and companies as well as their employees who have graciously provided access to their field installations over the last few years. Firstly I would like to thank the *Wasserverbund Region Bern (WVRB)*, for a fruitful collaboration spanning the whole thesis at the Aeschau field site. Many thanks to Philip Waldmann (*WVRB*), who always went out of his way to help us find solutions and give us a hand. Secondly, I would like to thank the *Rheinunternehmen (RU)* for providing long-term access to their field installations during the Rhine excavation experiment in 2021.

Many thanks to all my colleagues from the CHYN, too numerous to all list here. Laurent and Roberto, I'll never forget the time we've spent in the field together. None of the projects on which we've worked would have turned out as successfully without your help and commitment. A special thanks to Hugo for helping me find my footing with HydroGeoSphere and parallel computing; the last year of

my thesis would have been much more painful without your precious help. I'd also like to thank Alexis, Ludovic, and Valentin for helping me get the models up and running on the cluster, saving me from the headache of endless model restarts on the moribund computer in my office. To the others not named here, but nevertheless not forgotten, thank you for your support, for the laughs, coffee breaks, outings, and overall amazing time spent together since 2019.

Finally, my thoughts go out to my family. To my mother, sister and brother-in-law, thank you for always being supportive and letting me know I could count on you under any circumstance, even if we're usually an ocean apart. And lastly to my partner Myriam, thank you for your encouragement, love and patience over these last years, I can't imagine finishing this work without your unfailing support.

Contents

1. INTRODUCTION 1

1.1. BACKGROUND 1
1.2. APPROACHES TO STUDYING SW-GW INTERACTIONS : INSIGHTS FROM EMERGING TRACER TECHNIQUES 2
1.3. CLOSING A CONCEPTUAL GAP : THE EXPLICIT SIMULATION OF TRACER CONCENTRATIONS IN PHYSICS-BASED MASS TRANSPORT MODELS 4
1.3.1. Linking physics-based models to high-dimensional parameter spaces 6
1.4. OBJECTIVES OF THE THESIS 6
1.4.1. Research objectives and structure of the thesis 7

2. VARIABLE ^{222}RN EMANATION RATES IN AN ALLUVIAL AQUIFER: LIMITS ON USING ^{222}RN AS A TRACER OF SURFACE WATER – GROUNDWATER INTERACTIONS 11

2.1. ABSTRACT 11
2.2. INTRODUCTION 11
2.2.1. Radon occurrence and use as a tracer of SW-GW interactions 12
2.2.2. Issues with ^{222}Rn end-member(s) 12
2.3. MATERIALS AND METHODS 15
2.3.1. Study site 15
2.3.2. Core sampling 16
2.3.3. 2019 pumping test 16
2.3.4. Radon incubation experiments 18
2.4. RESULTS 22
2.4.1. Rn incubation experiments 22
2.4.2. Dissolved Rn in groundwater and surface water 24
2.5. DISCUSSION 27
2.5.1. ^{222}Rn emanation rates and emanation factors 27
2.5.2. Linking emanation rates to ^{222}Rn in GW 28
2.6. CONCLUSIONS 31
2.7. ACKNOWLEDGEMENTS 32
2.8. APPENDICES 33
2.8.1. Appendix A : Radon emanation experiments – full results 34

3. 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 37

3.1. ABSTRACT 37
3.2. INTRODUCTION 38
3.2.1. ^{222}Rn as a tracer of GW age 39
3.3. MATERIALS AND METHODS 41
3.3.1. Synthetic 2-D numerical model 41
3.3.2. Quantification of apparent age bias 45
3.4. RESULTS 48
3.4.1. Modelled ^{222}Rn end-member equilibrium activities 48
3.4.2. Apparent age bias in steady-state conditions 50
3.4.3. Apparent age bias in transient conditions 54
3.5. DISCUSSION 59
3.5.1. Bias of apparent ^{222}Rn ages 59

3.5.2. Implications for use of ^{222}Rn as a tracer of GW age *61*
3.5.3. Limitations *63*
3.6. CONCLUSIONS 64
3.7. ACKNOWLEDGEMENTS 65
3.8. APPENDICES 66
3.8.1. Appendix A *66*
3.8.2. Appendix B *69*
3.8.3. Appendix C *70*
3.9. ADDENDUM 71
3.9.1. Introduction *71*
3.9.2. Field site *72*
3.9.3. Field data *73*
3.9.4. 2-D model of the field site *74*
3.9.5. Parameter estimation *76*
3.9.6. Modelling relative mean and apparent ^{222}Rn ages *77*
3.9.7. Results *78*
3.9.8. Summary and conclusion *79*

4. BRIDGING NOVEL TRACER APPLICATIONS AND HIGHLY-PARAMETERIZED, PHYSICS-BASED MODELS : HOW CAN THE EXPLICIT SIMULATION OF TRACER CONCENTRATIONS INFORM MODEL PARAMETERS AND PREDICTIONS? 81

4.1. ABSTRACT 81
4.2. INTRODUCTION 82
4.3. MATERIALS AND METHODS 85
4.3.1. Field site and acquired dataset *85*
4.3.2. GW flow and transport simulator *89*
4.3.3. Model parameters and parameter estimation *93*
4.3.4. Data worth assessment and predictions *98*
4.4. RESULTS 99
4.4.1. Post history-matched flow and transport models *99*
4.4.2. Parameter fields *102*
4.4.3. Predictive performance *105*
4.5. DISCUSSION 107
4.5.1. Constraints on model parameters *108*
4.5.2. Reductions in predictive uncertainty *109*
4.5.3. Limitations and recommendations for future work *111*
4.6. CONCLUSION AND IMPLICATIONS 112
4.7. ACKNOWLEDGEMENTS 113
4.8. APPENDICES 114
4.8.1. Appendix A : Solute definition for mass-transport model *114*
4.8.2. Appendix B : Estimation of missing He data *115*
4.8.3. Appendix C: Construction of prior geological models *119*
4.8.4. Appendix D : Validation with fluorescein measurements *124*
4.9. SUPPLEMENTARY MATERIAL TO CHAPTER 4 127
4.9.1. Summary of observations for history matching *127*
Simulated heads and tracer concentrations (figures) *128*
4.9.2. Prior and post-history-matched model parameter fields *134*

5. CONCLUSIONS 137

5.1. SUMMARY 137
5.1.1. Study 1 (Chapter 2) 137
5.1.2. Study 2 (Chapter 3) 138
5.1.3. Study 3 (Chapter 4) 138
5.1.4. Studies A1 and A2 (see Appendix A) 139
5.2. OUTLOOKS AND RECOMMENDATIONS FOR FUTURE WORK 140

6. APPENDICES 143

6.1. APPENDIX A1 : NEW EXPERIMENTAL TOOLS TO USE NOBLE GASES AS ARTIFICIAL TRACERS FOR GROUNDWATER FLOW 144
6.1.1. Abstract 144
6.2. APPENDIX A2 : EFFICIENT INJECTION OF DISSOLVED GAS TRACERS INTO RIVERS: A TOOL TO STUDY SURFACE WATER – GROUNDWATER INTERACTIONS 145
6.2.1. Abstract (tentative) 145
6.3. APPENDIX A3 : EXPLICIT SIMULATION OF ENVIRONMENTAL GAS TRACERS WITH INTEGRATED SURFACE AND SUBSURFACE HYDROLOGICAL MODELS 146
6.3.1. Abstract 146

7. BIBLIOGRAPHY 147

1. Introduction

1.1. Background

Globally, shallow groundwater (GW) withdrawn from alluvial aquifer systems represents the single largest source of water for human consumption and irrigation (Margat and van der Gun, 2013). Owing to the intimate links between shallow groundwater and surface water (SW), these two components of the water cycle must be considered as a single integrated resource, and issues pertaining to water supply and quality of the one invariably affect the other (Boano et al., 2014; Brunner et al., 2017; Winter et al., 1998).

The study of the interactions between SW and GW has been gaining attention over the previous decades, as growing impacts from climatic and anthropogenic stressors have emphasized the need to ensure safe and sustainable water supply, and to preserve the ecological health of river corridors (Boano et al., 2014; Boulton et al., 2010; Sprenger et al., 2019). Climate change-induced variations in the availability and seasonality of water resources in rivers and GW are expected over the next decades; these issues are compounded by the anticipated increase in water demand stemming from population growth and intensification of agricultural practices (Blöschl et al., 2019; Milly et al., 2005). Changes in water use and availability, combined with changes in seasonal weather patterns will likely be associated with water quality issues, as sources and/or composition of recharge water may change. As an example, increased summer drought frequency in Europe will decrease average river discharge, which will in turn decrease the dilution of pollutant loads in river water (stemming e.g. from wastewater effluents) that could be transported to aquifers, and subsequently pumped for consumption or irrigation (Blöschl, 2019; Sprenger et al., 2011).

Securing sufficient high-quality water supply over the next several decades rests therefore on improving our understanding of the complex exchange processes between SW and GW. Although the fundamental processes governing the exchanges between SW and GW are well established, characterizing the complex interplay between SW and GW in a unified framework at water management-relevant scales has remained elusive, notably due to difficulties in acquiring highly-resolved field data through which the highly heterogeneous nature of riverbed and alluvial aquifer properties, as well as the dynamic spatial and temporal patterns of SW-GW exchange, can be efficiently addressed (Boano et al., 2014, Brunner et al., 2017, Sophocleus, 2002; Sprenger et al., 2019).

1.2. Approaches to studying SW-GW interactions : insights from emerging tracer techniques

Several complementary field approaches are routinely employed to characterize the spatial patterns, dynamics, and fundamental controls on SW-GW exchange across a range of spatial and temporal scales. Common methods include (hydro)geophysical, (hydro)geological, chemical, and remote sensing characterization techniques, each capable of providing unique insights into SW-GW exchange processes (Brunner et al., 2017; Fleckenstein et al., 2010; Kalbus et al., 2006).

Amongst these field approaches, hydrological tracers are some of the most powerful tools available to characterize alluvial systems, as they are uniquely capable of providing an integrated description of subsurface transport process, and can directly inform subsurface mass fluxes, reaction rates, and velocity fields (Cook and Herczeg, 2000; Suckow, 2014; Schilling et al., 2019; Sprenger et al., 2019; Thiros et al. 2021). In water management contexts, some of the most important management-relevant quantities, such as the areal extent of capture zones, travel times between streams and pumping wells, or SW-GW mixing ratios to name a few, are all directly related to streambed and subsurface mass transport properties; the value of acquiring high-quality and high-resolution data informing these properties is evident.

Recent developments in analytical and field technologies allow the autonomous acquisition of continuous and on-site measurements of a variety of physical, biological, or chemical tracers at unprecedented spatial and temporal resolutions (Brunner et al., 2017; Sprenger et al, 2019). These tracer methods can now be routinely deployed in a range of environmental conditions, effectively eliminating persistent analytical and logistical constraints associated with more traditional lab-based techniques. Of the available tracer technologies informing SW-GW exchange dynamics in alluvial GW systems, we can cite those allowing continuous on-site measurements of (i) dissolved (noble) gases (including the radiotracer ^{222}Rn; Brennwald et al., 2016; Burnett et al., 2001; Chatton et al., 2017), (ii) bacterial cell concentrations (Besmer et al., 2016), (iii) stable isotopes of water ($\delta^{18}O$, δD) (Berman et al., 2009), (iv) fluorescent dye tracers (Schnegg, 2002), and (v) water physicochemical parameters including electric conductivity (EC) (e.g. Cirpka et al, 2007) and water temperature (Anderson, 2005; Constantz, 2008).

The ever-increasing access to portable high-resolution tracer measurement technologies has the potential to provide previously unattainable insights into the spatial and temporal dynamics of environmental systems. However, the potentially rich information content of tracer measurements can only be extracted if they are interpreted within a conceptual framework that is representative of the environmental processes they are meant to inform. This ostensibly self-evident statement must

be considered within the broader context of the field of tracer hydrology in its present state, as a variety of tracer interpretation techniques are still commonly employed, ranging from simple lumped-parameter ("black-box") models to the integration of tracer observations in fully integrated surface-subsurface, physics-based models of (reactive) tracer mass transport (e.g., Turnadge and Smerdon, 2014, Schilling et al., 2019).

Lumped parameter models (LPMs) and apparent tracer age models are amongst the most widely used tracer interpretation techniques to infer groundwater ages and mixing between different water components (Amin and Campana, 1996; Kazemi and Perrochet, 2006; Maloszewski and Zuber, 1982; Turnadge and Smerdon, 2014). In their simplest form, these models relate measured tracer concentrations to the convolution of historic tracer boundary conditions ("input function") and a transfer function defined by at most a few parameters, representing the distribution of GW ages at a given measuring point, whilst accounting for possible tracer decay or production (Maloszewski and Zuber, 1982). From these models, estimates of "mean" or "apparent" GW age and mixing ratios typically result, which can in turn inform estimates of GW flow velocities, SW-GW exchange fluxes, GW recharge, and transient GW storage (Turnadge and Smerdon, 2014 and references therein). Whilst these models may provide a first-order approximation of the internal dynamics of the flow system under investigation, they can be plagued by several sources of uncertainty, stemming notably from assumptions concerning (i) the *a priori* shape of the distribution of GW ages, (ii) stationarity of hydraulic/mass-transport conditions, and (iii) spatially constant tracer sources and sinks (Bethke and Johnson, 2008; McCallum et al., 2015). Used on their own, LPMs and apparent age models may be representative of only the simplest hydrogeological systems, and can hardly capture the complex age distributions expected in alluvial aquifer systems characterized by complex and transient water exchange patterns. In certain contexts, some of the aforementioned issues can be addressed through inversion of non-parametric ("shape-free") and non-stationary transfer functions (e.g., Cirpka et al., 2007; Liao et al., 2014; Massoudieh and Ginn 2011; McCallum et al., 2014); however directly extending the outputs of these data-driven models towards predictive analysis and uncertainty quantification in water management contexts is seldom straightforward, in that these models do not contain a mechanistic foundation supporting the link between (predicted) external forcings and environmental system behaviour. Nevertheless, these age and mixing models remain popular amongst water management professionals, who may not have the time, resources, or ability to implement more sophisticated tracer interpretation approaches. Although efforts have been devoted to describing physical mechanisms that generate residence time distributions represented by these models (e.g. Leray et al., 2016), there is usually limited data justifying their use in real-world contexts. No universal guidelines concerning the applicability and limitations of tracer age and mixing models in specific field

contexts generally exist, potentially exacerbating the uncertainties and bias associated with these tracer interpretation methods.

The information content of tracer-derived age or mixing estimates can be expanded by linking them to physics-based numerical models of environmental flow systems. These models are amongst the most powerful tools available for the characterization of water flow and solute transport in complex environments, such as coupled stream - alluvial aquifer systems (e.g. Paniconi and Putti, 2015). Through these models, several hydrologically relevant processes such as variably-saturated flow in heterogenous systems, solute mass transport, or dynamic SW-GW exchanges can be explicitly considered. Tracer-derived quantities can be used in the model parameter estimation process by comparing the transformed observations to their simulated proxies. For example, tracer-derived estimates of mixing ratios or apparent GW ages may be integrated into a multivariate calibration dataset, and used as observations against which model outputs resulting e.g. form particle tracking approaches or mean GW age simulations are compared (Schilling et al. 2019 and references therein). Models conditioned to transformed tracer observations can then be used as tools for decision support and forecasting. This two-step approach whereby tracer measurements are first transformed and then integrated into the parameter estimation process of physics-based models has been successfully implemented in a range of contexts, allowing significant reductions in parameter non-uniqueness and commensurate improvement in model predictive ability (see e.g. Schilling et al., 2019 and references therein).

Despite enjoying some success in improving model performance, the assimilation of transformed tracer data in numerical models is grounded in a conceptual disconnection between the complexity of processes simulated in physics-based models and the simplifying assumptions underpinning the transformation of tracer concentrations into secondary quantities. Indeed, non-trivial relationships between transformed tracer observations and simulated proxies may introduce model bias, as parameters are improperly adjusted to reproduce tracer-derived quantities through modelled processes not encapsulated by the conceptual model and simplifying assumptions that underly the transformed observations.

1.3. Closing a conceptual gap : the explicit simulation of tracer concentrations in physics-based mass transport models

To avoid the negative impact on numerical model performance stemming from uncertainties inherited from LPMs and other tracer models, the explicit simulation of tracer concentrations in physics-based numerical models has been proposed as an alternative framework through which the information content of tracer measurements could be optimally extracted, without the simplifying assumptions

inherent to apparent age or mixing models (Bethke and Johnson 2008; Suckow, 2014; Troldberg et al., 2007; Turnadge and Smerdon, 2014). Through these models, complex physical, chemical and biological processes affecting the spatiotemporal distribution of tracer concentrations, such as transient environmental conditions, heterogeneous subsurface properties, tracer accumulation/decay, chemical reactivity, mixing, etc., can be explicitly accounted for. By informing model parameters which control the same natural processes leading to the observed tracer concentrations, this methodology offers, in principle, a sound conceptual framework to extract the information content of tracer measurements. This approach has been gaining traction over the last decade, and several studies have successfully illustrated how the explicit simulation of tracer concentrations and their inclusion within a multivariate calibration framework can inform model parameters and subsequent predictions (e.g. Pool et al., 2015; Schaper et al., 2022; Thiros et al., 2021; Thiros et al., 2022). Conversely, some authors (e.g., Knowling et al., 2020) argue against the uncritical assimilation of tracer concentrations with mass-transport models, notably in contexts where sufficiently complex models (in terms of discretization, process modelling, and parameters) cannot be used to appropriately model tracer transport. Indeed, numerical models only represent a gross simplification of natural systems, and the reliability of tracer simulations is subject to the approximations and assumptions encoded within them. Damaging and potentially undetectable model bias may arise when tracers (or any other data type in general) are integrated into imperfect models i.e., those unable to adequately represent the complexity of flow and mass-transport processes in natural systems (Knowling et al., 2020; Moore and Doherty, 2005; Thiros et al., 2022; White et al., 2020).

Adequately representing these processes in environmental models comes at a cost in terms of system conceptualization, acquisition of field data, and access to computational resources. Indeed, models faithfully representing tracer mass transport in heterogenous and dynamic SW-GW systems require several parameters, boundary conditions, and field observations that constrain and inform the relevant spatiotemporal scales of flow and transport process (Doherty and Hunt, 2009; Schilling et al., 2019; Troldborg et al., 2007; Turnadge and Smerdon, 2014). The recourse to high-dimensional parameter schemes through which the heterogeneity of subsurface system properties, known to fundamentally control solute transport (e.g. Dagan, 1989; Linde et al., 2015; Renard and Allard, 2013), can be expressed adds further constraints on model conceptualization and computational requirements.

1.3.1. Linking physics-based models to high-dimensional parameter spaces

The use of highly-parametrized, physics-based flow and transport models has traditionally been limited by the computational burden associated with widely-used parameter estimation (PE) and uncertainty quantification (UQ) techniques, which may become numerically prohibitive as the number of adjustable parameters grows. Recent developments in ensemble-based data assimilation (DA) techniques provide very promising avenues for computationally efficient PE and UQ with (very) high-dimensional, physics-based environmental models (van Leeuwen, 2015). Amongst these DA algorithms, iterative ensemble smoothers (iES) are well-adapted to complex and nonlinear inverse flow and transport problems, as they assimilate all data simultaneously and scale exceptionally well with an increasing number of parameters (Chen and Oliver, 2013; Emerick and Reynolds, 2013; White, 2018).

High-resolution tracer technologies, efficient DA techniques for high-dimensional models, and high-performance computing infrastructure have all reached the required level of maturity so that various tracer concentrations can now be directly included for PE and UQ through their explicit simulation in highly-parameterized, physics-based, field-scale flow and mass transport models. Within this framework, the uncertainties accrued through the transformation of tracer measurements into secondary quantities can be avoided, and tracer observations can directly inform model parameters – and subsequent predictions – to which they are sensitive. Despite its recognized potential, the fusion of tracer measurements, physically-based environmental models, and data assimilation techniques within an explicit tracer simulation framework is still a comparatively marginal practice. With the rapid development and increasing access to novel high-resolution tracer methods, there is a need to understand and document how various types of tracers can inform (or otherwise) different model parameters and ultimately management-relevant quantities and forecasts.

1.4. Objectives of the thesis

In this thesis, we investigate the applicability of established and emerging noble gas tracer methods towards the study of SW-GW interactions in alluvial aquifer systems. These tracers are particularly well-suited to the study of SW-GW interactions as they are biogeochemically inert, may exhibit distinct concentrations in SW and GW, contain integrated information on the flow paths and travel times of water, and can now be continuously quantified at very high analytical resolutions in a variety of field settings (Brennwald et al., 2016; Kipfer et al., 2002). The relevance and limitations of different (noble gas) tracer interpretation methods, ranging from simple apparent age models to the assimilation of untransformed tracer measurements in highly-parameterized mass transport models, are systematically assessed.

The first part of this thesis (Chapter 2) is devoted to exploring and challenging the assumptions, shortcomings, and applicability of one the most popular noble gas tracer methods for studies of SW-GW interactions, the radon-222 (^{222}Rn) apparent age model. By extending results from laboratory measurements to the explicit simulation of ^{222}Rn concentrations in synthetic mass-transport models (Chapter 3), the scope of applicability of the ^{222}Rn age model could be systematically tested and ultimately reframed, simultaneously highlighting the potential usefulness of simple apparent age models on the one hand, and the need for more robust tracer interpretation methods on the other. Finally, high-resolution tracer measurements, efficient DA techniques, and the explicit simulation of tracer concentrations in a highly-parameterized ISSHM are brought together (Chapter 4), highlighting the advantages of this methodological framework for PE and UQ, and ultimately model-based decision support.

The successful combination of novel tracer methods, the explicit simulation of tracer concentrations in highly-parameterized physics-based models, and DA techniques illustrate the potential of fusing cutting-edge field and modelling techniques employing noble gases as environmental and artificial tracers.

1.4.1. Research objectives and structure of the thesis

The objectives pursued in this thesis revolve around the use of noble gases as quantitative tracers of SW-GW interactions and GW dynamics in alluvial systems. More specifically, three main research questions are pursued within this work :

1. When can ^{222}Rn be reliably used as a tracer of SW infiltration?
    - The ^{222}Rn apparent age model has been used for decades and is still one of the most widely used tracer methods in riverbank filtration (RBF) contexts, and provides quantitative information on GW residence times at scales of days to weeks, which are highly relevant for the management of RBF systems. This tracer method is based on the assumption of purely advective mass transport and homogenous subsurface ^{222}Rn production. Understanding the reliability of the apparent ^{222}Rn age model in more complex and representative conditions (e.g., assuming variability in the spatial distribution of ^{222}Rn production rates, GW mixing,…) is of clear management interest.
2. Can high-resolution noble gas measurements be integrated into an explicit modelling framework and inform model parameters and predictive capacity?
    - The advent of novel tracer methods, efficient data assimilation algorithms, and high-performance computing support the assimilation of untransformed tracer

measurements with highly-parameterized, physically-based models of coupled SW-GW systems. It is still an open question whether, and the extent to which, the assimilation of high-resolution noble gas measurements can improve model performance in terms of reducing the non-uniqueness of subsurface parameter fields and subsequent model predictions.

3. Can access to high-resolution noble gas measurement technology support the development of new experimental gas tracer methods?

    - Labelling GW or SW by injecting an artificial tracer is widely a widely used method in hydro(geo)logical studies. Noble gases present some ideal characteristics as tracers, as they are inert, colourless, and do not impact the quality or appearance of water. Technology allowing the continuous on-site measurement of dissolved gases provides new opportunities to use noble gases as tracers of SW infiltration and subsurface flow and transport dynamics.

In Chapters 2 and 3 of the thesis, the applicability of the ^{222}Rn apparent age model is explored. We present high-resolution measurements of ^{222}Rn emanation rates from recovered aquifer material, and highlight how spatially variable ^{222}Rn production rates can lead to misleading and even contradictory estimates of GW age. The significance of these results was extended by including ^{222}Rn emanation measurements in a series of mass-transport models, in which the sensitivity of ^{222}Rn signatures in GW to hydraulic and mass-transport parameters as well as to the spatial variability of ^{222}Rn production was systematically addressed. It is shown that hydrodispersive mixing and variable ^{222}Rn production rates will often lead to highly biased estimates of GW age, greatly limiting the scope of application of the ^{222}Rn dating technique.

Chapter 4 of this thesis is dedicated to exploring how the assimilation of tracer data can be used to constrain physics-based models in terms of reducing parameter non-uniqueness and improving model forecast reliability. A unique tracer dataset acquired during a noble gas injection experiment was employed for history-matching of a mass-transport model developed with the integrated surface-subsurface hydrological model HydroGeoSphere (Aquanty, 2023). The inclusion of high-resolution and diverse tracer data in a highly-parameterized data assimilation framework simultaneously informs hydraulic and mass-transport parameters, leading to large discrepancies in post-calibration parameter distributions compared to those resulting from the assimilation of hydraulic data alone. Consequently, it is shown how the inclusion of tracer measurements in the data assimilation process can lead to higher accuracy and reductions in predictive uncertainties of model outputs related to GW travel times, SW-GW exchange fluxes, and SW-GW mixing ratios.

Chapter 5 presents the conclusions and outlooks resulting from the work presented in this thesis. This part concludes with recommendations for future research concerning the use of noble gases as tracers of SW-GW interactions and their assimilation in physics-based numerical models.

The final portion of this thesis (Appendix A) presents a short summary of other research work undertaken directly as a part of the present Ph.D. project. Three abstracts (two from peer-reviewed journal articles, and one from a manuscript in preparation at the time of writing) are presented, which summarize work on the development of novel applications of noble gases (e.g., He, Kr, Xe) as artificial tracers of SW-GW interactions. Robust and cost-efficient gas-injection schemes were developed to label both SW and GW with noble gases. The high analytical sensitivity of continuous dissolved gas measurements allows large volumes of water to be efficiently labelled with small amounts of gas, even for continuous gas injections lasting weeks to months. These results pave the way for the development of new clean and safe gas tracer techniques at a variety of spatial and temporal scales.

2. Variable ^{222}Rn emanation rates in an alluvial aquifer: limits on using ^{222}Rn as a tracer of surface water – groundwater interactions

Chapter 2 has been published as : 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, Chemical Geology*, 599, DOI 10.1016/j.chemgeo.2022.120829

2.1. Abstract

Radon-222 (^{222}Rn) is routinely used as a tracer of surface water (SW) – groundwater (GW) interactions. However, methods that rely on this tracer to quantify SW-GW exchange fluxes, GW residence time, SW-GW mixing ratios, or GW flow velocities usually depend on the assumption of spatially constant ^{222}Rn production within aquifers and/or hyporheic sediments. In the present study, we measure the natural variability of ^{222}Rn emanation rates in a highly instrumented alluvial aquifer and underline some important issues in assuming homogenous ^{222}Rn production in alluvial systems. A robust experimental setup was used to determine ^{222}Rn emanation rates from sediments. High-resolution measurements reveal that ^{222}Rn emanation rates are highest within the first few meters below the surface. These results have significant implications for the interpretation of ^{222}Rn activities in GW, notably in the vicinity of infiltration zones, as changes in hydraulic conditions can lead to variations in ^{222}Rn signatures that can only be resolved with proper knowledge of the spatial distribution of ^{222}Rn production rates. In such contexts, the assumption of constant ^{222}Rn production leads to an inadequate understanding of SW infiltration patterns, and biased ^{222}Rn-based estimates of SW-GW exchange.

Keywords: Radon emanation, surface water – groundwater interactions, environmental tracers, groundwater residence time

2.2. Introduction

Interactions between surface water (SW) and groundwater (GW) in alluvial systems have been the subject of study for many years, and are at the focus of a range of fundamental issues, such as ensuring safe and sustainable water supply, understanding river water quality, as well preserving the ecological health of river corridors (Boano et al., 2014; Boulton et al., 2010; Brunner et al., 2017; Sophocleous, 2002). Several physical, chemical, and modelling approaches exist to characterize the spatial patterns, magnitudes, and associated timescales of GW – SW exchanges (Brunner et al., 2017) Environmental tracers are routinely used to characterize interactions between surface water and groundwater, as they can provide information on sources, pathways, residence times, and mixing of surface- and

groundwater. Tracers such as electric conductivity (Cirpka et al., 2007; Vogt et al., 2010), water temperature (Kurtz et al., 2014; Vogt et al., 2012), stable isotopes of water (Meredith et al., 2009; Yang et al., 2012), and noble gases (Beyerle et al., 1999; Popp et al., 2021) have all proven to be excellent indicators of spatial and temporal patterns of SW-GW exchange.

The radioactive noble gas ^{222}Rn is well-suited to study SW-GW dynamics, as concentrations in groundwater are often orders of magnitude higher than those found in surface water bodies. Therefore, ^{222}Rn has been used for decades as a tracer of SW-GW interactions, such as for the characterization of stream water infiltration into adjacent alluvial aquifers (Bertin and Bourg, 1994; Hoehn and Cirpka, 2006; Hoehn and Von Gunten, 1989; Hoehn et al., 1992; Popp et al., 2021; Stellato et al., 2013; Vogt et al., 2010), GW discharge into streams, lakes, wetlands, and seas (Burnett and Dulaiova, 2006; Cartwright and Gilfedder, 2015; Cook et al., 2006; Cook et al., 2008; Dimova et al., 2013; Dulaiova et al., 2008; Kluge et al., 2012; Smerdon et al., 2012), as well as hyporheic exchange dynamics (Bourke et al., 2014; Cranswick et al., 2014) .

2.2.1. Radon occurrence and use as a tracer of SW-GW interactions

^{222}Rn is a ubiquitous radioactive tracer (half-life ~3.8 d), which occurs in groundwater primarily as the result of the decay of matrix-bound (and to a lesser extent dissolved) radium-226 (^{226}Ra) [e.g., Cecil and Green (2000)]. Several factors influence the magnitude of ^{222}Rn input into groundwater, such as ^{226}Ra activity of the mineral matrix, the location of ^{226}Ra in or on mineral grains, pore structure, aquifer mineralogy, grain size distribution, as well as water saturation (Cecil and Green, 2000). In particular, the specific surface area of unconsolidated sediments, which is controlled by the fraction of fine material, may exert a strong influence on ^{222}Rn release to groundwater, as a large proportion of ^{226}Ra tends to occur in mineral coatings or is adsorbed to mineral surfaces (Cecil and Green, 2000; Semkow, 1990; Tricca et al., 2001). Variations in the parameters described above can occur at any scale, which can lead to variations in ^{222}Rn production and ^{222}Rn activities. ^{222}Rn however remains a powerful tracer of SW-GW interactions, as measured activities in groundwater are commonly orders of magnitude higher than those measured in surface waters.

2.2.2. Issues with ^{222}Rn end-member(s)

Despite the breadth of settings in which ^{222}Rn can be used as a tracer, many studies share a common feature: that is the definition of a single ^{222}Rn end-member for GW, which is thought to be representative of water at radioactive equilibrium with the host aquifer matrix (Cecil and Green, 2000). The definition of a single end-member implicitly relies on the assumption of constant ^{222}Rn production at the scale of a given investigation. Definition of such an end-member is not always

straightforward, and needs to be justified by local hydrogeological and geochemical conditions. In practice, this means either sampling GW at a location where ^{222}Rn activities are at secular equilibrium with the aquifer matrix, or measuring the ^{222}Rn production rates of aquifer materials representative of the study zone. However, there exists a considerable body of literature indicating that ^{222}Rn activities in GW, and therefore the magnitude of ^{222}Rn production, can vary by an order of magnitude or more, even at small scales - e.g., at tens to hundreds of meters - within a given aquifer (Cecil and Green, 2000; Cook, 2013; Folger et al., 1996; Torgersen et al., 1990).

In studies where ^{222}Rn is used as a tracer of GW discharge to SW bodies (e.g., streams), ^{222}Rn emanation rates of streambed sediments are routinely measured in order to characterize hyporheic exchange and/or parafluvial flow (e.g., Bourke et al., 2014; Cartwright and Hofmann, 2016; Cook et al., 2006; Lamontagne and Cook, 2007). Conversely, ^{222}Rn emanation measurements of aquifer sediments are more seldom undertaken, in part because of logistical difficulties in acquiring aquifer material at sufficient spatial resolutions. Therefore, it is common practice to define either a single or zonally constant GW end-member(s) based on measured ^{222}Rn activities within the aquifer (e.g., using the mean or maximum ^{222}Rn activity measured in observation wells). This approach may lead to systematic biases if ^{222}Rn end-member activities are based on measurements made at some distance from the river; indeed, in contexts of both SW infiltration (i.e. losing rivers) and GW exfiltration (i.e. gaining rivers), ^{222}Rn activities in GW will in large part be determined by ^{222}Rn emanation of riparian sediments, which may differ from those elsewhere in the aquifer (Cook, 2013). Although many studies explicitly state the limitation in assuming a constant GW end-member, systematic exploration and quantification of the variability of ^{222}Rn production in GW near the stream/aquifer interface is not routinely undertaken. Unless riparian sediments systematically differ from those further away from the stream, estimates of GW discharge over large river reaches are only moderately sensitive to small-scale variations in GW ^{222}Rn activities, and uncertainties in the ^{222}Rn mass-balance model outputs can be minimized by sampling GW at different points along the river, and using a mean value over a large region (Cook, 2013; McCallum et al., 2012). However, neglecting local variations in ^{222}Rn emanation may be problematic when studying processes at smaller spatial scales, such as estimating GW residence times in bank filtration contexts, or quantifying GW discharge into streams over small reaches.

In this study, we measure the natural variability of ^{222}Rn emanation rates within an alluvial aquifer, and identify some pitfalls of assuming homogenous tracer input when using ^{222}Rn as an indicator of SW-GW exchange.

We first present a simple and robust method allowing the measurement of ^{222}Rn emanation rates from bulk aquifer sediments (> 4-5 kg sample size) with minimal sample preparation and sediment disturbance. High-resolution measurements of ^{222}Rn emanation rates were carried out on core samples recovered from the experimental Aeschau study site near Bern, Switzerland. Depth profiles of ^{222}Rn emanation rates within an alluvial aquifer were established at four locations in the immediate vicinity of a major wellfield used for drinking water production. Finally, we present measurements of ^{222}Rn activities in groundwater taken over the course of a large pumping experiment, which sought to characterize GW protection zones as well as stream-GW interactions near the wellfield. Changes in ^{222}Rn activities are linked to changes in hydraulic conditions at the field site (bank filtration fluxes and GW residence times), and are interpreted within the framework of measured ^{222}Rn emanation rates. We demonstrate that without proper knowledge of the distribution of ^{222}Rn emanation rates, ^{222}Rn signals may be ambiguous and lead to an inadequate characterization of SW-GW systems.

To our knowledge, this work represents the most thorough and high-resolution investigation of the distribution of ^{222}Rn emanation rates in an alluvial aquifer setting. Although this study is focused on measurements of ^{222}Rn emanation rates in a bank filtration context, the implications of heterogenous ^{222}Rn input can be extended to all fields where ^{222}Rn is used as a tracer of SW-GW interactions.

2.3. Materials and Methods

2.3.1. Study site

The Aeschau field site is located near the outlet of the Upper Emme Valley, a pre-alpine alluvial catchment near the northern edge of the Emmental Alps (see Figure 2.1). The catchment covers an area of ~200 km^2, and is drained by two rivers, the Emme and the Rötenbach tributary. These highly dynamic rivers are thought to be the main source of recharge to the alluvial aquifer of the Upper Emme Valley (Käser and Hunkeler, 2016). In the lower part of the catchment, the valley floor consists of quaternary deposits of gravel and sand with variable proportions of silt, which form a highly conductive unconfined alluvial aquifer (Käser and Hunkeler, 2016). At the study site, the aquifer has an average thickness of approximately 25 m (up to 46 m), and the alluvial plain is 200 to 400 m wide (Würsten, 1991). The average hydraulic conductivity of the aquifer is on the order of 500 m/d, with values above 1'000 m/d having been noted (Schilling et al., 2017; Würsten, 1991).

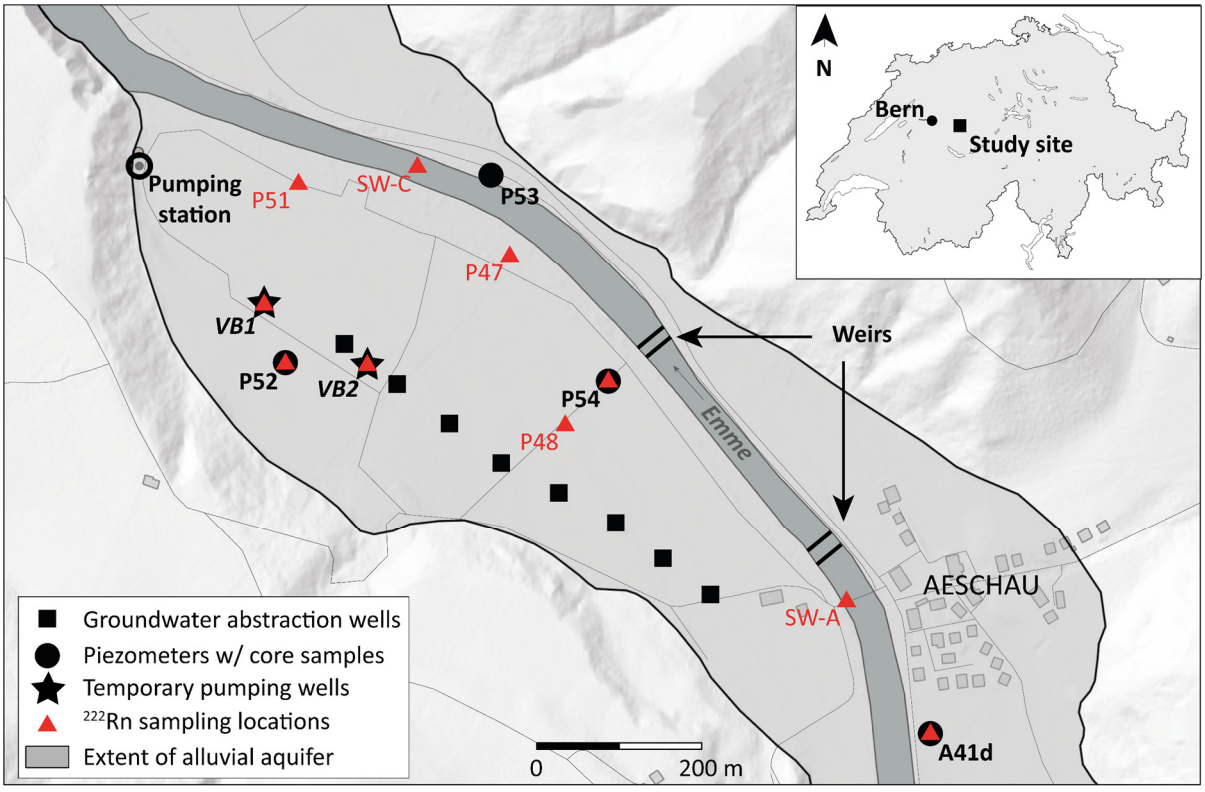

Figure 2.1 Location of the Aeschau study site (Canton of Bern, Switzerland), showing the eight GW abstraction wells, two temporary pumping wells (VB1 and VB2), piezometers from which core samples were recovered for incubation experiments (A41d, P52, P53, and P54), as well as ^{222}Rn sampling locations during the 2019 pumping experiment.

The alluvial valley has been extensively studied and is instrumented with a dense network of piezometers, SW gauging stations, and weather stations (Käser and Hunkeler, 2016). Over twenty piezometers, mostly screened within the first 10-15 meters below the surface, are located near Aeschau. Discharge of the Emme River is monitored by the Federal Office for the Environment (FOEN)

at 5-minute intervals at two gauging stations near the study site: the first is 5.5 km upstream [*Eggiwil* station (FOEN, 2023)], and the second is 5 km downstream, immediately after the confluence of the Emme and nearby Ilfis rivers (*Emmenmatt* station; FOEN, 2021).

A major wellfield is located at the Aeschau site, where eight abstraction wells pump groundwater at an average rate of 26 $m^3 min^{-1}$, which provides about 25% of the drinking water for the Swiss capital Bern. These pumping wells, coupled with highly dynamic river conditions and abrupt changes in hydraulic gradients next to the weirs control a complex pattern of SW-GW interactions with alternating zones of SW infiltration and GW discharge along the reach of the Emme River (Käser and Hunkeler, 2016; Schilling et al., 2017).

2.3.2. Core sampling

Sediment samples were retained from borehole drilling at four locations at the study site (P52, P53, P54, A41d; see Figure 2.1) for the ^{222}Rn incubation experiments (see section 2.3.4). 150 to 200 mm diameter cores were extracted with a percussion drill from the surface to depths between 8 and 25 meters (Table 2.1). In total, 55 metres of cored material were recovered from the study site. Cores were separated into 1-m subsamples and, owing to the unconsolidated nature of the aquifer material, had been thoroughly disturbed during packaging and transport. In-situ sediment structure could not be preserved and vertical resolution for the Rn incubation experiments is therefore limited to 1 meter.

Table 2.1 Summary of available core samples

Piezometer	Drilling date	Cored section (m below surface)	Core diameter (mm)	Number of 1-m subsections
P52	Dec. 2018	0 – 12	150	12
P53	Dec. 2018	0 – 10	150	10
P54	Dec. 2018	0 – 8	150	8
A41d	Oct. 2014	0 - 25	200	25

2.3.3. 2019 pumping test

2.3.3.1. ^{222}Rn measurements and ^{222}Rn age

A major six-week pumping test was carried out at the Aeschau site from January 15 to February 26, 2019, using the existing GW abstraction wells and two temporary pumping wells (VB1 and VB2, see Figure 1). The pumping test aimed to characterize SW-GW interactions at the study site, and to define a new protection zone related to the planned upgrades of the pumping facilities (Popp et al., 2021). Pumping rates were increased incrementally from 0 to 36 $m^3 min^{-1}$, before returning to normal operating conditions (i.e., ~26 $m^3 min^{-1}$). Over the course of the pumping experiment, discrete groundwater samples were collected twice per week at ten sampling locations (six piezometers, the

two temporary pumping wells, and two river sampling points – see Figure 2.1) and analysed for ^{222}Rn content. Note that three separate samples were taken at location A41d, as it is in fact a nest of three piezometers, each screened at different depths (A41d-1: 14-15 m; A41d-2: 19-20 m; A41d-3: 24-25 m). All water samples were collected in 250-ml glass vials after adequate purging of the observation wells, and analysed with the RAD H_20 accessory for the RAD7 within three days of sampling (Durridge, 2020b). The built-in WAT-250 protocol for the RAD7 was employed; after sample aeration, alpha decays of ^{222}Rn daughter ^{218}Po were counted for a period of 20 minutes (Durridge, 2020b). For each sample, ^{222}Rn activities and associated 2-σ uncertainties (derived from decay statistics) were subsequently corrected for radioactive decay between the times of sampling and analysis.

2.3.3.2. Estimation of residence time with ^{222}Rn

GW dating with ^{222}Rn is based on the first-order increase in ^{222}Rn activities as a function of residence time. If we assume spatially constant input of ^{222}Rn into groundwater, ^{222}Rn activities will vary according to the following equation:

$$\frac{d}{dt}C_{Rn} = P - \lambda_{Rn}C_{Rn} \qquad (1)$$

Where C_{Rn}[Bq l^{-1}] is the activity of ^{222}Rn in GW, P [Bq l^{-1} d^{-1}] the ^{222}Rn aquifer production rate per unit pore volume, and λ_{Rn} [d^{-1}] the ^{222}Rn decay constant. The general solution to eq. *[1]* is:

$$C_{Rn}(t) = C_{Rn,0}e^{-\lambda_{Rn}t} + C_{Rn,eq}\left(1 - e^{-\lambda_{Rn}t}\right) \qquad (2)$$

Where $C_{Rn,0}$ [Bq l^{-1}] is the ^{222}Rn activity of surface water at the time of infiltration, and $C_{Rn,eq}$ [Bq l^{-1}] the steady-state equilibrium ^{222}Rn activity of GW:

$$C_{Rn,eq} = \frac{P}{\lambda_{Rn}} \qquad (3)$$

Eq. *[2]* can be used to derive the apparent radiometric age a_{app} [d] of a water sample with activity C_{Rn}, assuming both $C_{Rn,eq}$ and $C_{Rn,0}$ are known:

$$a_{app} = t = -\frac{1}{\lambda_{Rn}} ln\left(\frac{C_{Rn,eq} - C_{Rn}}{C_{Rn,eq} - C_{Rn,0}}\right) \qquad (4)$$

Eq. *[4]* is frequently used for dating of recently infiltrated GW in bank filtration contexts (e.g., Cranswick et al., 2014; Frei and Gilfedder, 2021; Hoehn and Cirpka, 2006; Pittroff et al., 2017; Schilling et al., 2017), however strictly speaking is only valid if the following assumptions are met:

(i) ^{222}Rn input per unit volume of GW is spatially constant within the aquifer;

(ii) ^{222}Rn loss to the unsaturated zone is constant, and;

(iii) there is no mixing of waters of different ages within a given GW sample (due to intra-aquifer mixing or the interception of multiple flowlines during sampling).

In general, any of these assumptions is challenged, and need in principle to be addressed when using ^{222}Rn to estimate GW residence times.

2.3.4. Radon incubation experiments

2.3.4.1. Sample preparation

The ^{222}Rn incubation experiments follow a modified version of the method described in Corbett et al. (1998). Each 1-m core section contained approximately 10-20 kg of sediments, from which 6 to 7 kg were taken for analysis using a coning and quartering technique (Pitard, 1993). Each subsample was then oven-dried at 95 °C for at least 48 hours, until a constant weight was reached. Approximately 4 to 5 kg were further drawn from the dried sediments for the incubation experiments; the rest was put aside for subsequent grain size and ^{226}Ra content analysis (see section 2.3.4.3).

The dried sediment samples were then carefully placed in 5-litre borosilicate-glass bottles. As the mouths of these containers are rather small (ID 29.5 mm), some larger stones were cut into smaller pieces with a circular diamond saw. This technique was chosen rather than bulk-sample grinding to preserve as much as possible sediment texture, which can affect radon emanation.

Distilled water was added incrementally over a period of one week to allow full saturation of the core samples. Bottles were then filled to the brim and sealed with a PBT screw cap and a PTFE-coated septum. Sample bottles were then flipped upside-down and stored at room temperature for at least 30 days, at which point ^{222}Rn activities are effectively at secular equilibrium with matrix-bound ^{226}Ra. The mass of added sediment and volume of water were accurately measured by weighing the samples at each step. The relative masses of the empty bottle, of the bottle filled with water only, and of the water/sediment mixture were used to calculate the density of the sediment matrix.

Significant microbial growth had been observed over the course of some preliminary experiments (e.g., through the formation of microbial mats and gas bubbles in the sample bottles). To prevent excessive gas build-up and partitioning of ^{222}Rn into the gas and organic phases, subsequent samples were stabilized by adding small amounts of quaternary ammonium compounds (QACs) and isothiazolinones, two broad-spectrum biocides. This procedure effectively suppressed microbiological activity for the >30-day incubation period. The biocide mixture was diluted to well below its critical micelle concentration to avoid possible partitioning of ^{222}Rn into the micellar phase (see Vane and Giroux, 1999).

Including duplicates, a total of 61 samples were prepared from 54 of a possible 55 cored intervals. Analysis of one sample (A41d 2-3 m) was not carried out, as two sample bottles broke prior to analysis, and insufficient core material remained for preparation of a third.

2.3.4.2. Sample analysis

Radon in water was quantified by alpha spectrometry with the RAD7 detector (Durridge, 2020a). Following the incubation period, sealed bottles were gently shaken for approximately one minute and allowed to sit for an additional 30 minutes or until the fine sediments had settled. Samples were then unsealed, and approximately 1.5 litres of water were immediately transferred to a 1.5-l PET bottle with a small membrane pump (flow rate ≈ 75 ml s^{-1}), as shown in Figure 2.2. An 8 mm hole had been pierced near the bottom of the PET bottle, through which a small piece of PTFE tubing was inserted and connected to the pump. The rims of the hole were sealed with a two-component epoxy cement. The PET bottle was connected to the RAD7 with the Durridge Big Bottle System (Durridge, 2018) prior to water transfer. Once the PET bottle was sufficiently filled, the tubing at its bottom was clamped and a three-way valve at the RAD7 outlet, through which air displaced by the addition of water in the PET bottle could escape, was switched to create a closed air loop (see Figure 2.2).

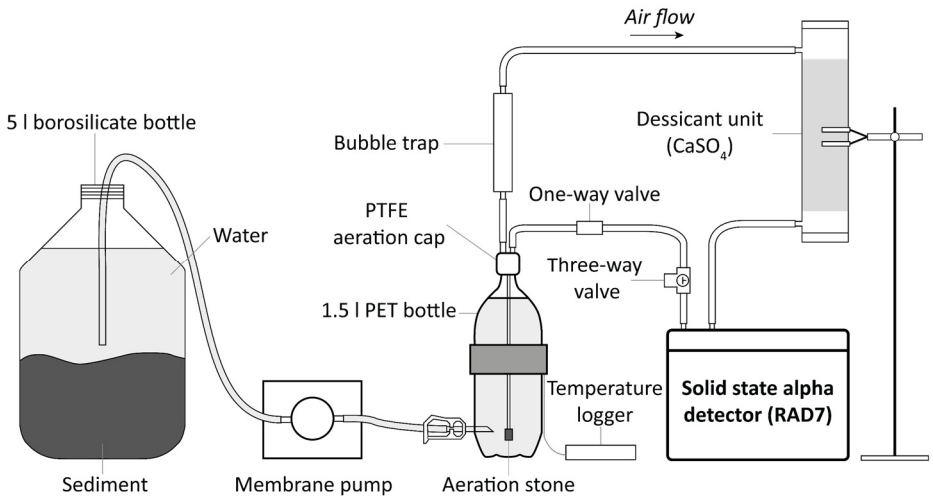

Figure 2.2 Schematic experimental setup for the measurement of dissolved ^{222}Rn concentrations in the water-sediment mixture.

The water sample was aerated for 45 minutes; ^{218}Po (^{222}Rn daughter) disintegrations were then counted by the RAD7 for a period of ~90 minutes. Once the test was completed, the PET bottle was disconnected from the experimental apparatus and weighed to determine the volume of water.

The ^{222}Rn in air activities provided by the RAD7 were converted to ^{222}Rn in water with the following mass-balance relationship (Schubert et al., 2006):

$$C_{water,i} = \frac{C_{air}}{kV_{sample}}\left[V_{sample} + kV_{air}\right] \quad (5)$$

Where $C_{water,i}$ [Bq l^{-1}] is the initial ^{222}Rn concentration in water *before* aeration, C_{air} [Bq l^{-1}] the measured ^{222}Rn-in-air activity, V_{sample} [l] the volume of water in the PET bottle, V_{air} [l] the volume of the air-loop, and k [-] the air-water partition coefficient of ^{222}Rn ($= C_{air}/C_{water}$ at equilibrium), calculated after Weigel (1978) and water temperatures recorded by the temperature logger. Assuming secular equilibrium between ^{226}Ra and ^{222}Rn activities, the sediment ^{222}Rn emanation rate E_m [Bq kg^{-1}] is related to $C_{water,i}$ by the following relationship:

$$E_m = C_{water,i}\frac{V_{water,tot}}{m_{grain}} \quad (6)$$

Where $V_{water,tot}$ [l] the total water volume in the 5-litre bottles and m_{grain} [kg] the total mass of the dried sediment sample. The emanation rate E_m is given in Bq kg^{-1} of dry sample. Uncertainties in the emanation rates were computed by error propagation of radioactive decay statistics, sensitivity of the RAD7 (~5%), total sediment mass, total water volume in the 5-litre containers, water volume in the PET bottle, volume of the air loop, and temperature during sample analysis. Conservative values for the above lead to 2-σ uncertainties of ^{222}Rn-in-water and ^{222}Rn emanation rates on the order of 10-15%.

Theoretical ^{222}Rn equilibrium activities in GW ($C_{Rn,eq}$ in eq. *[2]*) are related to the measured ^{222}Rn emanation rates E_m by the following relationship (e.g., Cook et al. (2006), eq. *[14]*):

$$C_{Rn,eq} = \frac{P}{\lambda_{Rn}} = E_m\rho_{grain}\frac{1-\varphi}{\varphi} \quad (7)$$

Where ρ_{grain} [kg l^{-1}] is the density of the mineral phase and φ [-] is the bulk sediment porosity. As the porosities of the recovered core samples could not be accurately measured, theoretical equilibrium ^{222}Rn activities in water were approximated by assuming a porosity of 0.2, typical of unconsolidated alluvial sediments (see section 2.4.1).

2.3.4.3. Sieve and ^{226}Ra analysis

In order to investigate any correlation between the grain size distribution and measured ^{222}Rn emanation rates, approximately 1 kg of dried sample material was sieved according to ASTM standard C136 / C136M-19 (ASTM, 2019). Mesh openings of 8 mm (5/16 in), 2 mm (No. 10), 1 mm (No. 18), 500 µm (No. 35), 250 µm (No. 60), 125 µm (No. 120) and 63 µm (No. 230) were used.

Finally, twenty-two samples were prepared for analysis of ^{226}Ra activity. The remainder of the dried sediment was crushed to a fineness of <20 µm with a vibratory disc mill. Approximately 10 g of

powdered sample were sealed in polycarbonate tubes and sent to the Eawag in Dübendorf, Switzerland, where ^{226}Ra activity was determined by gamma spectroscopy. For these samples, the emanation factor ε [-] (i.e., the ratio of ^{222}Rn released into water and matrix ^{226}Ra activity), was computed as:

$$\varepsilon = \frac{E_m}{A_{Ra}} \qquad (8)$$

Where A_{Ra} [Bq kg^{-1}] is the ^{226}Ra activity in sediment samples.

2.4. Results

2.4.1. Rn incubation experiments

^{222}Rn emanation rates and the fraction of fine material (< 63 µm) as a function of depth for all locations are shown in Figure 2.3. The *minimum* recorded depth of the water table is also shown to discriminate between unsaturated and saturated zones. In all four piezometers, the shallowest GW levels were recorded on Jan. 15, 2019, immediately prior to the start of the pumping test, at which point pumping had ceased for five days. Summary statistics of measured ^{222}Rn emanation rates and emanation factors are shown in Table 2.2 and Table *2.3*. Summary results of measured ^{222}Rn concentrations in observation wells, pumping wells, and surface water samples are shown in Table 2.4. Full results for the ^{222}Rn incubation experiments, the ^{226}Ra content, measured matrix densities, as well as the mass fractions of fine material f_{fines} (< 63 µm) are given in Appendix A. Results of the sieve analysis are provided in supplement S1. Finally, full results of field ^{222}Rn measurements are provided in supplement S2.

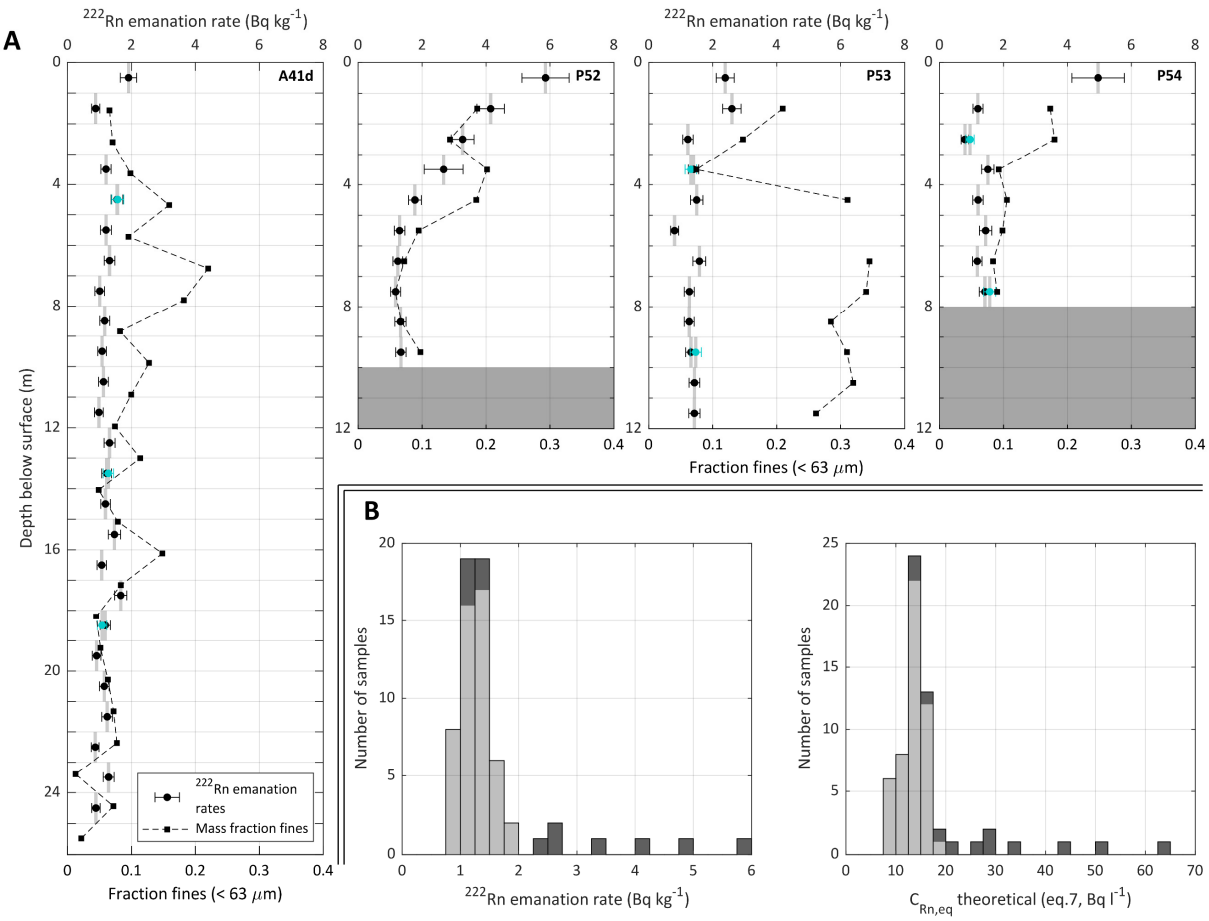

Figure 2.3 Results of the ^{222}Rn incubation experiments. ^{222}Rn emanation rates and mass fraction of fine material (< 63 µm) as a function of depth at locations A) A41d, P53, P52, and P54. Duplicate samples are shown in blue (n = 7); B) Histogram of measured ^{222}Rn emanation rates for all sediment samples and theoretical ^{222}Rn equilibrium activities in GW computed from the measured ^{222}Rn emanation rates and grain densities, assuming a constant porosity of 0.2. Darker colours represent samples taken from intervals above the minimum water table depth

Emanation rates are highest at shallow depths and tend to decrease to stable values within a few meters. At boreholes A41d, P53 and P54, higher emanation rates are present at depths up to two meters (E_m > 1.9 Bq kg^{-1}). At borehole P52 however, there is a regular decrease in emanation rates from 5.9 Bq kg^{-1} just below the surface to 1.3 Bq kg^{-1} at a depth of 5 m. Emanation factors at P52 follow a similar pattern, ranging from 0.2 near the surface to 0.1 at 5 m (see Appendix A). For samples from greater depths however, emanation rates are remarkably consistent both as a function of depth and across all four sampling locations (see Table 2.2). Mean emanation rates within the saturated zone at boreholes P52, P53, and P54 range from 1.34 to 1.39 Bq kg^{-1}, only slightly higher than those observed at borehole A41d (1.20 Bq kg^{-1}). The magnitude of ^{222}Rn emanation rates and emanation factors are consistent with values reported in other studies focused on soils in alluvial environments (Baretto et al., 1972; Mullinger et al., 2009).

For the calculation of theoretical ^{222}Rn equilibrium activities in GW (eq. *[7]*), a value of porosity (φ) of 0.20 ± 0.05 was chosen. This value is representative of sandy gravel aquifers (e.g., Fetter, 2001), and also consistent with previous field observations at the Aeschau site (Blau and Muchenberger, 1997; Schilling et al., 2017; Würsten, 1991)

Table 2.2 Summary of measured ^{222}Rn emanation rates for (i) all samples, and (ii) samples from intervals below the minimum water table (WT) depth. Theoretical ^{222}Rn equilibrium activities ($C_{Rn,eq}$) in the saturated zone were calculated using the mean emanation rates with eq. [7], assuming a porosity of 0.20 ± 0.05. $\mu_{1/2}$: median value

	^{222}Rn emanation rate (Bq kg^{-1})										
Well	*(i) All samples*					*(ii) Below minimum WT depth*					
	n	Mean ± σ	$\mu_{1/2}$	Min	Max	*n*	Mean ± σ	$\mu_{1/2}$	Min	Max	$C_{Rn,eq}$ [Bq l^{-1}]
P52	10	2.4 ± 1.6	1.6	1.2	5.9	7	1.4 ± 0.2	1.3	1.2	1.8	15 ± 4
P53	12	1.5 ± 0.4	1.4	0.8	2.6	10	1.3 ± 0.2	1.4	0.8	1.6	14 ± 4
P54	8	1.6 ± 1.2	1.3	0.8	5.0	5	1.4 ± 0.2	1.4	1.2	1.6	15 ± 4
A41d	24	1.2 ± 0.2	1.2	0.9	1. 9	22	1.2 ± 0.2	1.2	0.9	1.7	13 ± 4

Table 2.3 Summary of measured ^{222}Rn emanation factors (i) all samples, and (ii) samples from intervals below the minimum water table (WT) depth

	^{222}Rn emanation factor (-)							
Piezometer	*(i) All samples*				*(ii) Below minimum WT depth*			
	n	Mean (± σ)	Min	Max	*n*	Mean (± σ)	Min	Max
P52	6	0.12 ± 0.05	0.07	0.20	4	0.10 ± 0.03	0.07	0.13
P53	0	-	-	-	0	-	-	-
P54	7	0.07 ± 0.01	0.05	0.09	5	0.08 ± 0.01	0.06	0.09
A41d	9	0.08 ± 0.04	0.05	0.19	9	0.08 ± 0.04	0.05	0.19

The fraction of fine material f_{fines}, which is often thought to control ^{222}Rn emanation, is highly variable and ranges from 1% to 35%. However, with the notable exception of borehole P52, there is no apparent association between f_{fines} and the measured emanation rates. At P52, the five samples with f_{fines} > 10 % (depth 0 - 5 m) are those with the highest emanation rates.

2.4.2. Dissolved Rn in groundwater and surface water

Measured ^{222}Rn activities show distinct spatial and temporal trends at the study site. At both pumping wells (VB1 and VB2) and in observation wells located at a distance from the Emme River (P48, P52, P54), ^{222}Rn activities remained essentially stable over the course of the pumping experiment (See supplement S2). This was also the case in piezometer A41d, which is on the opposite bank of the river, and not expected to be influenced by riverbank filtration. ^{222}Rn activities in both A41d-3 and P52 are thought to be characteristic of GW at equilibrium with aquifer sediments, as GW residence times are well above 20 days at these locations (Schilling et al., 2017). Note that ^{222}Rn activities (mean values of 12.6 ± 0.6 and 14.4 ± 0.9 Bq l^{-1} at A41d-3 and P52 respectively) are in full agreement with the ^{222}Rn emanation rates measured within the saturated zone at these two locations (theoretical ^{222}Rn equilibrium activities of 13 ± 4 [A41d] and 15 ± 4 Bq l^{-1} [P52], see Table 2.2).

^{222}Rn activities in samples collected from piezometers located near the left bank of the Emme River (P47 and P51) are more sensitive to changes in pumping rates (Figure 2.4). ^{222}Rn activities at P51 tended to decrease at higher pumping rates, accurately tracking increased SW infiltration in the vicinity of the observation well. A temporary interruption in pumping caused by a power failure occurred between Feb. 03 and Feb. 04, 2019, and immediately resulted in an increase of ^{222}Rn activities to pre-pumping values. Finally, ^{222}Rn activities quickly returned to background values once the experiment had ended.

Conversely, ^{222}Rn activities at P47 show an inverse trend to what can be expected in a riverbank filtration context: after remaining stable for the first half of the pumping experiment, ^{222}Rn activities significantly increased in conjunction with pumping rates, before returning to near-background values once normal pumping was resumed. The highest activities measured during the entire 2019 sampling campaign were measured at P47 (16.6 ± 1.7 Bq l^{-1} and 16.1 ± 1.6 Bq l^{-1} on Feb. 21 and Feb. 25 respectively), at which time pumping rates had reached their maximum values (36 m^3 min^{-1}).

Table 2.4 Summary of measured ^{222}Rn activities in water. Measurements span from just before the beginning of the pumping experiment until one month after its end (Jan. 14 – Mar. 20, 2019).

Sampling location	Well depth (m)	Screened interval (m)	**Dissolved ^{222}Rn activities (Bq l^{-1})**			
			n	Mean (± σ)	Min	Max
A41	10	1-10	11	11.5 ± 0.6	10.6	12.5
A41d-2	20	14-15	15	12.1 ± 0.9	10.4	13.8
A41d-3	25	24-25	15	12.6 ± 0.6	10.9	13.8
P47	6.5	3.5-6.5	15	12.4 ± 1.8	10.1	16.6
P48	11	3.8-9.8	14	11.2 ± 0.6	10.2	12.1
P51	9.25	2.25-8.25	16	8.6 ± 1.9	5.8	11.5
P52	10	3-9	16	14.4 ± 0.9	12.3	15.6
P54	8	6-8	16	10.9 ± 0.9	9.6	12.3
VB1	43	6-41	13	13.0 ± 0.9	11.9	14.6
VB2	28	6-26	12	13.2 ± 0.9	11.7	14.9
SW-A	-	-	14	0.2 ± 0.1	0.0	0.5
SW-C	-	-	8	0.4 ± 0.2	0.1	0.7

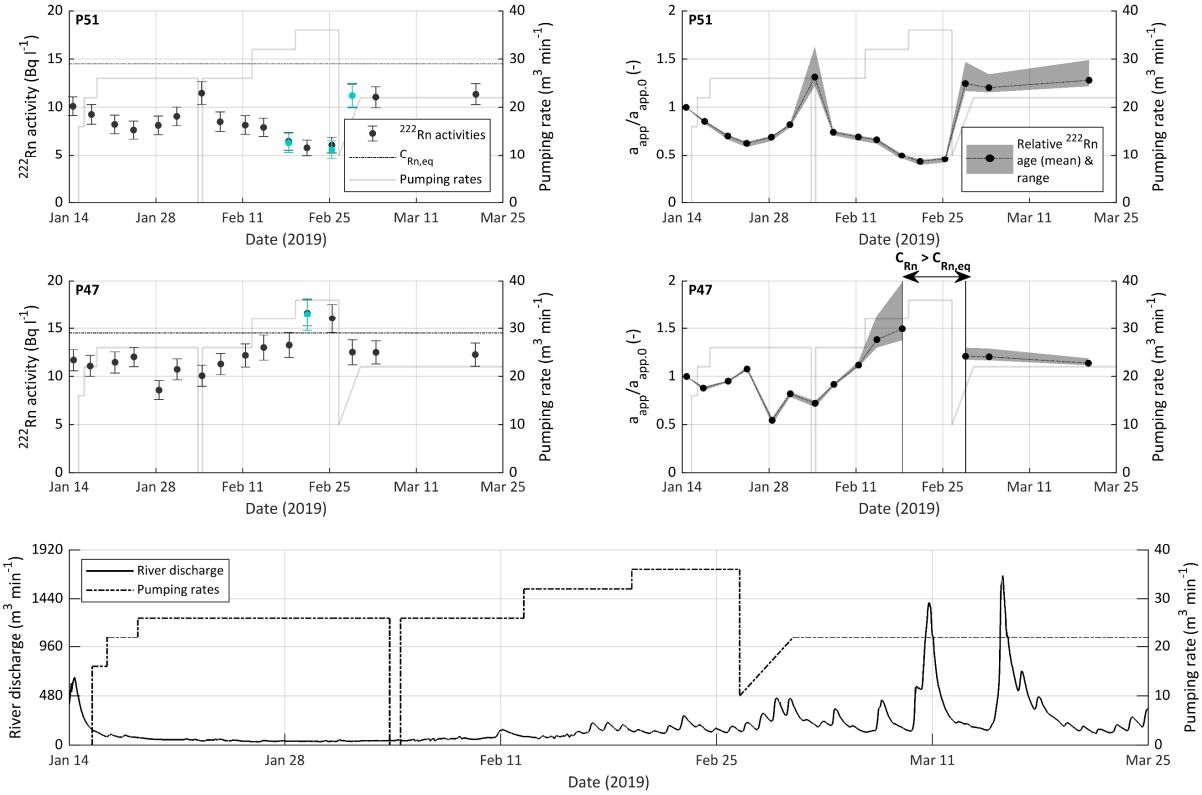

Figure 2.4 Results of ^{222}Rn measurements in piezometers P47 and P51 during the 2019 pumping experiment. Measured ^{222}Rn activities (with 2-σ uncertainties from radioactive decay statistics) and relative apparent ^{222}Rn ages in A) piezometer P51 and B) piezometer P47. Duplicate samples are shown in blue. Ages were computed using eq. [4], assuming that infiltrating water is radon-free ($C_{Rn,0} = 0$). Apparent ages were capped at 20 days and were normalized to the apparent age computed for the initial measurements (Jan. 14, 2019). Uncertainties in apparent ages are computed from uncertainties in measured ^{222}Rn activities and in the definition of the end-member. Note that measured ^{222}Rn activities in P47 were above those of the end-member on Feb. 21 and 25, 2019, and apparent ages could not be computed C) River discharge (from Eggiwil measurement station) and GW pumping rates.

For illustrative purposes, estimates of apparent ^{222}Rn ages (eq. *[4]*) relative to the apparent ages of measurements made on Jan. 14, 2019 (prior to the start of the pumping experiment) at P47 and P51 are also shown in Figure 2.4. Conversion of ^{222}Rn activities to apparent ages requires the definition of a ^{222}Rn end-member representative of the study site. This can be done in one of two ways: (i) by calculating theoretical equilibrium ^{222}Rn concentrations based on ^{222}Rn emanation measurements using eq. *[7]*; and (ii) by sampling GW thought to be at equilibrium with the aquifer sediments (i.e., with a residence time > 5 half-lives on ^{222}Rn, or about 20 days). Method (i) requires knowledge of sediment porosity, which was unavailable for the recovered core samples (see section 2.3.4.2). Therefore, the ^{222}Rn end-member was based on the mean ^{222}Rn activities measured at sampling location P52 (14.5 ± 1.0 Bq l^{-1}).

In P51, apparent ages declined by approximately 50% over the course of the pumping experiment (notwithstanding a sudden increase following the pump malfunction), before returning to values slightly above background. In contrast, at P47, following a moderate decline in the initial phases of the pumping test, apparent ages increased by more than 50%. In the latter part of the experiment, ^{222}Rn

activities rose to values above those of the end-member, and could not be computed. ^{222}Rn activities and thus apparent ages returned to near-background values once normal pumping was resumed, albeit at moderately higher values than those before the start of the pumping test.

Finally, ^{222}Rn activities in surface water were negligible compared to those in GW (0.2 ± 0.1 Bq l^{-1} [SW-A] and 0.4 ± 0.2 Bq l^{-1} [SW-C]). These low activities reflect the predominantly losing river conditions over the course of the pumping experiment.

2.5. Discussion

2.5.1. ^{222}Rn emanation rates and emanation factors

Two main features emerge from the high-resolution measurements of ^{222}Rn emanation rates at the Aeschau site: (i) emanation rates vary as a function of depth in the near subsurface, with highest rates in the topsoil, and (ii) decrease to stable values within a few meters; at depths below the water table, ^{222}Rn emanation rates are remarkably consistent both as a function of depth and sampling location. Although measurements presented here do not provide complete coverage of emanation rates at the Aeschau site, there is no indication of any large-scale heterogeneity in ^{222}Rn input (apart from somewhat larger emanation rates [~10%] within the Aeschau plain [P52, P53 and P54] compared to A41d). The observed decrease of emanation rates and emanation factors in shallow alluvial sediments has been noted elsewhere (e.g., Mullinger et al., 2009), and is likely related to the presence of weathering products in the topsoil (e.g., clays, Fe-Mn hydroxides, thorium precipitates), which exhibit favourable radiochemical and radionuclide sorption properties (Porcelli and Swarzenski, 2003; Tricca et al., 2001). The highest emanation rates within the first meter below the surface is likely related to higher proportions of organic mineral coatings, which can lead to high emanation rates/factors (Greeman and Rose, 1996).

It should be noted that these emanation rates were measured from sediments that were disturbed during packaging, transport, and sample preparation (see section 2.3.2). In other words, in-situ sediment structure could not be fully replicated in the 5-l glass bottles. In particular, we expect in-situ porosity to be slightly lower than in the sample bottles, owing to sediment compaction. This could lead to overestimates of true ^{222}Rn emanation rates, as larger pore spaces in the sample result in decreased embedding of recoiled ^{222}Rn atoms into adjacent grains, and higher release to the liquid phase (e.g.). Although in-situ porosity is likely only moderately lower than that of the samples, measurements likely mark an upper bound on true ^{222}Rn emanation rates.

No clear association can be made between measured emanation rates and the proportion of fines (< 63 µm) in samples from boreholes A41d, P53 and P54. Conversely, samples in the first 5-m below the

surface from P52 exhibit a strong association between emanation rates and proportion of fines. This suggests local variations in the magnitude of chemical weathering of the aquifer in the near subsurface; in the vicinity of P52 for example, weathering processes in the near subsurface could have led to higher amounts of surface-bound ^{226}Ra on finer grains, leading in turn to higher ^{222}Rn emanation factors and emanation rates. The consistency in emanation rates at greater depths suggests that aquifer sediments within the study area share common geochemical features regarding the distribution of ^{238}U-series isotopes, despite large variations in textural and mineralogical properties. This observation seems to reflect the common origin and depositional patterns of glaciofluvial sediments in the alluvial valley, or may also result from a relatively homogenous distribution of secondary deposits of ^{226}Ra-bearing minerals within the deeper aquifer or on grain coatings. Barring a more thorough characterization of sediment chemistry and mineralogy at the study site, linking ^{222}Rn emanation and aquifer properties remains nevertheless uncertain.

2.5.2. Linking emanation rates to ^{222}Rn in GW

^{222}Rn activities measured in groundwater are sensitive to both ^{222}Rn production and the prevailing hydraulic conditions. The characterization of ^{222}Rn emanation reduces the ambiguity in identifying some zones influenced by recently infiltrated surface water with low residence times (< 15 days; e.g., VB1, VB2, P48, P54), insofar as measured ^{222}Rn activities are lower than equilibrium activities as suggested by ^{222}Rn emanation measurements. ^{222}Rn activities measured in observation wells P52 and A41d are in full agreement with emanation rates at these locations, and, based on the consistency of ^{222}Rn input at depths of a few meters or more, are plausibly representative of the range of equilibrium ^{222}Rn activities at depth at the Aeschau site. Therefore, the modest but significant differences in ^{222}Rn activities from these observation wells likely stem from variations in emanation rates between these two locations. Mean ^{222}Rn activities at both pumping wells and observation wells P48 and P54 are somewhat lower than those found at P52 (mean of 13.0 ± 0.9 Bq l^{-1} [VB1], 13.2 ± 0.9 Bq l^{-1} [VB2], 11.2 ± 0.6 Bq l^{-1} [P48], and 10.9 ± 0.9 [P54], see Table 2.4), and, although mostly insensitive to changes in pumping rates and river discharge, may reflect the presence of a young GW component, i.e. recently infiltrated SW (Popp et al., 2021).

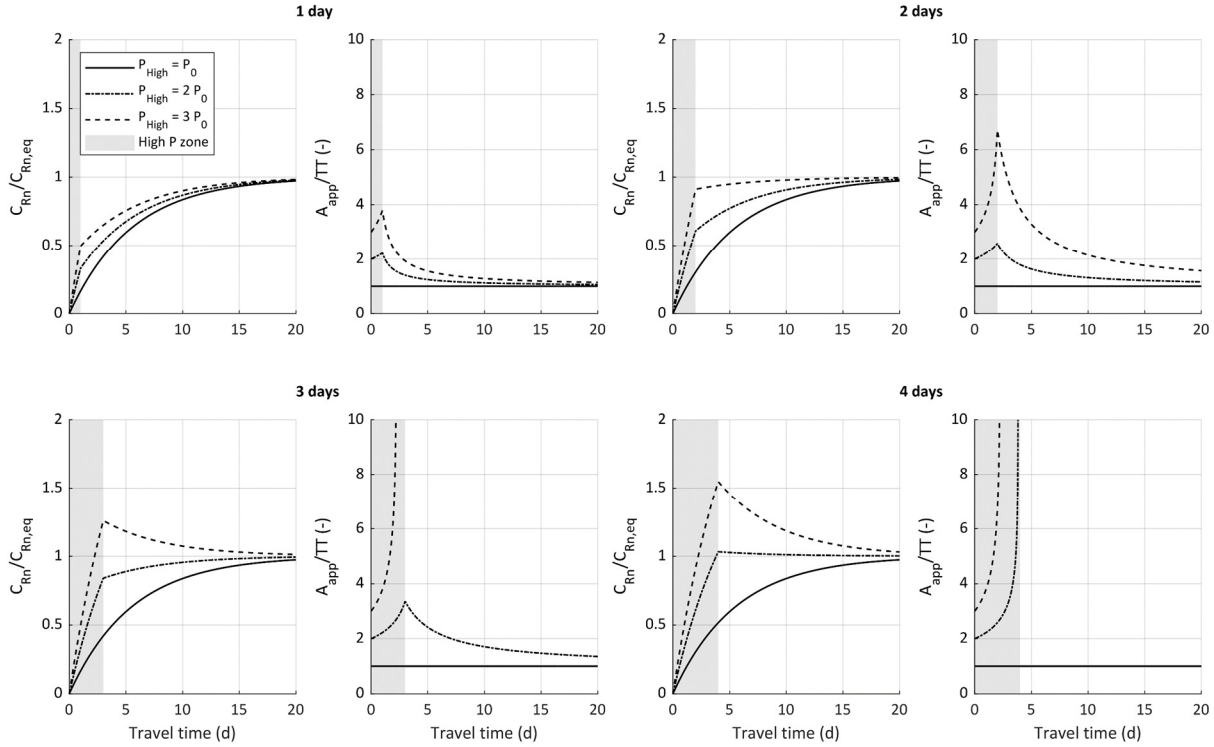

Figure 2.5 Effect of flow through a zone of high ^{222}Rn production (P_{High}) during infiltration on ^{222}Rn activities (C_{Rn}) and apparent ^{222}Rn ages (A_{app}). A hypothetical parcel of water flows for one to four days (1-D piston flow, no initial radon) through a zone where ^{222}Rn production is higher by a factor of 2 or 3 compared to that within the rest of the aquifer (P_0). Apparent ages were computed using eq. [4], assuming an end-member activity $C_{Rn,eq} = P_0/\lambda$. TT: advective travel time.

However, counterintuitive ^{222}Rn signals may arise if GW is influenced by higher emanation rates within the shallow subsurface. Indeed, infiltrating water may become quickly enriched in ^{222}Rn if it flows through a zone of high ^{222}Rn input even for periods as short as several hours, which will lead to biased estimates of GW age and infiltration patterns (Figure 2.5). This effect would impede GW dating with ^{222}Rn if a given sample is composed of water having flowed through a zone of high ^{222}Rn production for several hours or days.

Some puzzling effects of heterogenous ^{222}Rn input can be observed by comparing ^{222}Rn signals at piezometers P51 and P47, both located adjacent to the left bank of the Emme River. ^{222}Rn activities at P51 follow a typical trend in bank filtration contexts: increased pumping rates lead to increased infiltration of SW, higher GW flow velocities, and a commensurate decrease of residence times, measured ^{222}Rn activities, and computed apparent ages (Figure 2.4).

However, the opposite trend is observed at piezometer P47. Indeed, ^{222}Rn activities at this location rose to the highest values measured at the Aeschau site, suggesting that a portion at least of sampled water at P47 had recently travelled through a zone of particularly high ^{222}Rn input, as ^{222}Rn activities were above those measured in background piezometers A41d and P52. An uncritical appraisal of the increase in ^{222}Rn activities as a function of pumping rates could be interpreted as an effective increase in GW residence times at this location, e.g., as a result of local modifications of the GW flow field,

diverting more regional (and thus older) groundwater to sampling location P47. However, this interpretation is deemed unlikely based on the prevailing hydrogeological conditions at the site during the pumping test. Indeed, during normal pumping conditions, P47 is located in an area of predominant GW exfiltration towards the stream (Käser and Hunkeler, 2016; Schilling et al., 2017), and GW ^{222}Rn activities are representative of GW near equilibrium with aquifer sediments, as suggested by measured ^{222}Rn activities before and after the pumping experiment. However, during the period of highest pumping rates, measured GW heads as well as numerical modelling both suggest that GW in the vicinity of P47 is largely composed of recently infiltrated SW with residence times of a few days at most, likely originating from infiltration zones upstream of the nearest weirs (see Figure 1) (Popp et al., 2021). A simultaneous increase in both ^{222}Rn activities and in the fraction of recently infiltrated SW at P47 suggests that at least a portion of sampled GW had transited through a zone of high ^{222}Rn production. It is suggested that the combined effects of increased pumping rates and river discharge (see Figure 2.4C) lead to increased infiltration upstream of the weirs into the shallow parts of the aquifer, where emanation rates are likely significant (see Figure 2.6). A larger proportion of water sampled at P47 having transited through a zone of high ^{222}Rn input can lead to the observed increase in ^{222}Rn activities in the latter part of the experiment. Therefore, SW infiltration through zones of high ^{222}Rn emanation may occasionally lead to paradoxical observations of increased ^{222}Rn activities despite increased SW infiltration and decreased GW travel times.

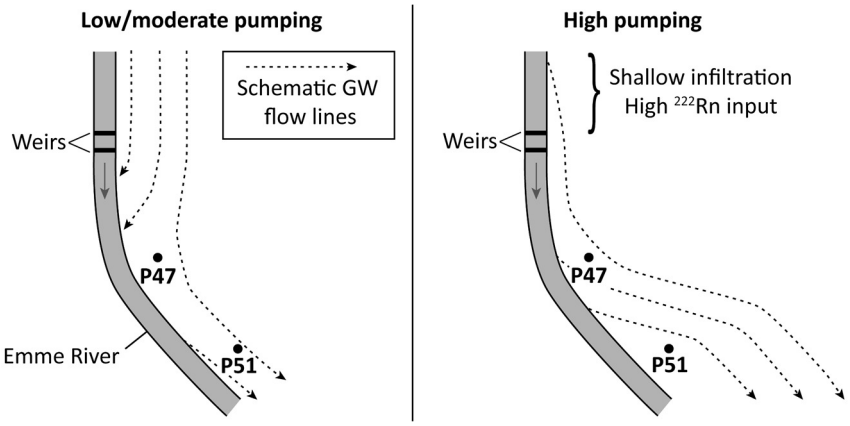

Figure 2.6 Conceptual model of GW infiltration and flow during low and high pumping conditions in the vicinity of observations wells P47 and P51 (see Figure 2.1 for location within the study site).

This interpretation is based on the general observation of high emanation rates near the surface, and decreasing ^{222}Rn emanation rates to stable values at depth. However, it is important to note that only a small portion of the aquifer was sampled, and information on the distribution of ^{222}Rn emanation rates within the alluvial system remains modest. Although no systematic variations in the distribution of emanation rates were observed at the Aeschau site, the presence of zones of higher or lower ^{222}Rn emanation can not be excluded. Moreover, there is no data relating aquifer properties (such as

hydraulic conductivity (K)) to observed ^{222}Rn emanation. Indeed, as GW will preferentially flow through transmissive zones, ^{222}Rn signatures in GW will be disproportionately affected by emanation rates within these portions of the aquifer. As K is intimately related to sediment properties (e.g., mineralogy, chemistry, grain-size distribution) it is reasonable to expect that emanation rates in high-K zones may systematically differ from those within the rest of the aquifer matrix. Nevertheless, these results highlight the importance of characterizing ^{222}Rn production of each component of SW-GW systems, be it hyporheic sediments, the riparian zone, as well as throughout the aquifer at greater distances from streams. Lack of proper knowledge on the heterogeneity of ^{222}Rn input marks a serious limitation on the use of ^{222}Rn as a tracer of SW-GW interactions and GW age, and may severely bias the understanding of the hydrogeological system under consideration.

2.6. Conclusions

The main goal of this study was to carry out a high-resolution characterization of the natural variability of ^{222}Rn emanation rates within an alluvial aquifer, and to explore some issues arising from variable ^{222}Rn input when investigating SW-GW interactions. To this end, a simple and robust experimental setup was presented, which allows accurate and consistent determination of ^{222}Rn emanation rates from bulk sediment samples. Some main advantages of this method are (i) minimal need for sample preparation and transformation (e.g. grinding), allowing preservation of sediment texture, (ii) the ability to measure ^{222}Rn emanation rates of samples > 5 kg for a large range of sediment compositions (fine- to coarse-grained unconsolidated sediments with clasts up to ~25 cm), (iii) high analytical sensitivity despite short counting times (~1.5 hours), and (iv) no deleterious microbial activity (e.g. leading to the formation of gas overpressures in sealed samples) for the >30 day incubation period by addition of safe and commercially-available biocides. If even larger stones are present in recovered sediments, such as may be the case in piedmont alluvial deposits, further sample preparation may be necessary. This could include removal of larger stones for incubation experiments: as they only contribute moderate amounts of ^{222}Rn, emanation rates of bulk sediment samples could be reliably determined from the finer sediment fractions provided the weight of both the larger stones and finer fractions are accurately measured.

Measurements of ^{222}Rn emanation rates were carried out on aquifer sediments recovered within the vicinity of a dynamic pre-alpine river (Emme River), which exhibits complex patterns of alternatively losing and gaining conditions. Emanation rates are consistent at depths greater than a few meters within the aquifer, whereas ^{222}Rn production is significantly higher at shallower depths. This observation greatly impacts the interpretation of measured ^{222}Rn activities in GW, particularly near SW/GW interfaces, where GW may be in contact with highly ^{222}Rn-emanating sediments. In such

contexts, changes in hydraulic conditions and flow patterns (e.g., caused by water table fluctuations, storm-flow events, GW pumping) can lead to locally high ^{222}Rn production in shallow GW. In such cases, the assumption of constant ^{222}Rn production (and the existence of a single ^{222}Rn end-member in GW) breaks down. Therefore, conceptual models based on defining a single ^{222}Rn end-member in GW may become inadequate to extract estimates of GW-SW exchanges, whether it be GW discharge to streams, SW infiltration fluxes, or GW residence time.

Due to its proximity to the river channel, the riparian zone will exert a significant influence on ^{222}Rn concentrations in GW, in contexts of both SW infiltration (losing rivers) and GW exfiltration (gaining rivers). Consequently, knowledge of ^{222}Rn emanation rates in riparian zone sediments is crucial to properly characterise the sources of ^{222}Rn in GW and SW and to understand how changes in measured ^{222}Rn activities are tied to changes in hydraulic conditions.

A viable alternative to the use of simple conceptual models could be to explicitly simulate of ^{222}Rn in physically-based numerical models informed by emanation measurements. These have the ability to realistically account for most processes affecting measured tracer concentrations without the need for some of the simplifying assumptions described above. Such exercises have seldom been undertaken, although they may offer powerful insights into assessments of GW age, SW/GW exchanges and mixing ratios.

In any case, the interpretation of ^{222}Rn in GW requires accounting for the variability of ^{222}Rn production, which is most critical as ^{222}Rn is one of very few environmental tracers sensitive to SW-GW exchange processes at timescales of days to a few weeks - often most relevant to water management in alluvial systems.

2.7. Acknowledgements

All data used in this study can be found in Appendix A and supplements S1 and S2. We would like to thank the *Wasserverbund Region Bern* (WVRB) for access to their installations and for providing pumping data. We would also like to thank Roberto Costa and Laurent Marguet for their valuable assistance in the field and laboratory. This study was funded by the Swiss National Science Foundation (grant number 200021_179017). The authors would like to thank Ian Cartwright as well as one anonymous reviewer for providing helpful comments that improved this manuscript.

2.8. Appendices

Supplementary data for this article, including raw laboratory and field data, can be accessed from its online version at doi.org/10.1016/j.chemgeo.2022.120829

2.8.1. Appendix A : Radon emanation experiments – full results

Table 2.5 Results of ^{222}Rn emanation experiments with 2-σ uncertainties. Duplicate samples are shown in italic. Dashes indicate that measurements were not undertaken

Sample ID	^{222}Rn emanation rate (Bq kg^{-1})	^{226}Ra activity (Bq kg^{-1})	Emanation factor (-)	Matrix density (g cm^{-3})	Mass fraction fines (≤ 63 µm)
P54 0-1 m	4.96 ± 0.83	-	-	2.60 ± 0.09	-
P54 1-2 m	1.20 ± 0.16	18.5 ± 2.7	0.07 ± 0.02	2.68 ± 0.02	0.17
P54 2-3 m	0.79 ± 0.11	17.3 ± 2.8	0.05 ± 0.01	2.68 ± 0.01	0.18
P54 2-3 m D	*0.96 ± 0.13*	-	-	*2.68 ± 0.01*	*0.18*
P54 3-4 m	1.51 ± 0.19	21.4 ± 2.2	0.07 ± 0.02	2.68 ± 0.02	0.09
P54 4-5 m	1.20 ± 0.16	16.1 ± 2.3	0.07 ± 0.02	2.68 ± 0.02	0.11
P54 5-6 m	1.44 ± 0.19	16.4 ± 2.3	0.09 ± 0.02	2.68 ± 0.02	0.10
P54 6-7 m	1.18 ± 0.14	18.8 ± 2.5	0.06 ± 0.01	2.68 ± 0.01	0.08
P54 7-8 m	1.42 ± 0.18	17.5 ± 2.5	0.08 ± 0.02	2.68 ± 0.02	0.09
P54 7-8 m D	*1.57 ± 0.19*	-	-	*2.68 ± 0.01*	*0.09*
P53 0-1 m	2.40 ± 0.28	-	-	2.62 ± 0.02	-
P53 1-2 m	2.60 ± 0.29	-	-	2.67 ± 0.02	0.21
P53 2-3 m	1.22 ± 0.17	-	-	2.68 ± 0.02	0.15
P53 3-4 m	1.39 ± 0.16	-	-	2.69 ± 0.01	0.07
P53 3-4 m D	*1.32 ± 0.18*	-	-	*2.68 ± 0.02*	*0.07*
P53 4-5 m	1.50 ± 0.19	-	-	2.68 ± 0.02	0.31
P53 5-6 m	0.81 ± 0.12	-	-	2.58 ± 0.02	-
P53 6-7 m	1.58 ± 0.20	-	-	2.67 ± 0.02	0.35
P53 7-8 m	1.27 ± 0.16	-	-	2.71 ± 0.01	0.34
P53 8-9 m	1.26 ± 0.15	-	-	2.70 ± 0.01	0.28
P53 9-10 m	1.32 ± 0.18	-	-	2.69 ± 0.01	0.31
P53 9-10 m D	*1.47 ± 0.18*	-	-	*2.69 ± 0.01*	*0.30*
P53 10-11 m	1.43 ± 0.17	-	-	2.71 ± 0.02	0.32
P53 11-12 m	1.42 ± 0.18	-	-	2.72 ± 0.01	0.26
P52 0-1 m	5.86 ± 0.75	-	-	2.64 ± 0.05	-
P52 1-2 m	4.16 ± 0.44	20.4 ± 2.4	0.20 ± 0.04	2.65 ± 0.02	0.19
P52 2-3 m	3.27 ± 0.35	26.7 ± 2.1	0.12 ± 0.02	2.67 ± 0.02	0.14
P52 3-4 m	2.67 ± 0.59	19.9 ± 2.6	0.13 ± 0.04	2.64 ± 0.11	0.20
P52 4-5 m	1.78 ± 0.20	16.5 ± 2.5	0.11 ± 0.02	2.69 ± 0.01	0.18
P52 5-6 m	1.30 ± 0.17	-	-	2.67 ± 0.02	0.09
P52 6-7 m	1.24 ± 0.15	13.4 ± 2.2	0.09 ± 0.02	2.69 ± 0.01	0.07
P52 7-8 m	1.17 ± 0.16	-	-	2.68 ± 0.02	0.06
P52 8-9 m	1.33 ± 0.18	-	-	2.69 ± 0.02	0.07
P52 9-10 m	1.34 ± 0.17	19.0 ± 2.3	0.07 ± 0.02	2.70 ± 0.01	0.10
A41d 0-1 m	1.91 ± 0.26	-	-	2.66 ± 0.02	-
A41d 1-2 m	0.88 ± 0.13	-	-	2.68 ± 0.02	0.07
A41d 2-3 m	***Sample broken***	-	-	2.68 ± 0.02	0.07
A41d 3-4 m	1.21 ± 0.16	-	-	2.68 ± 0.02	0.10
A41d 4-5 m	1.55 ± 0.18	18.7 ± 2.5	0.08 ± 0.02	2.70 ± 0.02	0.16
A41d 4-5 m D	*1.57 ± 0.19*	-	-	*2.69 ± 0.02*	*0.18*
A41d 5-6 m	1.21 ± 0.16	-	-	2.68 ± 0.02	0.10
A41d 6-7 m	1.31 ± 0.17	-	-	2.69 ± 0.02	0.22
A41d 7-8 m	1.01 ± 0.15	17.0 ± 2.6	0.06 ± 0.01	2.69 ± 0.02	0.18
A41d 8-9 m	1.17 ± 0.15	-	-	2.69 ± 0.02	0.08
A41d 9-10 m	1.08 ± 0.14	-	-	2.69 ± 0.01	0.13
A41d 10-11 m	1.13 ± 0.15	17.6 ± 2.13	0.06 ± 0.01	2.70 ± 0.02	0.10
A41d 11-12 m	0.98 ± 0.14	-	-	2.68 ± 0.02	0.07
A41d 12-13 m	1.32 ± 0.17	-	-	2.69 ± 0.02	0.11
A41d 13-14 m	1.22 ± 0.16	19 ± 2.3	0.06 ± 0.01	2.71 ± 0.01	0.05
A41d 13-14 m D	*1.28 ± 0.16*	-	-	*2.70 ± 0.02*	*0.05*
A41d 14-15 m	1.19 ± 0.15	-	-	2.68 ± 0.01	0.08
A41d 15-16 m	1.47 ± 0.19	7.7 ± 3.7	0.19 ± 0.10	2.70 ± 0.02	0.15

A41d 16-17 m	1.07 ± 0.15	-	-	2.69 ± 0.02	0.08
A41d 17-18 m	1.67 ± 0.19	16.4 ± 2.2	0.10 ± 0.02	2.69 ± 0.01	0.04
A41d 18-19 m	1.18 ± 0.15	-	-	2.68 ± 0.02	0.05
A41d 18-19 m D	*1.08 ± 0.14*	-	-	2.68 ± 0.01	*0.06*
A41d 19-20 m	0.91 ± 0.13	18.1 ± 2.6	0.05 ± 0.01	2.68 ± 0.02	0.06
A41d 20-21 m	1.15 ± 0.16	-	-	2.69 ± 0.02	0.07
A41d 21-22 m	1.24 ± 0.17	17.3 ± 2.3	0.07 ± 0.02	2.69 ± 0.02	0.08
A41d 22-23 m	0.87 ± 0.12	-	-	2.69 ± 0.01	0.01
A41d 23-24 m	1.29 ± 0.17	-	-	2.69 ± 0.02	0.07
A41d 24-25 m	0.89 ± 0.13	21.8 ± 2.4	0.04 ± 0.01	2.70 ± 0.02	0.02

3. 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

Chapter 3 has been published as : Peel M., Delottier, H., Kipfer, R., Hunkeler, D., Brunner, P. (2023), *Exploring the reliability of 222Rn as a tracer of groundwater age in alluvial aquifers : insights from the explicit simulation of variable 222Rn production*, *Water Research*, DOI 1016/j.watres.2023.119880

3.1. Abstract

Knowledge of groundwater residence times (GRT; the time elapsed since surface water infiltration) between losing rivers and pumping wells is crucial for the 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 heterogeneous environments.

3.2. 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 for the 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 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.

3.2.1. ^{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).

3.2.1.1. ^{222}Rn dating method

GW dating with ^{222}Rn is based on the 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). Equation *(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, the 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).

3.2.1.2. Apparent age bias and GW mixing

When tracer concentrations vary non-linearly with respect to time, any process that leads to the 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 the 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]):

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

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), the mixing of waters of different ages will lead to an apparent age younger than the mean age (McCallum et al., 2013).

3.3. Materials and methods

3.3.1. 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 the 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 riverbed and bank was included at the top-left of the model domain (channel depth: 4 m; width: 10 m; bank slope : 30 degrees). 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).

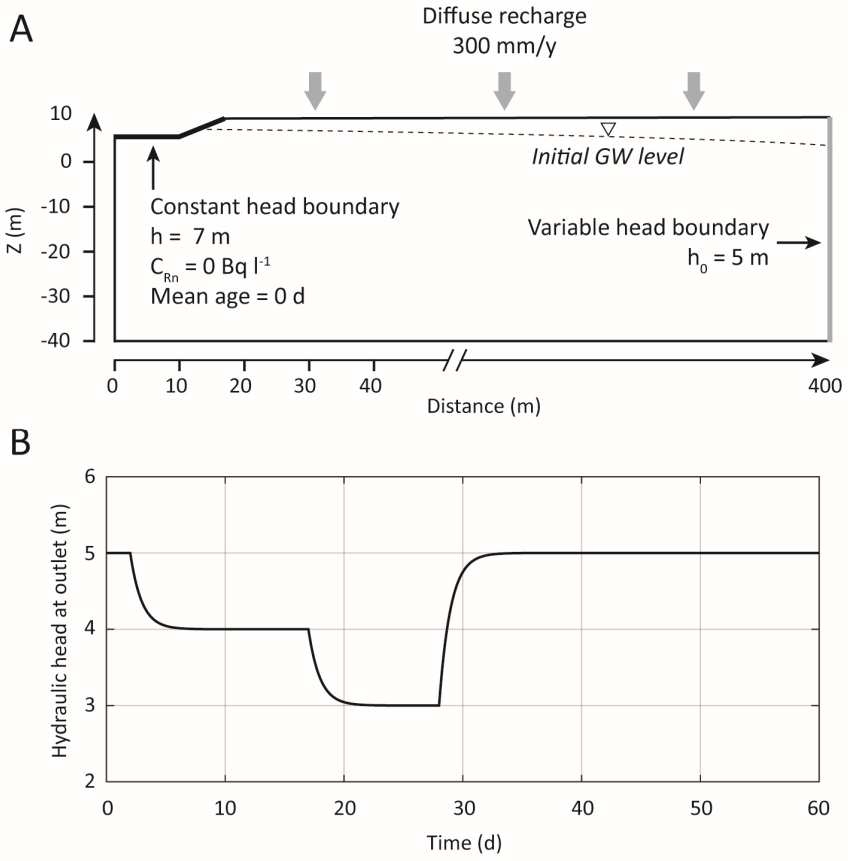

Figure 3.1 2D model setup. A: Geometry and boundary conditions; B: Variable head boundary condition at the model outlet

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 Figure 3.1). 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.

No-flow BCs were applied to the left (X = 0 m) and bottom (Z = -40 m) boundaries. The impermeable boundary 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 of 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)):

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

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 g cm^{-3} 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 the 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)):

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

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 (Figure 3.1B).

^{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 x 10 m up to 200 x 30 m in the X and Z directions respectively (see Table 3.1). 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 of 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} (Table 3.1).

Table 3.1 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 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

3.3.2. 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:

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

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 ($A_{Rn,eq}$). 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 the 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 hydraulic head, saturation, flow velocity, mean 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)*:

$$N_{Rn}^0 \approx N_{Rn} = a_{mean} \lambda_{Rn} \frac{S}{\alpha_L} \quad (9)$$

The amount of ^{222}Rn bias (eq. (7)) as a function of N_{Rn} may offer some insight into 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}.

3.4. Results

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

As described in section 3.3.2, 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 varies 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 3.3.1 is shown in Figure 3.2.

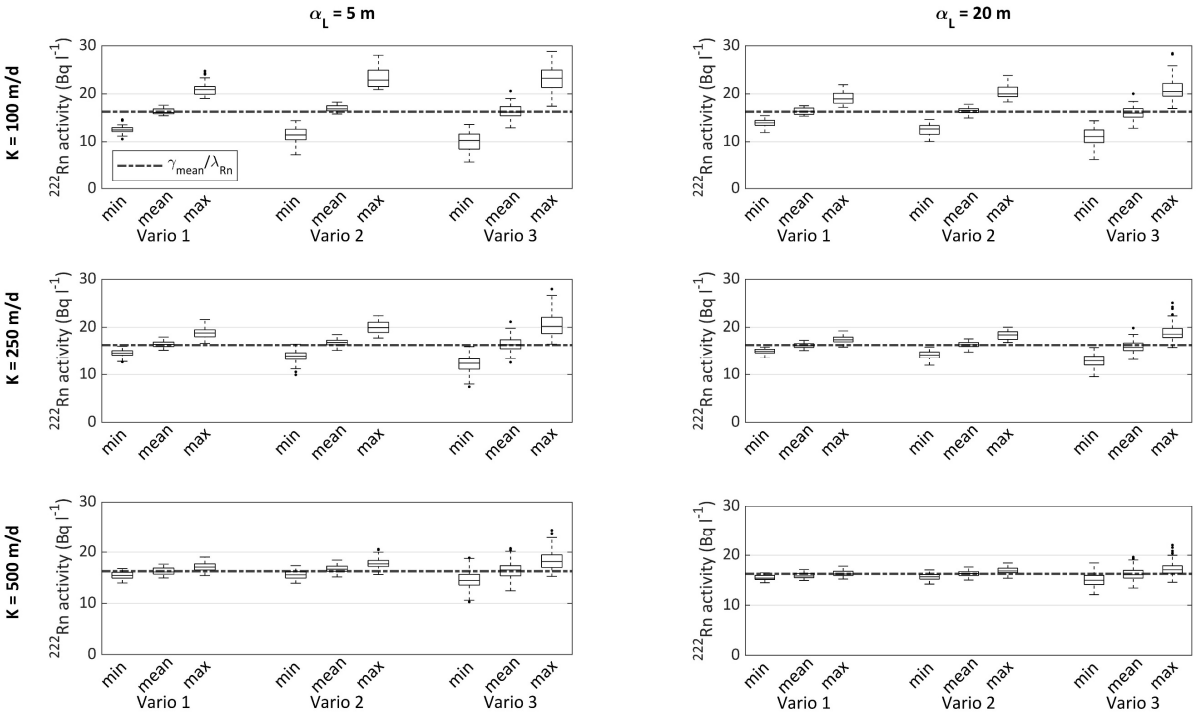

Figure 3.2 Boxplots of ^{222}Rn end-member activities for different combinations of hydraulic conductivity (K), longitudinal dispersivity (α_L), and variogram models for the distribution of ^{222}Rn production. Aggregated statistics from all simulations are shown here. Vario 1, 2, and 3 represent variogram models with ranges in the X and Z directions of 20 x 10 m, 50 x 15 m, and 200 x 30 m respectively (see Table 3.1) and maximum end-member activities as defined in section 3.3.2. The grey dashed line represents theoretical equilibrium ^{222}Rn activities assuming a spatially constant production rate equal to γ_{mean} ≈ 3.0 Bq l^{-1} s^{-1}, leading to $A_{Rn,eq}$ ≈ 16.5 Bq l^{-1} (see Table 3.1).

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 3.1). 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 is 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 α_L = 20 m, K = 500 m/d and smallest spatial correlation ("Vario 1") to 0.09 for the case α_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.

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 Figure 3.3, 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.

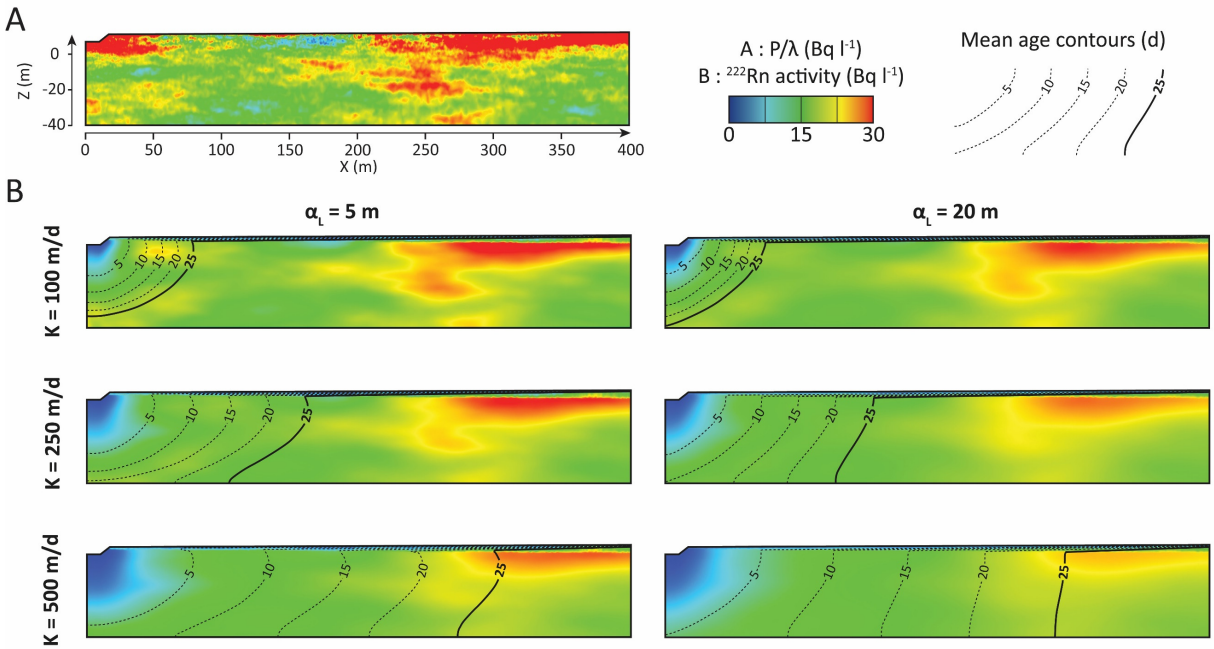

Figure 3.3 Effect of hydraulic conductivity (K) and longitudinal dispersivity (α_L) on simulated steady-state ^{222}Rn concentrations. Results from a single realization of the distribution of ^{222}Rn production rates and the six parameter combinations are shown for illustrative purposes. A : Equivalent equilibrium ^{222}Rn concentrations (=γ_{Rn} /λ_{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.

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 allow ^{222}Rn activities to equilibrate with the local aquifer ^{222}Rn production rates.

3.4.2. Apparent age bias in steady-state conditions

3.4.2.1. 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 Figure 3.4 for each combination of hydraulic conductivity, dispersivity, and choice of ^{222}Rn end-member activity (sections 3.3.1 and 3.3.2). 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 3.2.1); in these instances, additional age bias was therefore not computed. For readability, only results for one variogram model are shown in Figure 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 α_{L} = 5 m to approximately -0.5 for the case K = 100 m/d and α_{L} = 20 m (see Appendix C).

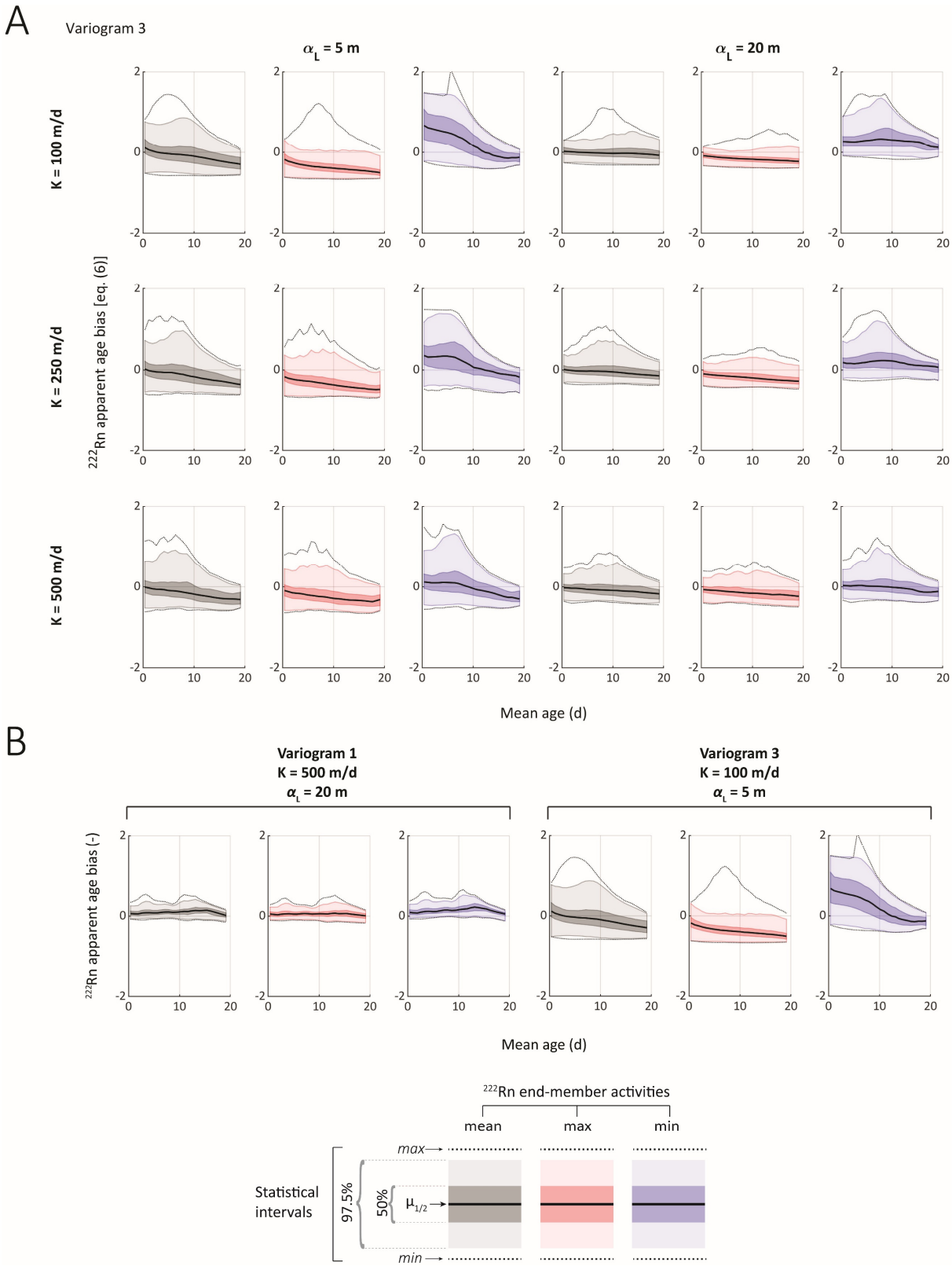

Figure 3.4 ^{222}Rn age bias (eq. (7)) in steady-state conditions as a function of mean age computed with the mean (grey), maximum (red), and minimum (blue) ^{222}Rn end-member activities (see section 3.3.2 for the definition of these end-member activities). The median ($\mu_{1/2}$), min and max biases, as well as the 50% and 97.5 % intervals are shown. A: Simulated ^{222}Rn bias as a function of mean age with variogram model 3 ("Variogram 3", highest correlation length for ^{222}Rn production). B: Comparison of two scenarios ((i) K = 500 m/d, α_L = 20, lowest correlation length; (ii) K = 100 m/d, α_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 K and α_L.

As shown in Figure 3.4A, 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 (Figure 3.4A), 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 are 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, α_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, α_L = 20 m, and the scale of variability is lowest ("Variogram 1", see Figure 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 3.4.1. 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 (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 Figure 3.2).

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 represent 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 are 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 Figure 3.4). 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 tend to approach even more rapidly than 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.

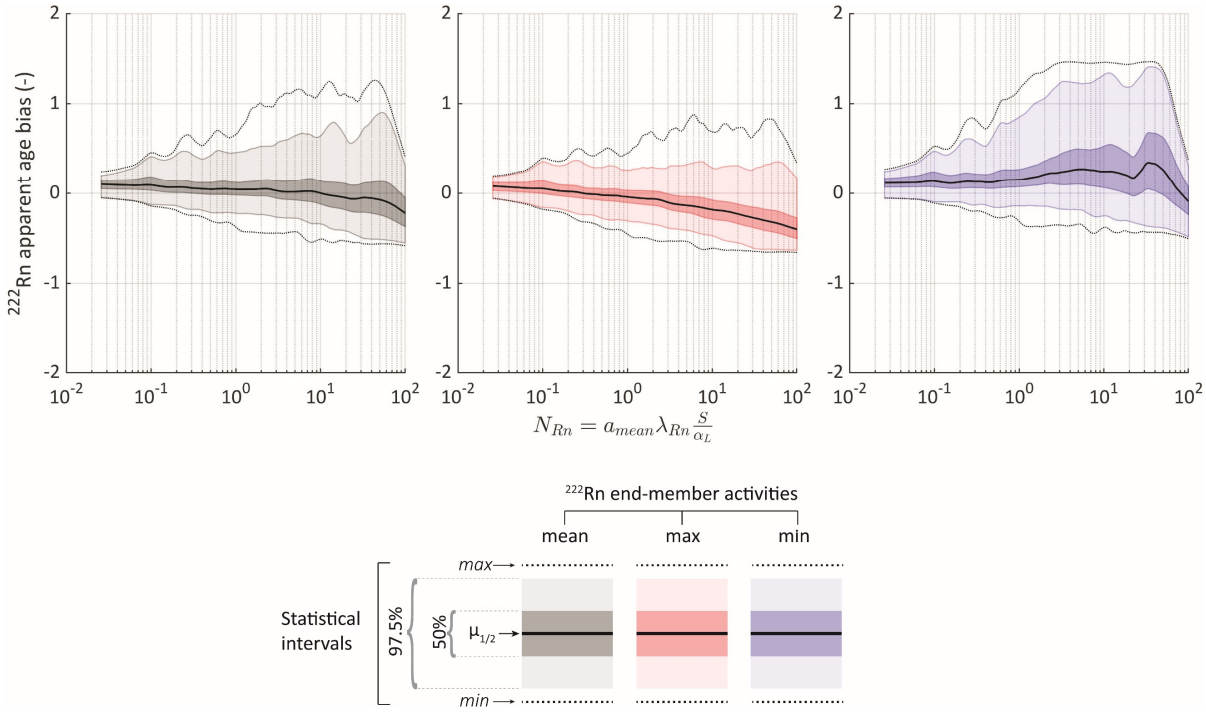

Figure 3.5 Additional apparent ^{222}Rn age bias in steady-state conditions as a function of N_{Rn} (eq. (9)). $\mu_{1/2}$ is the simulated median ^{222}Rn bias.

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 Figure 3.5, 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.

3.4.3. Apparent age bias in transient conditions

Figure 3.6 shows the range of transient simulated mean and apparent ages 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 Figure 3.1) will lead to lower mean ages at any given observation point.

As already highlighted in section 3.2.1.2, 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 Figure 3.6), 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 Figure 3.6 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.

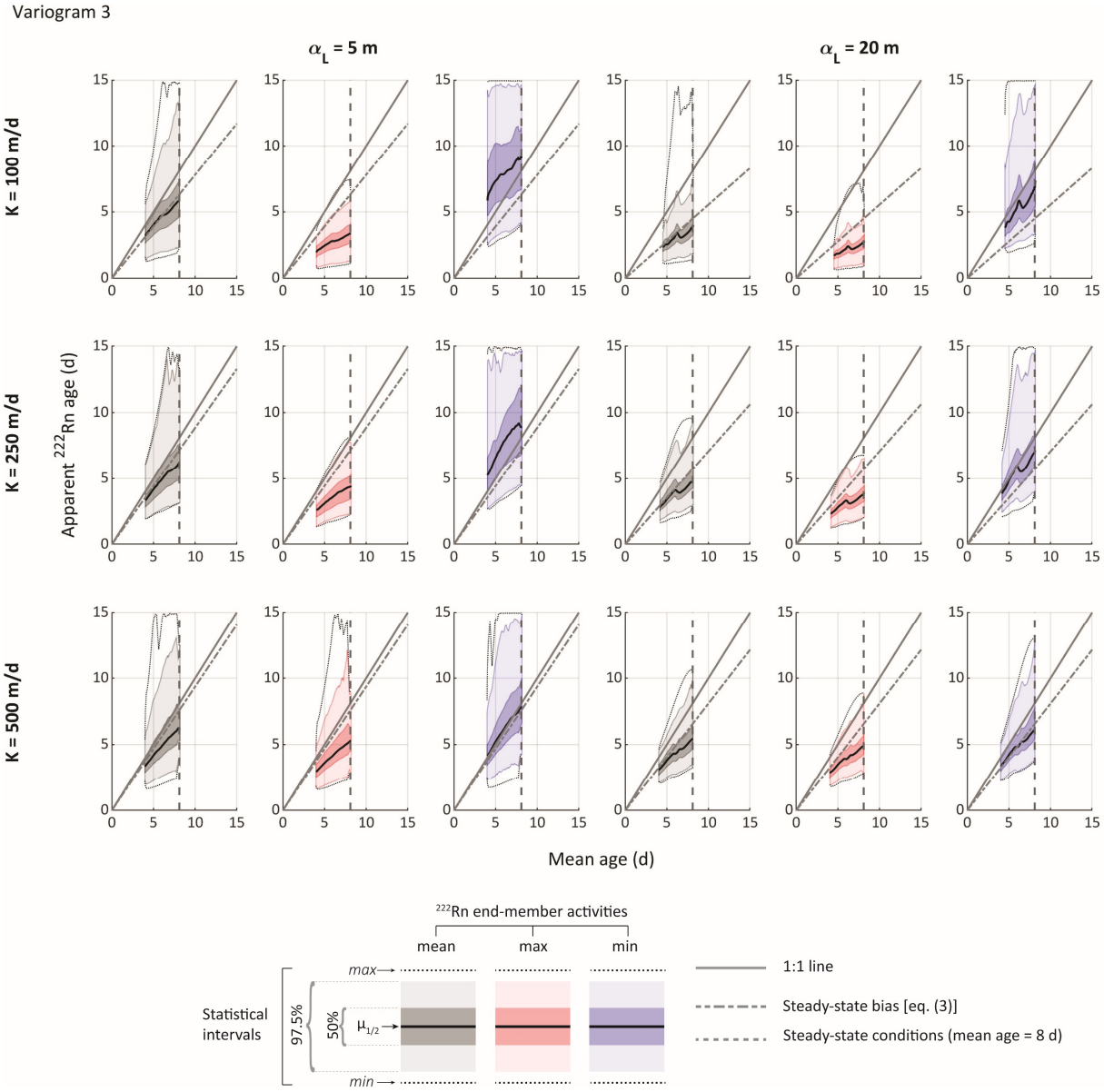

Figure 3.6 Aggregated apparent ^{222}Rn ages vs mean ages for the 500 transient simulations, for all nodes where the mean age in steady-state conditions is a_{mean} = 8 ± 0.1 d. Apparent ages along the mean age = 8 d line (vertical dashed line) represent steady-state conditions. The steady-state bias line represents the 1-D apparent age bias with constant ^{222}Rn production and the mean GW flow velocity in steady-state conditions.

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 :

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

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

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 Figure 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; 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.

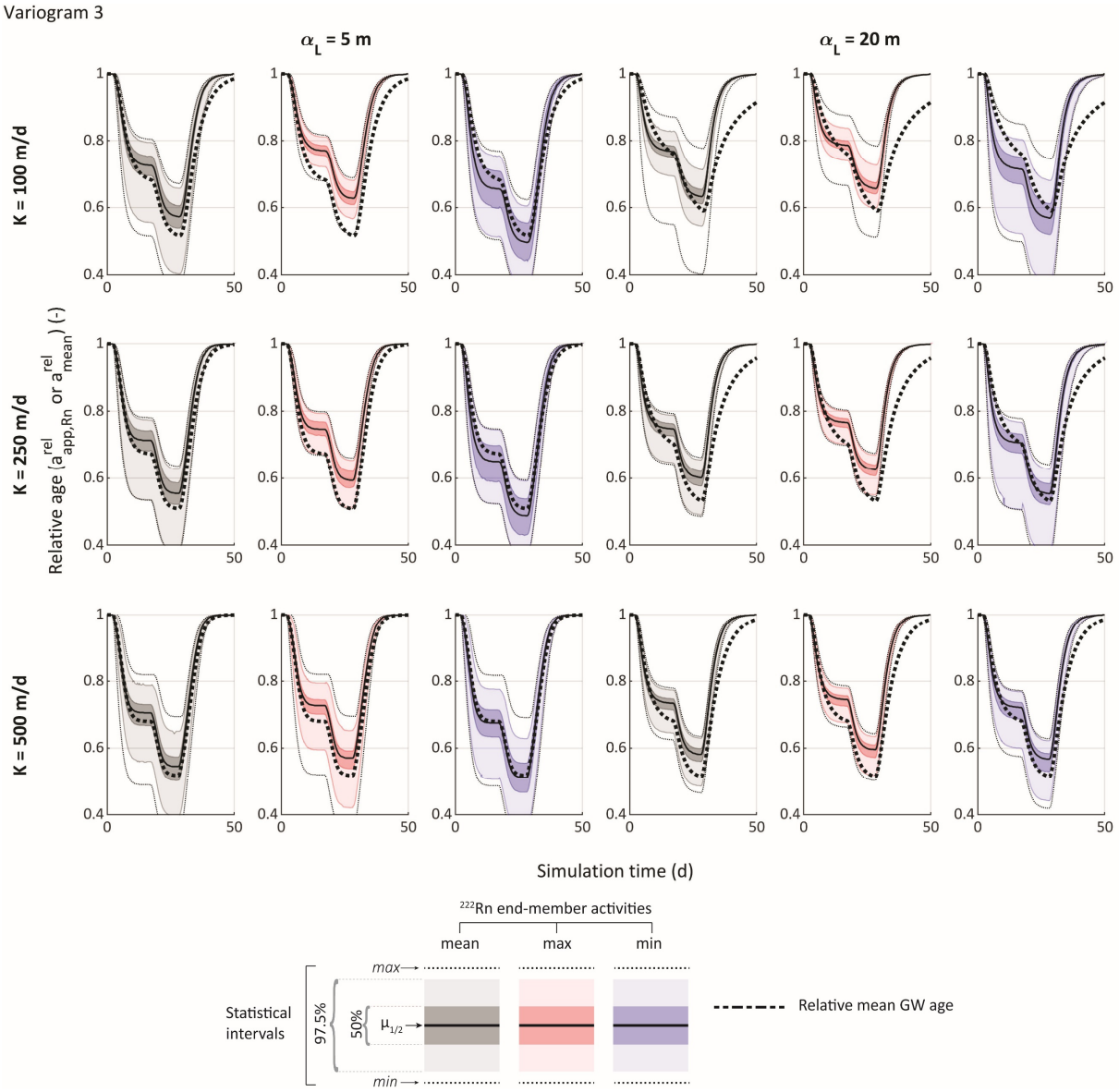

Figure 3.7 Relative apparent ^{222}Rn ($a^{rel}_{app,Rn}(\boldsymbol{x},t)$) and relative mean ages ($a^{rel}_{mean}(\boldsymbol{x},t)$) as a function of simulation time for all nodes where the mean age in steady-state conditions is a_{mean} = 8.0 ± 0.1 d. Apparent ages were computed using the mean (grey), maximum (red) and minimum (blue) ^{222}Rn end-member activities.

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 than dispersive processes, relative changes in mean apparent ^{222}Rn show excellent correlation with changes in mean GW age.

3.5. Discussion

3.5.1. 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 the 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.

3.5.2. 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 of the following conditions is 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 the conditions we have modelled, 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 limited 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 end-member activity. Indeed, as the sensitivity of the ^{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 heterogeneous environments.

In transient hydraulic conditions however, valuable information on 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 is 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 heterogeneous. 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 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.

3.5.3. 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 where 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 3.1, section 3.3.1); 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.

3.6. 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 heterogeneous.

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 the 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. Reduction of parameter uncertainty can be achieved 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.

3.7. 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).

3.8. Appendices

These appendices were published as supplementary material to the preceding manuscript. Several figures (33 in total) were also included in the appendices, and can be accessed from the online version of the article at doi:10.1016/j.watres.2023.119880.

3.8.1. Appendix A

3.8.1.1. Modelling the spatial distribution of ^{222}Rn production rate

Modelled ^{222}Rn production rates were based on a published dataset of ^{222}Rn emanation rates, measured on sediment core samples recovered from a bank filtration site (Aeschau site, Switzerland; Peel et al., 2022). Aggregated ^{222}Rn emanation rates as a function of depth are shown in Figure 1a. Emanation rates exhibit highest values within the first few meters (<5 m) below the surface, and are more stable at greater depths. The relationship between ^{222}Rn emanation and depth is likely prevalent in alluvial aquifer environments, as processes leading to ^{226}Ra enrichment in the near subsurface (e.g. adsorption of radionuclides on organic mineral coatings, chemical weathering) are ubiquitous in nature. The relationship between emanation and depth was therefore retained for all simulations.

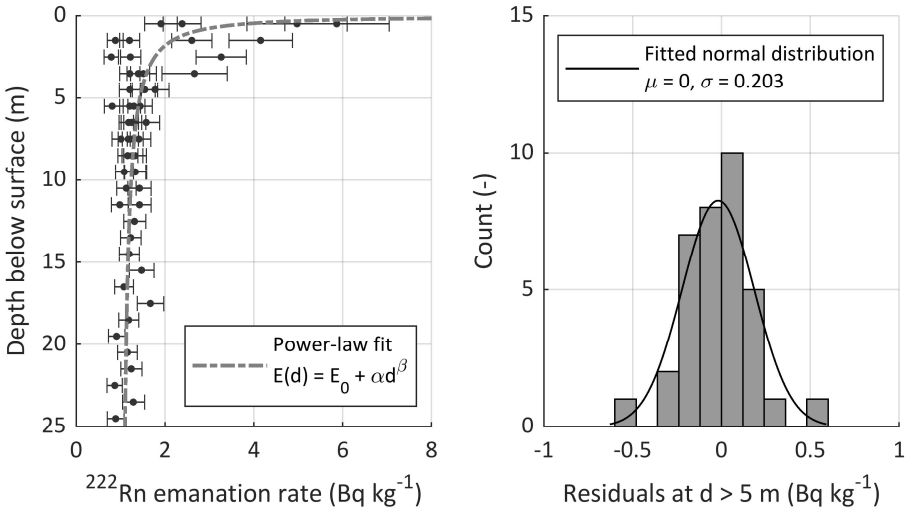

Figure 3.8 Emanation rates used to define elemental ^{222}Rn production rates in the 2-D model (data from Peel et al. (2022)) a) ^{222}Rn emanation rates as a function of depth, with power-law fit. E(d) is the emanation rate as a function of depth d *below the surface, E_0 is the emanation rate as d→ ∞, and α and β are fitting coefficients. Best fit yields values of E_0 = 0.969 α = 1.642, β = -0.783. b) Histogram of residuals of ^{222}Rn emanation rates after trend removal at depths > 5 m*

The spatial distribution of ^{222}Rn production rates in the synthetic aquifer was modelled with the Sequential Gaussian Simulation (SGS) algorithm as implemented in the GEONE python package (Straubhaar, 2020). SGS methods are widely used geostatistical tools that can produce multiple realizations of a continuous variable whilst honouring both spatial statistics and sample values (Goovaerts, 1997). These methods populate a grid with Gaussian random fields and are based on the

assumed spatial statistics and continuity of the targeted data, usually presented as a variogram model with or without hard conditioning data.

Data was first prepared for the stochastic simulations by detrending ^{222}Rn emanation rates as a function of depth. To this end, the association between ^{222}Rn emanation and depth was modelled as a power-law model of the form:

$$E(d) = E_0 + \alpha d^{\beta} \qquad (A1)$$

Where $E(d)$ is the ^{222}Rn emanation rate [Bq kg^{-1}] as a function of d, the depth below the ground surface [m], E_0 is the emanation rate [Bq kg^{-1}] as $d \rightarrow \infty$, and α and β are empirical fitting coefficients (see Figure 3.8). Data was detrended using the relationship described above, and only the statistical distribution of residuals (i.e. the difference between measured and fitted emanation rates) was used for subsequent geostatistical simulations.

No experimental variogram analysis was undertaken. Indeed, the objective of the modelling exercise is not to constrain the spatial distribution of ^{222}Rn production rates at the field site from which sediment samples were taken, but rather to make use of experimental data to explore the effect of both the magnitude and scale of variability of ^{222}Rn input on ^{222}Rn signatures in GW in general. Spatial distributions of ^{222}Rn production rates, representative of small-, intermediate- and large-scale variability, were constructed by employing an exponential variogram model with ranges of 20 x 10 m, 50 x 15 m, and 200 x 30 m in the X and Z directions respectively. Ranges in the horizontal (X) directions were larger than those in the vertical direction (Z), in part to reflect the horizontal anisotropy of alluvial sand and gravel aquifers. Indeed, such aquifers are formed by the consecutive layering of streambed and floodplain sediments, and are characterized by locally strong horizontal correlation and connectivity of material properties (e.g. mineralogical, chemical) (Bridge and Demicco, 2008), which could influence the spatial structure of ^{222}Rn emanation. No systematic trend was included in the geostatistical simulations, as none was observed between measured emanation rates and sample locations.

"True" variance in ^{222}Rn production rates at the field-scale are expected to be somewhat higher than that of experimental ^{222}Rn emanation rates for two main reasons: (i) for a given emanation rate, unresolved variations in sediment porosity at any scale will lead to commensurate variations in ^{222}Rn production rates (e.g. Thamer et al. (1981)), and (ii) only a minor portion of the aquifer is sampled, and the existence of zones of higher or lower ^{222}Rn production cannot be excluded. To partially account for these sources of uncertainty, the variance of simulated ^{222}Rn production rates was artificially increased compared to that of measured emanation rates by parameterizing the semivariogram with a sill equal to the variance of all ^{222}Rn emanation rates residuals at depths >5 m

(see Figure 3.8). A nugget effect equal to 10% of the sill was included in the simulations. To account for the increased variability of emanation rates at shallow depths (< 5 m), a variance equal to that of residuals at these depths was imposed by the geostatistical algorithm. This led to an overall more than threefold increase in the coefficient of variation of modelled production rates compared to measured emanation rates.

A reference grid covering the HGS model was used, with dimensions 400 x 50 m, divided into 800 x 100 square elements (edge size 0.5 m). Synthetic conditioning data were placed within the interpolation grid with spacing equal to half the variogram ranges in the X and Z directions. This data was randomly generated assuming normally distributed data with a mean of zero and variance equal to twice that of the residuals at depths > 5 m (Figure 3.8). For each variogram model, fifty sets of synthetic conditioning data were randomly generated. Ten equiprobable realizations of the spatial distribution of residuals were generated with the SGS algorithm in GEONE for each of these datasets (total of 500 realizations per variogram model).

Once the geostatistical simulations were completed, modelled data was re-trended to achieve the desired spatial distributions of ^{222}Rn emanation rates, from which ^{222}Rn production rates in GW could be calculated. Finally, production rates were assigned to each element in the triangular HGS grid with values from the nearest element in the rectangular GEONE grid.

3.8.2. Appendix B

3.8.2.1. Placement of observation wells for scenarios K = 100 m/d, K = 250 m/d, and K = 500 m/d

Placement of synthetic observation wells for the three different *K* scenarios is shown in Figure 3.9. Each well was placed from the surface to a depth of 20 m. From the model inlet, the horizontal distance between observation wells is 2.5 m for *K* = 100 m/d, 6.25 m for *K* = 250 m/d, and 12.5 m for *K* = 500 m/d. Output including hydraulic head, saturation, flow velocity, mean age, and ^{222}Rn activities was produced for all nodes located within 10 cm of these synthetic observation wells.

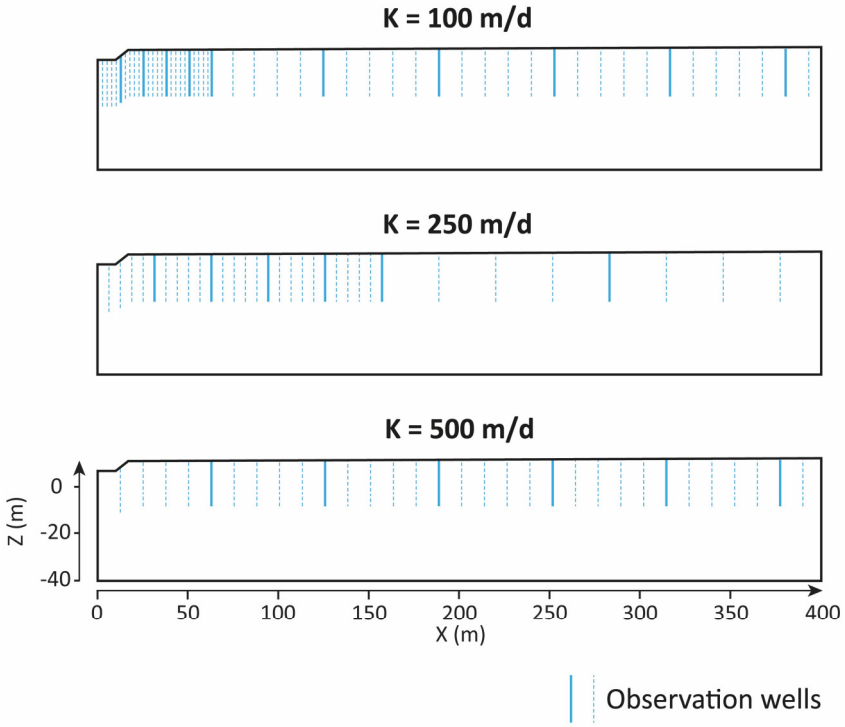

Figure 3.9 Synthetic observation well placement for K = 100 m/d (n = 158), 250 m/d (n = 63), and 500 m/d (n = 31). Each observation well is placed from the surface to a depth of 20 m. For readability, only every fifth observation well from the inlet to the left is highlighted as a thick continuous line.

3.8.3. Appendix C

3.8.3.1. Bias with constant ^{222}Rn input

Figure 3.10 shows the apparent age bias in steady-state conditions as a function of mean age with spatially constant ^{222}Rn production for the six combinations of hydraulic conductivity and dispersivity, simulated at the observation wells described in the accompanying paper. The ^{222}Rn end-member used for computation of apparent age was constant and simply defined as the ratio of the production rate to the ^{222}Rn decay constant.

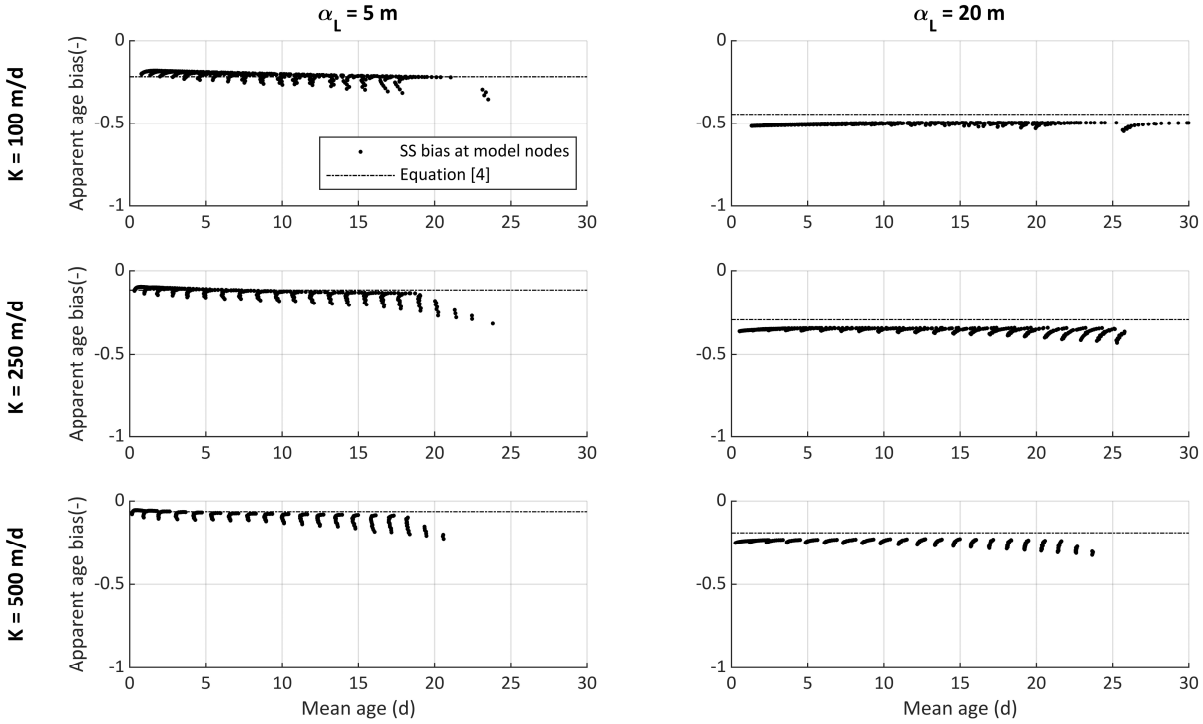

Figure 3.10 Apparent age bias as a function of mean age at observation nodes in steady-state conditions. ^{222}Rn production is spatially constant and age bias arises mostly from hydrodynamic dispersion, and to a lesser extent from ^{222}Rn partitioning in the unsaturated zone. The dashed line represents the apparent age bias computed with eq. (3), using the mean GW flow velocity in steady-state conditions.

Apparent ^{222}Rn ages are systematically smaller than mean ages (i.e. negative bias), with the magnitude of this age bias being directly related to hydraulic conductivity (K) and dispersivity (α_L). Age bias closely follows the theoretical relationship between mean and apparent ages in 1-D systems (eq. *(3)*) with some negative deviations at higher mean ages because of the additional effect of transverse dispersivity (α_{TV}).

When mass transport is dominated by advective processes (high K and low α_L), the age bias remains modest (between -0.05 and -0.2 for the case K = 500 m/d and α_L = 5 m); conversely, when dispersive processes become more important, large discrepancies between mean and apparent ages arise (bias < -0.5 for K = 100 m/d and α_L = 20 m).

3.9. Addendum

This section was initially included in the manuscript presented in Chapter 3 prior to its submission to *Water Research*. However, it was removed to avoid encumbering the research paper and diluting the main findings. It aimed to explore the ramifications of spatial heterogeneity of both ^{222}Rn production rates and hydraulic/mass transport parameters. Specifically, this section examined whether the correlation between mean and apparent ^{222}Rn ages observed in transient conditions with the synthetic models could be reproduced in a more realistic and heterogeneous setting. To this end, time series of ^{222}Rn activities measured during a field experiment at the Aeschau site were integrated into a simplified numerical model representative of field conditions. Spatial heterogeneity in (i) ^{222}Rn production and (ii) hydraulic and mass transport parameters was explicitly considered, and the validity and limits of the ^{222}Rn apparent age model under parameter uncertainty and variability were explored. Through the explicit simulation of ^{222}Rn, a set of heterogeneous parameter ensembles was estimated with which the ^{222}Rn measurements could be adequately reproduced. Despite a high degree of parameter uncertainty and heterogeneity, simulated relative mean and apparent ages consistently replicated apparent ages estimated from field data alone. These results, combined with the systematic association between simulated relative apparent and mean ages, further support the possibility of linking relative mean and apparent ^{222}Rn ages in transient hydraulic conditions.

3.9.1. Introduction

In this section, we attempt to determine if the principal findings from the synthetic models presented in section 3.4 can be extended to the interpretation of a time series of ^{222}Rn activities in a more realistic and heterogeneous setting. To this end, we integrate continuous field ^{222}Rn measurements taken at the experimental Aeschau site into a simplified 2-D HydroGeoSphere model representing flow and mass transport along a set of hypothesized flowlines in a losing river context. Heterogeneous spatial distributions of flow and transport parameters as well as ^{222}Rn production rates are introduced by placing pilot points within the model domain. Values of hydraulic conductivity, porosity, dispersivity as well as ^{222}Rn production are estimated by matching modelled ^{222}Rn activities to measurements made over the course of a large pumping experiment. The goal of this model is to create a range of realistic (and heterogenous) parameter ensembles that provide an adequate match between measured and modelled ^{222}Rn activities. From these parameter ensembles, mean and apparent GW ages can be computed and systematically compared as was done for the synthetic models. Although this 2-D model represents a simplification of the 3-D flow and transport processes operating at the field scale, it is meant to illustrate how GW ^{222}Rn and mean age signatures are affected by parameter heterogeneity. The objective of the model is thus to evaluate the consistency of mean and apparent

^{222}Rn GW ages under parameter uncertainty and variability in a representative field setting, whilst being constrained to the given ^{222}Rn dataset.

3.9.2. Field site

The Aeschau site is located near the outlet of a pre-alpine catchment (Upper Emme Valley), and harbours a major wellfield located next to the Emme River (Figure 1). At the field site, the valley floor forms a highly productive alluvial aquifer consisting of quaternary gravel and coarse to fine sand, with variable proportions of silt (Blau and Muchenberger, 1997; Käser and Hunkeler, 2016). The alluvial aquifer is on average approximately 25 m thick (up to 45m), and is underlain by the freshwater molasse sandstone, which acts as an aquitard (Würsten, 1991). Saturated hydraulic conductivities of the aquifer range between approximately 200 and >1500 m/d, and effective porosity is on the order of 0.1 – 0.3 (Blau and Muchenberger, 1997; Schilling et al., 2017a; Würsten, 1991).

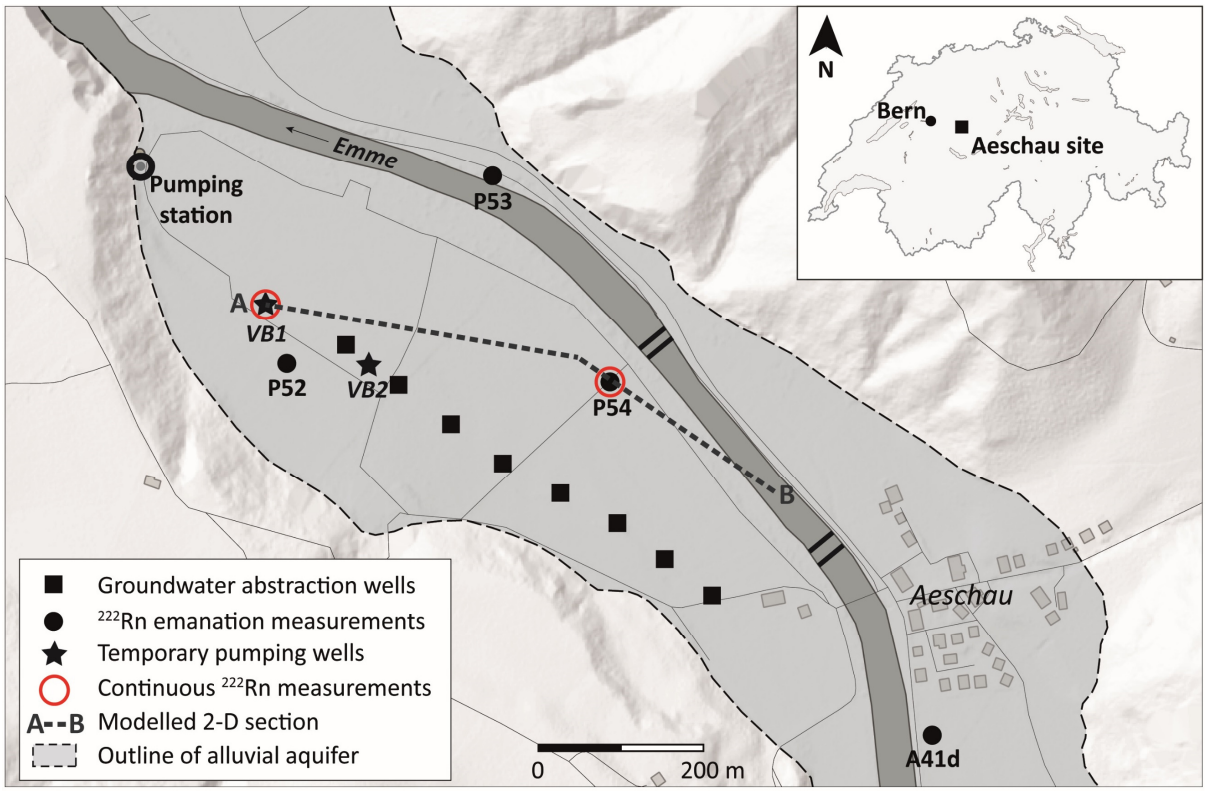

Figure 3.11 Location of the Aeschau study site (Bern, Switzerland), showing the eight GW abstraction wells, two temporary pumping wells for the 2019 pumping experiment (VB1 and VB2) and locations of continuous ^{222}Rn monitoring. Modified from Peel et al. (2022).

The field site is instrumented with a dense monitoring network, and has been used as a test site for many studies related to SW-GW interactions (e.g. Figura et al., 2013; Käser and Hunkeler, 2016; Peel et al., 2022; Popp et al., 2021; Schilling et al., 2017; Tang et al., 2018). A major wellfield is located at this site, where eight abstraction wells pump drinking water at an average rate of approximately 26 m^3/min (WVRB, 2022). In anticipation of planned infrastructure upgrades, a major six-week pumping

test was carried out from January 15 to February 26, 2019, using the eight existing abstraction wells and two temporary pumping wells (*VB1* and *VB2*, see Figure 1). The eight permanent wells are drilled to depths of approximately 10-15 m, whereas the two temporary wells were drilled to depths greater than 30 m, from the surface to the impermeable molasse substrate. Pumping rates were gradually increased from zero to 36 m^3/min, before returning to normal pumping conditions (i.e. 26 m^3/min). ^{222}Rn activities were continuously monitored at sampling locations P54 and VB1 according to the procedures described in Popp et al. (2021), and are briefly described in the following section. These continuous measurements of ^{222}Rn during pumping conditions will be used to support the parameter estimation.

3.9.3. Field data

Continuous measurements of dissolved ^{222}Rn activities in GW were undertaken at sampling locations P54 (from Jan 22 – Feb 26, 2021) and VB1 (from Jan 23 – Feb 25, 2021) by installing two RAD7 solid-state alpha detectors (Durridge, 2020). The Piezometer P54 is drilled to a depth of 8 m, and only the bottom two meters are screened. A submersible pump was placed at a depth of approximately 6 meters, and provided a continuous flow of ~3 l/min through a silicone membrane module for air-water gas exchange (3M, 2020), to which the RAD7 was connected. At pumping well VB1, the RAD7 was attached to the headspace of the pipe through which pumped GW circulated. In both cases, ^{222}Rn activities were measured at 30-minute intervals, and radon-in-air activities were converted to radon-in-water with the Weigel equation (Weigel, 1978), using water temperatures provided by loggers installed at both sampling locations. These data loggers also provided continuous measurements of GW levels. All ^{222}Rn measurements made at P54 after Feb. 26 were discarded, as the relative humidity rose to unacceptable values in both RAD7 measurement cells.

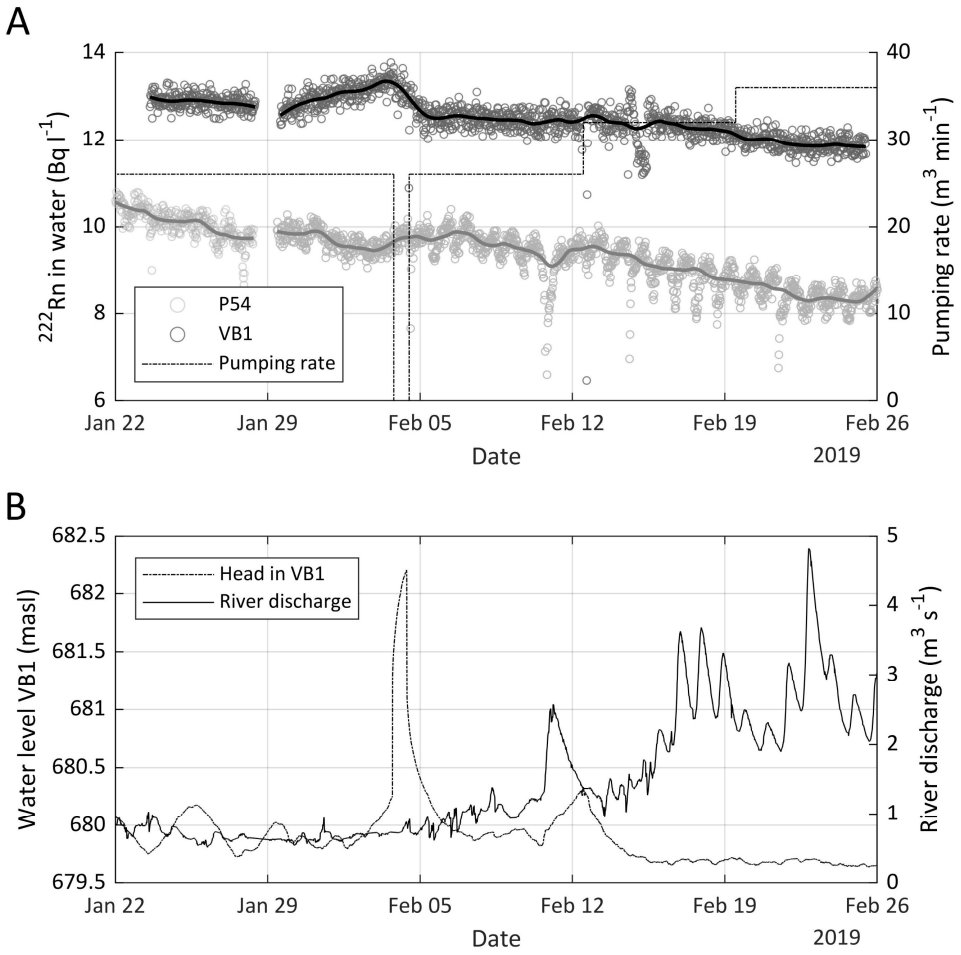

Figure 3.12 A) Measured ^{222}Rn concentrations at piezometer P54 (light grey) and pumping well VB1 (dark grey - data from Popp et al. (2021)), as well as total pumping rates. Analytical errors for the ^{222}Rn measurements (not shown) are on the order of 3%. The grey zone represents a data gap when both RAD7 devices were not measuring. The grey data points represent actual measurements, whereas the solid lines are smoothed signals B) Water levels at pumping well VB1 (dotted line) and Emme river discharge (solid black line) over the course of the pumping experiment. The discharge was recorded at the Heidbuehl-Eggiwil gauging station, 8 km upstream of the study site.

As described in Popp et al. (2021) ^{222}Rn activities in both wells show an overall decreasing trend with time, and range from 10.6 Bq l^{-1} on Jan. 22, to 8.2 Bq l^{-1} on Feb. 26 at P54, and a more modest decrease from 13.0 Bq l^{-1} to 11.9 Bq l^{-1} on Feb. 26 at VB1. The decrease at both locations is attributed to increased river water infiltration as a result of increased pumping rates and river discharge, although their relative contribution to the observed signals is ambiguous (Popp et al., 2021). Note that the synoptic variations in ^{222}Rn activities in P54 are most likely related to changes in instrument sensitivity as a function of daily air temperature fluctuations, and do not represent true variations in GW ^{222}Rn activities.

3.9.4. 2-D model of the field site

A simplified 2-D model representing a vertical cross-section from the Emme River to VB1 through P54 was constructed to simulate flow and mass transport for the duration of the pumping experiment (see Figure 1). The location of this cross-section is roughly parallel to a set of flowlines from the Emme

River to pumping well VB1 during the experiment, as simulated by a fully coupled SW-GW flow model in Popp et al. (2021).

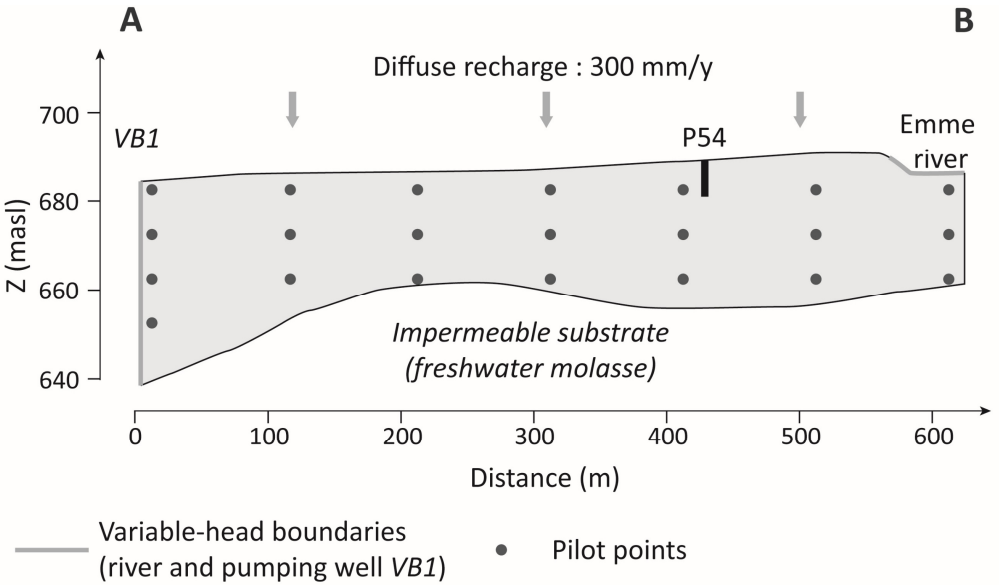

Figure 3.13 2D model domain along transect A-B (location shown in Figure 3.11). Pilot points were placed at regular intervals of 100 m x 10 m in the X and Z directions respectively, equal to half the variogram range.

The model geometry is approximately 620 m in the horizontal (X) and between 20 and 40 m in the vertical (Z) directions. The top and bottom elevations were extracted from high-resolution digital elevation models of the surface and of the molasse basement (AWA, 2013). The model domain was discretized into approximately 22'500 triangular elements, with element sizes ranging from 25 cm near the river and the top of the model to 2 m near the model outlet (location A along the cross-section).

River stage at the model inlet (location B) was estimated from river water height provided by the SW component of the 3-D flow model in Popp et al. (2021), and was used as a specified head boundary condition for all nodes within the river channel. Groundwater levels measured in VB1 served as a specified head boundary condition for all nodes along the model outlet. A constant recharge corresponding to the approximate mean recharge rate during the experiment (~300 mm/year) was applied to the top boundary (Popp et al., 2021). No-flow conditions were set to all other model boundaries. The model was run for a period of 45 days (Jan. 12 to Feb. 26, 2019) with 4-hour stress periods. The model was spun up for 2000 days with constant forcing representing hydraulic conditions on Jan. 12, 2019, prior to the start of the pumping experiment.

^{222}Rn and mean age were explicitly simulated in an identical manner as described in 3.3.1. A zero-concentration boundary condition was applied for ^{222}Rn and mean age to the inlet and top boundaries.

The zero-concentration BC for ^{222}Rn reflects the negligible ^{222}Rn activities measured in the Emme River during the pumping experiment (Popp et al., 2021).

Values of hydraulic conductivity (K), longitudinal dispersivity (α_L), porosity (n) and ^{222}Rn emanation rates (E_m) were distributed within the model domain through twenty-two regularly spaced pilot points (PPs; see Figure 3.13). Spatial interpolation of parameter values was achieved by kriging the Z-scores of values assigned to PPs (log-transformed in the case of hydraulic conductivity) with a single exponential variogram (range of 200 x 20 m in the X and Z directions respectively, sill equal to 1, and nugget equal to 0.05). Parameter values were then back-calculated from the interpolated Z-scores. In the case of ^{222}Rn emanation, the same power-law trend as described in Appendix A was applied. The ^{222}Rn production rates were computed from interpolated emanation rates and porosity with equation *(4)*. A homogenous grain density of 2.65 g cm^{-3} was assumed, consistent with laboratory measurements of sediment core samples (Peel et al., 2022).

3.9.5. Parameter estimation

Estimates of the posterior distribution of all 88 parameters (22 pilot points x 4 parameter groups) were achieved through the use of a Gauss-Levenberg-Marquardt form of the iterative ensemble smoother (iES; Chen and Oliver, 2013) algorithm as implemented in the PEST++ software suite (PEST++IES, White et al. (2020a)). In total, 60 realizations of parameter ensembles were randomly drawn from the uniform prior parameter ranges (see Table 3.2), and three parameter upgrade iterations were considered. The size of the ensemble was only slightly smaller than the parameter dimension to ensure a consistent estimation of posterior parameter distributions while minimizing computational burden (Chen and Oliver, 2017).

Initial values for hydraulic conductivity and porosity were based on reported values for the experimental field site (data from Schilling et al. (2017a); Würsten (1991)), whereas the longitudinal dispersivity was based on typical length scales of 100 to 1000 m (Schulze-Makuch, 2005). Finally, emanation rates were based on those reported by Peel et al. (2022).

Table 3.2 Initial parameter values and prior parameter ranges. ^{222}Rn production rates were computed at each model cell from eq. (4).

Parameter	Initial value	Lower bound	Upper bound	Source
K (m d^{-1})	500	50	1500	Schilling et al. (2017a)
α_L (m)	10	1	100	Schulze-Makuch (2005)
n (-)	0.2	0.1	0.4	Schilling et al. (2017a); Würsten (1991)
E_m (Bq kg^{-1})	1.2	0.5	6	Peel et al. (2022)

The continuous ^{222}Rn and water level measurements at location P54 (averaged every 3 hours), and the ^{222}Rn activity corrected for mixing with regional GW at VB1 for the duration of the experiment were considered as observations. As the 2-D model does not consider regional (and thus older) GW components introduced through convergent flow at the pumping well, the ^{222}Rn contribution from the regional GW component was subtracted from ^{222}Rn measurements at VB1 using tracer-based estimates of the fraction of regional GW at VB1, as described in Popp et al. (2021). For the simulated period, the fraction of regional GW at VB1 varied between 30 and 50% (Popp et al., 2021); the ^{222}Rn signal from the regional component was subtracted from measured ^{222}Rn at VB1, assuming regional GW is at radioactive equilibrium with the aquifer matrix with a ^{222}Rn end-member activity of 14.5 Bq l^{-1} (Peel et al., 2022). To further counteract the conceptual simplification brought by simulating pumping conditions with a 2-D model, a significant 2-σ uncertainty of 2.0 Bq l^{-1} was attributed to observations at VB1 for model calibration. Finally, an uncertainty of 1.0 Bq l^{-1} was attributed to observations at P54.

3.9.6. Modelling relative mean and apparent ^{222}Rn ages

^{222}Rn activities and mean GW ages were modelled at P54 by integrating the flux-weighted values at all nodes intercepted by the well screen (6-8 m depth). ^{222}Rn activities and mean ages modelled from the 60 posterior parameter ensembles were used to compute relative apparent and mean ages from the beginning of the ^{222}Rn measurements (Jan. 22, 2019), in the same manner as presented in section 3.4.3. For each simulation, a single ^{222}Rn end-member was retained for computation of apparent ^{222}Rn ages at P54, corresponding to the *mean* ^{222}Rn value at nodes where the mean age in steady-state conditions (i.e., after model spin-up) was above 20 days.

3.9.7. Results

The range of simulated ^{222}Rn activities as well as simulated mean and apparent GW ages from the 60 posterior parameter ensembles is shown in Figure 3.14. Statistics of the posterior parameter ensembles including the mean of estimated parameter values, the standard deviation (SD) of mean ensemble values, as well as the mean inter-ensemble coefficient of variation (CV) of parameters are shown in Table 3.3.

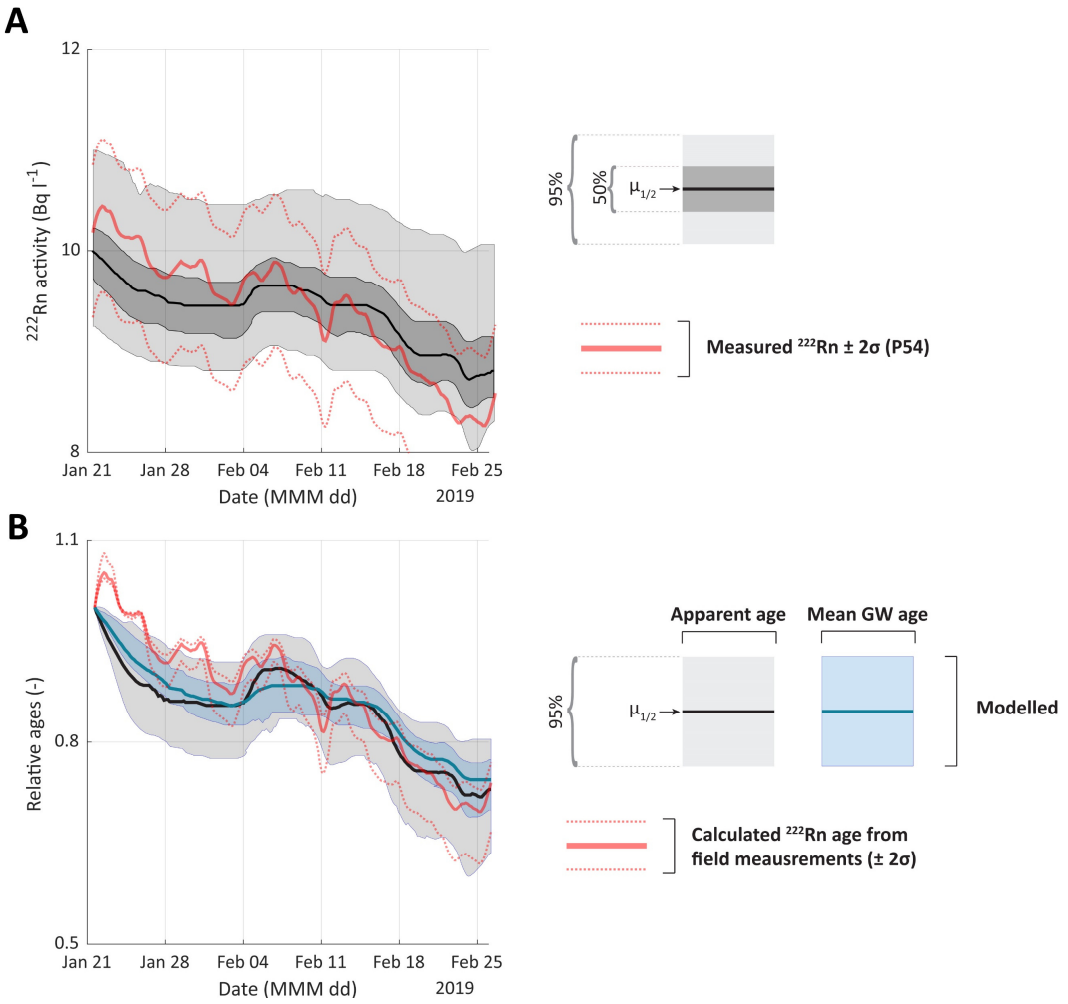

Figure 3.14 Results of the explicit simulation of ^{222}Rn and mean age along transect A-B. A) Range of simulated ^{222}Rn activities at location P54, overlain by measured activities. B) Range of simulated relative apparent and mean ages at location P54 (mean age in blue). The apparent ages from observed ^{222}Rn at P54 were computed with a representative end-member activity of 14.5 ± 2.5 Bq l^{-1} (uncertainty shown by the dotted red line) at the Aeschau site (Peel et al., 2022). The 50% and 95% statistical intervals of simulated ^{222}Rn activities and mean/apparent ages were aggregated from the 60 ensembles. $\mu_{1/2}$ is the ensemble median.

The simulations capture the overall decreasing trend in ^{222}Rn activities, and are mostly within the ± 1.0 Bq l^{-1} uncertainty range attributed to the observations (see Figure 3.14A). Some small systematic differences can however be noted. Indeed, simulated ^{222}Rn activities are on average lower than measurements for the first half of the simulation period, whereas the contrary can be noted for the second half. This leads to more modest changes in simulated ^{222}Rn activities than those measured

(average simulated decrease of 1.4 Bq l^{-1} vs. measured decrease of 1.7 Bq l^{-1}). Moreover, high-frequency variations in measured ^{222}Rn activities could not be reproduced. These observations can mostly be attributed to simplified boundary conditions and the inability of the model to consider 3-D flow and transport. Posterior inter-ensemble parameter variability is large, with mean coefficients of variation ranging from ~30 % for porosity to ~70% for hydraulic conductivity (see Table 3.3). Despite high parameter heterogeneity, simulated relative apparent ages closely track those computed from measured ^{222}Rn activities, albeit with slight underestimations for the first part of the simulation and overestimations for the second half. Relative apparent ages from field observations are moderately sensitive to uncertainties in the ^{222}Rn endmember at the Aeschau site ($A_{Rn,eq} = 14.5 \pm 2.5$ Bq l^{-1}).

Most interestingly, the range of simulated relative mean and apparent ages are tightly constrained and consistently track one another, both decreasing on average by about 25% by the end of the simulation period (see Figure 3.14B). As noted for the synthetic models (section 3.4.3), abrupt or high-frequency variations in apparent ages are more abrupt than those for mean ages, as ^{222}Rn rapidly re-equilibrates with the aquifer matrix even in transient conditions owing to its short half-life. However, apparent and mean GW ages follow very similar trends for lower-frequency and more gradual changes in hydraulic conditions. Finally, the simulated variations in relative mean and apparent age very closely reproduce those stemming from observed ^{222}Rn activities.

Table 3.3 Statistics of posterior parameter ensembles. SD : standard deviation of mean parameter values. CV : mean coefficient of variation within the 60 parameter ensembles. Statistics of the ^{222}Rn production rates were calculated from statistics of porosity (n) and ^{222}Rn emanation rates (E_m) with eq. (4)

Parameter	Posterior ensemble mean	Posterior ensemble SD	Mean inter-ensemble CV (no units)
K (m d^{-1})	742	99	0.70
α_L (m)	11.05	1.38	0.50
n (-)	0.21	0.02	0.32
E_m (Bq kg^{-1})	1.24	0.07	0.09
γ_{Rn} (Bq l^{-1} d^{-1})	*3.03*	*0.28*	*0.41*

3.9.8. Summary and conclusion

Results from the 2-D mass-transport model presented here allow us to extend some of the findings from the homogeneous parameter fields modelled in section 3.4 to heterogeneous settings. Indeed, with the simplified model of field conditions, the consistent association between changes in mean and apparent ages can be further highlighted, even in settings characterized by significant heterogeneity in both ^{222}Rn production rates and hydraulic/mass-transport parameters. By combining these results with those presented in section 3.4.3**,** we have illustrated how the correlation between mean and apparent ages in transient conditions is less sensitive to the spatial heterogeneity of hydraulic

conductivity and ^{222}Rn production rates than to the degree of system transience and to the relative influence of advective and dispersive mass-transport processes. Taken together, these results support linking variations of apparent ^{222}Rn and mean ages provided the conditions listed in 3.5.2 apply.

Confidence in associating mean and apparent ^{222}Rn ages in field settings should however be tempered by the typical lack of data constraining variability in aquifer hydraulic, mass-transport, and radiochemical parameters in real-world settings. Even the heterogenous models presented in this section only sample an infinitesimal range of plausible field conditions, and conclusions are not expected to be universally applicable to bank filtration settings. We rather hope that these findings motivate further research towards understanding how the information content of ^{222}Rn measurements can be best extracted in a larger range of field conditions. We refer readers to section 3.5, where these matters are discussed more thoroughly.

4. Bridging novel tracer applications and highly-parameterized, physics-based models : how can the explicit simulation of tracer concentrations inform model parameters and predictions?

Chapter 4 is in preparation as : Peel M., Delottier, H., Schilling, O.S., Blanc, T., Brennwald, M.S., Kipfer, R., Brunner, P., *Bridging novel tracer applications and highly-parameterized, physics-based models : how can the explicit simulation of tracer concentrations inform model parameters and predictions?*

4.1. Abstract

The assimilation of hydrological tracer observations in coupled surface – subsurface flow and transport models has been gaining traction as a means to effectively improve model performance in terms of constraining model parameters and reducing predictive uncertainty. Owing to the rapid development of novel tracer measurement techniques, a variety of observation types can now routinely be acquired in field settings at very high temporal and spatial resolutions, potentially providing unprecedented insights into the characterization and conceptualization of environmental systems. Different methodologies are employed to integrate tracer observations into model parameter estimation frameworks, ranging from transforming tracer measurements into secondary quantities with simple mathematical models, to fully simulating (reactive) tracer transport with the advection-dispersion equation. Although the explicit simulation of tracer mass-transport in principle offers a conceptually sound framework through which the information content of tracers can be explored, recourse to simpler "black-box" models is still widespread; the extent to which models can be constrained by the assimilation of various untransformed tracer measurements in an explicit modelling framework remains in large part an open question. In this study, we illustrate how the assimilation of noble gas observations (helium and ^{222}Rn) and their explicit simulation in a highly-parameterized, integrated surface-subsurface hydrological model (ISSHM) of a riverbank filtration site can decrease model parameter uncertainty and improve predictive performance. Specifically, we show how the joint assimilation of hydraulic and tracer observations constrains parameters and several predictions of water management relevance, including river infiltration fluxes, surface water (SW) - groundwater (GW) mixing ratios, and mean groundwater age. Compared to models only assimilating "classical" hydraulic observations, significant reductions in parameter variance and predictive uncertainty are systematically achieved through the assimilation of tracer data. Moreover, this work serves as a proof-of-concept for the explicit simulation of the noble gas helium when used as an artificial tracer of SW-GW interactions. The insights accrued through the assimilation of tracer concentrations in this study support the relevance of including untransformed tracer observations in

an explicit modelling and parameter estimation framework, thus avoiding the additional layer of conceptual simplification introduced by transforming tracer measurements into secondary quantities.

4.2. Introduction

Physically-based flow models are routinely employed as tools for water resources characterization and management (Anderson et al., 2015). As it is practically impossible to fully characterize the complexity and heterogeneity of natural system properties through field investigations, it is common practice to estimate model parameters through calibration against measurements of system states, which typically include hydraulic heads and boundary fluxes (e.g. river discharge, spring flow) (Anderson et al., 2015; Hrachowitz and Clark, 2017). It has been recognized that model calibration against hydraulic data alone is insufficient for many modelling needs, as it may lead to high parameter non-uniqueness and low model predictive capacity (Doherty and Hunt, 2009; Schilling et al., 2019). The inclusion of more diverse observation types such as solute concentrations, water temperature, residence time information, or exchange fluxes in the parameter adjustment process has been gaining traction over the last several years as a means to further constrain model parameters and predictions of water management interest (e.g. Delsman et al. (2016); Hunt et al. (2006); Kurtz et al. (2014); Schilling et al. (2022); Thiros et al. (2021)). The documented benefits of including such "unconventional" observations in flow and transport models have led to a general push within the hydro(geo)logical community to more systematically have them implemented in model calibration frameworks (Schilling et al., 2019; Sprenger et al., 2019; Suckow, 2014; Turnadge and Smerdon, 2014).

Of the many diverse observation types typically measurable in field settings, natural tracers are frequently collected as they provide unique insights on groundwater (GW) origins, flow pathways, residence times, quality, and mixing with surface physicochemical parameters such as electric conductivity (EC) or water temperature. water (SW) (Cook and Herczeg, 2000; Leibundgut et al., 2009). Several types of tracer measurements can now be routinely acquired in field settings at unprecedented temporal and spatial resolutions. These include amongst others continuous on-site measurements of (i) dissolved (noble) gases (Brennwald et al., 2016; Chatton et al., 2017), (ii) radiochemical species such as ^{222}Rn (Burnett et al., 2001; Hofmann et al., 2011), (iii) bacterial concentrations (Besmer et al., 2016) (iv) stable isotopes of water ($\delta^{18}O$, $\delta^{2}H$) (Berman et al., 2009), (v) chemical dye tracers (Schnegg, 2002), and (vi) water physicochemical parameters including electric conductivity (EC) (e.g. Cirpka et al, 2007) and water temperature (Anderson, 2005; Constantz, 2008).

Different strategies may be adopted to integrate tracer measurements into flow and transport models. Some of the most popular tracer interpretation methods are based on lumped parameter models (LPMs), which rely on transforming tracer measurements into secondary quantities such as

apparent GW ages and end-member mixing ratios (Kazemi and Perrochet, 2006; Małoszewski and Zuber, 1982); the estimated quantities can then be compared to simulated model outputs, such as advective travel times, mean GW age, or mixing ratios. However, LPMs and age models often rely on a set of highly simplifying assumptions, such as predefined distributions of GW age, stationary hydraulic conditions, spatial homogeneity of subsurface properties, etc., which are rarely representative of the complexity of natural systems. In contrast, the explicit simulation of tracers in physics-based numerical models offers a well-founded conceptual framework through which the information content of tracer measurements can be extracted, as complex flow and mass transport processes (e.g. transient environmental conditions, tracer accumulation/decay, chemical reactivity, mixing, etc..) can be explicitly considered, and uncertainty in hydraulic and mass transport parameters and boundary conditions accounted for. Recently, several studies have sought to quantify how the combination of tracer measurements and their explicit simulation in mass-transport models can inform model parameters and predictions of management interest (e.g. Knowling et al. (2020); Pool et al. (2015); Schaper et al. (2022); Thiros et al. (2021); Thiros et al. (2022)). Although results from the assimilation of raw tracer data in an explicit modelling framework show some promise in reducing uncertainty in model parameters and subsequent predictions, there is yet to be a consensus on the way in which tracer measurements should be integrated into environmental models. Damaging and potentially undetectable bias may arise when data assimilation is employed with models incapable of adequately representing the complexity of flow and mass-transport processes [i.e. “imperfect” models, (Knowling et al., 2020; Moore and Doherty, 2005; Thiros et al., 2022; White et al., 2020b)].

The requirements of increased model complexity for the assimilation of untransformed tracer measurements naturally come at a cost in terms of system conceptualization, data acquisition and computational needs. Physically-based models which allow faithful representation of (reactive) tracer mass-transport typically require a large number of parameters, boundary conditions, and ultimately observations which constrain the underlying processes at the required spatial and temporal scales (Doherty and Hunt, 2009; Schilling et al., 2019; Troldborg et al., 2007; Turnadge and Smerdon, 2014). For field-scale solute modelling, this practically entails combining multiple sources of field data into large, finely discretized, physics-based numerical models, all within a highly-parameterized framework (e.g. Hunt et al. (2007); Knowling et al. (2020); Schilling et al. (2019)). Although they are usually associated with significant computational burden, high-dimensional parameter schemes are doubly warranted for solute modelling in that they allow robust quantification of uncertainty in model inputs and predictions (Dausman et al., 2010; Knowling et al., 2019; Moore and Doherty, 2005), whilst allowing the expression of spatial heterogeneity of system properties (Aster et al., 2018; White et al., 2014). The latter consideration demands particular attention in the case of solute transport modelling,

as complex superpositions of multiscale subsurface heterogeneity, from small-scale material variabilities to large-scale geological structures, exert fundamental controls on solute transport, notably through the existence of preferential flow pathways (Linde et al., 2015; Renard and Allard, 2013; Webb and Anderson, 1996). Assimilation of tracer measurements in numerical models is therefore an exercise in balancing the needs of increased model complexity (in terms of parameter dimensionality, representation of heterogeneity and preferential flow, realistic modelling of natural processes), the limited resolution of field data, partial knowledge of subsurface geology, and finite computing resources.

Recourse to highly-parameterized, physics-based flow models is constrained by the numerical cost associated with some of the most common parameter estimation and uncertainty quantification algorithms, which may become prohibitive as the number of adjustable model parameters grows. This constraint often requires model simplification through parameter reduction and modified representation of physical processes, which in turn may lead to model error and predictive bias (Doherty and Christensen, 2011; White et al., 2014). This issue can partially be alleviated through the implementation of efficient ensemble-based data assimilation (DA) schemes, which relax the computational burden associated with placing history-matching constraints on (very) high-dimensional numerical models (e.g. van Leeuwen et al. (2015)). Amongst these DA methods, iterative ensemble smoothers (iES) are particularly well-suited for large and nonlinear inverse problems, as they assimilate all data simultaneously (i.e., no need for frequent model restarts) and scale exceptionally well with increasing number of model parameters; indeed provided the number of realizations in the ensemble is sufficiently large in comparison to the size of the solution space, the numerical burden is effectively insensitive to the number of parameters (Chen and Oliver, 2013; Emerick and Reynolds, 2013; White, 2018).

Access to novel field technologies, efficient DA methods for high-dimensional problems, as well as high-performance computing resources, is such that various (untransformed) tracer observations can now be included for history matching through the explicit simulation of tracer concentrations in highly-parameterized, field-scale mass transport models, even with complex physics-based numerical models. Although the potential of assimilating tracer measurements within this modelling framework has been acknowledged, the practice remains comparatively marginal. There is a need to further contextualize and document how the explicit simulation and assimilation of different types of tracer measurements with physics-based models can inform various model parameters as well as management-relevant quantities and predictions.

In this work, we illustrate how modern, state-of-the-art tracer techniques can be combined with the explicit simulation of tracer concentrations in a highly-parameterized, physics-based flow and transport model to inform model parameters and reduce the uncertainty of certain predictions of groundwater management interest. Specifically, we quantify the contribution of simulating tracer concentrations and including tracer observations in an automatic and multivariate parameter estimation framework towards (i) reducing uncertainties of selected management quantities, (ii) reducing the non-uniqueness of model parameters, and (iii) identifying the location of preferential flow structures. To this end, high-resolution hydraulic and noble gas data (helium and 222-radon), acquired over the course of a novel helium injection experiment at a riverbank filtration site, are combined with an integrated surface-subsurface hydrological model (ISSHM) of the field site. An iterative ensemble smoother (PEST++IES, White (2018)) is employed for history matching in a highly-parameterized framework, with values of hydraulic conductivity (*K*) estimated for each model cell. We show how the inclusion of high-resolution measurements of tracers sensitive to SW-GW exchange processes within an explicit modelling framework provides unique insights on important quantities of water management relevance, such as GW age, SW infiltration fluxes, and SW-GW mixing ratios. Moreover, this study provides a proof-of-concept for the reliable integration of dissolved noble gas measurements in physics-based models when these species are employed as artificial tracers of SW infiltration into unconfined alluvial aquifers. To our knowledge, this work describes the first field-scale study exploring the ability of artificially injected (noble) gases to inform and constrain the parameters and predictive ability of coupled SW-GW models of alluvial aquifer systems.

4.3. Materials and methods

4.3.1. Field site and acquired dataset

The experimental Aeschau site is located at the outlet of the Upper Emme catchment, in Bern, Switzerland (Figure 4.1). The field site and its surroundings have been thoroughly described in previous works (e.g. Käser and Hunkeler (2016); Peel et al. (2022); Popp et al. (2021); Schilling et al. (2022); Tang et al. (2018)). Readers are referred to these references for more information on the hydro(geo)logical context of the study site.

Eight abstractions wells (Br1 – Br8, see Figure 4.1) continuously pump GW at an average rate of 26 m^3/min, providing approximately 25% of the drinking water for the Swiss capital Bern. Large pumping rates coupled with highly dynamic river conditions lead to complex patterns of SW-GW interactions at the study site, with alternating reaches of SW infiltration and GW exfiltration (Käser and Hunkeler, 2016; Schilling et al., 2022). The presence of two sets of weirs (next to observation wells A32 and P54,

Figure 4.1) add an additional layer of complexity to the overall SW-GW dynamics in the area, as sudden drops in river level lead to zones of focused GW exfiltration (Schilling et al., 2017a).

4.3.1.1. 2021 tracer experiment and dataset

A novel helium tracer test was carried out at the Aeschau site in 2021, during which the noble gas helium (He) was injected into river water over several weeks, and regularly monitored in GW (Blanc et al., in prep). The tracer test aimed to showcase a new gas tracer injection technique and demonstrate its potential in tracking SW infiltration into an alluvial aquifer. Dissolved 222-radon (^{222}Rn) concentrations were also continuously monitored in the upstream pumping well (Br1) during the experiment (Figure 4.1). This radioactive noble gas, which is continuously produced by the aquifer matrix as a result of alpha decay of matrix-bound ^{226}Ra, is routinely used as a tracer of GW age and SW/GW interactions (Bertin and Bourg, 1994; Hoehn and Von Gunten, 1989; Peel et al., 2023). It is sensitive to GW residence times at scales of days to weeks, which are those most relevant at the study site (Peel et al., 2022; Popp et al., 2021). These two gas tracers provide complementary information on SW/GW exchange processes relevant to typical water management needs: (i) artificially induced variations in dissolved helium in GW reflect the level of connectivity between the stream and adjacent alluvial aquifer, whereas ^{222}Rn is most sensitive to the time elapsed since the infiltration of SW ("GW age").

Helium (He) gas was continuously injected into the Emme River at the southern edge of the field site for a period of several weeks (Feb 12 – March 23, 2021) using a diffusion-injection setup (Blanc et al., in prep). Two portable gas-equilibrium membrane-inlet mass spectrometers ("miniRUEDI" GE-MIMS, Brennwald et al. (2016)) were installed for the duration of the experiment to continuously monitor dissolved He concentrations in the Emme River and in pumping well Br1. The temporal resolution of dissolved gas measurements was approximately 5 minutes. Moreover, a GE-MIMS device was used to measure dissolved He concentrations in six observation wells at regular intervals (Figure 4.1). Note that at two locations (A32 and A17), GW was sampled with a peristaltic pump owing to the small diameter of the piezometer tubes (1'' inner diameter). Negative pressure pumps could favour He gas exsolution and escape before reaching the air/water exchange membrane, and possibly yield negatively biased He concentrations. GW was sampled with a submersible pump at all other locations.

For the measurement of ^{222}Rn in pumping well Br1, GW was continuously pumped through a silicone air-water exchange module (3M, 2023), to which a solid-state alpha detector (RAD7, Durridge (2020)) was connected. ^{222}Rn-in-air activities were measured at 30-minute intervals, and were converted to radon-in-water with the Weigel equation (Weigel, 1978), using continuous water temperature measurements provided by the GE-MIMS module.

Several fluorescein injections into the Emme River were also carried out during the study period, with concentrations being continuously monitored in the river and in pumping well Br1; the experiments and results are detailed in Appendix D. The acquired fluorescein data is used as an independent tracer not included in history matching, and meant to validate the reliability of posterior model parameters and subsequent predictive capacity.

Tracer measurements were complemented by a suite of hydraulic observation data. GW levels were monitored at 15-min intervals in eight observation wells (Figure 4.1). Finally, river discharge was monitored at 5-minute intervals at a national gauging station located a few kilometres upstream of the study site (*Eggiwil-Heidbüel* station (FOEN, 2023)).

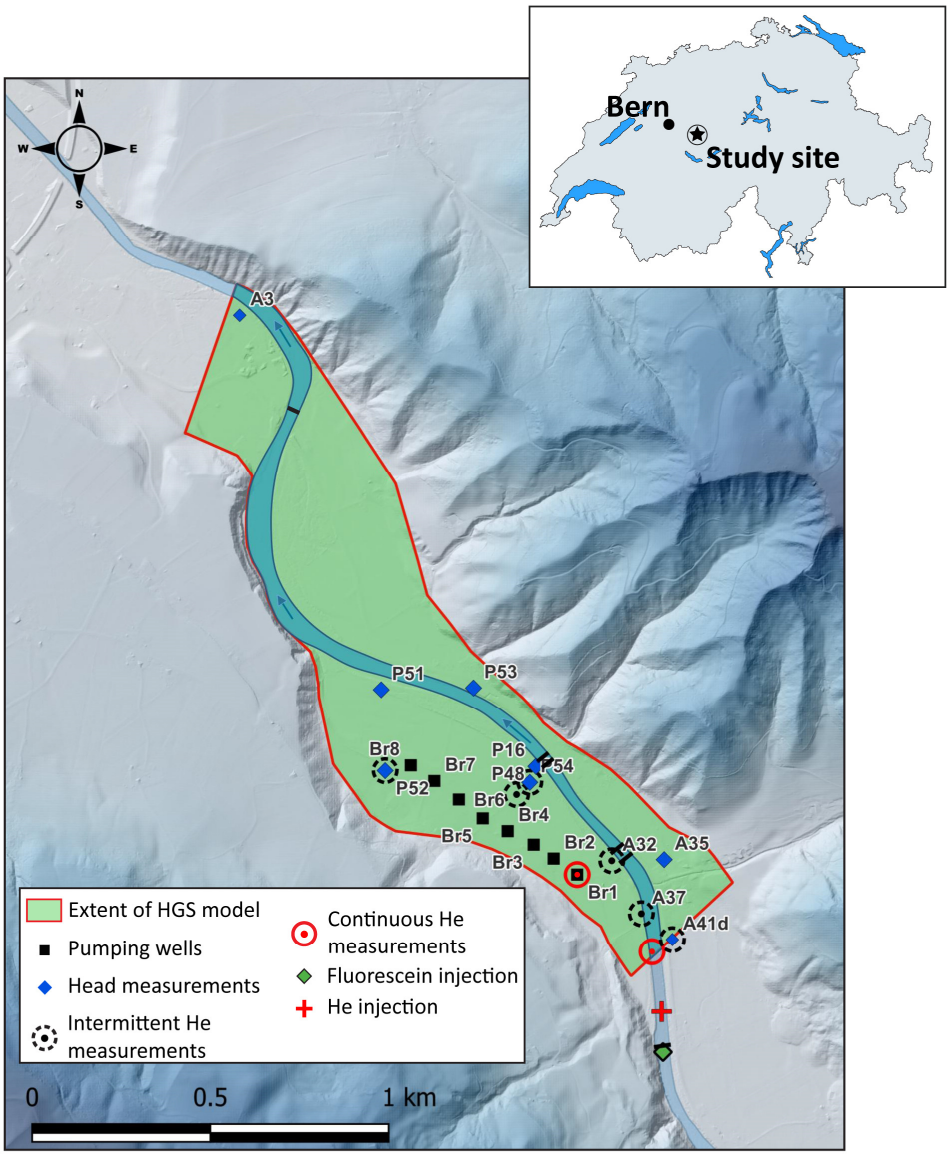

Figure 4.1 Map of study site with measurement locations. Step weirs are shown as black lines perpendicular to the river, adjacent to observation points A32 and P54.

Measured He and ^{222}Rn concentrations are shown in Figure 4.2. Technical difficulties with field equipment (frozen pumps, material damage following a flood event) led to gaps in He measurements in the Emme River. Missing data were reconstructed from estimates of diffusive He injection rates and from river discharge, as detailed in Appendix B. He concentrations in river water were also continuously measured for approximately one week ~100 m downstream of the river sampling location shown in Figure 4.2; no significant differences in He concentrations could be measured between both locations, suggesting minimal He degassing along the upper reach of the river (Blanc et al., in prep).

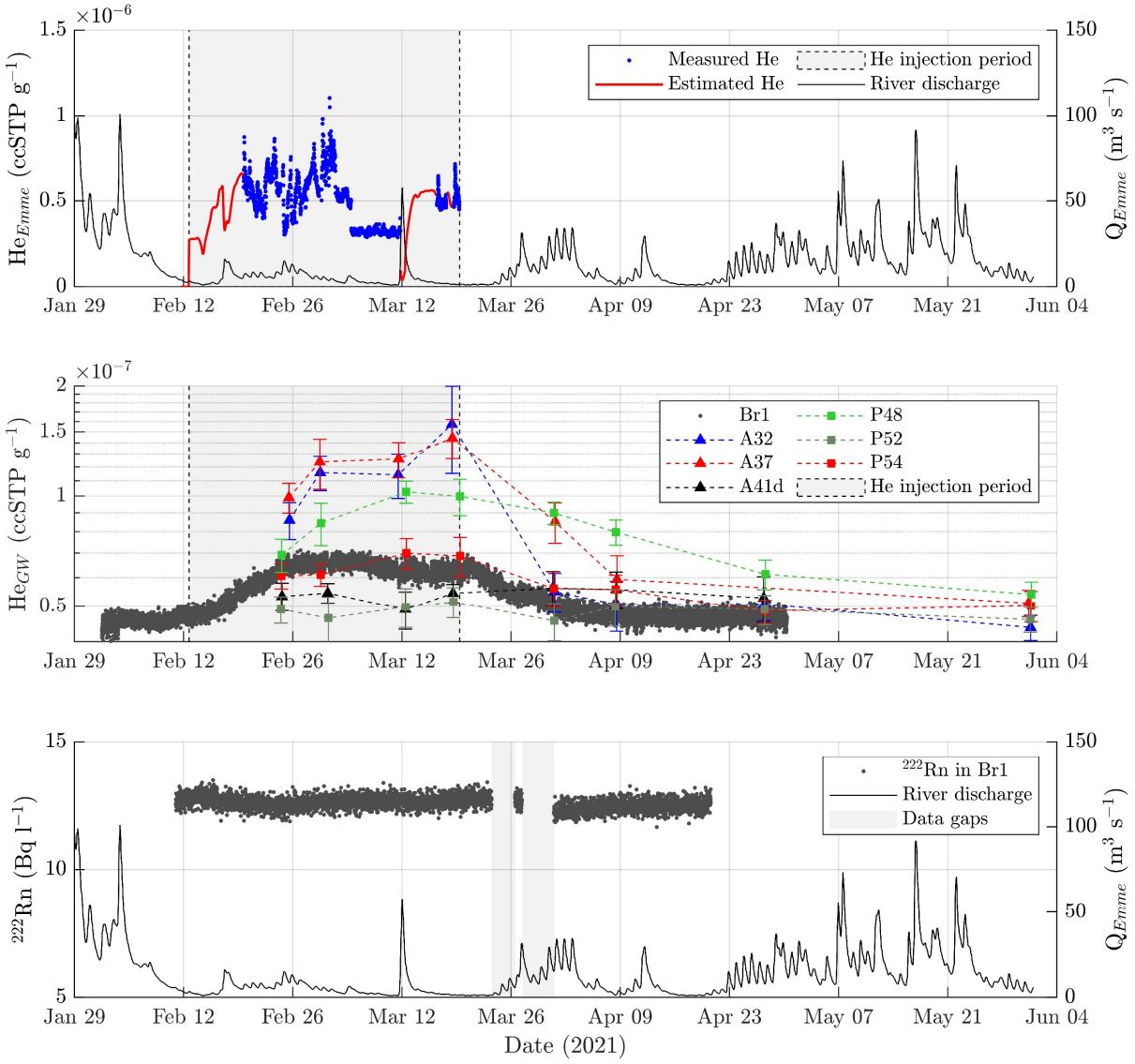

Figure 4.2 Helium and ^{222}Rn measurements during the 2021 tracer experiment (Blanc et al., in prep). Top : Measured and estimated He concentrations in the Emme river, and Emme discharge. Centre : Measured He concentrations in pumping well (Br1) and observation wells. Bottom : Continuous ^{222}Rn activities in pumping well (Br1).

Dissolved He concentrations in river water were supersaturated by a factor of 5 – 20 compared to natural concentrations for the duration of the experiment. A strong He signal was observed in pumping well Br1 as well as in observation wells A32, A37, P48, and P54. No significant He signal could be discerned in observation wells A41d and P52. Finally, ^{222}Rn was very stable at pumping well Br1 (12.5 ± 0.5 Bq l^{-1}) for the duration of the experiment, despite significant variations in SW discharge. Measured ^{222}Rn activities are below expected equilibrium activities at the study site (~14.5 Bq l^{-1}, Peel et al. (2022)), suggesting a consistent contribution of recently infiltrated SW (days to weeks) at this location.

4.3.2. GW flow and transport simulator

4.3.2.1. Simulator setup

For the simulation of flow and mass transport, the ISSHM HydroGeoSphere (HGS) was used (Aquanty, 2023). HGS simulates variably-saturated GW flow with a modified form of the Richards equation, whereas SW is simulated with the diffusion wave approximation of the Saint-Venant equations (Therrien and Sudicky, 1996). Interaction between surface and subsurface domains was implemented with the dual node approach (Ebel et al., 2009). Solute transport is simulated with the Fickian advection-dispersion equation (ADE), which also accounts for solute production and decay in variably-saturated conditions (Aquanty, 2023).

The HGS model of the Aeschau site used in the present study is based on the one described in Schilling et al. (2017), to which readers are referred for more detailed information on model setup and conceptualization. The finite element mesh was refined from originally ~11,000 elements to ~17,000 elements per layer (total ~250,000 elements) to ensure the stability of the mass transport solution (max grid Peclet number < 2, e.g. Zienkiewicz et al. (2014)). The mass-transport model was developed to simulate transport of measured solutes (^{222}Rn and He), SW/GW mixing ratios and mean GW age.

4.3.2.2. Boundary and initial conditions

A visual depiction of boundary conditions used for the flow and mass transport model is shown in Figure 4.3. The flow simulator was set up with the same overall boundary conditions as those described in Schilling et al. (2017), updated with conditions (i.e. precipitation, hydraulic heads, and river discharge) for the simulation period. A summary of solute properties used in the mass transport model is shown in Table 4.1.

Radon was modelled as a chemically inert solute with a decay constant of 0.1814 d^{-1} (half-life ~ 3.82 d). Zero ^{222}Rn concentrations were assigned to all surface nodes, as ^{222}Rn concentrations in SW in the southern portion of the Aeschau site are negligible compared to those found in GW (Peel et al., 2022;

Popp et al., 2021). This echoes the general tendency of any dissolved ^{222}Rn in SW bodies to be quickly volatilized when in contact with the atmosphere (Cecil and Green, 2000). ^{222}Rn in GW at the inlet were assigned a constant concentration of 12.5 Bq l^{-1}, in accordance with the stable values previously measured at background piezometer A41d (Peel et al., 2022; Popp et al., 2021). Aquifer ^{222}Rn production rates were considered spatially homogenous at the study site. Although some degree of spatial heterogeneity in ^{222}Rn production rates within the alluvial aquifer has been highlighted in a previous study, high GW flow velocities and hydrodynamic dispersion effectively blur the small-scale heterogeneities in ^{222}Rn production at the study site (Peel et al., 2023; Peel et al., 2022). A ^{222}Rn production rate of 2.63 Bq l^{-1} s^{-1} was assigned to all subsurface nodes, leading to a ^{222}Rn equilibrium activity of 14.5 Bq l^{-1}, equivalent mean ^{222}Rn activities measured in observation well P52, which is considered to be water at radioactive equilibrium with the local aquifer matrix (Peel et al., 2022). ^{222}Rn production rates in the unsaturated zone were adjusted to account for air/water partitioning of produced ^{222}Rn, using the method described in Delottier et al. (2022).

Helium (He) was simulated as a stable and conservative solute. As baseline He concentrations in observation wells are quite consistent and show little natural variations compared to the influence of the He injection (see Figure 4.2), only departures from baseline concentrations at each observation point were considered for the modelling study. That is, natural He in both SW and GW were set to zero, and any changes in He concentrations in the model domain are due to the infiltration of spiked river water into the aquifer. This simplifies the modelling of He, as complex phenomena such as excess air formation do not need to be considered. For the simulated He injection period, riverbed (RB) nodes were assigned specific He concentrations according to their location in relation to the two set of weirs at the study site. Nodes located on the upper reach of the river (i.e., upstream of the first weirs) were assigned He concentrations equal to the measured (or estimated, see Figure 4.2) He excess in river water. To account for He degassing downstream of the first set of weirs, He concentrations were set as equal to 25% of those on the upper reach. This value was estimated from subsequent point measurements of He concentrations immediately up- and downstream of the weirs. Finally, He concentrations in downstream riverbed nodes did not need to be constrained and were set to zero, as they are too far downgradient to have had an effect on simulated He concentrations in observation wells.

No helium gas exchange between the saturated and unsaturated portions of the aquifer is considered, although this effect has been documented in previous He tracer tests undertaken in variably-saturated systems (e.g. Brennwald et al. (2022); Carter et al. (1959); Gupta et al. (1994)). However, we consider the resulting loss of tracer mass and apparent He retardation (through gas entrapment) to be minimal

for the short timescales modelled here (i.e., days to weeks; see sections 4.5.2 and 4.5.3 for a more detailed discussion concerning this assumption).

Fluorescein was simulated as a stable and conservative solute. All riverbed nodes in the model domain were assigned a specific concentration equal to measurement values from the river monitoring location. Although this compound is known to be light-sensitive, its transit time along the relevant river reaches is too short (on the order of minutes) to lead to substantial degradation. Moreover, for the modelled period, SW-GW exchange fluxes are marginal compared to river discharge, leading to insignificant dilution of this compound from GW exfiltration.

SW/GW mixing ratios were modelled by simulating a virtual stable and conservative solute ("*Mixing*"), and assigning distinct concentrations to surface and GW inlet nodes (i.e. orange region in Figure 4.3). All surface and riverbed nodes were assigned a concentration of zero, whereas GW inlet nodes were assigned a concentration of unity. The resulting subsurface solute concentrations directly provide the fraction of GW at any point in the model domain.

Mean GW age was directly simulated through the advection-dispersion transport equation, and modelled as a conservative solute ("*Age*"); 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)). As no reliable information on mean GW age near the model inlet exists, an arbitrary value of zero was applied to all inlet GW nodes. A zero-age BC was also applied to all surface nodes. Disentangling the mean age of both SW and GW components was achieved by modelling an additional unstable, chemically inert solute (herein referred to as "*MG"*), according to the method outlined in Massoudieh and Ginn (2011) (see Appendix A for detailed information).

Free-solution diffusion coefficients for both ^{222}Rn and age solutes were set to 10^{-9} m^2/s (Goode, 1996; Ishimori et al., 2013), to $5 \cdot 10^{-9}$ m^2/s for He (Verhallen et al., 1984), and $3.5 \cdot 10^{-10}$ m^2/s for fluorescein (Sabatini, 2000).

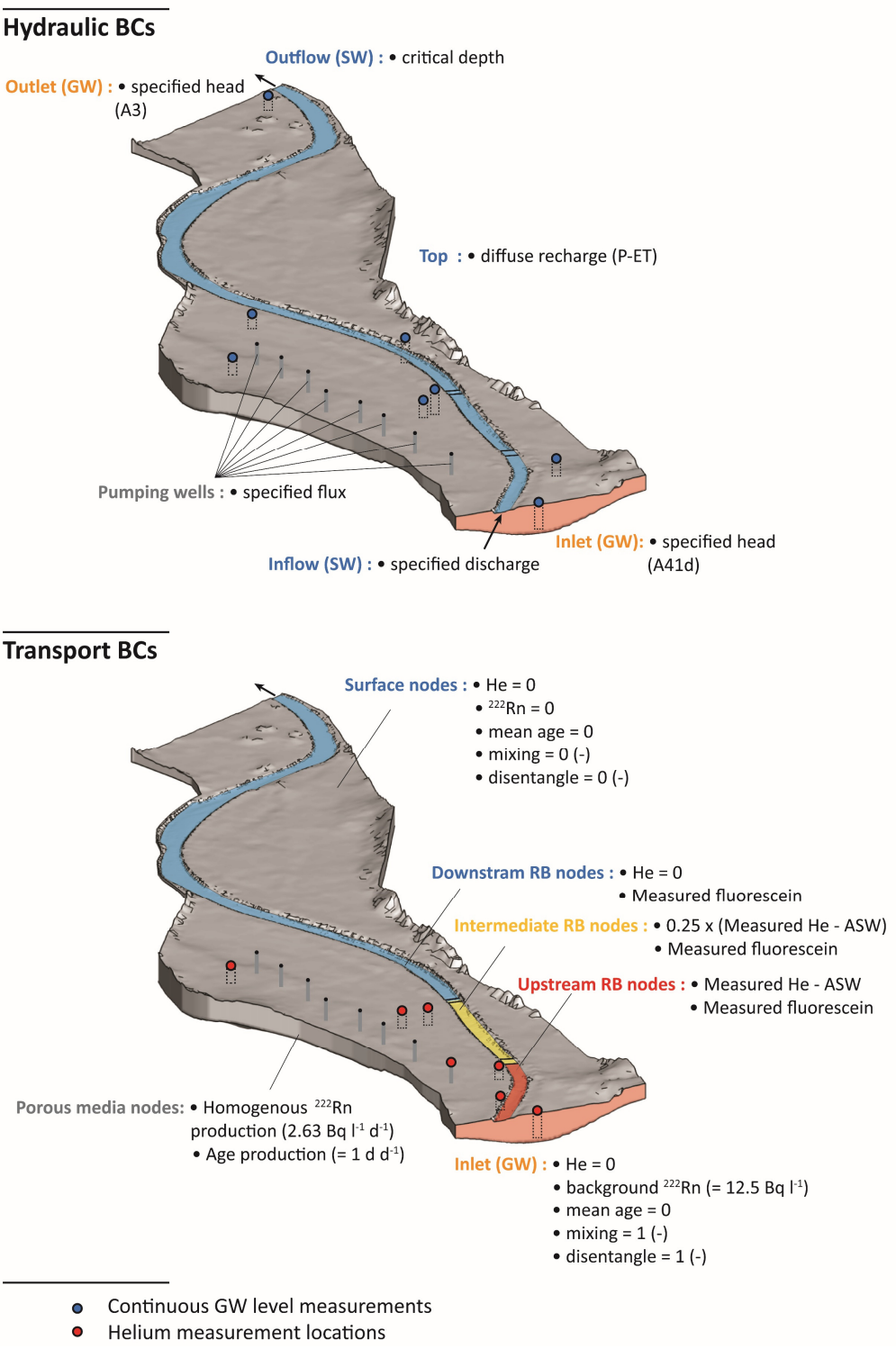

Figure 4.3 Flow and transport boundary conditions for the numerical model of the Aeschau site (see text for details on their implementation). Vertical exaggeration ~2.5x. Measurement locations for GW levels (blue dots, top figure) and helium concentrations in GW (bottom figure) are shown. RB : riverbed. Weirs are shown as dark lines perpendicular to the river.

The flow model was forced by six-hour averaged observations of streamflow from the upstream *Eggiwil-Heidbüel* national gauging station (FOEN, 2023), six-hour averaged observations of hydraulic heads as measured in observation wells A41d and A3, daily averaged values of precipitation from the

nearby *Langnau i.E.* national meteorological station (MeteoSwiss, 2023) minus potential evapotranspiration, as well as daily averaged pumping rates from wells Br1 to Br8 (see Figure 4.3).

Before the transient simulation of the helium injection experiment, every model instance was subjected to a 2,000-day spin-up period with hydraulic forcings corresponding to the stable conditions observed on Jan 06, 2021, approximately one month before the start of the injection experiment. Mass transport was forced for the spin-up with boundary conditions shown in Figure 4.3, with zero He concentration throughout the model domain. For the gas injection period, the mass-transport model was forced with hourly averaged observations of He concentrations (adjusted for expected He volatilization downstream of the first weirs). Finally, the model was forced with 15-minute averaged observations of fluorescein concentrations during the nine modelled dye tracer injections.

Table 4.1 Modelled solute properties. λ_{Rn}: ^{222}Rn decay constant; λ_{MG} decay constant for hypothetical solute "MG", used to disentangle mean age of regional and recently infiltrated GW (see text for details); $\gamma_{Rn,sat}$: zero-order source for ^{222}Rn (in saturated conditions); γ_{Age}: unit source term for age ; D_0^i: free-solution diffusion coefficient for given solute i (^{222}Rn, Age, Mixing, MG, He, or Fluorescein)

Property	Assigned value
$\boldsymbol{\lambda_{Rn}}$	0.1814 (d^{-1})
$\boldsymbol{\gamma_{Rn,sat}}$	2.63 (Bq l^{-1} d^{-1})
$\boldsymbol{\gamma_{Age}}$	1 (d d^{-1})
$\boldsymbol{\lambda_{MG}}$	10^{-3} (d^{-1})
$\boldsymbol{D_0^{Rn,Age,Mixing,MG}}$	10^{-9} (m^2 s^{-1})
$\boldsymbol{D_0^{He}}$	$5 \cdot 10^{-9}$ (m^2 s^{-1})
$\boldsymbol{D_0^{Fluorescein}}$	$3.5 \cdot 10^{-10}$ (m^2 s^{-1})

4.3.3. Model parameters and parameter estimation

4.3.3.1. Hydrogeological framework and simplified (hydro)geological models

Fluvioglacial aquifers such as the Aeschau aquifer are characterized by complex and heterogenous sediment structures and hydraulic properties, which reflect the dynamic and changing interplay between depositional and erosional processes during their formation (Heinz et al., 2003; Huber and Huggenberger, 2015). These aquifers typically exhibit several connected high-conductivity (*K*) features, which can lead to the channelling of flow and mass transport along preferential pathways. Although the geometry and connectivity of sedimentological units in the Aeschau sector are poorly constrained, the existence of large-scale (100s of meters) preferential GW flow pathways at the study site had already been strongly suggested from the first recorded hydrogeological investigations of the Aeschau sector in the first part the 20th century (Arbenz and Peter, 1925; Gubelmann, 1930). This hypothesis has been strongly corroborated by subsequent modelling and tracer investigations (Popp et al., 2021; Schilling et al., 2022). However, the typical lack of sedimentological, geological and

geophysical data at the study site implies that the geometry, extent, connectivity, and hydraulic properties of these preferential flow paths remain largely unresolved. It follows that any model of flow and mass transport at the field site needs to reconcile the lack of geological and structural subsurface data with the importance (or need) of representing connected high-*K* structures. The framework presented below attempts to balance these two features and favour the emergence of preferential flow structures in post-history-matched parameter ensembles by non-restrictive conditioning to simplified sedimentological models of the aquifer.

To this end, a simplified conceptual model consisting of two hydrofacies which represent (i) connected and highly-conductive gravel and sandy gravel trough and channel fill deposits, and (ii) less conductive gravel sheet deposits is proposed for prior model parameter ensembles. Estimates of relative proportions and representative hydraulic properties of both hydrofacies are based on geological well-log descriptions and results from mini-pumping tests, as thoroughly described in Appendix C. We recognize that the discrimination of glaciofluvial deposits into two categories based on the visual description of borehole logs grossly simplifies the true sedimentological and hydrogeological complexity of the aquifer. However, the objective here is not to provide a detailed model of subsurface structure and heterogeneity, but rather to estimate plausible ranges of the connectivity and volumetric fraction of sediments potentially acting as preferential flow pathways.

The stochastic event-based modelling software ALLUVSIM (Pyrcz et al., 2009) was used to create 100 (prior) realizations of shallow (<15 m) subsurface structure based on the above conceptual model. As there is little to no data constraining the hydraulic and structural properties of deeper structures, homogenous aquifer properties were assumed at greater depths (> 15 m). Conductive facies, representing preferential flow structures, were modelled with ALLUVSIM as “channel fill” facies; other units are merged into a single hydrofacies (“matrix” facies). Parameters controlling the shape (e.g., sinuosity, width, depth) of the former were assigned typical values for braided river alluvial aquifers (e.g. Huber and Huggenberger (2015); Huggenberger and Regli (2006)), albeit with high variance, in order to sample a large range of possible preferential flow structures. A detailed workflow describing the construction of geological models can be found in Appendix C.

4.3.3.2. Prior parameter ensembles

Data from several different sources were combined to set the prior bounds on estimated model parameters (i.e. horizontal hydraulic conductivity (K_H), porosity (φ), and longitudinal dispersivity α_L), see Table 4.2). Preliminary calibration of the flow model against hydraulic heads was carried out with the classic Gauss-Levenberg-Marquardt (GLM) algorithm within PEST++ (Welter et al., 2015), and used as a basis to produce prior estimates of mean aquifer and riverbed K_H and φ. These prior simulations

also constrained values of vertical anisotropy of hydraulic conductivity (K_H/K_V, the ratio of horizontal and vertical hydraulic conductivities) for both the riverbed and aquifer, respectively at 4 and 12 (-). The estimated values of K_H/K_V for both the riverbed and aquifer are consistent with published values for layered sediments in similar alluvial settings (e.g. Chen (2000); Freeze and Cherry (1979); Lu et al. (2012)). Prior estimates of dispersivity are based on the size of the model domain, and representative of sampling scales between 1 and 1500 m (Gelhar et al., 1992; Schulze-Makuch, 2005). Transverse dispersivities (horizontal and vertical) were set as equal to 1/10th of longitudinal dispersivity.

As the emphasis in this study in terms of model parameters is on characterizing subsurface (i.e. aquifer) heterogeneity, the riverbed was assigned spatially constant hydraulic properties (although distinct in each realization). Riverbed hydraulic conductivity was assumed stationary for the study period. This assumption is supported by the lack of discernible hydraulic or tracer signals of riverbed unclogging (e.g. preferential infiltration) following high-discharge events. Moreover, river water temperature was relatively stable during the study period (3 – 6 °C`), leading to marginal effects of water temperature on streambed *K*. The width of the riverbed layer was artificially increased to prevent focused infiltration of SW at the interface between the riverbed and aquifer during high-flow events. Aquifer porosity and dispersivity were also considered homogenous, and only hydraulic conductivity was allowed to vary spatially. Although spatial variations in porosity and dispersivity obviously impact solute transport, the scale of variability of hydraulic conductivity is much larger than that for other parameters. We therefore consider spatial variations in *K* as the most uncertain and impactful on simulated solute concentrations.

For each prior model, the three facies produced with ALLUVSIM ("matrix", "channel fill", and "deeper aquifer") were assigned distinct but homogenous values of K_H. Differences in K_H between channel and matrix facies were based on vertical conductivity profiles from two wells in the study area, as described in Appendix C. Hydraulic conductivity of the deeper aquifer facies was assigned a value equal to the harmonic mean of matrix and channel conductivities, weighted by their relative abundances (or net-to-gross ratio, see Appendix C). In summary, for each of the 100 prior realizations, a total of six parameters (three values of K_H, two of φ, and one of α_L) were randomly drawn from the prior parameter distributions given in Table 4.2.

Table 4.2 Prior parameter values and bounds for aquifer and riverbed hydraulic and mass-transport properties (see text for details). $K_{H,RB}$, $K_{H,MT}$, $K_{H,CH}$: horizontal hydraulic conductivities of the riverbed (RB), aquifer matrix (MT) and channel fill (CH) facies; α_L : longitudinal dispersivity; $\varphi_{MT,CH}$, φ_{RB} : porosity of the riverbed (RB), aquifer matrix (MT) and channel fill (CH) facies.

Property	Prior mean ± 1σ (bounds)
$\boldsymbol{logK_{H,RB}}$ **(K in m d^{-1})**	0.18 ± 0.5 (-2 ≤ $logK_{H,RB}$ ≤ 2)
$\boldsymbol{logK_{H,MT}}$ **(K in m d^{-1})**	2.5 ± 0.18(-1 ≤ $logK_{H,MT}$ ≤ 3.3)
$\boldsymbol{logK_{H,CH}}$ **(K in m d^{-1})**	3.0 ± 0.13 (-1 ≤ $logK_{H,CH}$ ≤ 3.3)
$\boldsymbol{(K_H/K_V)_{RB}}$ **(-)**	4 (-)
$\boldsymbol{(K_H/K_V)_{MT,CH}}$ **(-)**	12 (-)
$\boldsymbol{\varphi_{RB}}$ **(-)**	0.3 ± 0.1 (0.1 – 0.45)
$\boldsymbol{\varphi_{MT,CH}}$ **(-)**	0.2 ± 0.1 (0.05 – 0.3)
$\boldsymbol{\alpha_L}$ **(m)**	10 ± 5 (0. 1 – 50)

4.3.3.3. Objective function and parameter estimation

Parameter estimation was carried out with an iterative ensemble smoother (iES) implementation of the Gauss-Levenberg Marquard (GLM) algorithm, as implemented in PEST++IES (Chen and Oliver, 2013; White, 2018). Data used for history matching included continuous measurements of hydraulic heads in six observation wells, continuous helium (He) and ^{222}Rn measurements in pumping well Br1, and regular He measurements in six observation wells (see *Supplementary material* for a summary of observations). The final observation dataset used for calibration consisted of 1624 distinct observation points representing three observation types (i.e., heads, He, and ^{222}Rn).

To evaluate the added contributions of tracer measurements to "classical" hydraulic head observations, three distinct history matching runs are undertaken using the same data assimilation scheme and initial parameter ensemble: the first is only updated using head observations, the second includes head and dissolved He, whereas the third includes hydraulic head, He, and ^{222}Rn observations. As each observation type provides important information on model parameters and prediction-relevant processes, we adopt a weighting scheme that allows each observation type to inform model parameters in a balanced fashion according to the data used for history matching (Doherty and Welter, 2010; Schilling et al., 2019). Base weights for each observation are set equal to the inverse of measurement uncertainty; final weights result from adjusting base weights so their relative contribution to the objective function follows the desired balancing between observation types. Relative contributions of different observation types to the objective function are shown in Table 4.3.

Table 4.3 Relative contributions to the objective function assigned to different observation groups for different assimilation schemes (assimilation of head measurements, of heads + He, and of heads + He + Rn).

Relative weights	***Assimilated observations***		
	Heads	**Heads + He**	**Heads + He + ^{222}Rn**
Heads (all)	1	0.5	0.35
He (Br1)	-	0.3	0.25
He (OW)	-	0.2	0.25
^{222}Rn (Br1)	-	-	0.15

Owing to the uncertainty arising from the use of peristaltic pumps on measured dissolved gas concentrations, He observations from sampling points A32 and A37 were assigned a weight of ½ the weight of He observations in other piezometers. To avoid unreasonable compensation of data mismatches through extreme parameter values (and subsequent predictive bias), measurements not encompassed by the simulations run with model priors were not considered during the parameter estimation process.

The high contribution of tracer observations to the objective function reflects (i) the excellent resolution of tracer measurements (notably in observation well Br1), (iii) the rich information content of acquired tracers in terms of informing the quantities of management interest considered here, and (iii) the expected sensitivity of tracers to geological heterogeneity and structure (i.e., preferential pathways).

Estimation of *aquifer* hydraulic conductivity was carried out on an elemental basis (i.e. for each aquifer FEM element), whereas single parameter values were estimated for aquifer porosity ($\varphi_{MT/CH}$), longitudinal dispersivity (α_L), vertical anisotropy $(K_H/K_V)_{MT,CH}$, as well riverbed hydraulic conductivity ($K_{H,RB}$), porosity (φ_{RB}) and anisotropy $(K_H/K_V)_{RB}$. Drastic departures from prior model parameters fields during the parameter upgrade process are dampened through singular value decomposition truncation (TSVD), which acts as a form of subspace regularization (Lawson and Hanson, 1995; Weiss and Smith, 1998). The number of singular components used for parameter upgrades was limited to 80 i.e., 80% of the ensemble size. No additional enforcement of preferred parameter values or structural parameter continuity was imposed during parameter upgrades. This approach represents a balance between preserving prior model structural elements, whilst allowing the local emergence of discontinuities and departures from prior parametrization as commanded by field data. The degradation of prior structure through parameter upgrade is expected and even encouraged through this approach; its goal is to favour the preservation of preferential pathways controlling solute transport which are known to exist at the field site (Popp et al., 2021; Schilling et al., 2022), without overly constraining posterior parameter fields to inevitably questionable subsurface structure expressed through the prior models. Although the latter are meant to represent as much as

possible expert knowledge of subsurface properties, they are subject to incomplete and fallible description and high levels of uncertainties.

4.3.4. Data worth assessment and predictions

We quantify the relative worth of hydraulic and tracer observations (^{222}Rn and He) through their ability to inform model parameters as well as several quantities (referred to as ”predictions” herein) of possible management interest. Within the ensemble data assimilation framework presented here, we measure the data worth of a particular calibration dataset as the relative decrease in the inter-ensemble variance of parameter values or of specific predictions achieved through data assimilation.

Specifically, the following management-relevant quantities are assessed : (i) the fraction of river water in pumping well Br1, (ii) the mean age of river water pumped at this location, and (iii) the river infiltration flux along the upper reach of the Emme river, upstream of the first set of weirs (see Figure 4.1). Rather than running pre- and post-calibration parameter ensembles through a separate “prediction” modelling phase, the quantities listed above were simulated for the entirety of the calibration period, as detailed in section 4.3.2.2. Field conditions during this period were highly variable and sampled a wide range of expected environmental conditions at the field site (see Figure 4.2). Alternating periods of low-flow, snowmelt-dominated high-flow, and storm flood events serve well to illustrate system behaviour in dynamic environmental conditions.

4.4. Results

4.4.1. Post history-matched flow and transport models

For all data assimilation schemes considered in the present study, a single iteration of the iES algorithm led to an almost identical and adequate level of fit between modelled and observed transient groundwater heads (posterior global RMSE ≈ 23-24 cm after history matching, see Table 4.4). In all cases, hydraulic head dynamics were very well captured by the models (see Supplement S1), albeit with some systematic bias in observation wells P51 (mean error (ME) ≈ + 40 cm) and P53 (ME ≈-50 cm)). No further iteration was conducted, and model outputs may be compared to easily quantify the added contribution of tracer measurements, in that all post-history-matched model ensembles provide similar performance in terms of reproducing classical head observations.

Table 4.4 Summary of RMSE of simulated heads, helium (He) and ^{222}Rn achieved through history matching of different combinations of observation types. "He Br1" represents continuous He measurements in pumping well Br1, whereas He OW represents measurements in all other observation wells. The lowest RMSE for each observation is shown in boldface.

RMSE	***Assimilated observations***			
	Prior	**Heads**	**Heads + He**	**Heads + He + ^{222}Rn**
Heads (cm)	49	24	**23**	24
He Br1 (10^{-9} ccSTP g^{-1})	49	23	**3**	4
He OW (10^{-9} ccSTP g^{-1})	83	67	35	**32**
^{222}Rn (Bq l^{-1})	0.7	0.3	0.4	**0.3**

Prior model parameter fields as well as those assimilating only hydraulic data tend to significantly overestimate the amount of dissolved He in GW, often by a factor of 3 or more in all observation wells (see Figure 4.4). Only through the assimilation of He data (with or without ^{222}Rn, see Table 4.4) could He observations approximately be reproduced, despite some significant discrepancies at earlier simulation times. Indeed, simulated values at locations A32, A37 and P54 tend to overestimate measured He concentrations, most notably in the first weeks of the injection experiment. This can be linked to high levels of uncertainty in river He concentrations at the beginning of the experiment, which had to be estimated from river discharge and assumed injection efficiency (see section 4.3.1.1 and Appendix D). Furthermore, low measured values in piezometers A32 and A37 may be related to the use of peristaltic pumps to sample water in these piezometers (see section 4.3.1.1). Finally, the mismatch between observations and simulations at P54 near the start of the injection experiment may be linked to uncertainty in the Dirichlet He BC along the intermediate reach of the river, where He concentrations were estimated from measurements made after the injection experiment. As detailed in section 4.3.3.3, He measurements in GW which were incompatible with model priors were discarded from the calibration dataset to avoid unreasonable parameter adjustments and subsequent predictive bias.

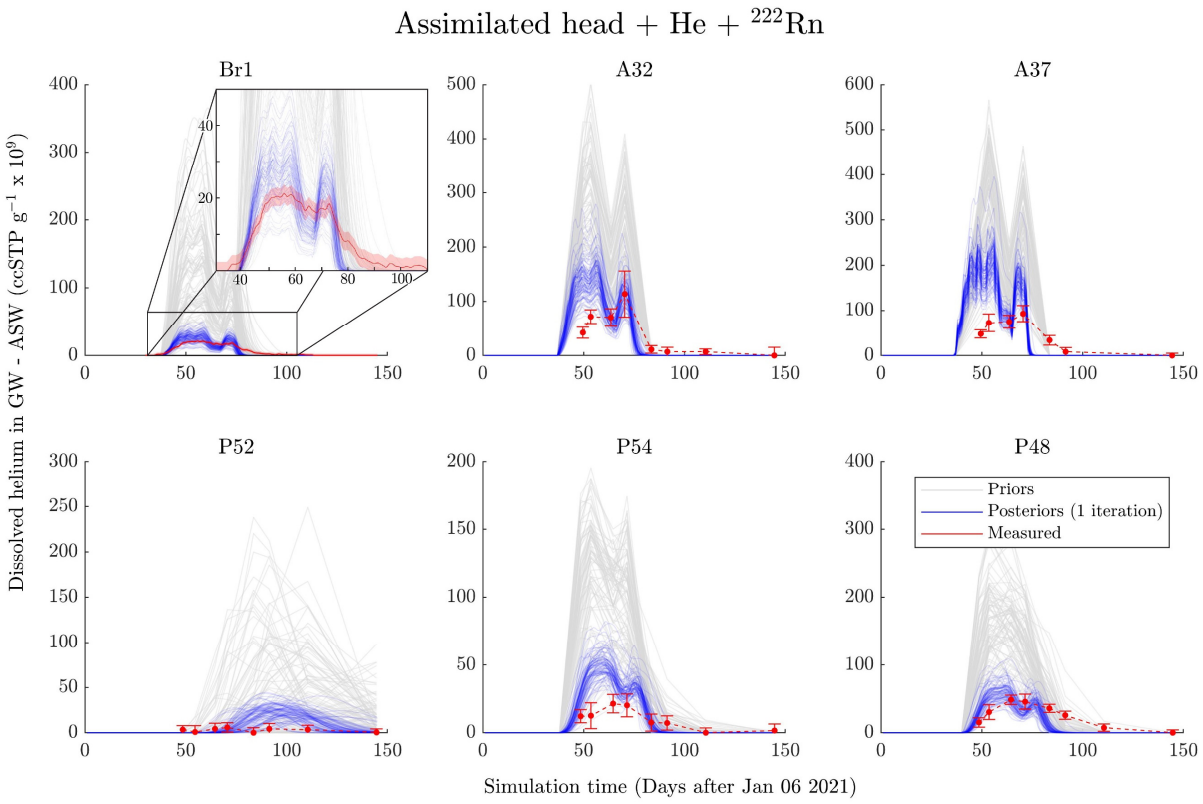

Figure 4.4 Simulated (after assimilation of He and ^{222}Rn) and observed He concentrations in pumping well Br1 and the five targeted observation wells. The grey lines represent simulated He concentrations using prior parametrizations, whereas the dark blue lines are outputs from models updated with head and He measurements.

At later times and in other observation wells however, simulations successfully reproduced observed He concentrations as well as transient dynamics (RMSE ≈ 30·10^{-9} ccSTP g^{-1}in observation wells and ≈ 4·10^{-9} ccSTP g^{-1} in Br1), although signals are less smoothed than those observed (Figure 4.4). This is most evident in pumping well Br1, where despite an excellent overall agreement between observed and simulated signals, both modelled He peak breakthrough and tailing (following cessation of He injection) are steeper than measured responses. This likely results from at least two overlapping effects : (i) He tracer retardation and loss by interaction with the unsaturated zone, and (ii) insufficient levels of modelled subsurface heterogeneity and the non-representation of (very) low-*K* areas in the subsurface, in which solute retention for longer timescales favourably occurs (e.g. Bouchaud and Georges (1990)). As mentioned in section 4.3.2.2, the former has been documented in other unconfined systems and is likely a relevant process in the Aeschau system, as discussed in the following paragraph. As for the latter effect, the inversion framework adopted herein does not favour the emergence of zones exhibiting sharp contrasts in hydraulic properties in posterior parameter ensembles (see section 4.3.3.3), although such areas are likely interspersed throughout the alluvial plain. A final observation relates to the earliest tracer arrival measured in Br1, which is poorly captured by the models, highlighting the existence of local areas of strong preferential flow that remain unresolved in prior and updated model parameter fields. This is illustrated in Figure 4.5, which shows

the distribution of "fast" He arrival times in pumping well Br1 (t_{Q10}, defined as arrival time of the 10th percentile of peak He breakthrough in the well from He injection onset). Although post-history-matched models encompass measured t_{Q10}, they tend to systematically overestimate early tracer arrival times by an average of ~0.5-1 day (average t_{Q10} ≈ 3.4-3.9 days for history-matched models, vs ~2.9 days measured).

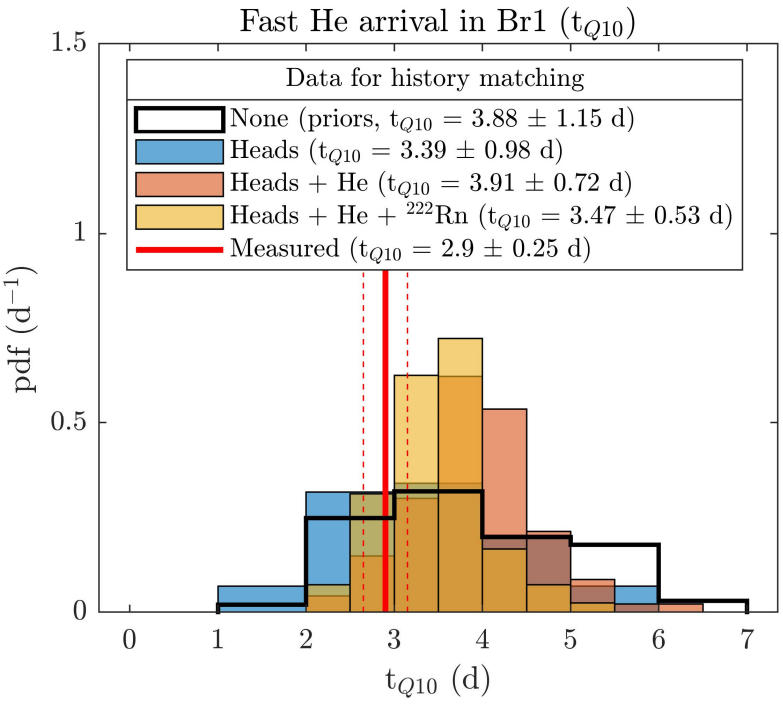

Figure 4.5 "Fast" He arrival times in pumping well Br1. t_{Q10} is the 10th percentile of peak He breakthrough times in the well from He injection onset (in days).

Generally similar trends emerge when comparing outputs of updated mass-transport models to fluorescein measurements in pumping well Br1 (see Appendix D). Indeed, only through the assimilation of tracer (He and ^{222}Rn) measurements are models capable of adequately reproducing independent fluorescein observations (Appendix D), as models conditioned to heads alone tend to overestimate solute concentrations. Nevertheless, the assimilation of He data leads to moderate but significant underestimates of modelled fluorescein concentrations and tracer recovery rates compared to field measurements - an effect dampened by the further assimilation of ^{222}Rn. This could result from a number of factors, including (i) infiltration of fluorescein between dye tracer and He injection points, which is not considered by the model, (ii) biased conceptualization of mass-transport BCs along the riverbed (e.g., variable levels of lateral tracer mixing, constant tracer concentrations along river reaches), and (iii) parameter compensation and bias resulting from unaccounted He loss during infiltration and transport, e.g. through interactions with the unsaturated zone. Based on our understanding of field conditions and He tracer behaviour (e.g., apparent retardation, tailing), we believe the latter factor is most impactful on low modelled dye tracer recovery. The resulting

parameter compensation can be illustrated by considering measured and modelled relative He and dye tracer recovery rates in pumping well Br1 (i.e. the ratio of He and fluorescein recovery rates): measured relative recovery is 0.70 ± 0.25, whereas modelled values are higher at 0.85 ± 0.02. Assuming model boundary conditions are properly conceptualized, this would suggest He tracer loss of up to 20 ± 10 % between the river and the pumping well. Reduced He concentrations will favour posterior parameter fields which strongly limit SW infiltration (i.e., through lower riverbed conductivity, leading to lower modelled He concentrations in the subsurface); modelled concentrations of other infiltrating solutes will therefore tend to be negatively biased, i.e., lower than measured. Despite the expected biased posterior estimates of riverbed conductivity, differences in measured and modelled recovery rates remain modest, and the robust reproduction of fluorescein measurements by these models greatly substantiates their reliability in terms of capturing relevant mass-transport processes.

4.4.2. Parameter fields

Ensemble mean $\log_{10}K$ fields as well as the ensemble variance of $\log_{10}K$ are shown in Figure 4.6 for (i) prior parameter fields, as well as post history-matched models updated with (ii) hydraulic head data, (iii) head + He observations, and (iv) head + He+ ^{222}Rn observations. Results from randomly selected realizations are also shown in section 4.9.2. Finally, histograms of prior and posterior parameter values for porosity (φ_{RB}, $\varphi_{MT,AQ}$), longitudinal dispersivity (α_L), and riverbed permeability are shown in Figure 4.7.

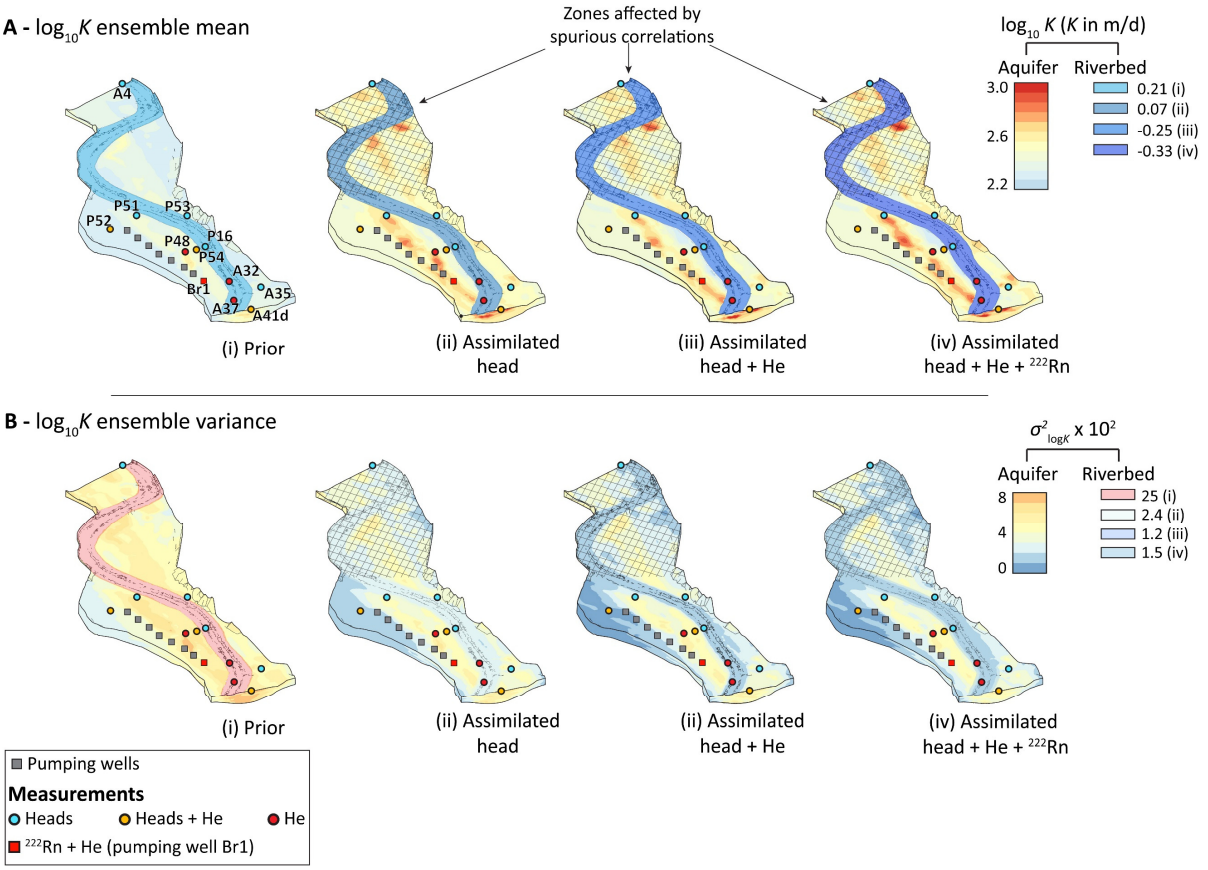

Figure 4.6 Ensemble mean prior and posterior parameter fields. Note that the riverbed zone was artificially widened to prevent any focused infiltration at the aquifer/streambed interface during high-flow events.

Overall, prior mean aquifer hydraulic conductivity is homogenous, albeit slightly elevated in the central portion of the domain owing to the higher density of "channel" facies in prior models (Figure 4.6A). The values and spatial distributions of aquifer *K* resulting from both data assimilation strategies are largely similar, with mean *K* being slightly higher in posterior parameter fields (prior mean $\log_{10}K$ ≈2.54, and of ≈2.78 for posteriors). The added contribution of tracer data in informing aquifer *K* is however manifest by the noticeably sharper contrasts between zones of varying aquifer *K* when assimilating He + ^{222}Rn observations, most notably the area between the pumping wells and the left bank of the river. All data assimilation schemes result in significantly lower riverbed conductivity, with the specific contribution of tracer measurements being much more noticeable (mean $\log_{10}K$ ≈0.21 for priors, ≈0.07 with assimilation of heads, and ≈-0.25 and -0.33 with assimilation of heads + He and heads + He + ^{222}Rn respectively, see Figure 4.6).

Comparable reductions in the variance of aquifer hydraulic conductivity (σ^2_{logK}) result from the assimilation of head and head + tracer data (mean reductions of ~30% and 40 % of σ^2_{logK} respectively, see Figure 4.6B and Supplement S1). Riverbed conductivity ($K_{H,RB}$) is much more strongly constrained, with reductions of parameter variance of over 90% in all three cases (Table 4.5).

Of the other adjustable parameters, no reduction in parameter uncertainty is achieved through the assimilation of hydraulic head observations (Table 4.5). Riverbed porosity (φ_{RB}) is not constrained by any dataset, whereas uncertainties of aquifer mass transport parameters (porosity [$\varphi_{MT,CH}$] and longitudinal dispersivity [α_L]) are unsurprisingly reduced by the assimilation of tracer data.

Table 4.5 Parameter variance and reductions in parameter uncertainty (Red σ^2) compared to prior parameter fields. $K_{H,RB}$: riverbed hydraulic conductivity; φ_{RB} : riverbed porosity; $\varphi_{MT,CH}$: aquifer porosity; α_L: longitudinal dispersivity. The lowest values of σ^2 are indicated in boldface.

Parameter	**Prior**	***Assimilated observations***					
		Head		**Head + He**		**Head + He + ^{222}Rn**	
	σ^2_{prior}	σ^2	Red σ^2 (%)	σ^2	Red σ^2 (%)	σ^2	Red σ^2 (%)
$logK_{H,RB}$	$24 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$	90%	$\mathbf{1.2 \cdot 10^{-2}}$	**95%**	$1.5 \cdot 10^{-2}$	94%
φ_{RB}	$\mathbf{4.5 \cdot 10^{-3}}$	$4.8 \cdot 10^{-3}$	-6%	$4.5 \cdot 10^{-3}$	0%	$4.7 \cdot 10^{-3}$	-5%
$\varphi_{MT,CH}$	$1.3 \cdot 10^{-2}$	$1.2 \cdot 10^{-2}$	9%	$0.7 \cdot 10^{-2}$	50%	$\mathbf{0.5 \cdot 10^{-2}}$	**60%**
$log\alpha_L$	$10 \cdot 10^{-2}$	$9.6 \cdot 10^{-2}$	4%	$2.8 \cdot 10^{-2}$	72%	$\mathbf{2.4 \cdot 10^{-2}}$	**76%**

Connected high-*K* structures tend to emerge in posterior parameter fields in the down-valley direction, parallel to the eight pumping wells (Figure 4.6). Although largely similar in all cases, these structures are more clearly defined when He and ^{222}Rn data is also assimilated. Differences in *K* between these structures and the surrounding aquifer matrix are of approximately half an order of magnitude, and will result in focused flow and transport from the model inlet and Emme River towards the first set of pumping wells, and further downgradient in the centre of the plain towards piezometer P51. The more heterogeneous and discontinuous structures that emerge further downgradient of the wellfield are likely artefacts resulting from the simplified inversion framework employed here, as little data constrains aquifer hydraulic properties in this area. Indeed, estimated aquifer *K* in the northern portion of the model materialize from prior parameter covariances and, as no localization was employed, spurious correlations with measurements in the southern part of the model domain. Estimated aquifer parameters in the northern part of the model are therefore highly uncertain and should not be regarded as reliable.

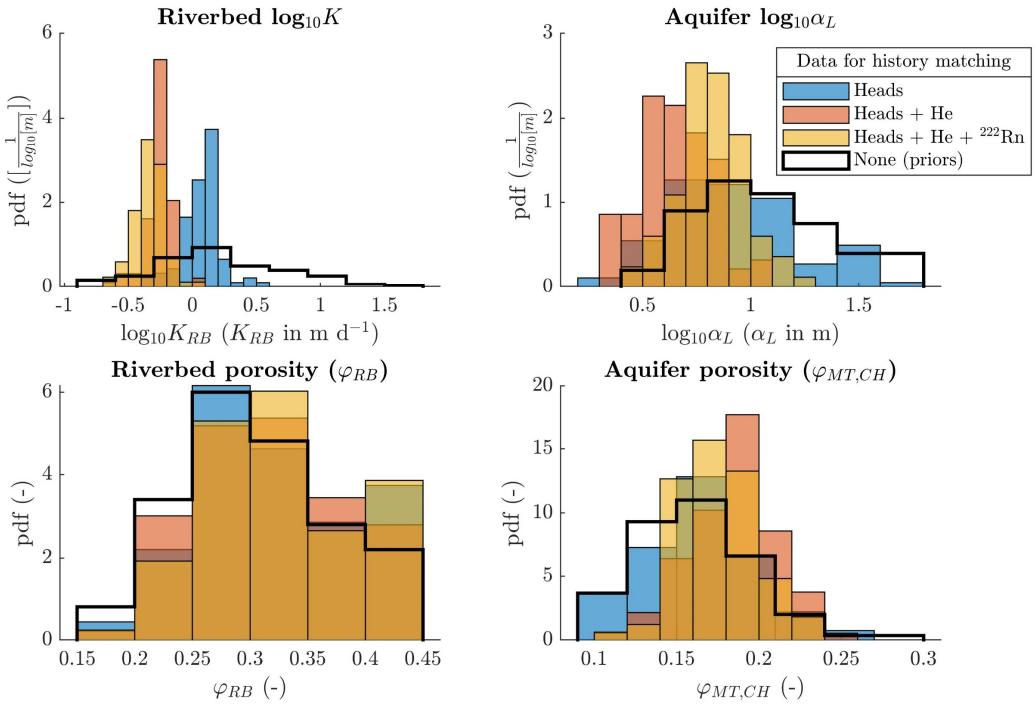

Figure 4.7 Prior and post history-matched parameter distributions. Note that these parameters are spatially constant for each ensemble member, and only vary between ensembles.

4.4.3. Predictive performance

Values of the selected management-related quantities for the simulation period, namely (i) mean GW age at pumping well Br1, (ii) the fraction of recently infiltrated SW at Br1, and (iii) infiltration flux on the upper reach of the Emme river, are illustrated in Figure 4.8 for prior and posterior parameter ensembles (with assimilated head and head + He + ^{222}Rn).

Table 4.6 "Predictions" at simulation time t = 145 days. Ensemble mean and ensemble variance. Values with the lowest standard deviation are shown in boldface.

Prediction	**Prior**	***Assimilated observations***		
		Heads	**Head + He**	**Heads + He + ^{222}Rn**
		(mean ± 1σ)		
$a_{mean,SW,Br1}$ (d)	5.9 ± 2.2	4.6 ± 1.5	5.3 ± 1.2	**4.5 ± 0.8**
$f_{SW,Br1}$ (d)	0.31 ± 0.19	0.17 ± 0.09	0.07 ± 0.02	**0.07 ± 0.02**
$Q_{inf,top}$ (l s^{-1})	105 ± 111	53 ± 19	**25 ± 6**	**21 ± 6**

All selected predictions are strongly constrained by the assimilation of hydraulic and tracer data, as illustrated by the significantly lower posterior ensemble spread at all simulation times, with the lowest degree of predictive variance associated with the simultaneous assimilation of both head and tracer (He + ^{222}Rn) measurements (Table 4.6). Compared to prior parameter fields, predictions from calibration-constrained models exhibit overall decreases in (i) mean age of the recently infiltrated portion of SW ($a_{mean,SW,Br1}$) at Br1, (ii) the fraction of recently infiltrated SW pumped at well Br1

($f_{SW,Br1}$), and (iii) the infiltration flux from along the upper reach of the river ($Q_{inf,top}$). These observations can be illustrated by comparing predictions given by prior and calibration-constrained models at the end of the simulation period (t = 145 days): estimates of $a_{mean,SW,Br1}$ range from 5.9 ± 2.2 days for priors to 4.4 ± 0.8 days after assimilation of head and tracer (He + ^{222}Rn) data; the same prior and posterior ensembles yields estimates of $f_{SW,Br1}$ of 0.31 ± 0.19 and 0.07 ± 0.02 respectively; finally, simulated values of $Q_{inf,top}$ range from 105 ± 111 l s^{-1} to 21 ± 6 l s^{-1} (see Table 4.6). These trends reflect the sensitivity of simulated quantities to both riverbed and aquifer hydraulic properties. Indeed, the overall decrease in calibration-constrained riverbed permeability compared to prior parametrizations effectively limits infiltration of SW, and leads to lower estimates of $f_{SW,Br1}$ and $Q_{inf,top}$. Conversely, increased aquifer hydraulic conductivity in posterior parameter fields leads to higher subsurface flow velocities of infiltrated SW and consequently decreased GW age.

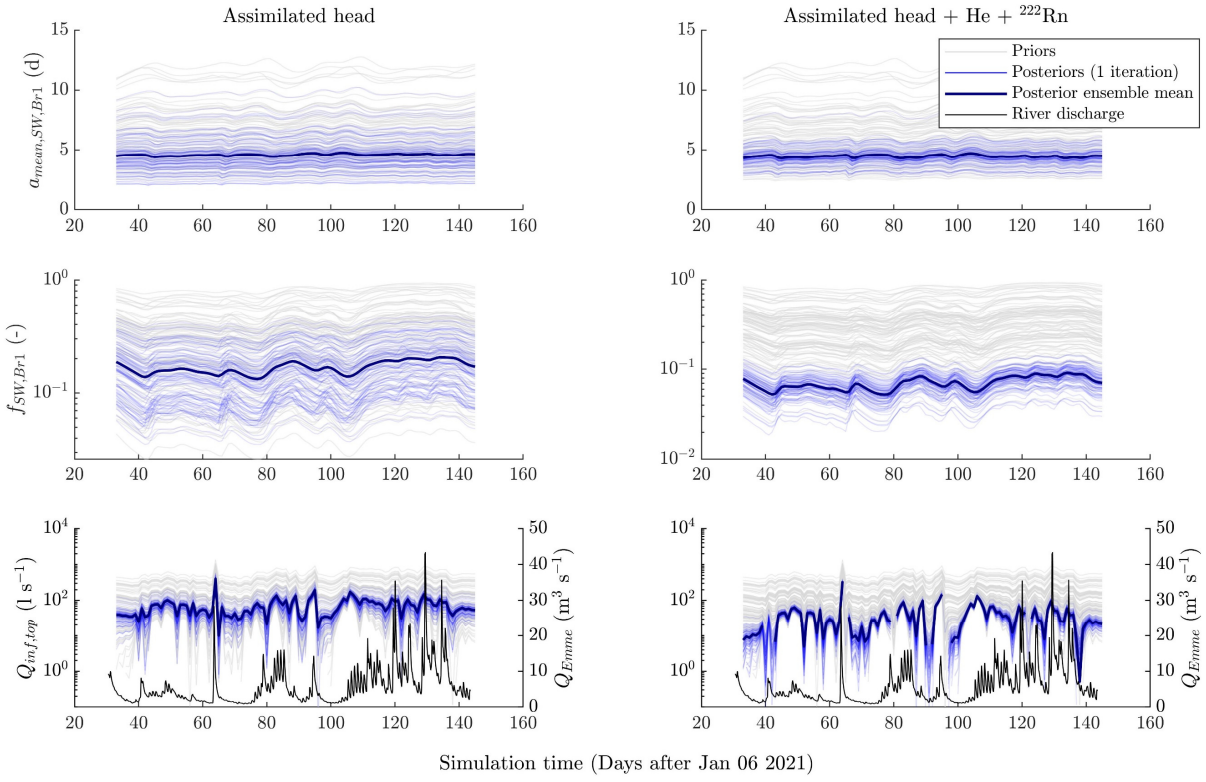

Figure 4.8 Predictions of mean GW age at pumping well Br1 (top), the fraction of recently infiltrated SW at Br1 (centre), and river infiltration flux on the upper reach of the Emme (bottom). Measured river discharge is also shown in the bottom graphs.

As stated in section 4.3.4, we can relate the prediction-specific data worth of different combinations of observation types to the relative decrease in predictive uncertainty (i.e., empirical ensemble variance) resulting from their inclusion in the calibration dataset. Figure 4.9 shows the relative average SD (i.e., square root of the average ensemble variance) of the three management quantities using different combinations of observation types for history matching. As noted above, these quantities are all constrained by head and tracer measurements, with the lowest predictive uncertainty achieved

with the joint assimilation of both observation types. Predictions related to SW infiltration, that is SW/GW mixing ratios at Br1 ($f_{SW,Br1}$) and infiltration flux ($Q_{inf,top}$), are already well-constrained by the assimilation of head measurements (resp. 60% and 80% decrease of average RMSE compared to priors); compared to calibration against hydraulic data alone, the inclusion of both types of tracer measurements leads to an *additional* decrease of average RMSE of 80% and 70% for $f_{SW,Br1}$ and $Q_{inf,top}$ respectively. Decreases in uncertainty related to mean GW age estimates ($a_{mean,SW,Br1}$) are more modest but nevertheless significant, with decreases of average RMSE of 30% by including head measurements, of 45% by including both head and He measurements, and of ~65% by including all hydraulic and tracer measurements.

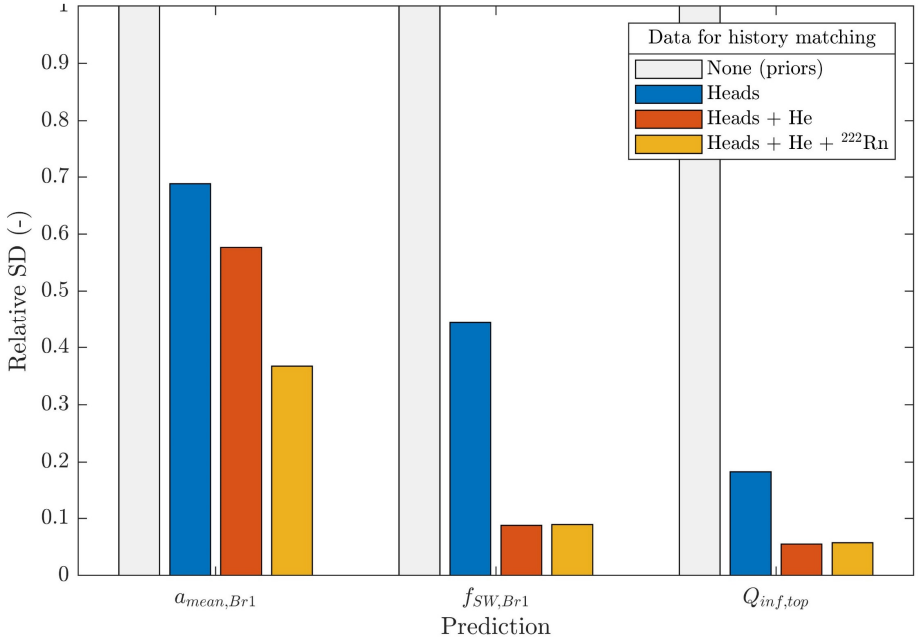

Figure 4.9 Relative average standard deviation (square root of the average ensemble variance) of the three selected predictions for the simulation period.

4.5. Discussion

The use of different observational datasets for history matching illustrates the contribution of tracer measurements towards informing model parameters and quantities of water management interest. The information content of tracers in informing key system characteristics related to SW-GW exchanges is made evident by differences in post-history-matched parameter fields and reductions in parameter and predictive uncertainty resulting from joint assimilation of head and tracer data compared to the assimilation of "classical" head observations alone. The assimilation of different combinations of observation types leads to comparable levels of performance in terms of reproducing hydraulic head measurements, whilst the assimilation of noble gas tracer data significantly improves the ability of posterior models to characterize SW-GW exchange and subsurface transport processes,

which is corroborated through the robust reproduction of independent dye tracer concentrations (Appendix D). The latter observation is highly relevant in that it substantiates the reliability of assimilating dissolved (noble) gas observations in a relatively simple modelling framework, without resorting to more complex multiphase flow and transport models.

4.5.1. Constraints on model parameters

Unsurprisingly, only tracer measurements significantly constrain aquifer mass-transport parameters (aquifer porosity and dispersivity), reflecting the insensitivity of hydraulic head measurements to aquifer dispersivity, and the strong correlation between *K* and porosity towards reproducing hydraulic measurements (Anderson et al., 2015; Schilling et al., 2017b). The constraints placed on riverbed permeability by tracer measurements echo the unique insights offered by the selected tracers to SW-GW exchanges. Indeed, subsurface distributions of ^{222}Rn and artificially introduced He strongly depends on the level of connectedness between the surface water body and the subsurface, which is essentially controlled by the riverbed.

In terms of subsurface parameter fields, all data assimilation schemes lead to the emergence of connected preferential flow structures in posterior parameter fields. These structures are most obviously delineated along the transect linking the river to the first pumping well, and further towards observation point P48 (see Figure 4.6), which is consistent with previously postulated locations of buried paleochannels at the field site (Arbenz and Peter, 1925; Gubelmann, 1930; Schilling et al., 2022). The degree to which tracer observations inform subsurface parameters (in terms of *K* fields and preferential flow paths) is not as obvious as for the riverbed, although moderate decreases in posterior variance of *K* and clearer delineation of high *K* structures result from the joint assimilation of head and tracer data. Parameter fields that appear in the northern part of the model are unsupported or constrained by field observations in this area; apparently connected structures emerge mostly from prior parameter covariances and spurious correlations between observations and parameters. These artefacts could be minimized or eliminated through the use of localization (Chen and Oliver, 2016), which was not employed in the present work. We expect the deleterious effects of not implementing localization to marginally impact the southern portion of the model, where posterior parameter fields are supported by a high-density measurement network.

Even though the framework for inverting subsurface parameter fields adopted in this study does favour the local emergence of preferential flow pathways as commanded by the hydraulic and tracer data, the connectivity of posterior parameter fields is not enforced during the parameter adjustment process. Posterior parameter fields will not exhibit the same large-scale connectivity patterns as in model priors, although a significant degree of spatial parameter correlation is retained. This will limit

the ability of posterior models to represent some important mass-transport processes, such as zones of highly focused flow and early tracer arrival times. Note that the degradation of large-scale structural continuity is an expected outcome of the iES parameter adjustment process. Indeed, the theoretical underpinnings of iES algorithms are based on the assumption of multigaussian parameter distributions (Chen and Oliver, 2013), which is not met by the categorical prior models considered here. As a result, the iES algorithm will (and should) not ensure the preservation of the large-scale connectivity of parameter fields whenever they are not described by multigaussian distributions. Moreover, the piecewise homogeneity of prior parameter fields combined with truncated SVD regularization effectively limits the expression of small- to medium-scale subsurface heterogeneity, which in turn limits the ability of the history-matched models to represent some relevant mass-transport processes, including tailing of tracer breakthrough and (very) early tracer arrival times (e.g. Dagan (1989)). Indeed, despite parameter adjustments operating on an element-by-element basis, the SVD-based inversion limits deviations from prior parameter values; whilst this prevents unreasonable overfitting of parameters to observations, it also largely conditions posterior parameter heterogeneity to that encountered in model priors. In summary, posterior parameter fields partially retain the large-scale spatial correlations from the model priors, at the cost of underrepresentation of small-scale heterogeneity. This methodology strikes a balance between informing subsurface structure and parameter heterogeneity whilst maintaining relative conceptual simplicity.

The coherence of calibration-constrained parameter fields and the overall performance of posterior mass-transport models are supported by the excellent reproduction of independent dye tracer observations by models assimilating tracer concentrations (Appendix D). Nevertheless, and as touched upon in section 4.4.1, the systematic underestimation of dye tracer concentrations and recovery rates suggest some degree of parameter compensation and model bias resulting from unaccounted loss of helium tracer during recharge and transport. Gas tracer exchange phenomena are expected to occur in unconfined systems (e.g. (Carter et al., 1959; Gupta et al., 1994)), especially with low-solubility gas species such as He. In the present study, this leads to negatively biased estimation of riverbed conductivity, and subsequently to underestimates of SW infiltration fluxes and solute movement from the river to the subsurface; this is discussed in the following section.

4.5.2. Reductions in predictive uncertainty

The reductions of parameter and subsequent predictive uncertainty accrued through history matching against field data are systematically highest when hydraulic and tracer data are jointly assimilated. This is an expected and intuitive outcome of the assimilation of diverse observation data, which contain rich and varied information related to environmental processes being modelled. In this study,

we have selected tracers that are uniquely sensitive to SW-GW interactions specifically to the infiltration of river water to the subsurface; it is therefore natural that parameters controlling SW infiltration – most notably riverbed permeability - are those best constrained by the assimilation of tracer observations. Associated predictions, such as SW-GW mixing ratios, SW infiltration fluxes, and GW age are therefore all strongly informed by tracer observations. Importantly, these quantities are some of the most relevant in water management contexts.

The complementary insights provided by He and ^{222}Rn as tracers of SW infiltration and GW age respectively are reflected in how they inform different predictions. Measurements of artificially injected He strongly constrain measures of SW infiltration, whereas little additional information is gained through the assimilation of ^{222}Rn. Conversely, and although the temporal dynamics of He measurements provide some constraints on mean GW age, this latter quantity is uniquely informed by the radioactive tracer ^{222}Rn. The lower relative decrease in predictive uncertainty for mean GW age compared to SW infiltration and mixing ratios results from the lower constraints placed on subsurface conductivity fields compared to riverbed permeability. Modelled GW residence times being comparatively small in the riverbed, mean GW age estimates will in large part reflect subsurface flow velocities, which depend on aquifer hydraulic parameters. ^{222}Rn, in fact only moderately constrains subsurface parameters. This observation can be related to (i) the low proportion of SW at observation point Br1 to which ^{222}Rn is sensitive, and (ii) the fact that the assimilation of hydraulic head measurements already captures ^{222}Rn observations within measurement uncertainty, i.e. the information content of ^{222}Rn is partially redundant with that already provided by head measurements. We would expect higher information content from ^{222}Rn at locations more strongly influenced by active SW infiltration, e.g. from observation points near the stream.

The reductions in model predictive uncertainty achieved through the assimilation of tracer data need to be confronted with the identification of model bias resulting from the incomplete conceptualization of He subsurface transport. The unaccounted loss of He mass, presumed to stem from gas volatilization towards the unsaturated zone, leads to parameter compensation through which modelled He concentrations in the subsurface are reduced to reproduce He measurements. This compensation manifests as exaggeratedly low estimates of riverbed permeability, which in turn will lead to negatively biased estimates of SW infiltration and associated predictions. We therefore expect biased first and second moments of predictions considered in this study, although their magnitude is likely modest, as testified by the excellent reproduction of the independent dye tracer observations. These results underscore the importance of independent tracer measurements to test and validate model performance, and to recognize otherwise unidentifiable bias accumulated from inaccurate process modelling, i.e., model structural error.

Notwithstanding, these results illustrate the great potential in explicitly simulating the concentrations of tracers sensitive to SW-GW exchange towards informing model parameters and predictions of management interest in bank filtration contexts. We have not addressed how the assimilation of tracer concentrations through their explicit simulation compares to assimilating transformed tracer quantities, such as tracer ages or apparent SW-GW mixing ratios. Indeed the tracers employed in this study could theoretically have been transformed into secondary measures of age (through He breakthrough or with the ^{222}Rn age model) or SW-GW mixing ratios (from the ratio of river and GW He concentrations), and used as calibration targets for parameter adjustment. While improvements in model performance are likely to arise through the assimilation of transformed tracer data, we note that, in general, producing model outputs against which these quantities would be compared (such as mean GW age or mixing ratios), requires the development of a mass-transport model or otherwise resorting to computationally expensive model post-processing (e.g. modified hydraulic mixing-cell (HMC) approaches (Partington et al., 2011)); the added cost of explicitly simulating tracer concentrations are therefore often marginal. Rather than deriving secondary quantities conceptually removed from the processes being modelled, the explicit simulation of tracer concentrations can directly inform model parameters while accounting for non-stationary flow conditions, complex tracer sources and sinks, etc... without the need of producing often questionable tracer-derived quantities.

4.5.3. Limitations and recommendations for future work

Several simplifying assumptions were implemented to ensure the tractability of the modelling and data assimilation framework presented in this study. This includes simplified model parameter fields, conceptualization of BCs and mass-transport processes, as well as relaxed constraints on the data assimilation algorithm.

The assumed homogeneity of riverbed conductivity and porosity limits the ability of models to faithfully represent as well as small- and medium-scale phenomena, such as focused GW infiltration and zones of preferential SW-GW exchange (Irvine et al., 2012; Tang et al., 2018). Inasmuch as the riverbed manifestly acts as a very low *K* interface between the river and the alluvial aquifer, we expect the spatially-averaged contrasts between the riverbed and aquifer to provide a reasonable characterization of large-scale SW-GW interaction patterns. Nevertheless, a compensatory role may be locally assumed by aquifer parameters adjacent to the riverbed, and emulate riverbed heterogeneity wherever commanded by the field data.

Another source of uncertainty and potential bias is related to the conceptualization of mass-transport processes, including piecewise constant solute concentrations along riverbed nodes, as well as conservative subsurface He transport. The former issue is mostly relevant for the intermediate river

reach (yellow in Figure 4.3), where river concentrations were not measured during the experiment. The assumed constant degassing ratio after the first set of weirs (measured after the main experiment) is therefore at best representative of mean He concentrations along the intermediate reach. The effect on subsequent model parameter fields is likely marginal, in that most observation points are chiefly influenced by river water infiltrating from the upstream reach. More generally the issue of atmospheric exchange and turbulent degassing needs to be addressed whenever gaseous tracers are used as indicators of SW infiltration, ideally through regular high-resolution measurement of dissolved gas concentrations in surface waters. Moreover, dissolved gas behaviour during recharge and subsurface transport cannot always be represented as that of a conservative solute. Indeed, indications of tracer loss through interaction with the unsaturated zone may limit the applicability of He as a tracer in variably-saturated media (Carter et al., 1959; Gupta et al., 1994). If extensive interaction with the unsaturated zone is expected, e.g. in unconfined systems with long GW residence times, He gas exchange with the unsaturated zone would likely need to be explicitly accounted for.

Finally, while the adopted parameter inversion framework does favour the emergence of connected high *K* structures through the imprint of spatial correlations expressed in prior parameter fields, the enforcement and preservation of continuous subsurface structures are not ensured. Indeed, the methodology presented in this study represents a balance between the lack of high-resolution subsurface characterization with a priori knowledge of the existence of connected preferential flow structures at the field site. Stricter enforcement of (pseudo)categorical *K* fields could be achieved through the use of data assimilation methods better-adapted for handling non-multigaussian parameter distributions, such as the normal-score ensemble smoother (NS-iES, e.g. Li et al. (2018)). More faithfully representing the hydraulic and structural properties of a heterogenous, multi-layered glaciofluvial aquifer system could in principle be achieved with more advanced geological and geostatistical modelling tools (Pirot et al., 2015; Renard and Allard, 2013); whether these approaches would be warranted with the available field data (or lack thereof) is beyond the scope of this work.

4.6. Conclusion and implications

ISSHMs are some of the most powerful tools currently available for the characterization and management of coupled SW-GW systems. The reliability of some of the most management-relevant quantities derived from these models depends on their ability to faithfully represent mass transport processes. In turn, constraining model parameters through which these processes are simulated requires the assimilation of field data that directly inform these parameters. In this study, we illustrate how the joint assimilation of high-resolution hydraulic head and noble gas (He and ^{222}Rn) observations, coupled with the explicit simulation of tracer concentrations in a highly-parameterized field-scale

ISSHM, can greatly constrain model parameters and reduce predictive uncertainty of several important water management quantities far beyond what is achieved with hydraulic data alone.

This modelling framework aims to directly bridge tracer concentrations with the model parameters (and subsequent predictions) they inform, without the addition of an intermediate level of conceptual simplification resulting from the calculation of possibly questionable tracer-derived estimates of GW age or mixing ratios. The improved model performance achieved through the assimilation of noble gas tracer data in this study provides further insights into how the information content of tracer measurements may be extracted through the explicit simulation of tracer concentrations, echoing results from a mounting number of studies documenting the benefits of assimilating diverse observation types in physically-based flow and transport models.

The significance of this study also lies in the illustration of how dissolved gas measurements can be integrated into an explicit modelling framework when these species are used as artificial tracers of surface water – groundwater interactions. Although we identify some caveats in assuming conservative transport with the noble gas helium, the results presented herein strongly support the continued development of artificial gas tracer methods for quantitative studies of SW-GW interactions. Noble gas tracer methods offer several advantages warranting their widespread use in environmental systems, as they can be applied continuously for long durations (weeks to months), are colourless, non-toxic, and can be applied with minimal regulatory constraints.

4.7. Acknowledgements

This study was funded by the Swiss National Science Foundation (grant number 200021_179017).

4.8. Appendices

4.8.1. Appendix A : Solute definition for mass-transport model

4.8.1.1. Simulation of fraction recently infiltrated surface water (SW)

To simulate the fraction of recently infiltrated SW f_{SW}, a theoretical conservative tracer ("*Mixing*") was assigned a constant concentration of unity at all inlet GW nodes, and a concentration of zero at all surface nodes. As any dilution of this tracer is due to the mixing of SW and GW components, f_{SW} is simply calculated as :

$$f_{SW}(\mathbf{x}, \boldsymbol{t}) = 1 - f_{GW}(\mathbf{x}, \boldsymbol{t}) = 1 - C_{Mixing}(\mathbf{x}, \boldsymbol{t}) \qquad (A1)$$

Where $f_{SW}(\mathbf{x}, \boldsymbol{t})$ [-] and $f_{GW}(\mathbf{x}, \boldsymbol{t})$ [-] are respectively the fractions of recently infiltrated surface water and regional GW, and $C_{Mixing}(\mathbf{x}, \boldsymbol{t})$ [-] the concentration of tracer "*Mixing*" at location $\mathbf{x}$ and time $\boldsymbol{t}$.

4.8.1.2. Simulation of mean groundwater age for both water components

Mean GW age was simulated as a conservative solute ("*Age*") using the age-mass concept developed by Goode (1996). All elements were assigned a constant source term equal to the product of porosity and water saturation, effectively simulating a theoretical tracer with a growth rate of unity (i.e., 1 d d^{-1}). All GW inlet and surface nodes were assigned an initial concentration of zero. This method only provides a weighted average of the mean ages of both GW and SW components at each node, without however discriminating their relative ages:

$$C_{Age}(\mathbf{x}, \boldsymbol{t}) = f_{SW} a_{mean,SW}(\mathbf{x}, \boldsymbol{t}) + f_{GW} a_{mean,GW}(\mathbf{x}, \boldsymbol{t}) \qquad (A2)$$

Where $C_{Age}(\mathbf{x}, \boldsymbol{t})$ [T] is the concentration of tracer "*Age*" corresponding to the mean age of GW at location $\mathbf{x}$ and time $\boldsymbol{t}$, and $a_{mean,SW}(\mathbf{x}, \boldsymbol{t})$ [T] and $a_{mean,GW}(\mathbf{x}, \boldsymbol{t})$ [T] are similarly the mean ages of the SW ang GW components in the water mixture.

As suggested by *Massoudieh and Ginn* [2011], mean GW age may also be estimated through the introduction of an unstable, non-chemically reactive solute ("*MG*") with constant concentration $C_{MG,0}$ [-] at the model inlet. Insofar as the residence times of this solute are low in comparison to its half-life, the decay equation may be linearized and, assuming no mixing of GW components, mean GW age is given by (*Massoudieh and Ginn* [2011], eq. 20):

$$a_{mean}(\mathbf{x}, \boldsymbol{t}) \approx \frac{1}{\lambda_{MG}} \left(1 - \frac{C_{MG}(\mathbf{x}, \boldsymbol{t})}{C_{MG,0}}\right) \qquad (A3)$$

Where $a_{mean}(\mathbf{x}, \boldsymbol{t})$ [T] is the mean GW age at location $\mathbf{x}$ and time $\boldsymbol{t}$, λ_{MG} [T^{-1}] is the solute first-order decay constant, and $C_{MG}(\mathbf{x}, \boldsymbol{t})$ is the concentration of tracer *"MG*".

For the present model, a constant concentration $C_{MG,0}$ of unity was applied to inlet GW nodes, and a concentration of zero to all surface nodes. By combining equations *(A2)* and *(A3)* and correcting $C_{MG}(\mathbf{x}, \boldsymbol{t})$ for mixing of SW and GW components, the mean age of recently infiltrated surface water can finally be computed from solute concentrations as :

$$a_{mean,SW}(\mathbf{x}, \boldsymbol{t}) \approx \frac{1}{f_{SW}(\mathbf{x}, \boldsymbol{t})}\left(C_{AGE}(\mathbf{x}, \boldsymbol{t}) - \frac{f_{GW}(\mathbf{x}, \boldsymbol{t})}{\lambda_{MG}}\left[1 - \frac{C_{MG}(\mathbf{x}, \boldsymbol{t})}{f_{GW}(\mathbf{x}, \boldsymbol{t})C_{MG,0}}\right]\right) \qquad (A4)$$

Where $a_{mean,SW}(\mathbf{x}, \boldsymbol{t})$ [T] is the mean age of GW at location $\mathbf{x}$ and time $\boldsymbol{t}$. The choice of an adequate value for λ_{MG} depends on the problem at hand; solute decay must be large enough to adequately represent GW ageing, but small enough for the hypothesis of linear decay to remain reasonable for the timescales modelled. For the present study, a value of 10^{-3} d^{-1} led to satisfactory results.

4.8.2. Appendix B : Estimation of missing He data

4.8.2.1. Gas diffusion through a semi-permeable membrane

Assuming constant membrane physical properties, the diffusive flux gas flux $I_{G_i,diff}$[L^3 T^{-1}] of a gas species G_i across a semi-permeable membrane can be written as [*Daynes*, 1920]:

$$I_{G_i,diff} = k_{G_i}\frac{S}{e}\delta P_{G_i} \qquad (A5)$$

Where k_{G_i} [M^{-1} L^2 T] is the areal permeability of the membrane to gas G_i, S [L^2] is the diffusive surface area (i.e. membrane area), e [L] is the membrane thickness, and δP_{G_i} [M L^{-1} T^{-2}] is the partial gas pressure differential of species G_i across the membrane.

Assuming the membrane is tubular with constant diameter and thickness, the diffusive surface is proportional to the tube length L:

$$I_{G_i,diff} = \frac{k_{G_i}\pi d}{e}L\delta P_{G_i} = \alpha_{Gi}L\delta P_{G_i} \qquad (A6)$$

With

$$\alpha_{Gi} = \frac{k_{G_i}\pi d}{e} \qquad (A7)$$

Where d [L] is the *diffusive* diameter, approximately equal to the inner diameter of the tube, and α_{Gi} = [M^{-1} L^2 T] the linear permeability of a semi-permeable tube (i.e., the diffusive gas flux per length of tube for a given partial gas pressure differential).

The permeability α_{Gi} of the silicone tubes used to inject helium gas into the river was estimated empirically during the first part of the injection experiment, when no gas leaks were assumed to be present. From the monitored decrease in helium tank pressures during the first week of the

experiment, a value of α_{Gi} of approximately 2.4 ± 0.2 x 10^{-3} ccSTP s^{-1} m^{-1} bar^{-1} was found for helium (α_{He}).

When He gas is into different lines of silicone tubing at different pressures, the total diffusive flux at any given moment is simply the weighted sum of fluxes from each line:

$$I_{He,diff} = \alpha_{He} \sum_i L_i \delta P_{He,i} \qquad (A8)$$

Where L_i [m] and $\delta P_{He,i}(t)$ [bar] are respectively the lengths and He injection pressures for each line i.

4.8.2.2. Estimating gas injection efficiency

The efficiency of the injection scheme, that is the proportion of diffusive flux that effectively dissolves into water, can be assessed from He concentrations over the course of the injection experiment. Assuming perfect mixing and no losses between injection and sampling points (e.g. through atmospheric exchange), the efficiency η [-] can be approximated by the ratio of the observed He flux in the river $I_{He,Emme}$ and the diffusive He flux through the semi-permeable tubing:

$$\eta = \frac{I_{He,Emme}}{I_{He,diff}} = \frac{Q_{Emme} C_{He,meas}}{\alpha_{He} \sum_i L_i \delta P_{He,i}} \qquad (A9)$$

Where Q_{Emme} [M T^{-1}] is the river discharge, and $C_{He,meas}$ [M^{-1} L^3] is the measured He concentrations in the river.

The gas injection rate was estimated from regular measurements (1 -3 days interval) of helium gauge pressures at the He tank outlets. The discharge in the Emme River is from an upstream river gauging system maintained by the federal office for the environment [*FOEN*, 2023]. The river discharge, He injection rate, measured He concentrations, as well as estimated injection efficiencies are shown in Figure 4.10.

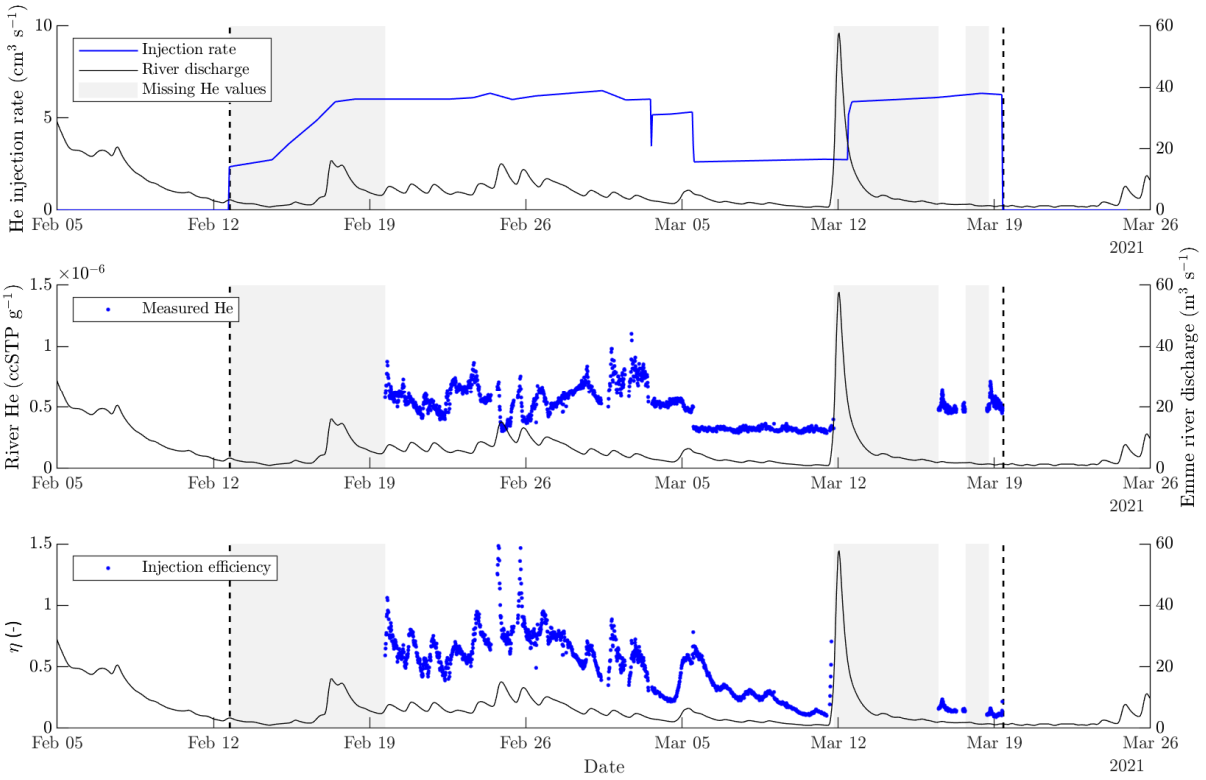

Figure 4.10 Results from the 2021 helium injection experiment. Diffusive injection of He began on Feb. 12, 2021, and lasted 35 days until Mar. 19, 2021. Top : He injection rates, estimated from equation (A8). Centre : Measured He concentrations at the river sampling location. Bottom : Diffusive injection efficiency η, estimated from equation (A9).

The injection efficiency varies as a function of time, and reacts strongly to changes in river discharge. Indeed, efficiencies are highest at higher discharges, and decrease significantly in low-flow conditions. This can be seen by plotting the injection efficiency η as a function of river discharge (Figure 4.11):

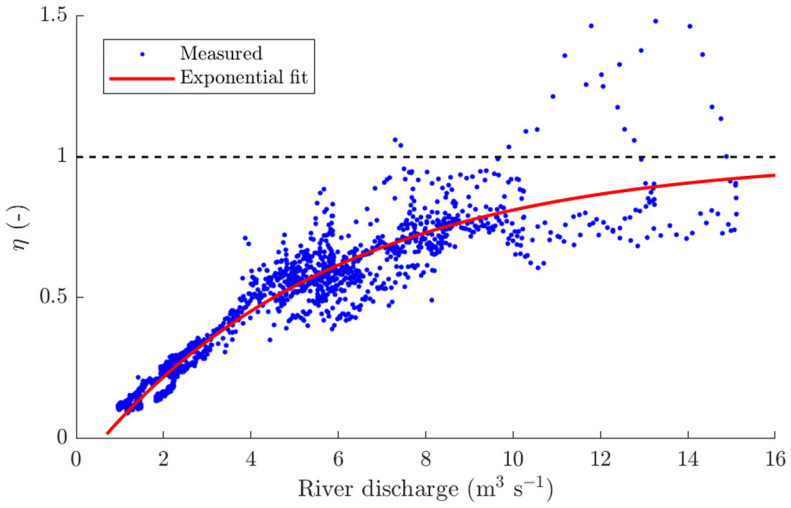

Figure 4.11 Helium injection efficiency η (equation [A9]) as a function of river discharge. The red line represents an adjusted exponential relationship of the form $\eta = 1 - a\ exp(-b * Discharge)$, with coefficients a = 1.114 and b = 0.1778 (see eq.11)

Above discharges of approximately 8 $m^3 s^{-1}$, the efficiency stabilises to values between 70% and 100%. Outliers above 100% efficiency are related to high-discharge events, during which abrupt changes in river discharge intermittently outpace the temporal resolution of He measurements.

The relationship between discharge and injection efficiency is likely related to the formation of gas bubbles on the exterior of membrane surfaces in low-energy environments. In periods of low discharge, nucleation and growth of He bubbles becomes more prevalent; these bubbles may then detach and escape to the atmosphere, reducing the proportion of injected gas dissolved in river water. Other factors such as varying degrees of lateral mixing of injected He may influence measured concentrations, although there is no obvious mechanism leading to the observed signal. Finally, several measurements of dissolved He further downstream of the main measurement location show that degassing of dissolved He between the injection and measurement locations is marginal, and not a likely factor in the smaller injection efficiencies at low discharge.

4.8.2.3. Estimating missing He values

Issues with field equipment (frozen water pump, damage from flood event), prevented river He measurements for the first week of the injection experiment (Feb. 12 – Feb. 19), then from Mar. 11 – Mar. 16, and from Mar. 17 – Mar. 18 (Figure 4.10). As measured river He values are used as a specified concentration boundary condition for the model, values need to be estimated for periods where measurements are missing.

Equation (A9) can be rewritten to relate river He concentrations C_{He} to the diffusive He injection rate $I_{He,diff}$, the river discharge Q_{Emme}, and the injection efficiency η:

$$C_{He} = \eta \frac{I_{He,diff}}{Q_{Emme}} \qquad (A10)$$

Estimates of η, and therefore of dissolved He concentrations, were calculated for periods without measurements from the observed relationship between η and Q_{Emme}, which is well approximated by an equation of the form:

$$\eta = 1 - a\, exp(-bQ_{Emme}) \qquad (A11)$$

Equation (A11) with coefficients a = 1.114 and b = 0.1778 provides a reasonable fit to observed data ($R^2 \approx 0.87$, see Figure 4.11). Dissolved He concentrations estimated with equation *(A11)* are shown in Figure 4.12.

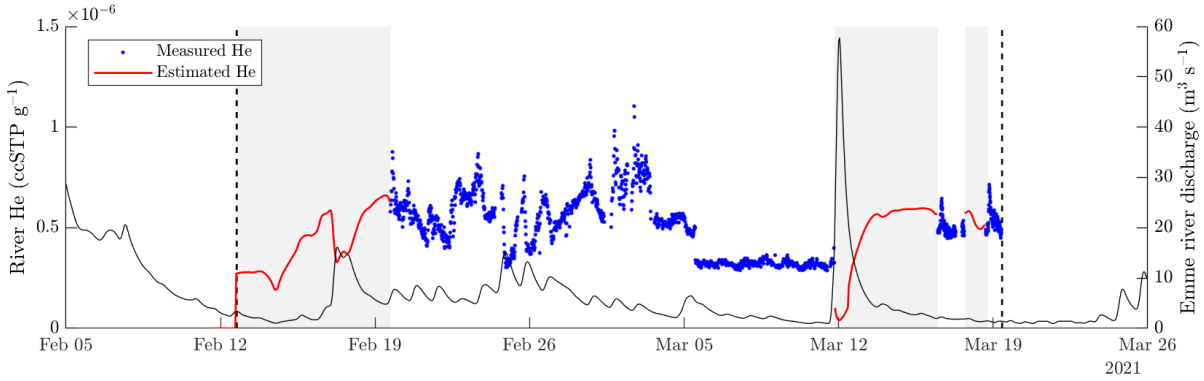

Figure 4.12 Estimated (red, equations (A10) and (A11)) and measured (blue) dissolved He concentrations in the Emme River.

4.8.3. Appendix C: Construction of prior geological models

4.8.3.1. General geology of the Aeschau sector

The general (hydro)geology of the Aeschau sector has been thoroughly described in previous works, to which readers are referred for a more general description (e.g. *Käser and Hunkeler* [2016]; *Peel et al.* [2022]; *Popp et al.* [2021]; *O. S. Schilling et al.* [2022]; *Oliver S. Schilling et al.* [2017]; *Tang et al.* [2018]). Here we examine only aspects most relevant to the creation of prior hydrogeological models fed to the iterative ensemble smoother (iES) algorithm.

The bulk of the Upper Emmental alluvial aquifer is formed by glaciofluvial gravel-sand deposits, with several clay/silt intercalations [*Blau and Muchenberger*, 1997; *Würsten*, 1991]. The average thickness of these glaciofluvial deposits is on the order of 20 m, and up to 40 m in the centre of the valley. These unconsolidated sediments were deposited in braided river systems as glacial outwash during the late Würm glaciation [*Würsten*, 1991]. These deposits are overlain by a fluvial cover between 0.5 and 5 m thick composed of clayey loam, silt overflow sediments, and sandy gravel deposits [*Würsten*, 1991]. The lower and lateral limits of the aquifer are bound by either low-permeability basal till deposits, or by sandstones of the Upper Marine Molasse (OMM).

4.8.3.2. Representing geometry and hydraulic properties of deposits at the Aeschau site

While many sedimentological units typical of fluvioglacial aquifers may exhibit favourable hydraulic properties (e.g. high porosity and conductivity), gravel and sandy gravel trough fill deposits (i.e. filling of scours) are those most likely to act as pathways for preferential flow and transport [*Huber and Huggenberger*, 2015]. The lateral and longitudinal extent of preserved trough fill elements is usually difficult (or impossible) to ascertain in most field contexts; however they may form continuous structures over significant distances (up to several hundred meters) in the overall direction of (paleo)river flow, especially in river systems characterized by highly dynamic discharge conditions,

where the extent of scouring during high-discharge events, and subsequent filling during the waning of flow, is large [*Huber and Huggenberger*, 2015; *Huggenberger and Regli*, 2006].

As stated in the preceding section, the Aeschau aquifer can be divided into two main sections, based on the depositional setting in which they formed. The inferior, glaciofluvial body forms the main aquifer in the sector, whereas the importance of the overlying fluvial cover is mostly related to its capacity to transmit river water towards the main aquifer [*Blau and Muchenberger*, 1997].

As is the case for most similar aquifers, the geometry of sedimentological units within the Aeschau sector is poorly known and constrained. Several studies [*Gubelmann*, 1930; *O. S. Schilling et al.*, 2022; *Würsten*, 1991] as well as tracer tests carried out in the area [*Popp et al.*, 2021] strongly suggest the existence of preferential flow pathways in the subsurface. However, clearly describing these structures in terms of geometry and hydraulic properties remains elusive. Therefore any (hydro)geological model of the Aeschau area has to balance the need to represent preferential flow structures with the lack of data allowing characterization of three-dimensional subsurface heterogeneity.

Several well logs from boreholes in the Aeschau area provide some information on the abundance of different sediment types in the subsurface. Two boreholes in the sector [BF04 and BF05 [*Würsten*, 1991]] additionally provide detailed vertical profiles of hydraulic conductivity (K) and of vertical anisotropy of K. Naturally this data only minimally constrains the identification, extent, connectivity, and hydraulic properties of depositional units. It nevertheless provides a basis upon which a simplified conceptual model of the abundance and hydraulic properties of preferential flow structures can be constructed.

As stated in the previous section, the presence of gravel and sandy-gravel trough-fill deposits are most likely to act as well-connected flow pathways in the subsurface. Although they cannot be unambiguously identified as such on well-logs they are usually composed of well-sorted gravels and sandy gravels, and typically exhibit very high K [*Huber and Huggenberger*, 2015]. Other major depositional units expected in glaciofluvial aquifers, such as gravel sheets and channel-fill sediments, tend to be more poorly sorted, exhibit higher proportions of fine-grained sediment (sand and silt), and overall lower K values [*Bayer et al.*, 2011; *Huber and Huggenberger*, 2015; *Huggenberger and Regli*, 2006].

Based on the above considerations, a conceptual model consisting of two principal hydrofacies which represent (i) connected and highly-conductive trough fill gravels, and (ii) less-conductive gravel sheet and channel fill sediments, may capture the most salient features of the flow system. Relative

proportions of these two hydrofacies can be approximated by interpreting well logs from the study area and assigning one of two facies based on sediment description. Only clean and sandy gravels were attributed to trough fill deposits, whereas other sediment types such as poorly sorted sandy gravels, sands, and silts were attributed to the second facies. In total well logs from 13 boreholes (aggregated length > 320 m) were analysed, from which summary statistics on facies proportion could be computed (see Table 4.7). Finally, results from mini-pumping tests with flowmeter measurements conducted in boreholes BF04 and BF05 [*Würsten*, 1991] provide some constraints on the hydraulic properties (*K* and vertical anisotropy of *K*) of the two defined hydrofacies.

Table 4.7 Estimates of relative facies proportions from borehole logs from the Aeschau sector. CG : clean gravel/sandy gravel, attributed to the trough fill hydrofacies. The "TOTAL" row provides weighted mean / standard deviation values. FG interval : interval of fluvioglacial deposits identified in well logs.

Well ID	**FG interval** (m)	**Facies** %		***K* (± 1-σ)** (m/d)		**Anisotropy of *K*** mean [range] (-)	
		CG	*Other*	*CG*	*Other*	*CG*	*Other*
P52	10	30	70	-	-	-	-
P53	12	8	92	-	-	-	-
P54	8	35	65	-	-	-	-
RB2/P	23	65	35	-	-	-	-
RB1	38	52	48	-	-	-	-
RB2	47	27	73	-	-	-	-
RB3	27	80	20	-	-	-	-
B1	30	20	80	-	-	-	-
B2	15	45	55	-	-	-	-
B3	23	22	78	-	-	-	-
B4	28	21	79	-	-	-	-
BF04	20	52	48	900 ± 270	170 ± 120	10 [7 - 35]	
BF05	39	23	77	680 ± 270	160 ± 120		
WP 789B	13	38	62	-	-	-	-
TOTAL	321	37 ± 40	63 ± 25	750 ± 270	165 ± 120	10 [7 - 35]	

4.8.3.3. Event-based geological models

The stochastic event-based modelling software ALLUVSIM [*Pyrcz et al.*, 2009] was used to create several realizations of subsurface structure based on the geological description given in the previous sections. This software allows the construction of realistic pseudo-process-based models of alluvial deposits, and allows flexibility in defining the geometry of various architectural elements (e.g. channel fills, flood plains, crevasse splays, etc…).

In the case of the Aeschau model, only "channel fills", meant to represent connected gravel and sandy gravel trough fill deposits, were modelled with ALLUVSIM. Indeed, for this work, all other sedimentological units are amalgamated into a single facies, and not considered individually. The

longitudinal (i.e. down-valley) extent of each clean gravel body is not known a priori, although several previous field and modelling investigations (e.g. *Popp et al.* [2021]; *O. S. Schilling et al.* [2022]) highlight the likely existence of large-scale preferential flow structures (at least several hundred meters) in the aquifer under consideration. Nevertheless, the paucity of volumetric and geometrical constraints on the distribution of such flow paths in the study area encourages building prior geological models which sample a large range of preferential flow structures. This entails modelling various structure geometries (width, depth, sinuosity, connectedness) and their overall volumetric fraction in the modelled aquifer.

Preferential flow structures were represented in ALLUVSIM as single-layered channel systems with varying geometric properties. Available field data does not allow the inversion and delineation of complex 3-D subsurface structures, and only structures in the near subsurface (depth <15 m) were represented in the HGS model, as these are expected to most impact SW/GW interactions and transport of recently infiltrated SW towards observation and pumping wells. Model parameters controlling channel properties (i.e. channel depth, width, sinuosity, volumetric fraction) were assigned typical values [see e.g. *Huber and Huggenberger* [2015]; *Huggenberger and Regli* [2006]], albeit with a high level of variance (see Table 4.8). A 15-m thick channel system level was stacked on top of a homogenous 55-m matrix level. Therefore channel systems were only represented in the uppermost 15 m of the system.

Table 4.8 Parameters used for ALLUVSIM simulations. See Pyrcz et al. [2009] for the implementation of these parameters in ALLUVSIM.

Parameter	Value	Description
d_T (m)	5 ± 2	Mean channel depth (± 1-σ)
w_T (m)	20 ± 5	Mean channel width (± 1-σ)
φ_T (-)	1.2 ± 0.2	Mean channel sinuosity (± 1-σ)
NTG (-)	0.35 ± 0.35	Net-to-gross ratio (relative abundance of trough fill / other hydrofacies) in alluvial system level
n_{level} (-)	2	Number of stacked levels

The ALLUVSIM grid was aligned with the HydroGeoSphere (HGS) model domain, and encompassed the entire 3-D extent of the hydrogeological model (Figure 4.13). A structured orthogonal grid with dimensions 1001 x 2499 x 70 m was constructed for ALLUVSIM simulations, with element sizes of 7 x 7 x 1 m in the X, Y, and Z directions, respectively (> 3.5×10^6 elements total).

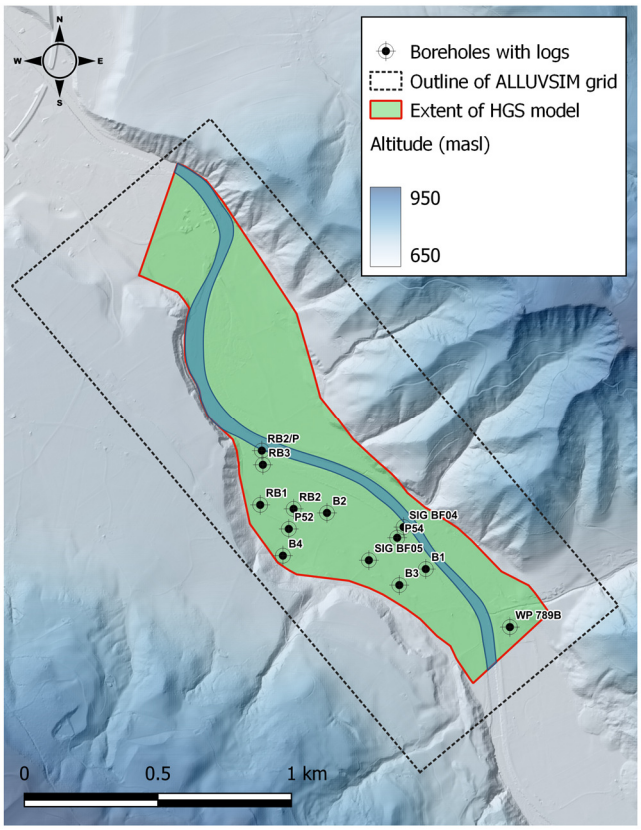

Figure 4.13 Lateral extent of both ALLUVSIM and HydroGeoSphere (HGS) model domains. Locations of boreholes from which geological logs are available are also shown.

In total, 100 (prior) 3-dimensional facies models were generated with ALLUVSIM. As ALLUVSIM and HGS grids were aligned, facies were then assigned to each element in the hydrogeological model on a nearest centroid basis (i.e., nearest neighbour). The irregular geometry of the HGS grid leads to many channelized structures being discontinuous (see Figure 4.14). This result is desirable, in that it limits the systematic emergence of overly large preferential flow structures.

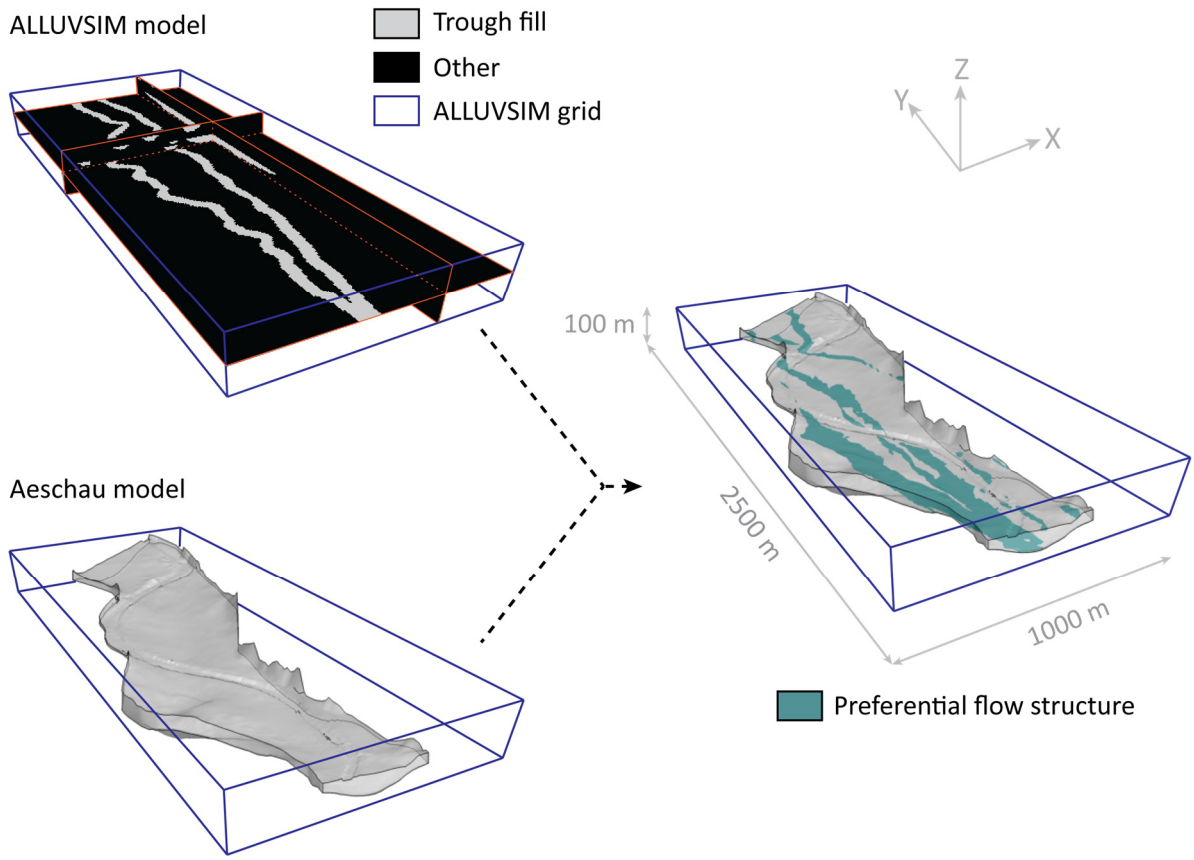

Figure 4.14 Overview of facies assignment from ALLUVSIM facies outputs to the HGS grid. Trough fill facies is assigned to the HGS grid on an elemental basis using nearest neighbour interpolation. One realization of channelized structures is shown for illustrative purposes.

4.8.4. Appendix D : Validation with fluorescein measurements

4.8.4.1. Fluorescein injections and measurements

For each fluorescein injection experiment, approximately 20 litres of concentrated fluorescein solution (~40 g/l) was injected into the Emme river upstream of the He injection location (Figure 4.1) using a Mariotte's bottle at an average rate of ~ 0.5 ml/s – 1 ml/s. Constant injection rates could not be ensured, as there were several issues with the freezing of the solution and clogging of the bottle outlet. Injections lasted on average approximately 6 – 12 hours. Fluorescein concentrations were monitored at 1-minute intervals with a portable field fluorometer (GGUN-FL30, Schnegg [2002]) at the same river monitoring location where He measurements were taken. Fluorescein was also continuously measured in pumping well Br1 at 1-minute intervals for the duration of the experiment. The GGUN-FL30 fluorometer has a typical detection limit of 0.01 – 0.02 ppb, and a resolution of 0.01 ppb for fluorescein [Schnegg, 2002]. Results are shown in Figure 4.15.

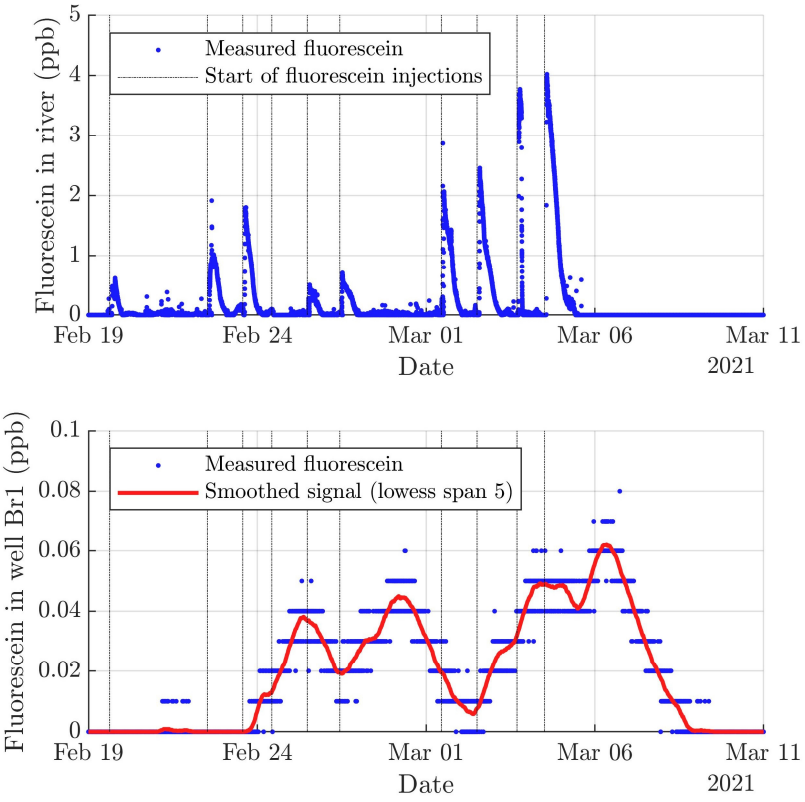

Figure 4.15 Measured fluorescein concentrations in the Emme River (top) and pumping well Br1 (bottom). Vertical dashed lines represent the start of each injection (duration 6-12 hours on average).

First arrival of fluorescein in the pumping well occurred within 1.5 days of injection in the river, with peak breakthrough occurring after 2-3 days (Figure 4.15). Measured concentrations in the well were only slightly above the instrument detection limit, even though clearly distinguishable from background.

4.8.4.2. Simulation results

The consistency of posterior model performance is verified by comparing history-matched models against the fluorescein tracer dataset. Measured and simulated (post-history-matched) fluorescein concentrations in pumping well Br1 are shown in Figure 4.16. Models assimilating head and tracer (He + ^{222}Rn) data are capable of reliably reproducing measured tracer concentrations and temporal trends, albeit with a moderate systematic negative bias (average ensemble RMSE ≈ 0.006 ppb; ME ≈ -0.001 ppb, R^2 ≈ 0.76). Conversely, models assimilating only head measurements exhibit the same trends presented in section 4.4.1, in that they tend to overestimate tracer concentrations (RMSE ≈ 0.012 ppb; ME ≈ +0.005 ppb, R^2 ≈ 0.14).

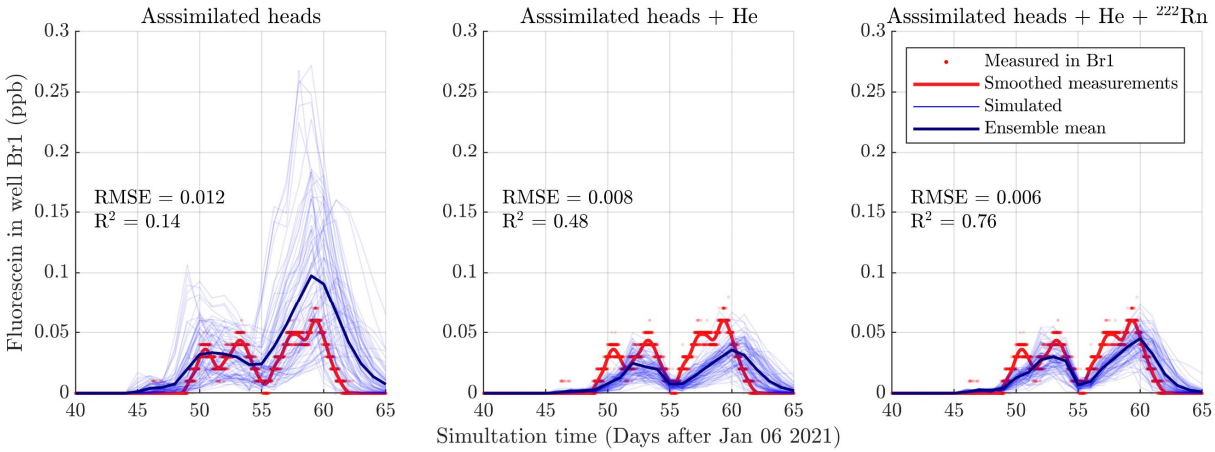

Figure 4.16 Measured and modelled fluorescein concentrations in pumping well Br1. Left figure : models after assimilation of head measurements (one iteration); right; models assimilating head and tracer (He + ^{222}Rn measurements).

The same trends are evident by comparing tracer recovery rates in pumping well Br1; indeed only models assimilating tracer data are capable of adequately reproducing observed quantities, whereas those assimilating only hydraulic measurements tend to significantly overestimate tracer recovery (see Table 4.9). In all cases, the ratio of modelled He and fluorescein recovery rates in Br1 is systematically higher than that measured (average modelled ratio ≈ 0.85 vs measured ≈ 0.70, Table 4.9). Although the estimate of measured fluorescein recovery in Br1 is uncertain due to measured concentrations being near the instrument detection and resolution limits, the systematically higher modelled He recovery suggests some degree of model structural error. This most likely results from the possible misrepresentation of modelled river boundary conditions (e.g., underestimated He concentrations along the intermediate reach of the river, see section 4.5.3), or He loss during subsurface transport in field conditions, which is not considered in the model.

Table 4.9 Measured and simulated tracer recovery rates (recovered fraction of tracer mass injected in the river) in pumping well Br1. The uncertainties in measured recovery rates result from the propagation of uncertainty in measured tracer concentrations (in SW and GW), river discharge, and GW pumping rates. Uncertainty in modelled recoveries represents the ensemble simulated standard deviation.

Recovery in Br1	**Measured**	**Prior**	**Heads**	**Heads + He**	**Heads + He + ^{222}Rn**
			Mean ± 1σ		
Fluorescein	$13 \pm 4 \cdot 10^{-4}$	$48 \pm 31 \cdot 10^{-4}$	$25 \pm 16 \cdot 10^{-4}$	$8.6 \pm 4 \cdot 10^{-4}$	$10 \pm 3 \cdot 10^{-4}$
He	$9 \pm 2 \cdot 10^{-4}$	$41 \pm 27 \cdot 10^{-4}$	$21 \pm 14 \cdot 10^{-4}$	$7.3 \pm 3 \cdot 10^{-4}$	$8.8 \pm 2 \cdot 10^{-4}$

4.9. Supplementary material to Chapter 4

4.9.1. Summary of observations for history matching

Table 4.10 Number of observations included in the calibration dataset for different observation types.

Location	Heads	He	^{222}Rn
A32	-	8	-
A35	231	-	-
A37	-	7	-
P16	231	-	-
P48	-	8	-
P51	231	-	-
P52	231	8	-
P53	231	-	-
P54	231	8	-
Br1	-	177	22
Total	***1386***	***216***	***22***

Simulated heads and tracer concentrations (figures)

4.9.1.1. Simulated and observed hydraulic heads

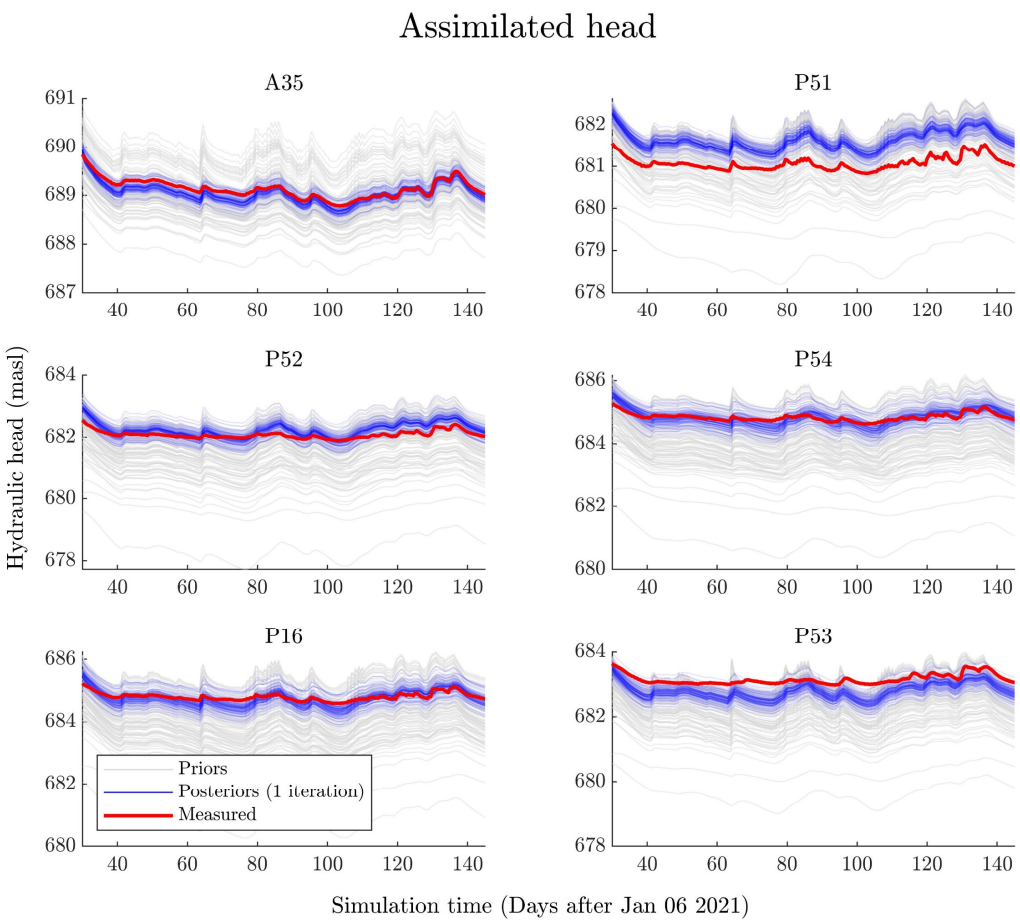

Figure 4.17 Simulated and observed hydraulic heads in the six observation wells used for calibration. Only hydraulic head data was assimilated with the model.

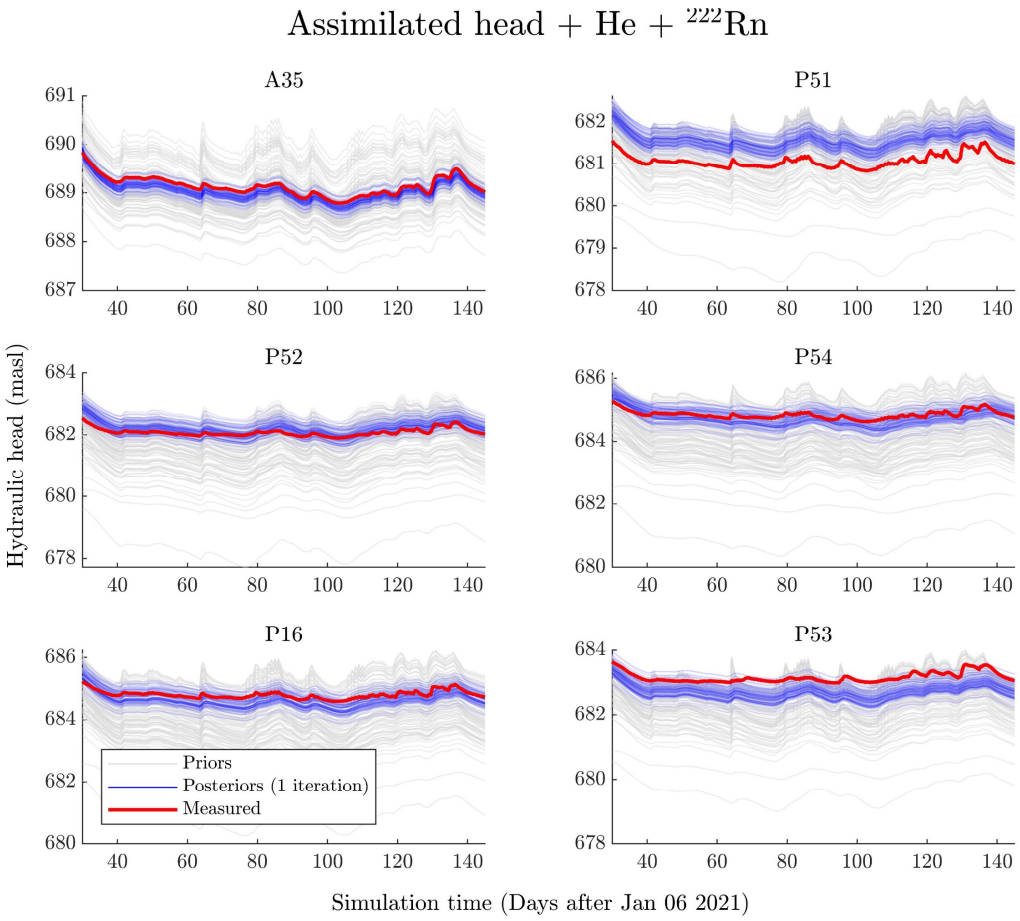

Figure 4.18 Simulated and observed hydraulic heads in the six observation wells used for calibration, with assimilation of hydraulic head, helium, and ^{222}Rn data.

4.9.1.2. *Simulated and observed helium (He)*

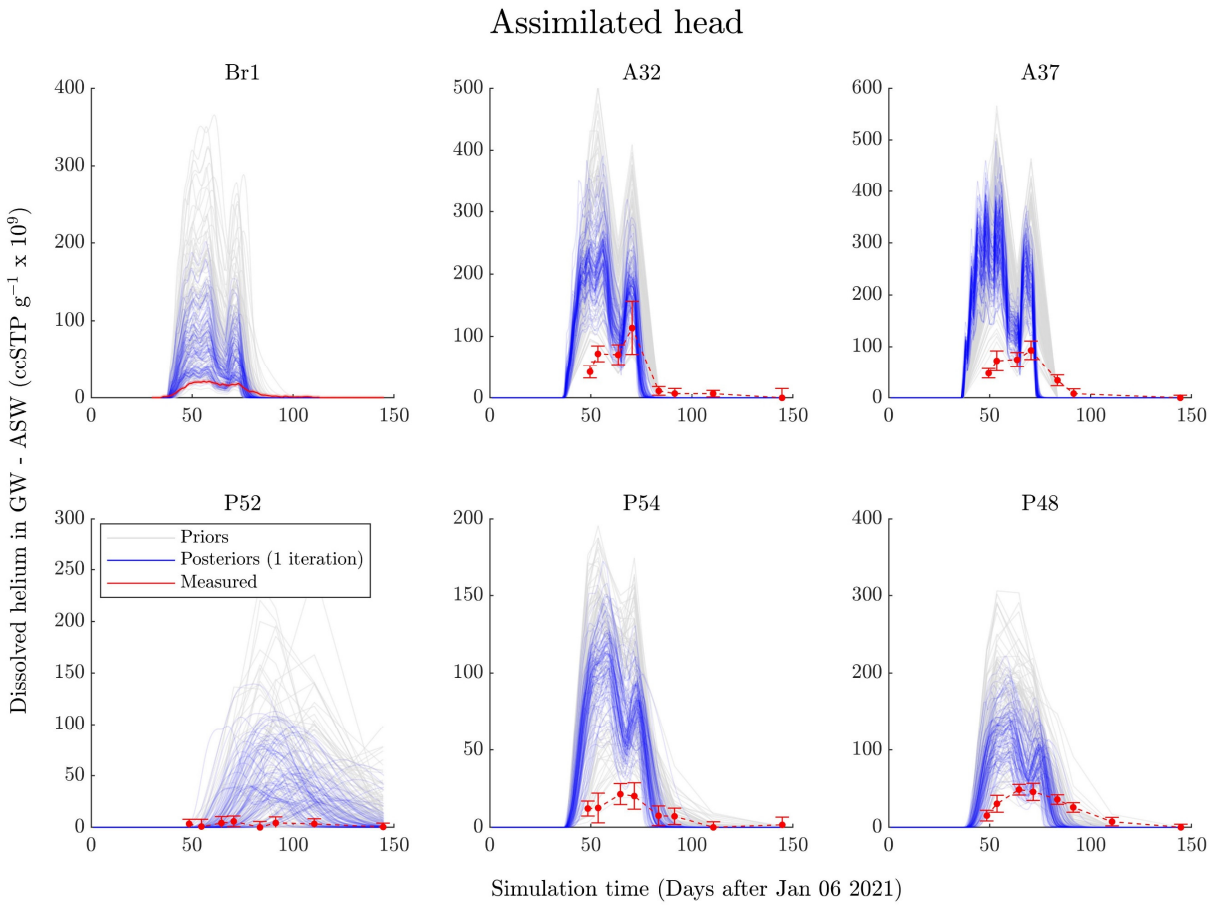

Figure 4.19 Simulated and observed helium (He) in the pumping well and five observation wells, with the assimilation of hydraulic head data.

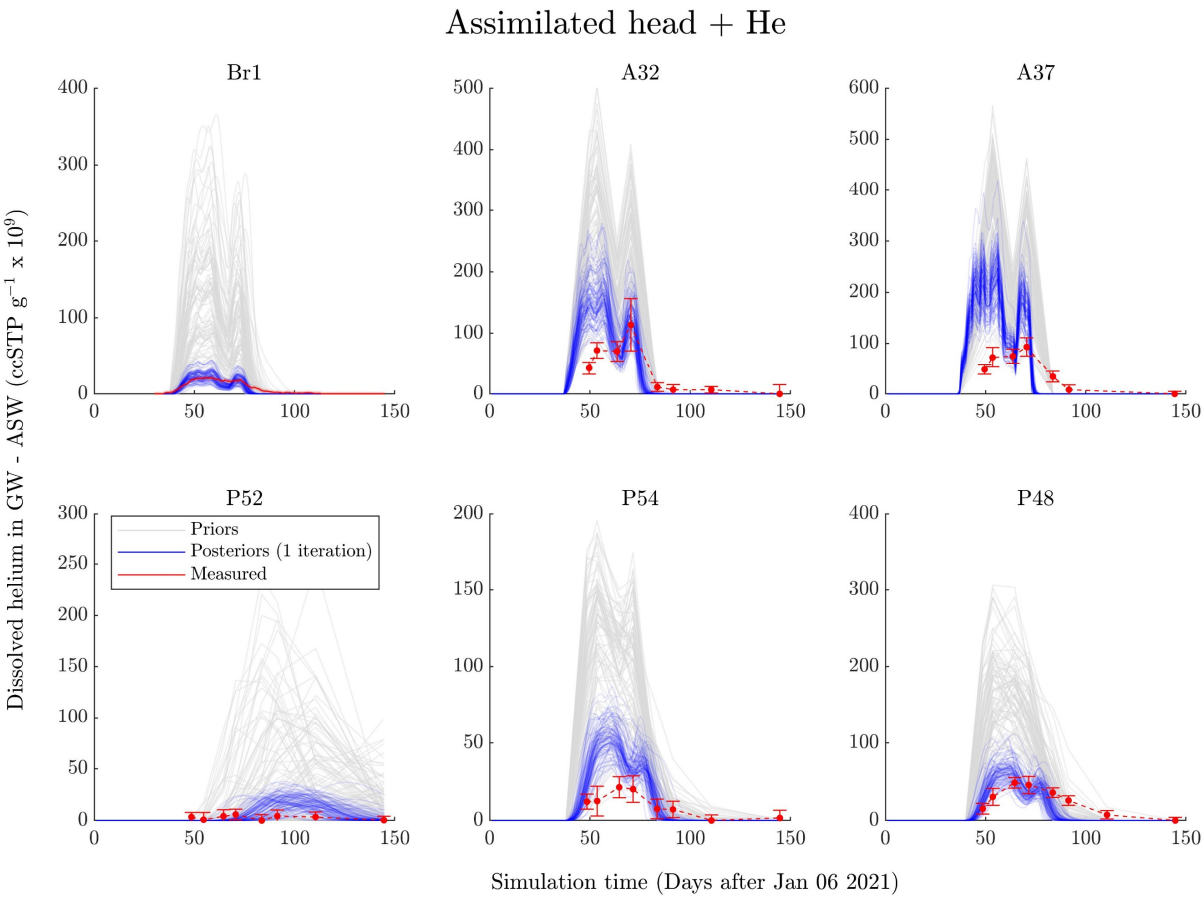

Figure 4.20 Simulated and observed helium (He) in the pumping well and five observation wells, with the assimilation of hydraulic head, helium, and ^{222}Rn data.

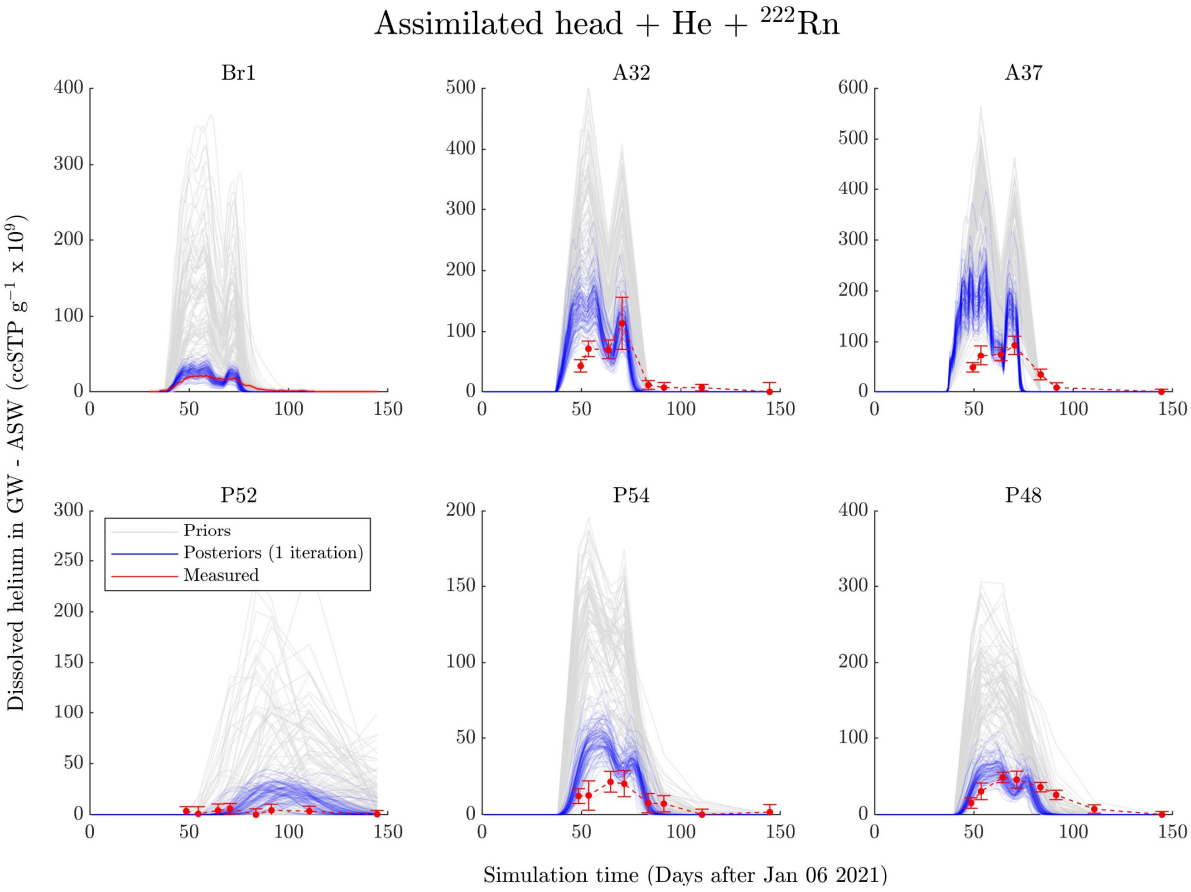

Figure 4.21 Simulated and observed helium (He) in the pumping well and five observation wells, with the assimilation of hydraulic head, helium, and ^{222}Rn data.

4.9.1.3. *Simulated and observed ^{222}Rn*

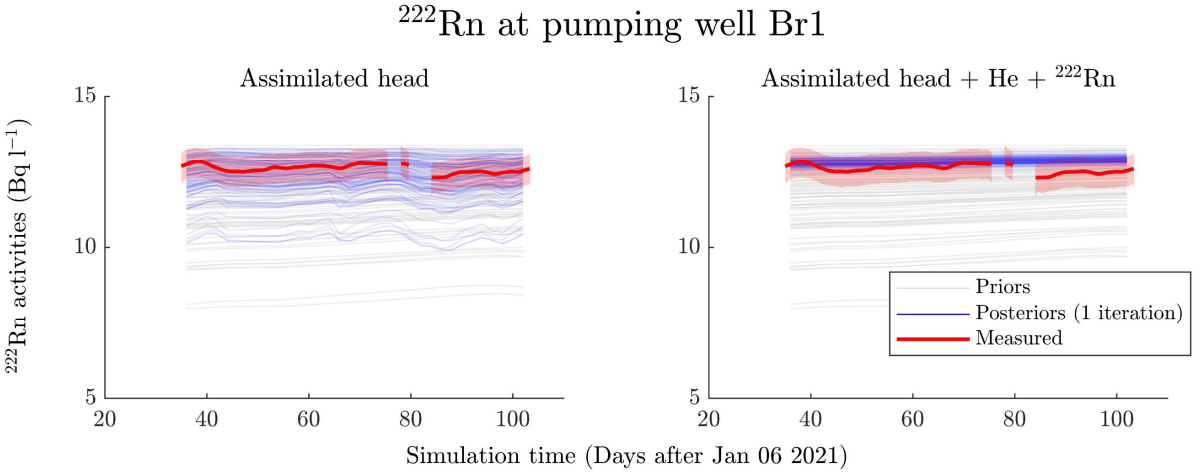

Figure 4.22 Simulated and observed ^{222}Rn activities in pumping well Br1. Results from different calibration datasets are shown, as indicated in the figure.

4.9.2. Prior and post-history-matched model parameter fields

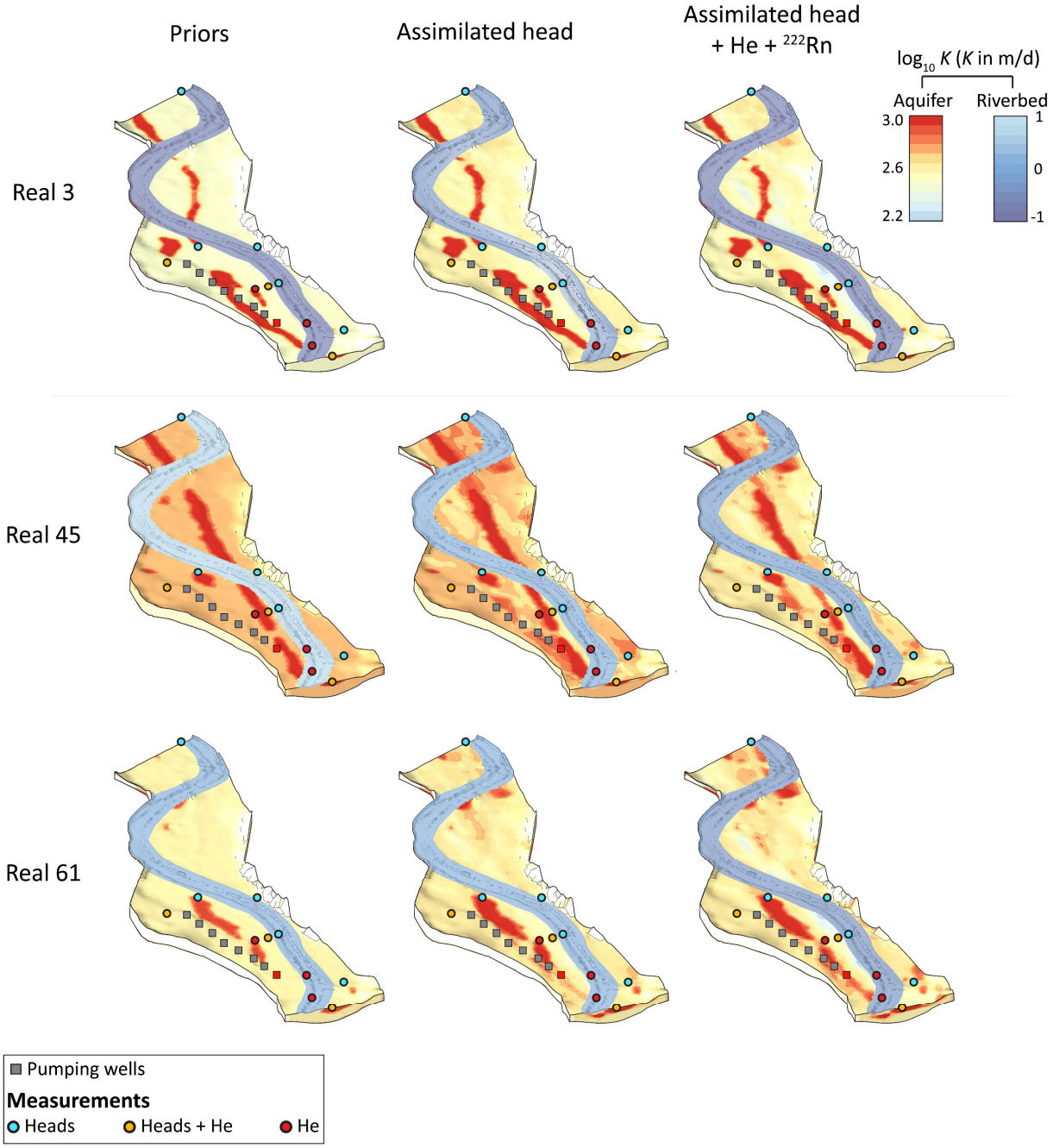

Figure 4.23 Prior and posterior parametrizations for three randomly selected realizations (realizations 3, 45, and 61) from the 100-member parameter ensemble.

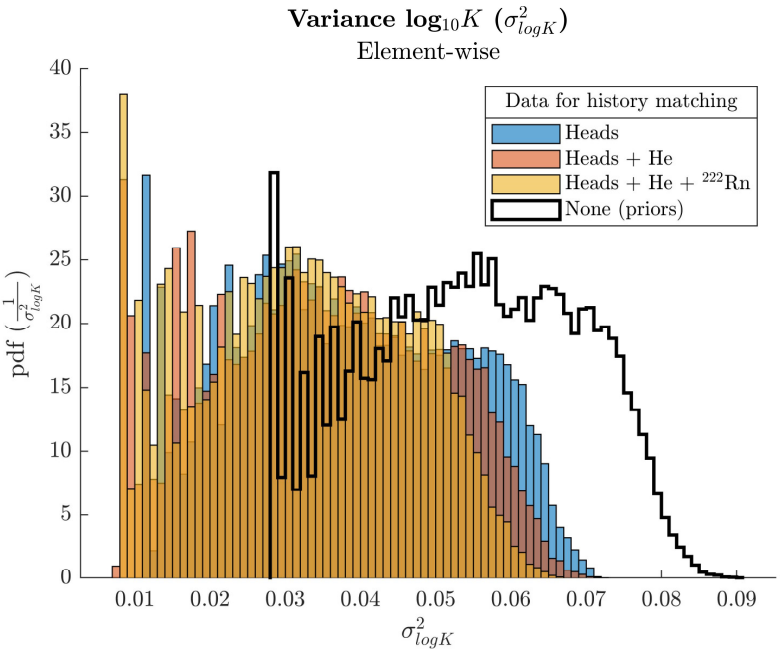

Figure 4.24 Prior and post-history-matched distribution of variance of aquifer $log_{10}K$ (calculated for each model cell)

5. Conclusions

5.1. Summary

The work presented in this thesis aimed at investigating how existing and newly developed (noble) gas tracer methods could be optimally used for the study of SW-GW interactions in alluvial aquifer systems. By addressing these questions through a range of novel laboratory, field, and numerical modelling approaches, the applicability of various noble gas tracer methods in studies on SW-GW interactions has been challenged, expanded, and when necessary reframed. Although the first two works presented in this thesis emphasize the (sometimes severe) limitations of the ^{222}Rn apparent age model, results from subsequent studies highlight the great potential of noble gases to provide reliable quantitative insights into complex alluvial systems, and strongly support further development of (noble) gas tracer methods as artificial tracers of SW – GW exchange.

This section aims to summarize some of the key findings resulting from the work presented in this thesis. The following section lists some important open questions which should be addressed in future research, as well as recommendations for further methodological developments of gas tracer methods.

5.1.1. Study 1 (Chapter 2)

A robust experimental design allows quick and accurate analysis of ^{222}Rn emanation rates of bulk sediment samples, contributing to the high-resolution characterization of spatially variable ^{222}Rn emanation rates in an alluvial aquifer; the relationship between ^{222}Rn emanation rates and depth can lead to counterintuitive ^{222}Rn signals in transient hydraulic conditions.

A simple, robust, and analytically sensitive experimental setup was developed to measure ^{222}Rn emanation rates from large (> 5kg), minimally disturbed sediment samples. This setup was employed to carry out a high-resolution characterization of the spatial distribution of ^{222}Rn emanation rates at a riverbank filtration site (Aeschau site, Switzerland). Significant heterogeneity in the spatial distribution of ^{222}Rn emanation rates could be identified, with direct consequences for the use of the ^{222}Rn age method. Higher emanation rates identified in the near subsurface can be linked to counterintuitive ^{222}Rn signals in groundwater, which cannot be interpreted without proper knowledge of the spatial and (in this case depth-dependant) distribution of ^{222}Rn emanation rates. In such cases, simple conceptual models based on defining a single ^{222}Rn equilibrium end-member for estimates of GW age, SW infiltration, or GW discharge to stream may become inadequate.

5.1.2. Study 2 (Chapter 3)

Spatially-variable ^{222}Rn production rates and GW mixing may lead to significant bias in tracer-derived estimates of GW age; temporal variations in ^{222}Rn activities in GW may however serve as a powerful tracer of changes in mean GW age.

The explicit simulation of ^{222}Rn activities accounting for spatially variable ^{222}Rn production, transport, production and decay in a synthetic 2-D river-aquifer system was carried out. It is shown that variable ^{222}Rn production rates, coupled with hydrodispersive mixing of GW can lead to highly biased estimates of GW age. Only in very restrictive circumstances (mass transport is advection dominated, sampling scale larger than scale of variability of ^{222}Rn production rates, no GW mixing at sampling points), can the ^{222}Rn apparent age model provide reliable estimates of GW age. As these circumstances can seldom be justified in field contexts, computation of ^{222}Rn apparent ages must be done with caution, or avoided altogether. Time series of ^{222}Rn activities in GW may however provide reliable insights into changes in mean GW age in a broader range of conditions, inasmuch as GW flow paths remain stable between infiltration and observation points. These results may even extend to conditions where both hydraulic/mass transport properties and the spatial variability of ^{222}Rn production rates are highly heterogeneous.

5.1.3. Study 3 (Chapter 4)

Assimilation of noble gas measurements and the explicit tracer simulation with highly-parameterized ISSHMs can lead to vastly improved model performance compared to the assimilation of hydraulic observations, paving the way for the further development and application of noble gases as artificial tracers of SW-GW interactions.

By jointly assimilating hydraulic and noble gas measurements acquired over the course of a novel tracer experiment, parameter and predictive uncertainties of a highly-parameterized ISSHM of a real-world bank filtration site were reduced far beyond what could be achieved by assimilating hydraulic head measurements only. The sensitivity of the employed gas tracers (^{222}Rn and artificially injected He) to SW-GW exchange processes marks them as excellent tools to provide quantitative information on several management-relevant quantities, such as GW age, SW-GW mixing ratios, in SW infiltration fluxes. The reliability of the novel tracer method presented in this work was verified through the ability of the post-history-matched models to reproduce an independent fluorescent dye tracer dataset. Artificially injected noble gases are capable of providing similar quantitative insights into SW-GW exchanges compared to established dye tracer methods (in this case fluorescein), without the constraints associated with the latter. This study is the first to integrate and illustrate the value of

high-resolution noble gas tracer measurements in an explicit tracer simulation framework. These results strongly support the notion of jointly assimilating hydraulic and tracer observations through the explicit simulation of tracer concentrations in physically-based models, thereby removing the need to introduce a layer of conceptual simplification through the transformation of tracer measurements. Finally, the unique insights offered by the noble gas tracer dataset warrant further research into the continued development of gas tracer methods for studies of SW-GW interactions.

5.1.4. Studies A1 and A2 (see Appendix A)

Large volumes of SW and GW can be efficiently labelled with noble gases using simple and efficient gas injection techniques, and continuously monitored with portable field equipment, expanding the range of available tracer methods for studies of SW and GW flow systems.

In two separate studies, the applicability of artificially injected noble gases as hydrological tracers was assessed. In the first study (Blanc et al, in prep), helium (He) gas was injected into a pre-alpine river for a period of 35 days using a diffusion-injection setup. He was continuously monitored in-river and in a pumping well placed in the adjacent alluvial using two portable gas-equilibrium membrane inlet mass spectrometers (GE-MIMS, Brennwald et al., 2016). River He concentrations were supersaturated by a factor of ten or more above natural concentrations in surface waters (ASW) for the duration of the experiment, despite relatively high river discharge (> 4m^3/s). He breakthrough in the adjacent alluvial aquifer could clearly be observed, identifying SW infiltration in a straightforward and unambiguous manner. The efficiency of the diffusion-injection system makes this option economically manageable; moreover, this method is compatible with long-term tracer injections (durations of weeks to months), which are in general logistically unfeasible using traditional fluorescent dye tracer methods.

In a second study (Brennwald et al., 2022) three different noble gases (He, Kr, Xe) were injected into three piezometers at locations upgradient of a temporary pumping well as Dirac pulses using a bubbling injection method. Dissolved gas concentrations were continuously monitored at the pumping well with a GE-MIMS instrument, where gas tracer breakthrough following tracer injections could be clearly observed. Breakthrough curves were compared between traditional fluorescent dye tracers and noble gas tracers, revealing comparable patterns.

Together, these two studies illustrate how the analytical sensitivity and simplicity of continuous on-site measurements of dissolved gas species combined with efficient gas injection techniques provide new opportunities to assess the internal dynamics of complex flow systems.

5.2. Outlooks and recommendations for future work

Although the research carried out over the course of this thesis has allowed to expand knowledge on the limitations, scope of applicability, and untapped potential of some noble gas tracer methods, several knowledge gaps and avenues for future research have been identified. By addressing these questions, further insights can be gained into the optimal use of noble gas tracer methods in a range of field conditions.

1. Understanding the link between ^{222}Rn emanation/production and depth in alluvial systems

    - Measurements of ^{222}Rn emanation rates at the Aeschau site highlighted the existence of zones of high ^{222}Rn emanation in the near subsurface. This observation is linked to the presence of organic mineral coatings and other weathering products with favourable radionuclide sorption properties in the topsoil. This trend has been noted in other similar contexts (e.g. Mullinger et al., 2009), and whether it is a general property of alluvial aquifers or rather site-specific is an open question. If not accounted for, vertical variability in ^{222}Rn emanation rates may lead to ambiguous signals in cases of water table fluctuations, and in turn biased estimates of ^{222}Rn-derived estimates of GW age, or GW exfiltration fluxes (Mullinger et al., 2019).

2. Assessing the role of hydraulic parameter heterogeneity on ^{222}Rn signatures in GW in 3-D systems

    - In this work, the reliability of the apparent ^{222}Rn age method assuming heterogeneity in both hydraulic subsurface properties and the spatial distribution of ^{222}Rn production rates was only illustrated with a simple 2-D model, and by no means systematically assessed. It is unknown whether the insights gained with this 2-D model can be scaled to highly heterogeneous 3-D systems. Moreover, complex relationships may exist between subsurface structures, hydraulic parameters, aquifer geochemistry, and ^{222}Rn production rates, which were not considered in this study. As ^{222}Rn is one of the few environmental tracers sensitive to timescales relevant to water management needs in bank filtration contexts, further research is warranted into understanding the information content of ^{222}Rn observations in heterogenous environments towards informing GW age.

3. Linking tracers to the parameters and predictions they inform

- As different tracers inform different processes and spatiotemporal scales, data acquisition strategies should aim at collecting measurements of the specific tracers which inform relevant model parameters and predictions. However, there is still much uncertainty as to which data types can most inform different model parameters and predictions, and under which hydraulic forcing their information context is maximised. These questions can be systematically addressed through synthetic modelling studies, in which the worth of individual or combinations of observation types towards informing model parameters and predictions can be quantified. These synthetic modelling approaches should be complemented by field-scale modelling studies, where the "real" value of tracer measurements towards informing management-relevant quantities in a model-based decision support framework is illustrated.

4. Representing gas tracer transport in alluvial systems

    - In this work, gas tracer transport was considered as that of a conservative solute. Although this hypothesis is reasonable for the transport of dissolved noble gases in closed (i.e. single-phase) systems, interactions between GW and the gas phase in the unsaturated zone in unconfined aquifers may lead to tracer exsolution and loss of dissolved tracer mass (Carter, 1959; Gupta, 1994). This phenomenon was identified as the likely cause of lower than expected He tracer recovery rates during the gas tracer injection experiment described in this work. The simplified conceptualization of tracer transport only led to marginal bias for the Aeschau model; however, the same cannot be guaranteed if gas tracer methods are deployed in other alluvial systems. Quantitative estimates of SW-GW exchanges supported by dissolved gas measurements may require more advanced conceptualization of gas exchanges between GW and the unsaturated zone, although it is still not well understood in which hydrogeological contexts, for which tracers, and for which experimental timescales this is necessary. For example, more soluble gases (e.g. Kr, Xe) could be expected to be much less affected by exchange with the unsaturated zone, and subsequently less prone to introducing bias through their assimilation in numerical models.

5. Development of artificial gas tracer methods

    - The two gas injection and measurement methods presented in this work should be seen as "proof-of-concepts" rather than fully-developed tracer techniques. Further

refinements are envisaged for these experimental tools, which would allow (i) improved quantification of injected tracer quantities and tracer recovery rates, (ii) more efficient alternatives for instantaneous injection of large gas volumes compared to bubbling methods, and (iii) finding suitable alternative gas species to nobles gases. The last point has become particularly relevant over the last years, as the price of noble gases has consistently increased and is reaching prohibitive levels. Candidate gas species include light-chain alkanes such as methane (CH_4), ethane (C_2H_6), or propane (C_3H_8), which are safe and minimally reactive in environmental systems. Assessment of the suitability of these gas species (and others) to studies of SW-GW interactions is ongoing.

6. Appendices

On top of the research presented in the main body of this thesis, several collaborations were undertaken with colleagues over the course of my PhD studies, and have resulted in articles being submitted or published in academic journals. The titles and abstracts of the three publications most relevant to the work presented herein are presented in this section. They document work that was done directly within the scope of the Ph.D. project, and tie in with the project described in the previous sections. A short summary of my contributions is outlined below each abstract.

6.1. Appendix A1 : New Experimental Tools to Use Noble Gases as Artificial Tracers for Groundwater Flow

This paper has been published as : Brennwald, M.S., **Peel, M**, Blanc, T., Tomonaga, Y., Kipfer, R., Brunner, P., Hunkeler, D. (2022), *New Experimental Tools to Use Noble Gases as Artificial Tracers for Groundwater Flow*, Frontiers in Water 4, DOI 10.3389/frwa.2022.925294

6.1.1. Abstract

Labelling groundwater by injecting an artificial tracer is a standard and widely used method to study groundwater flow systems. Noble gases dissolved in groundwater are potentially ideal artificial tracers, as they are not subject to biogeochemical transformations, do not adsorb onto the aquifer matrix, are colourless, and have no negative impact on the quality of groundwater resources. In addition, combining different noble-gas species in multi-tracer tests would allow direct analysis of the spatio-temporal heterogeneity of groundwater flow systems. However, while the handling of noble gases is safe and straightforward for injection into groundwater, conventional methods to analyse dissolved noble gases tend to be impractical for groundwater tracer tests. The sampling and subsequent lab-based analysis of dissolved noble gases are laborious, expensive and time intensive. Therefore, only researchers with access to specialized noble-gas labs have attempted such tracer tests. The recently developed gas-equilibrium membrane-inlet mass spectrometers (GE-MIMS) allow efficient on-site analysis of dissolved gases at high temporal resolution. The GE-MIMS instruments thereby eliminate most of the analytical and logistical constraints of conventional lab-based techniques and therefore provide new opportunities for groundwater tests using artificially injected gases. We used a GE-MIMS to systematically test the applicability of He, Kr, and Xe as artificial groundwater tracers. These gas species were injected into groundwater as Dirac-like pulses at three piezometers located at various locations upstream of a pumping well, where dissolved gas concentrations were continuously monitored with the GE-MIMS instrument. The groundwater travel times observed in these tracer tests ranged from a few hours to several weeks, and were consistent with the groundwater flow field at the experimental test site. Travel times determined from the noble gas tracer tests were also consistent with those obtained with traditional dye tracers.

Main contributions

Conceptualization; Development of gas injection method and protocol; Installation of field instruments; Fieldwork; Draft writing

6.2. Appendix A2 : Efficient injection of dissolved gas tracers into rivers: a tool to study surface water – groundwater interactions

This paper is in preparation as : Blanc, T., **Peel, M**., Brennwald, M.S., Kipfer, R., Brunner, P. (2023), Efficient injection of dissolved gas tracers into rivers: a tool to study surface water - groundwater interactions

6.2.1. Abstract (tentative)

Surface water (SW) and groundwater (GW) interactions exhibit complex spatial and temporal patterns that are typically studied using tracers. However, most natural and artificial tracers have limitations in studying SW-GW interactions, particularly when there is no significant contrasts in concentrations between SW and GW exist or can be maintained for long durations. To address the lack of robust tracer methods for labelling infiltrating river water for long durations, (noble) gases have emerged as a promising alternative. With the recent development of portable mass spectrometers, it is now possible to continuously monitor the dissolved concentrations of gases directly in the field. However, injecting gases for long durations into river water presents logistical challenges, particularly in rivers exhibiting highly variable discharge patterns. To overcome this limitation, we present an efficient and robust method for injecting large amounts of gas into a river with a diffusion-injection apparatus. Our setup allows fine, real-time control of the gas injection rate, and is suitable for different gas species and extended injection durations. To illustrate the effectiveness of our method, we present a case study where helium is used as an artificial tracer to study river water infiltration into an alluvial aquifer. This experiment yields valuable information on travel times from the river to a pumping well, as well as mixing ratios between freshly infiltrated river water and regional groundwater.

Main contributions

Conceptualization; Development of gas injection method and protocol; Installation of field instruments; Fieldwork; Draft writing

6.3. Appendix A3 : Explicit simulation of environmental gas tracers with integrated surface and subsurface hydrological models

This paper has been published as : 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*, Frontiers in Water 4, DOI 10.3389/frwa.2022.980030

6.3.1. Abstract

Environmental gas tracers allow inferring groundwater travel times and mixing ratios. Their concentrations are commonly interpreted with simplified and indirect approaches that are conceptually at odds with the high degree of complexity found in natural systems. However, the information content of the tracers can potentially be fully explored through the explicit simulation of an advection-dispersion transport equation, for example using integrated surface-subsurface hydrological models (ISSHMs). These integrated models can be used to explicitly simulate environmental tracers in complex environments. ISSHMs are usually variably saturated flow models. However, these models do not explicitly simulate gas partitioning with the aqueous phase, restricting explicit simulation of gas tracers to fully saturated conditions or to tracers with very low solubilities. We propose a mathematical formulation for the production of environmental gas tracers that are emanated in the subsurface. The production is scaled according to gas/water partitioning and water saturation, which is already computed by the model. Therefore, ISSHMs can now be used to their full potential to explicitly simulate tracer concentrations under variably saturated and dynamic conditions. The new formulation has been successfully verified against reference simulations provided with a multi-phase flow and transport model. In addition, explicit simulation of 222Rn and 37Ar groundwater concentrations in a synthetic alluvial river-groundwater system was demonstrated, for the first time, with an ISSHM.

Main contributions

Derivation of instant partitioning mass-balance equation; Conceptualization of study; Methodology; Draft writing

7. Bibliography

3M 2023. 3M™ Liqui-Cel™ MM Series Membrane Contactor. URL (accessed Jul 11, 2023) : http://multimedia.3m.com/mws/media/1412485O/3m-liqui-cel-membrane-contactors-liquid-degasgaslc-1096-pdf.pdf

Amin, I.E. and Campana, M.E. 1996. A general lumped parameter model for the interpretation of tracer data and transit time calculation in hydrologic systems. Journal of Hydrology 179(1), 1-21.

Anderson, M.P. 2005. Heat as a Ground Water Tracer. Groundwater 43(6), 951-968.

Anderson, M.P., Woessner, W.W. and Hunt, R.J. 2015. Applied Groundwater Modeling (Second Edition), Academic Press, San Diego.

Aquanty 2023. HGS Reference Manual, 366 pp., Waterloo, ON.

Arbenz, P. and Peter, H. 1925. Bericht an die Baudirektion des Kantons Berns betreffend Wasserableitung aus dem Emmental durch die Stadt Bern, 36 pp.

Asher, M.J., Croke, B.F.W., Jakeman, A.J. and Peeters, L.J.M. 2015. A review of surrogate models and their application to groundwater modeling. Water Resources Research 51(8), 5957-5973.

Aster, R.C., Borchers, B. and Thurber, C.H. 2018. Parameter estimation and inverse problems, Elsevier, Cambridge.

Baretto, P.M.C., Clark, R.B. and Adams, J.A.S. 1972. The Natural Radiation Environment II, Symposium Proceedings. Adams, J.A.S., Lowder, W.M. and Gesell, T.F. (eds), pp. 731-740, Houston, TX.

Bayer, P., Huggenberger, P., Renard, P. and Comunian, A. 2011. Three-dimensional high resolution fluvio-glacial aquifer analog: Part 1: Field study. Journal of Hydrology 405(1-2), 1-9.

Bear, J. 1972. Dynamics of fluids in porous media, American Elsevier, New York.

Berman, E.S.F., Gupta, M., Gabrielli, C., Garland, T. and McDonnell, J.J. 2009. High-frequency field-deployable isotope analyzer for hydrological applications. Water Resources Research 45(10).

Bertin, C. and 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. Environmental Science & Technology 28(5), 794-798.

Besmer, M.D., Epting, J., Page, R.M., Sigrist, J.A., Huggenberger, P. and Hammes, F. 2016. Online flow cytometry reveals microbial dynamics influenced by concurrent natural and operational events in groundwater used for drinking water treatment. Scientific Reports 6(1), 38462.

Bethke, C.M. and Johnson, T.M. 2008. Groundwater Age and Groundwater Age Dating *In* R. Jeanloz, A.A., K. Burke, & K. Freeman (eds). Annual Review of Earth and Planetary Sciences., pp. 121-152.

Beyerle, U., Aeschbach-Hertig, W., Hofer, M., Imboden, D.M., Baur, H. and Kipfer, R. 1999. Infiltration of river water to a shallow aquifer investigated with 3H/3He, noble gases and CFCs. Journal of Hydrology 220(3-4), 169-185.

Blau, R.V. and Muchenberger, F. 1997. Grundlagen für Schutz und Bewirtschaftung der Grundwasser des Kantons Bern: Nutzungs-, Schutz- und Überwachungskonzept für die Grundwasserleiter des obersten Emmentals, zwischen Emmenmatt, Langnau und Eggiwil, Synthesebericht, Wasser- und Energiewirtschaft des Kantons Bern, Bern, Switzerland.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R.A.P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G.T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G.B., Claps,

P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T.R., Kohnová, S., Koskela, J.J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J.L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K. and Živković, N. 2019. Changing climate both increases and decreases European river floods. Nature 573(7772), 108-111.

Boano, F., Harvey, J.W., Marion, A., Packman, A.I., Revelli, R., Ridolfi, L. and Wörman, A. 2014. Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. Reviews of Geophysics 52(4), 603-679.

Bouchaud, J.-P. and Georges, A. 1990. Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. Physics Reports 195(4), 127-293.

Boulton, A.J., Datry, T., Kasahara, T., Mutz, M. and Stanford, J.A. 2010. Ecology and management of the hyporheic zone: stream–groundwater interactions of running waters and their floodplains. Journal of the North American Benthological Society 29(1), 26-40.

Bourke, S.A., Cook, P.G., Shanafield, M., Dogramaci, S. and Clark, J.F. 2014. Characterisation of hyporheic exchange in a losing stream using radon-222. Journal of Hydrology 519, 94-105.

Brennwald, M.S., Peel, M., Blanc, T., Tomonaga, Y., Kipfer, R., Brunner, P. and Hunkeler, D. 2022. New Experimental Tools to Use Noble Gases as Artificial Tracers for Groundwater Flow. Frontiers in Water 4.

Brennwald, M.S., Schmidt, M., Oser, J. and Kipfer, R. 2016. A Portable and Autonomous Mass Spectrometric System for On-Site Environmental Gas Analysis. Environmental Science & Technology 50(24), 13455-13463.

Bridge, J. and Demicco, R. 2008. Rivers, alluvial plains, and fans In Bridge, J. and Demicco, R. (eds) Earth Surface Processes, Landforms and Sediment Deposits., pp. 365-461, Cambridge University Press, Cambridge.

Brunner, P., Therrien, R., Renard, P., Simmons, C.T. and Franssen, H.-J.H. 2017. Advances in understanding river-groundwater interactions. Reviews of Geophysics 55(3), 818-854.

Burnett, W.C. and Dulaiova, H. 2006. Radon as a tracer of submarine groundwater discharge into a boat basin in Donnalucata, Sicily. Continental Shelf Research 26(7), 862-873.

Burnett, W.C., Kim, G. and Lane-Smith, D. 2001. A continuous monitor for assessment of 222Rn in the coastal ocean. Journal of Radioanalytical and Nuclear Chemistry 249(1), 167-172.

Carter, R.C., Kaufman, W.J., Orlob, G.T. and Todd, D.K. 1959. Helium as a ground-water tracer. Journal of Geophysical Research (1896-1977) 64(12), 2433-2439.

Cartwright, I. and Gilfedder, B. 2015. Mapping and quantifying groundwater inflows to Deep Creek (Maribyrnong catchment, SE Australia) using 222Rn, implications for protecting groundwater-dependant ecosystems. Applied Geochemistry 52, 118-129.

Cartwright, I. and Hofmann, H. 2016. Using radon to understand parafluvial flows and the changing locations of groundwater inflows in the Avon River, southeast Australia. Hydrology and Earth System Sciences 20(9), 3581-3600.

Cecil, L.D. and Green, J.R. 2000. Radon-222. *In*: Cook, P.G., Herczeg, A.L. (eds) Environmental Tracers in Subsurface Hydrology. Springer, Boston, MA, pp. 175-194.

Chatton, E., Labasque, T., de La Bernardie, J., Guiheneuf, N., Bour, O. and Aquilina, L. 2017. Field Continuous Measurement of Dissolved Gases with a CF-MIMS: Applications to the Physics and Biogeochemistry of Groundwater Flow. Environmental Science & Technology 51(2), 846-854.

Chen, X. 2000. Measurement of streambed hydraulic conductivity and its anisotropy. Environmental Geology 39(12), 1317-1324.

Chen, Y. and Oliver, D.S. 2013. Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. Computational Geosciences 17(4), 689-703.

Chen, Y. and Oliver, D.S. 2016. Localization and regularization for iterative ensemble smoothers. Computational Geosciences 21(1), 13-30.

Cirpka, O.A., Fienen, M.N., Hofer, M., Hoehn, E., Tessarini, A., Kipfer, R. and Kitanidis, P.K. 2007. Analyzing bank filtration by deconvoluting time series of electric conductivity. Groundwater 45(3), 318-328.

Constantz, J. 2008. Heat as a tracer to determine streambed water exchanges. Water Resources Research 44(4).

Cook, P.G. 2013. Estimating groundwater discharge to rivers from river chemistry surveys. Hydrological Processes 27(25), 3694-3707.

Cook, P.G. and Herczeg, A.L. 2000. Environmental Tracers in Subsurface Hydrology, Springer New York, NY.

Cook, P.G., Lamontagne, S., Berhane, D. and Clark, J.F. 2006. Quantifying groundwater discharge to Cockburn River, southeastern Australia, using dissolved gas tracers 222Rn and SF6. Water Resources Research 42(10).

Cook, P.G., Wood, C., White, T., Simmons, C.T., Fass, T. and Brunner, P. 2008. Groundwater inflow to a shallow, poorly-mixed wetland estimated from a mass balance of radon. Journal of Hydrology 354(1-4), 213-226.

Corbett, D.R., Burnett, W.C., Cable, P.H. and Clark, S.B. 1998. A multiple approach to the determination of radon fluxes from sediments. Journal of Radioanalytical and Nuclear Chemistry 236(1-2), 247-253.

Cranswick, R.H., Cook, P.G. and Lamontagne, S. 2014. Hyporheic zone exchange fluxes and residence times inferred from riverbed temperature and radon data. Journal of Hydrology 519, 1870-1881.

Dagan, G. 1989. Flow and transport in porous formations, Springer-Verlag Berlin Heidelberg, 465 pp.

Dausman, A.M., Doherty, J., Langevin, C.D. and Sukop, M.C. 2010. Quantifying data worth toward reducing predictive uncertainty. Groundwater 48(5), 729-740.

Daynes, H.A. 1920. The process of diffusion through a rubber membrane. Proceedings of the Royal Society of London, A 97(685), 286-307.

Delottier, H., Peel, M., Musy, S., Schilling, O.S., Purtschert, R. and Brunner, P. 2022. Explicit simulation of environmental gas tracers with integrated surface and subsurface hydrological models. Frontiers in Water 4.

Delsman, J.R., Winters, P., Vandenbohede, A., Oude Essink, G.H.P. and Lebbe, L. 2016. Global sampling to assess the value of diverse observations in conditioning a real-world groundwater flow and transport model. Water Resources Research 52(3), 1652-1672.

Dimova, N.T., Burnett, W.C., Chanton, J.P. and Corbett, J.E. 2013. Application of radon-222 to investigate groundwater discharge into small shallow lakes. Journal of Hydrology 486, 112-122.

Doherty, J. and Christensen, S. 2011. Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. Water Resources Research 47(12).

Doherty, J. and Hunt, R.J. 2009. Two statistics for evaluating parameter identifiability and error reduction. Journal of Hydrology 366(1-4), 119-127.

Doherty, J. and Welter, D. 2010. A short exploration of structural noise. Water Resources Research 46(5).

Dulaiova, H., Gonneea, M.E., Henderson, P.B. and Charette, M.A. 2008. Geochemical and physical sources of radon variation in a subterranean estuary — Implications for groundwater radon activities in submarine groundwater discharge studies. Marine Chemistry 110(1-2), 120-127.

Durridge 2020. Big Bottle System- High Sensitivity Radon in Water Accessory for the RAD7 With Aerator Cap Revision B. User manual. URL (accessed Jul 11, 2023): https://durridge.com/documentation/Big%20Bottle%20System%20Manual%20Aerator%20Cap%20Rev%20B.pdf

Durridge 2020. RAD7 Electronic Radon Detector User Manual. URL (accessed Jul 11, 2023) : https://durridge.com/documentation/RAD7%20Manual.pdf

Durridge 2020 RAD H2O - Radon in Water Accessory for the RAD7. URL (accessed Jul 11, 2023) : https://durridge.com/documentation/RAD_H2O_Manual.pdf

Ebel, B.A., Mirus, B.B., Heppner, C.S., VanderKwaak, J.E. and Loague, K. 2009. First-order exchange coefficient coupling for simulating surface water–groundwater interactions: parameter sensitivity and consistency with a physics-based approach. Hydrological Processes 23(13), 1949-1959.

Emerick, A.A. and Reynolds, A.C. 2013. Ensemble smoother with multiple data assimilation. Computers & Geosciences 55, 3-15.

Fetter, C.W. 2001. Applied hydrogeology, Prentice Hall, Upper Saddle Creek, NJ.

Fleckenstein, J.H., Krause, S., Hannah, D.M. and Boano, F. 2010. Groundwater-surface water interactions: New methods and models to improve understanding of processes and dynamics. Advances in Water Resources 33(11), 1291-1295.

FOEN 2021. Emme - Emmenmatt, nur Hauptstation (Station 2070). URL (accessed Sep 20, 2021) : https://www.hydrodaten.admin.ch/de/2070.html

FOEN 2023. Emme - Eggiwil, Heidbüel (Station 2409). URL (accessed Mar 15, 2023) : https://www.hydrodaten.admin.ch/de/2409.html

Folger, P.F., Poeter, E., Wanty, R.B., Frishman, D. and Day, W. 1996. Controls on 222Rn Variations in a Fractured Crystalline Rock Aquifer Evaluated Using Aquifer Tests and Geophysical Logging. Groundwater 34(2), 250-261.

Freeze, R.A. and Cherry, J.A. 1979. Groundwater, Englewood Cliffs, NJ.

Frei, S. and Gilfedder, B.S. 2021. Quantifying residence times of bank filtrate: A novel framework using radon as a natural tracer. Water Research 201, 117376.

Gardner, W.P., Hammond, G. and Lichtner, P. 2015. High performance simulation of environmental tracers in heterogeneous domains. Groundwater 53 Suppl 1, 71-80.

Gelhar, L.W., Welty, C. and Rehfeldt, K.R. 1992. A critical review of data on field-scale dispersion in aquifers. Water Resources Research 28(7), 1955-1974.

Gilfedder, B.S., Cartwright, I., Hofmann, H. and Frei, S. 2019. Explicit Modeling of Radon-222 in HydroGeoSphere During Steady State and Dynamic Transient Storage. Groundwater 57(1), 36-47.

Ginn, T.R., Haeri, H., Massoudieh, A. and Foglia, L. 2009. Notes on Groundwater Age in Forward and Inverse Modeling. Transport in Porous Media 79(1), 117-134.

Goode, D.J. 1996. Direct Simulation of Groundwater Age. Water Resources Research 32(2), 289-296.

Goovaerts, P. 1997. Geostatistics for natural resources evaluation, Oxford University Press.

Greeman, D.J. and Rose, A.W. 1996. Factors controlling the emanation of radon and thoron in soils of the eastern U.S.A. Chemical Geology 129(1-2), 1-14.

Gubelmann, H. 1930. Die neue Grundwasser-Fassungsanlage der Wasserversorgung der Stadt Bern, Schweizerischer Verein von Gas- und Wasserfachmänner, Bern, Switzerland.

Gupta, S.K., Lau, L.S. and Moravcik, P.S. 1994. Ground-Water Tracing with Injected Helium. Groundwater 32(1), 96-102.

Heinz, J., Kleineidam, S., Teutsch, G. and Aigner, T. 2003. Heterogeneity patterns of Quaternary glaciofluvial gravel bodies (SW-Germany): application to hydrogeology. Sedimentary Geology 158(1-2), 1-23.

Hermans, T. 2017. Prediction-Focused Approaches: An Opportunity for Hydrology. Groundwater 55(5), 683-687.

Hiscock, K.M. and Grischek, T. 2002. Attenuation of groundwater pollution by bank filtration. Journal of Hydrology 266(3-4), 139-144.

Hoehn, E. and Cirpka, O.A. 2006. Assessing residence times of hyporheic ground water in two alluvial flood plains of the Southern Alps using water temperature and tracers. Hydrology and Earth System Sciences 10(4), 553-563.

Hoehn, E. and Von Gunten, H.R. 1989. Radon in groundwater: A tool to assess infiltration from surface waters to aquifers. Water Resources Research 25(8), 1795-1803.

Hoehn, E., Von Gunten, H.R., Stauffer, F. and Dracos, T. 1992. Radon-222 as a groundwater tracer. A laboratory study. Environmental Science & Technology 26(4), 734-738.

Hrachowitz, M. and Clark, M.P. 2017. HESS Opinions: The complementary merits of competing modelling philosophies in hydrology. Hydrology and Earth System Sciences 21(8), 3953-3973.

Huber, E. and Huggenberger, P. 2015. Morphological perspective on the sedimentary characteristics of a coarse, braided reach: Tagliamento River (NE Italy). Geomorphology 248, 111-124.

Huggenberger, P., Hoehn, E., Beschta, R. and Woessner, W. 1998. Abiotic aspects of channels and floodplains in riparian ecology. Freshwater Biology 40(3), 407-425.

Huggenberger, P. and Regli, C. .2006. A Sedimentological Model to Characterize Braided River Deposits for Hydrogeological Applications In I. Jarvis, G.H. Sambrook Smith, J.L. Best, C.S. Bristow and G.E. Petts (eds) Braided Rivers, pp. 51-74.

Hunt, R.J., Doherty, J. and Tonkin, M.J. 2007. Are models too simple? Arguments for increased parameterization. Groundwater 45(3), 254-262.

Hunt, R.J., Feinstein, D.T., Pint, C.D. and Anderson, M.P. 2006. The importance of diverse data types to calibrate a watershed model of the Trout Lake Basin, Northern Wisconsin, USA. Journal of Hydrology 321(1), 286-296.

Irvine, D.J., Brunner, P., Franssen, H.-J.H. and Simmons, C.T. 2012. Heterogeneous or homogeneous? Implications of simplifying heterogeneous streambeds in models of losing streams. Journal of Hydrology 424-425, 16-23.

Ishimori, Y., Lange, K., Martin, P., Mayya, Y.S. and Phaneuf, M. 2013 Measurement and Calculation of Radon Releases from NORM Residues,IAEA Technical Reports Series No. 474, Vienna, Austria.

Kalbus, E., Reinstorf, F. and Schirmer, M. 2006. Measuring methods for groundwater - surface water interactions: a review. Hydrology and Earth System Sciences 10(6), 873-887.

Käser, D. and Hunkeler, D. 2016. Contribution of alluvial groundwater to the outflow of mountainous catchments. Water Resources Research 52(2), 680-697.

Kazemi, G.A. and Perrochet, P. 2006. Groundwater Age, John Wiley and Sons, Hoboken, New Jersey, USA.

Kipfer, R., Aeschbach-Hertig, W., Peeters, F. and Stute, M. 2002. Noble Gases in Lakes and Ground Waters. Reviews in Mineralogy and Geochemistry 47(1), 615-700.

Kluge, T., von Rohden, C., Sonntag, P., Lorenz, S., Wieser, M., Aeschbach-Hertig, W. and Ilmberger, J. 2012. Localising and quantifying groundwater inflow into lakes using high-precision 222Rn profiles. Journal of Hydrology 450-451, 70-81.

Knowling, M.J., White, J.T. and Moore, C.R. 2019. Role of model parameterization in risk-based decision support: An empirical exploration. Advances in Water Resources 128, 59-73.

Knowling, M.J., White, J.T., Moore, C.R., Rakowski, P. and Hayley, K. 2020. On the assimilation of environmental tracer observations for model-based decision support. Hydrology and Earth System Sciences 24(4), 1677-1689.

Kurtz, W., Hendricks Franssen, H.-J., Kaiser, H.-P. and Vereecken, H. 2014. Joint assimilation of piezometric heads and groundwater temperatures for improved modeling of river-aquifer interactions. Water Resources Research 50(2), 1665-1688.

Lamontagne, S. and Cook, P.G. 2007. Estimation of hyporheic water residence time in situ using222Rn disequilibrium. Limnology and Oceanography: Methods 5(11), 407-416.

Lawson, C.L. and Hanson, R.J. 1995. Solving Least Squares Problems, Society for Industrial and Applied Mathematics.

Leibundgut, C., Maloszewski, P. and Külls, C. 2009. Tracers in Hydrology, John Wiley & Sons, Ltd.

Leray, S., Engdahl, N.B., Massoudieh, A., Bresciani, E. and McCallum, J. 2016. Residence time distributions for hydrologic systems: Mechanistic foundations and steady-state analytical solutions. Journal of Hydrology 543, 67-87.

Li, L., Stetler, L., Cao, Z. and Davis, A. 2018. An iterative normal-score ensemble smoother for dealing with non-Gaussianity in data assimilation. Journal of Hydrology 567, 759-766.

Liao, Z., Osenbrück, K. and Cirpka, O.A. 2014. Non-stationary nonparametric inference of river-to-groundwater travel-time distributions. Journal of Hydrology 519, 3386-3399.

Linde, N., Renard, P., Mukerji, T. and Caers, J. 2015. Geological realism in hydrogeological and geophysical inverse modeling: A review. Advances in Water Resources 86, 86-101.

Lu, C., Chen, X., Cheng, C., Ou, G. and Shu, L. 2012. Horizontal hydraulic conductivity of shallow streambed sediments and comparison with the grain-size analysis results. Hydrological Processes 26(3), 454-466.

Małoszewski, P. and Zuber, A. 1982. Determining the turnover time of groundwater systems with the aid of environmental tracers: 1. Models and their applicability. Journal of Hydrology 57(3), 207-231.

Manning, A.H., Solomon, K.D. and Thiros, S.A. 2005. 3H/3He age data in assessing the susceptibility of wells to contamination. Groundwater 43(3), 353-367.

Margat, J. and van der Gun, J. 2013. Groundwater around the World - A Geographic Synopsis, CRC Press, Boca Raton, FL.

Massoudieh, A. and Ginn, T.R. 2011. The theoretical relation between unstable solutes and groundwater age. Water Resources Research 47(10).

McCallum, J.L., Cook, P.G., Berhane, D., Rumpf, C. and McMahon, G.A. 2012. Quantifying groundwater flows to streams using differential flow gaugings and water chemistry. Journal of Hydrology 416-417, 118-132.

McCallum, J.L., Cook, P.G. and Simmons, C.T. 2015. Limitations of the Use of Environmental Tracers to Infer Groundwater Age. Groundwater 53(S1), 56-70.

McCallum, J.L., Cook, P.G., Simmons, C.T. and Werner, A.D. 2013. Bias of apparent tracer ages in heterogeneous environments. Groundwater 52(2), 239-250.

McCallum, J.L., Engdahl, N.B., Ginn, T.R. and Cook, P.G. 2014. Nonparametric estimation of groundwater residence time distributions: What can environmental tracer data tell us about groundwater residence time? Water Resources Research 50(3), 2022-2038.

Meredith, K.T., Hollins, S.E., Hughes, C.E., Cendón, D.I., Hankin, S. and Stone, D.J.M. 2009. Temporal variation in stable isotopes (18O and 2H) and major ion concentrations within the Darling River between Bourke and Wilcannia due to variable flows, saline groundwater influx and evaporation. Journal of Hydrology 378(3-4), 313-324.

MeteoSwiss 2023 Langnau i.E. weather station (LAG).URL (accessed Jul 06 2023) : https://www.meteosuisse.admin.ch/services-et-publications/applications/valeurs-mesurees-et-reseaux-de-mesure.html#param=messnetz-automatisch&lang=fr&station=LAG&chart=hour

Milly, P.C.D., Dunne, K.A. and Vecchia, A.V. 2005. Global pattern of trends in streamflow and water availability in a changing climate. Nature 438(7066), 347-350.

Moore, C. and Doherty, J. 2005. Role of the calibration process in reducing model predictive error. Water Resources Research 41(5).

Mullinger, N.J., Pates, J.M., Binley, A.M. and Crook, N.P. 2009. Controls on the spatial and temporal variability of 222Rn in riparian groundwater in a lowland Chalk catchment. Journal of Hydrology 376(1-2), 58-69.

Oldham, C.E., Farrow, D.E. and Peiffer, S. 2013. A generalized Damköhler number for classifying material processing in hydrological systems. Hydrology and Earth System Sciences 17(3), 1133-1148.

Paniconi, C. and Putti, M. 2015. Physically based modeling in catchment hydrology at 50: Survey and outlook. Water Resources Research 51(9), 7090-7129.

Partington, D., Brunner, P., Simmons, C.T., Therrien, R., Werner, A.D., Dandy, G.C. and Maier, H.R. 2011. A hydraulic mixing-cell method to quantify the groundwater component of streamflow within spatially distributed fully integrated surface water–groundwater flow models. Environmental Modelling & Software 26(7), 886-898.

Peel, M., Delottier, H., Kipfer, R., Hunkeler, D. and Brunner, P. 2023. Exploring the reliability of 222Rn as a tracer of groundwater age in alluvial aquifers : insights from the explicit simulation of variable 222Rn production. Water Research 235.

Peel, M., Kipfer, R., Hunkeler, D. and Brunner, P. 2022. Variable 222Rn emanation rates in an alluvial aquifer: Limits on using 222Rn as a tracer of surface water – Groundwater interactions. Chemical Geology 599.

Pirot, G., Straubhaar, J. and Renard, P. 2015. A pseudo genetic model of coarse braided-river deposits. Water Resources Research 51(12), 9595-9611.

Pitard, F.F. 1993. Pierre Gy's sampling theory and sampling practice: heterogeneity, sampling correctness, and statistical process control, CRC Press, Boca Raton, FL.

Pittroff, M., Frei, S. and Gilfedder, B.S. 2017. Quantifying nitrate and oxygen reduction rates in the hyporheic zone using 222Rn to upscale biogeochemical turnover in rivers. Water Resources Research 53(1), 563-579.

Pool, M., Carrera, J., Alcolea, A. and Bocanegra, E.M. 2015. A comparison of deterministic and stochastic approaches for regional scale inverse modeling on the Mar del Plata aquifer. Journal of Hydrology 531, 214-229.

Popp, A.L., Pardo-Álvarez, Á., Schilling, O.S., Scheidegger, A., Musy, S., Peel, M., Brunner, P., Purtschert, R., Hunkeler, D. and Kipfer, R. 2021. A Framework for Untangling Transient Groundwater Mixing and Travel Times. Water Resources Research 57(4).

Porcelli, D. and Swarzenski, P.W. 2003. The Behavior of U- and Th-series Nuclides in Groundwater. Reviews in Mineralogy and Geochemistry 52(1), 317-361.

Pyrcz, M.J., Boisvert, J.B. and Deutsch, C.V. 2009. ALLUVSIM: A program for event-based stochastic modeling of fluvial depositional systems. Computers & Geosciences 35(8), 1671-1685.

Renard, P. and Allard, D. 2013. Connectivity metrics for subsurface flow and transport. Advances in Water Resources 51, 168-196.

Sabatini, D.A. 2000. Sorption and Intraparticle Diffusion of Fluorescent Dyes with Consolidated Aquifer Media. Groundwater 38(5), 651-656.

Schaper, J.L., Zarfl, C., Meinikmann, K., Banks, E.W., Baron, S., Cirpka, O.A. and 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 Resources Research.

Schilling, O.S., Cook, P.G. and 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. Reviews of Geophysics 57(1), 146-182.

Schilling, O.S., Gerber, C., Partington, D.J., Purtschert, R., Brennwald, M.S., Kipfer, R., Hunkeler, D. and Brunner, P. 2017. Advancing Physically-Based Flow Simulations of Alluvial Systems Through Atmospheric Noble Gases and the Novel 37Ar Tracer Method. Water Resources Research 53(12), 10465-10490.

Schilling, O.S., Partington, D.J., Doherty, J., Kipfer, R., Hunkeler, D. and Brunner, P. 2022. Buried Paleo-Channel Detection With a Groundwater Model, Tracer-Based Observations, and Spatially Varying, Preferred Anisotropy Pilot Point Calibration. Geophysical Research Letters 49(14).

Schnegg, P.A. 2002 An inexpensive field fluorometer for hydrogeological tracer tests with three tracers and turbidity measurement. Bocanegra E, M.D., Massone H (ed), Mar del Plata, Argentina.

Schubert, M., Buerkin, W., Peña, P., Lopez, A.E. and Balcázar, M. 2006. On-site determination of the radon concentration in water samples: Methodical background and results from laboratory studies and a field-scale test. Radiation Measurements 41(4), 492-497.

Schulze-Makuch, D. 2005. Longitudinal dispersivity data and implications for scaling behavior. Groundwater 43(3), 443-456.

Semkow, T.M. 1990. Recoil-emanation theory applied to radon release from mineral grains. Geochimica et Cosmochimica Acta 54(2), 425-440.

Smerdon, B.D., Payton Gardner, W., Harrington, G.A. and Tickell, S.J. 2012. Identifying the contribution of regional groundwater to the baseflow of a tropical river (Daly River, Australia). Journal of Hydrology 464-465, 107-115.

Sophocleous, M. 2002. Interactions between groundwater and surface water: the state of the science. Hydrogeology Journal 10(1), 52-67.

Sprenger, C., Lorenzen, G., Hulshoff, I., Grutzmacher, G., Ronghang, M. and Pekdeger, A. 2011. Vulnerability of bank filtration systems to climate change. Science of The Total Environment 409(4), 655-663.

Sprenger, M., Stumpp, C., Weiler, M., Aeschbach, W., Allen, S.T., Benettin, P., Dubbert, M., Hartmann, A., Hrachowitz, M., Kirchner, J.W., McDonnell, J.J., Orlowski, N., Penna, D., Pfahl, S., Rinderer, M., Rodriguez, N., Schmidt, M. and Werner, C. 2019. The Demographics of Water: A Review of Water Ages in the Critical Zone. Reviews of Geophysics 57(3), 800-834.

Stellato, L., Terrasi, F., Marzaioli, F., Belli, M., Sansone, U. and Celico, F. 2013. Is 222Rn a suitable tracer of stream–groundwater interactions? A case study in central Italy. Applied Geochemistry 32, 108-117.

Straubhaar, J. 2020. GEONE, Python package for geostatistical and multiple-point statistics modeling, University of Neuchâtel.

Suckow, A. 2014. The age of groundwater – Definitions, models and why we do not need this term. Applied Geochemistry 50, 222-230.

Tang, Q., Schilling, O.S., Kurtz, W., Brunner, P., Vereecken, H. and Hendricks Franssen, H.J. 2018. Simulating Flood-Induced Riverbed Transience Using Unmanned Aerial Vehicles, Physically Based Hydrological Modeling, and the Ensemble Kalman Filter. Water Resources Research 54(11), 9342-9363.

Thamer, B.J., Nielson, K.K. and Felthauser, K. 1981 Effects of moisture on radon emanation including the effects on diffusion. US DOI, R.B.O. (ed).

Therrien, R. and Sudicky, E.A. 1996. Three-dimensional analysis of variably-saturated flow and solute transport in discretely-fractured porous media. Journal of Contaminant Hydrology 23(1), 1-44.

Thiros, N.E., Gardner, W.P. and Kuhlman, K.L. 2021. Utilizing Environmental Tracers to Reduce Groundwater Flow and Transport Model Parameter Uncertainties. Water Resources Research 57(7).

Thiros, N.E., Gardner, W.P., Maneta, M.P. and Brinkerhoff, D.J. 2022. Quantifying subsurface parameter and transport uncertainty using surrogate modelling and environmental tracers. Hydrological Processes 36(11), e14743.

Torgersen, T., Benoit, J. and Mackie, D. 1990. Controls on groundwater Rn-222 concentrations in fractured rock. Geophysical Research Letters 17(6), 845-848.

Tricca, A., Wasserburg, G.J., Porcelli, D. and Baskaran, M. 2001. The transport of U- and Th-series nuclides in a sandy unconfined aquifer. Geochimica et Cosmochimica Acta 65(8), 1187-1210.

Troldborg, L., Refsgaard, J.C., Jensen, K.H. and Engesgaard, P. 2007. The importance of alternative conceptual models for simulation of concentrations in a multi-aquifer system. Hydrogeology Journal 15(5), 843-860.

Tufenkji, N., Ryan, J.N. and Elimelech, M. 2002. The promise of bank filtration, ACS Publications, pp. 423-428A.

Turnadge, C. and Smerdon, B.D. 2014. A review of methods for modelling environmental tracers in groundwater: Advantages of tracer concentration simulation. Journal of Hydrology 519, 3674-3689.

van Leeuwen, P.J., Cheng, Y. and Reich, S. 2015. Nonlinear Data Assimilation, Springer Cham.

Vane, L.M. and Giroux, E.L. 1999. Henry's Law Constants and Micellar Partitioning of Volatile Organic Compounds in Surfactant Solutions. Journal of Chemical & Engineering Data 45(1), 38-47.

Varni, M. and Carrera, J. 1998. Simulation of groundwater age distributions. Water Resources Research 34(12), 3271-3281.

Verhallen, P.T.H.M., Oomen, L.J.P., Elsen, A.J.J.M.v.d., Kruger, J. and Fortuin, J.M.H. 1984. The diffusion coefficients of helium, hydrogen, oxygen and nitrogen in water determined from the permeability of a stagnant liquid layer in the quasi-s. Chemical Engineering Science 39(11), 1535-1541.

Vogt, T., Hoehn, E., Schneider, P., Freund, A., Schirmer, M. and Cirpka, O.A. 2010. Fluctuations of electrical conductivity as a natural tracer for bank filtration in a losing stream. Advances in Water Resources 33(11), 1296-1308.

Vogt, T., Schirmer, M. and Cirpka, O.A. 2012. Investigating riparian groundwater flow close to a losing river using diurnal temperature oscillations at high vertical resolution. Hydrology and Earth Systems Science 16, 473-487.

Waugh, D.W., Hall, T.M. and Haine, T.W.N. 2003. Relationships among tracer ages. Journal of Geophysical Research 108(C5).

Webb, E.K. and Anderson, M.P. 1996. Simulation of Preferential Flow in Three-Dimensional, Heterogeneous Conductivity Fields with Realistic Internal Architecture. Water Resources Research 32(3), 533-545.

Weigel, F. 1978. Radon. Chem Ztg 102(9), 287-299.

Weiss, R. and Smith, L. 1998. Parameter space methods in joint parameter estimation for groundwater flow models. Water Resources Research 34(4), 647-661.

Welter, D.E., White, J.T., Hunt, R.J. and Doherty, J.E. 2015. Approaches in highly parameterized inversion—PEST++ Version 3, a Parameter ESTimation and uncertainty analysis software suite optimized for large environmental models, Reston, VA.

White, J.T. 2018. A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. Environmental Modelling & Software 109, 191-201.

White, J.T., Doherty, J.E. and Hughes, J.D. 2014. Quantifying the predictive consequences of model error with linear subspace analysis. Water Resources Research 50(2), 1152-1173.

White, J.T., Knowling, M.J. and Moore, C.R. 2020. Consequences of Groundwater-Model Vertical Discretization in Risk-Based Decision-Making. Groundwater 58(5), 695-709.

Winter, T.C., Harvey, J.W., Franke, O.L. and Alley, W.M. 1998. Ground water and surface water: A single resource. U.S. Geological Survey, Denver, CO

Würsten, M. 1991. GWB—Hydrogeologische Untersuchungen Aeschau: Schlussbericht, Geoteschnisches Institut, Zürich, Switzerland.

Yang, L., Song, X., Zhang, Y., Han, D., Zhang, B. and Long, D. 2012. Characterizing interactions between surface water and groundwater in the Jialu River basin using major ion chemistry and stable isotopes. Hydrology and Earth System Sciences 16(11), 4265-4277.

Zienkiewicz, O.C., Taylor, R.L. and Nithiarasu, P. 2014. The Finite Element Method for Fluid Dynamics (Seventh Edition), Butterworth-Heinemann, Oxford, 544 pp.